Title
Dynamic Demand Input Preparation for Planning Applications

Author
Jintanakul, Klayut

Publication Date
2009

University of California Transportation Center
UCTC Dissertation UCTC-DISS-2011-02

Dynamic Demand Input Preparation for Planning Applications

Klayut Jintanakul
University of California, Irvine
2011

UNIVERSITY OF CALIFORNIA,

IRVINE

Dynamic Demand Input Preparation for Planning Applications

DISSERTATION

submitted in partial satisfaction of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in Civil Engineering

by

Klayut Jintanakul

Dissertation Committee:

Professor R. Jayakrishnan, Chair

Professor Wilfred W. Recker

Professor Michael G. McNally

2009

The dissertation of Klayut Jintanakul

is approved and is acceptable in quality and form for

publication on microfilm and in digital formats:

_____

_____

_____

Committee Chair

University of California, Irvine

2009

DEDICATION

To my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

APPENDICES

ACKNOWLEDGEMENTS

I would like to acknowledge many people who collaborated either directly or indirectly. I would especially like to thank my esteemed advisor, Professor R. Jayakrishnan, for his invaluable support throughout my doctoral work at the University of California, Irvine. He gave me the freedom to focus my research on the area I wished and was always available to refine my ideas. His insightful comments were always delivered with clarity and wit. Without his guidance, and financial support, this dissertation would not have been possible.

I am very grateful for having an exceptional doctoral committee and wish to thank Professor Wilfred Recker and Professor Michael McNally for reviewing this dissertation and providing constructive comments. I would like to thank Professor Stephen Ritchie and Professor Amelia Regan for their teaching and encouragement. Appreciation is also extended to Professor Wenlong Jin and Professor Jean-Daniel Saphores who provided input on my research. I would also like to thank Professor James Moore and Professor Hanh Dam Le Griffith, my academic advisors at the University of Southern California for their continual support.

I am very thankful to Dr. Younshik Chung, Dr. James Marca, Dr. Craig Rindt, Dr. Hyunmyung Kim, and Dr. Shin-Ting (Cindy) Jeng who were always available to discuss research ideas especially during my early doctoral study at UCI. I am indebted to Dr. Lianyu Chu who was very patient in teaching me many traffic simulation tools. I also owe many thanks to Dr. Hee-Kyung Kim for introducing me to Bayesian statistics and Professor Hal Stern for revising and co-authoring my Bayesian modeling papers. I thank Inchul Yang for assisting me with some computer programming and Ying Jun (Joseph) Chow for offering to read this dissertation and providing very useful comments.

I have been very fortunate to be surrounded by a very productive and cheerful research group at the Institute of Transportation Studies. My graduate student life would have been different without ITS friends, and I wish to thank them all. I am also thankful to Chavinee Poonpipat, Navin Poonpipat, and all friends in Thailand whose affection, encouragement, and blessing helped me overcome several difficulties. My special thanks to Yafen Tseng for her patience and for helping me keep my life in proper perspective and balance.

I would like to express my deepest appreciation to my father and mother for their endless love and unconditional support. No words can express my gratitude to them for the sacrifices, without which this work would not have been completed. I am also very grateful to my brothers for their enthusiasm and encouragement.

This research was partially funded by the University of California Transportation Center. I extend thanks to Elizabeth Deakin for doing an excellent job with overseeing and managing the award.

CURRICULUM VITAE

EDUCATION

2009	Ph.D. in Civil Engineering (Transportation Systems Engineering) University of California, Irvine, USA *(GPA: 3.96)*
2003	M.S. in Civil Engineering (Transportation Engineering) University of Southern California, USA *(GPA: 4.00)*
2001	B.S. in Civil Engineering King Mongkut's University of Technology, Thonburi, Thailand *(GPA: 3.69)*

HONORS, FELLOWSHIPS, AND GRANTS

2009	Finalist of 2009 Best Paper Award, Transportation Research Board (TRB), Committee on Statistical Methodology and Statistical Computer Software in Transportation Research
2007	Ph.D. Dissertation Grants, University of California Transportation Center (UCTC)
2005	Golden Key Medal of Honor in recognition of the most outstanding academic performance, University of California, Irvine
2003	First Class Honors, University of Southern California
2001	First Class Honors, King Mongkut's University of Technology, Thonburi
2001	Medal of Honor from the Engineering Institute of Thailand, presented by the King of Thailand

PUBLICATIONS

Jintanakul, K., Chu, L., Jayakrishnan, R., 2009. Bayesian Mixture Model for Estimating Freeway Travel Time Distributions Using Small Probe Samples from Multiple Days. Transportation Research Record, Journal of the Transportation Research Board, *In press (Nominated for best paper award).*

Jintanakul, K., Chu, L., Jayakrishnan, R., 2008. Hierarchical Bayesian Estimation of Freeway Travel Time Distributions with Small Samples of Non-Identical Probes in Successive Days. Proceeding of the 15th World Congress on Intelligent Transport Systems, ITS World Congress, New York, NY.

Jintanakul, K., Jayakrishnan, R., Stern, H., An Estimation Framework for Time-Dependent Origin-Destination Trip Tables with Sampled Vehicle Trajectories and Observed Link Counts. In preparation for submission to the Transportation Research Part B.

Jintanakul, K., Jayakrishnan, R., Chu, L., Dynamic Demand Input Preparation for Microscopic Traffic Simulation Analysis under Planning Applications. In preparation for submission to the 2010 meeting of the Transportation Research Board.

PRESENTATIONS

Jintanakul, K., Chu, L., Jayakrishnan, R., 2009. Bayesian Mixture Model for Estimating Freeway Travel Time Distributions Using Small Probe Samples from Multiple Days. Presented at the 88th Annual Meeting of the Transportation Research, Washington, D.C.

Jintanakul, K., Chu, L., Jayakrishnan, R., 2008. Hierarchical Bayesian Estimation of Freeway Travel Time Distributions with Small Samples of Non-Identical Probes in Successive Days. Presented at the 15th World Congress on Intelligent Transport Systems, ITS World Congress, New York, NY.

Jayakrishnan, R., Kim, H., Jintanakul, K., 2006. Large Network Simulation: Model Validation Issues. Presented at the 2nd International Symposium of Transport Simulation 2006 EPFL, Lausanne, Switzerland.

TEACHING AND RESEARCH EXPERIENCE

12/2004 - 03/2009	Graduate Research Assistant, Civil and Environmental Engineering Department, the Henry Samueli School of Engineering, University of California, Irvine
10/2003 – 12/2003	Teaching Assistant for “Principles of Transportation Engineering”, School of Civil Engineering, University of Southern California

PROFESSIONAL SERVICE

2007 to Present	Member, Transportation Research Board (TRB)
2007 to Present	Member, Institute of Transportation Engineers (ITE)
2001 to Present	License of Engineer, Civil Engineering in Thailand

ABSTRACT OF THE DISSERTATION

Dynamic Demand Input Preparation for Planning Applications

By

Klayut Jintanakul

Doctor of Philosophy in Civil Engineering

University of California, Irvine, 2009

Professor R. Jayakrishnan

A spectrum of traffic engineering and modern transportation planning problems requires the knowledge of the underlying trip pattern, commonly represented by dynamic Origin-Destination (OD) trip tables. In view of the fact that direct survey of trip pattern is technically problematic and economically infeasible, there have been a great number of methods proposed in the literature for updating the existing OD tables from traffic counts and/or other data sources. Unfortunately, there remain several common theoretical and practical aspects which impact the estimation accuracy and limit the use of these methods from most real-world applications. This dissertation itemizes and examines these critical issues. Then, the dissertation presents the developments, evaluations, and applications of

two new frameworks intended to be used with the current and near-future data, respectively.

The first framework offers a systematic and practical procedure for preparing dynamic demand inputs for microscopic traffic simulation under planning applications with an estimation module based solely on traffic counts. Under this framework, the traditional planning model is augmented with a filter traffic simulation step, which captures important spatial-temporal characteristics of route and traffic patterns within a large surrounding network, to improve the flow estimates entering and leaving the final microscopic simulation network. A new bounded dynamic OD estimation model and a solution algorithm for solving a large problem are also proposed.

The second framework utilizes additional information from small probe samples collected over multiple days. There are two steps under this framework. The first step includes a suite of empirical and hierarchical Bayesian models used in estimating time-dependent travel time distributions, destination fractions, and route fractions from probe data. These models provide multi-level posterior parameters and tend to moderate extreme estimates toward the overall mean with the magnitude depending on their precision, thus overcoming several problems due to non-uniform (over time and space) small sampling rates. The second step involves a construction of initial OD tables, an estimation of route-link fractions via a Monte Carlo simulation, and an updating procedure using a new dynamic OD estimation formulation which can also take into account the stochastic properties of the assignment matrix.

CHAPTER 1 INTRODUCTION

1.1 BACKGROUND

Delay due to traffic congestion is one of the major costs to motorists in metropolitan areas. In 2003, the 85 largest U.S. metropolitan areas experienced 3.7 billion vehicle-hours of delay, amounting to $63 billion in wasted time and fuel (Schrank and Lomax, 2005). Congestion is expected to only get worse. Many economic factors and population growth continue to raise demand for travel, but construction of new highways has not kept pace and it is also unclear if we can build our way out of the congestion problem. During 1980 to 1999, the total length of highway lane-mile in the U.S. increased by only 1.5 percent whereas the total number of traveled miles increased by 76 percent (USDOT, 2000). Since the conventional transportation planning process which has been used for several decades is generally considered suitable only for evaluating and selecting capacity expansion projects, there has been an identified need for transport agencies to include operational analysis into the traditional planning process, so that it is possible to effectively evaluate the efficiency of existing systems and potential alternatives.

Recently, many transportation agencies have considered microscopic traffic simulation as a possible alternative for making decisions on different network improvements. For instance, the California Department of Transportation (Caltrans) has chosen micro-simulation for comprehensive corridor planning for over a dozen California corridors, under a large multi-year program (CCIT, 2006). Unfortunately, the results

from many such early attempts to incorporate *more dynamic* analysis into the planning process have not always been satisfactory. Though there is a fairly correct common perception that the dynamic models may have certain deficiencies, many of which come from the lack of sufficiently well-calibrated models of vehicle and driver behavior, years of improvements on these models have not helped the situation much, and it is increasingly clear that a more serious issue is the well-known *garbage-in, garbage-out* problem. Poor demand inputs into any dynamic analysis for real-world applications will only produce poor results.

A suitable form of transport demand input depends on the kind of problem being considered. In evaluating control strategies for isolated signalized intersections, traffic counts on approach links can typically be considered sufficient. That is, the traffic counts are assumed to be independent of the strategies being evaluated. Clearly, this assumption is not valid for applications under the context of network analysis due to the strong interaction between network performance and individual routing behavior. Therefore, most transport planning and operation tasks require the knowledge of the underlying trip pattern in the network, commonly represented by an Origin-Destination (OD) table. Each cell in an OD table represents the number of trips between a particular origin and destination pair departing during a given time period. Similarly, the use of an OD table as an independent demand input requires the assumption that the strategies being evaluated do not alter the origins and destinations of individuals. In the short term, such an assumption is usually justified since changes in the primary factors affecting the demand for use of the transportation network, such as land use, residential and employment locations are relatively slow compared to the analysis time scale.

OD tables can be further categorized into 2 types: *static* and *dynamic*. A static OD table assumes that all trips in the table are to be completed within one single analysis time slice. The condition within each time slice is further assumed to be uniform and homogeneous. Due to these assumptions, static OD tables are only derived for a relatively large time slice (such as one hour, peak period, and one day). Since all vehicles assigned to the network as flows in the static traffic assignment (STA) are viewed as existing simultaneously on all links of their route, applications based on a static OD table ignore traffic dynamics altogether. On the other hand, dynamic OD tables do not require vehicles to reach their destination within one single time slice and thus can be constructed for a smaller time slice (say, 5 to 30 minute). In addition, dynamic OD tables are to be used with the dynamic traffic assignment (DTA) which explicitly considers the temporal characteristics of traffic, congestion, travel time, and route flow, and they require such shorter time intervals to be meaningful.

In general, there are three major approaches used for obtaining OD demand estimates. The first approach involves some form of direct survey, such as license plate matching technique, household interview, and roadside interview. Although these methods offer valuable information regarding the underlying travel patterns in the network, direct survey of OD demand is usually very costly and time-consuming. Moreover, estimates from these methods are subject to various sources of sampling and factoring-up errors. The second approach infers OD demand using various land use attributes, typically via the traditional four-step model. Under this model, the numbers of trip production and attraction of each zone are estimated by means of a previously calibrated multiple regression model. These trip ends are then distributed into an OD

demand table via some form of the gravity model. The use of the planning model is widely accepted in conventional static analyses both for estimating base-year demand and forecasting future demand. However, the zonal and temporal aggregation required for this approach precludes its use for inferring dynamic OD tables. The last approach is to estimate OD demand tables from a set of observed link counts. The models under this approach typically seek an OD table that once assigned to the network replicates observed link counts as closely as possible subject to some assumptions on the behavior of trip makers and performance of the network. As link counts are readily available from most existing traffic surveillance systems, this approach provides a very cost-effective way to update OD tables. In addition, the use of link counts allows for the estimation of dynamic OD tables through a DTA model which explicitly considers the inter-relationship between dynamic OD demands and traffic flows. Hence, there have been a great number of models of this kind in the literature, both for static and dynamic OD tables. Unfortunately, there remain several theoretical and practical aspects which impact the accuracy of the estimates and limit the use of these methods in real-world applications.

With the recent advances in Automatic Vehicle Identification (AVI) and Global Positioning System (GPS) technologies, vehicle probes are expected to be a new traffic information source. Increasing use of car navigation and GPS-embedded mobile devices also potentially enable road users to share traffic data. According to the Consumer Electronics Association (CEA), the estimated number of GPS-enabled mobile devices sold in the United States in 2006 is 2.28 million which doubles the amount sold in 2005 (Global Ecology Corporation, 2008). It was also forecasted that the market for GPS-

enabled mobile devices in North America will exceed 80 million units in 2010. In Japan, 34 million units have been sold in recent years (Road Bureau MLIT, 2008), with about 57 million total vehicles. These statistics highlight the potential for the use of trajectory information in estimating dynamic OD tables. Thus far, theoretical and algorithmic developments of dynamic OD estimation models with this new kind of data have not been sufficiently accomplished.

Since traffic counts and vehicle probes hold the promise for the cost-effective ways in estimating dynamic OD tables, this dissertation aims to investigate the current developments, itemize and examine critical problems, as well as propose new frameworks and necessary mathematical models under these data approaches.

1.2 DISSERTATION OBJECTIVES

The primary purpose of this dissertation is to develop two new frameworks for preparing dynamic demand inputs for planning applications along with the necessary mathematical models and solution algorithms which can be used with different data sources. The first framework focuses on a traffic simulation-based approach and the use of information that is currently available. The second framework focuses on incorporating additional information from small probe samples of path-wise data from non-identical vehicles collected over multiple days, expected to be obtainable in the near future, with the necessary device and sensor deployment of various kinds already occurring around the world. The objectives under these two frameworks are described separately as follows.

1.2.1 Traffic simulation-based approach with current data

The primary objectives under the first framework are listed as follows:

1. To develop a systematic and implementable procedure for preparing dynamic OD demand inputs for microscopic traffic simulation under a planning context without assuming the availability of prior OD tables, route-choice, or travel time data;
2. To develop a link count based dynamic OD estimation model that incorporates bound constraints for regularizing the estimation and weights to reflect relative reliabilities of the data obtained from different sources and modeling done at different levels of fidelity;
3. To develop a corresponding memory efficient solution algorithm that is suitable for solving a large-scale problem.

1.2.2 Modeling framework with small random samples of probes

The main objectives under the second framework are as follows:

1. To develop a systematic and flexible procedure for preparing dynamic OD demand inputs for planning applications that is statistically sound without

assuming the availability of prior OD tables, route-choice, travel time data, or using traffic simulation models;

2. To develop a suite of data-driven models for estimating time-dependent link travel time distributions, destination fractions, and route-choice fractions from small probe samples collected over multiple days;
3. To develop a method for estimating time-dependent route-link fractions, used in mapping dynamic route flows to a set of observed link counts, from travel time distributions;
4. To develop a method for estimating initial OD flows from destination fractions for different zonal characteristics;
5. To develop a constrained dynamic OD estimation model that can also take into account the stochastic elements in estimated route-link and route-choice fractions;
6. To develop a corresponding memory efficient solution algorithm that is capable of solving a large-scale problem.

1.3 OVERVIEW OF THE PROPOSED FRAMEWORKS

1.3.1 Traffic simulation-based approach with current data

The first framework is a procedure for preparing dynamic demand inputs for microscopic traffic simulation under planning applications, using a new estimation module based

solely on traffic counts. Under this framework, the traditional 4-step planning model is augmented with a filter traffic modeling step, which captures important spatial-temporal characteristics of route and traffic patterns within a large surrounding network, to improve the flow estimates entering and leaving the final microscopic simulation network. Figure 1-1 shows the abstract flowchart of this framework.

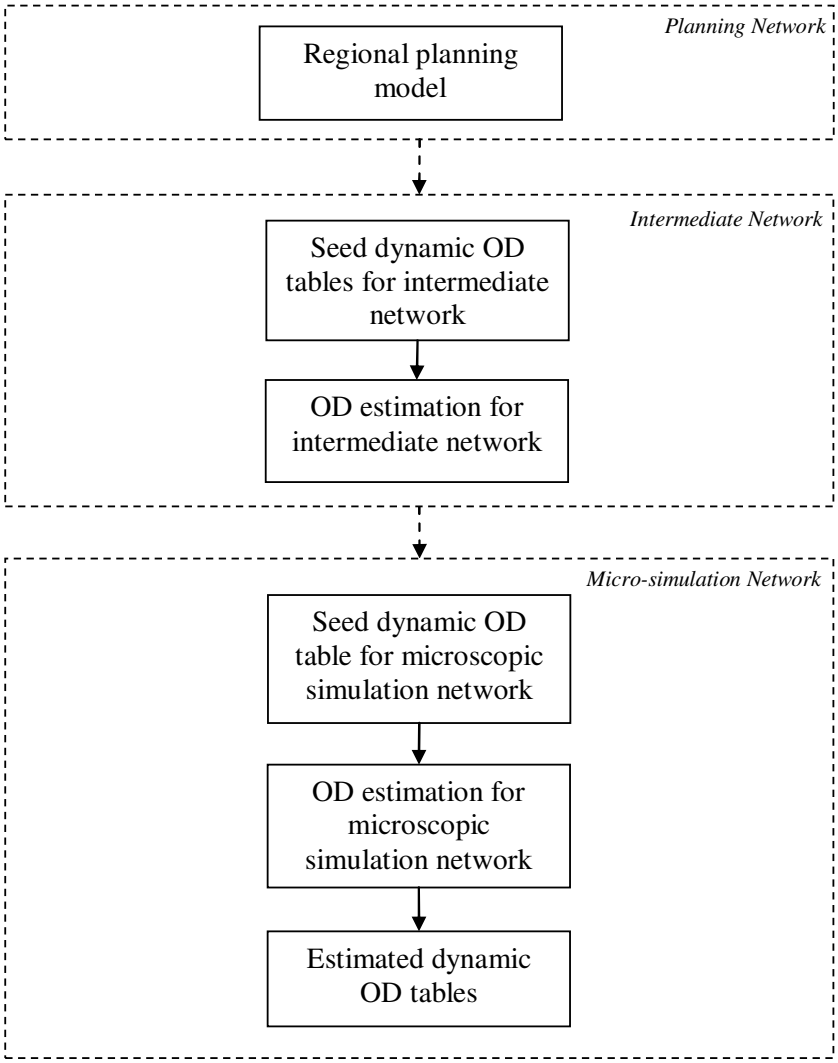


Figure 1-1 Flowchart of the proposed framework for preparing dynamic OD tables for micro-simulation

A bi-level OD estimation scheme is employed both within the filtering step and the micro-simulation step, using a new bounded dynamic OD estimation model which minimizes the deviations of observed and estimated counts and the deviations of seed and estimated OD flows. Since the study networks, especially at the filtering level, usually involve several hundred OD pairs with several departure intervals, the Frank-Wolfe algorithm, a memory-efficient method, is suggested.

1.3.2 Modeling framework with random small samples of probes

The second framework proposed in this dissertation utilizes additional information from small probe samples collected over multiple days. There are two main steps under this procedure as shown in Figure 1-2.

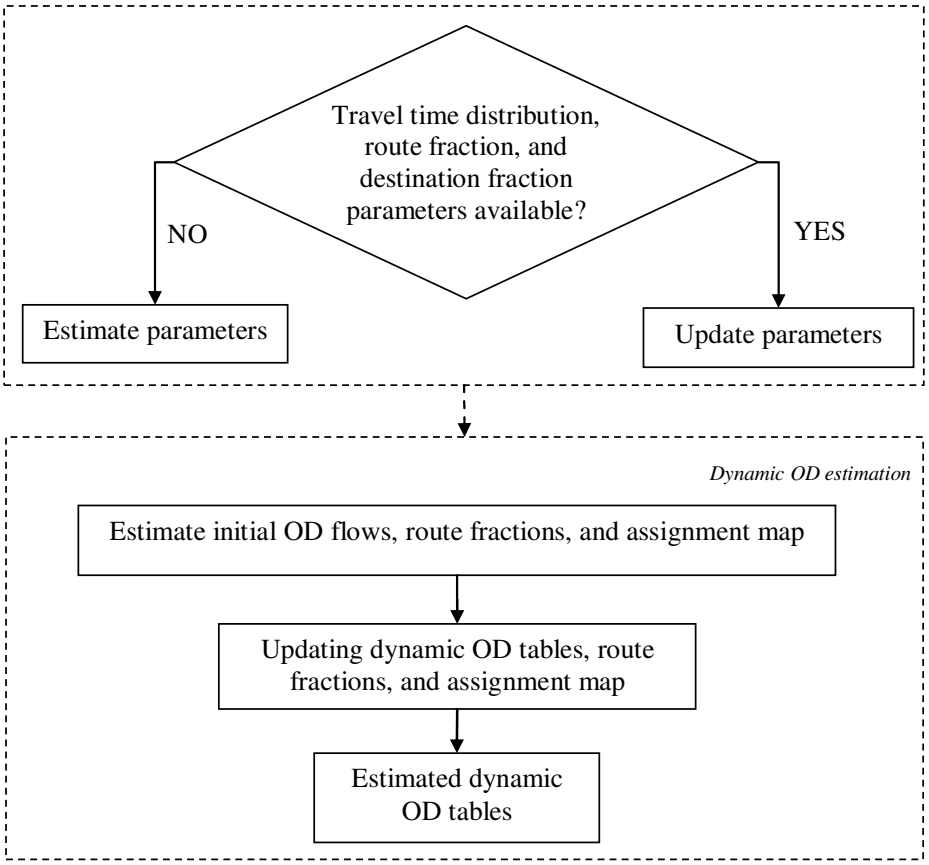


Figure 1-2 Flowchart of the modeling framework with random small probe samples

The first step encompasses a suite of empirical Bayesian (EB) and hierarchical Bayesian (HB) models used in estimating destination fractions, time-dependent travel time distributions, and route fractions from the probe data. These models provide multi-level posterior parameters and tend to moderate extreme estimates toward the overall means with the magnitude depending on their precision, thus overcoming several problems due to non-uniform (over time and space) small sampling rates.

The second step involves a construction of initial OD tables, an estimation of route-link fractions using a Monte Carlo simulation technique, and an updating procedure via a new estimation formulation which simultaneously adjusts OD flows, route fractions, and

route-link fractions. Since the formation is non-convex and involves a large number of variables, a specially designed algorithm based on the Block Coordinate Descent method is proposed.

1.4 DISSERTATION OUTLINE

The overall dissertation outline is shown in Figure 1-3 below.

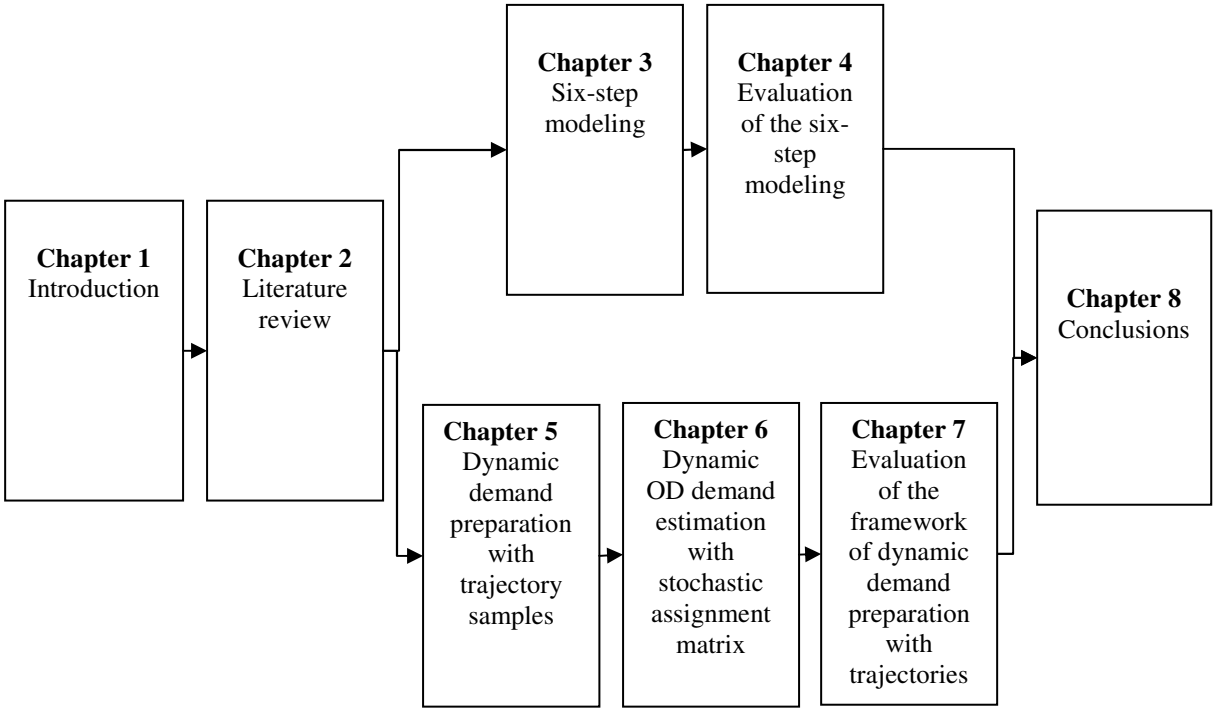


Figure 1-3 Outline of the dissertation

The dissertation consists of 8 chapters. CHAPTER 2 provides a comprehensive review and discussion on both static and dynamic OD estimation models in the literature.

Then, it provides an assessment of the current state-of-practice of schemes implemented in real-world applications. Related theoretical and practical difficulties are indentified, and some commonly overlooked issues are pointed out. Finally, the chapter discusses the reasons why some ad-hoc procedures are being used by many practitioners despite the existence of many dynamic OD estimation models in the literature.

In CHAPTER 3, a six-step modeling framework, a new analytical formulation of the key component (OD estimation model), and a solution algorithm are presented. Some practical considerations regarding dynamic profiles, OD bounds, and weights are discussed. Then, the chapter evaluates several related characteristics of different traffic simulation models which are the critical components of the proposed framework using a set of simulation studies.

In CHAPTER 4, additional experimental studies are conducted. In addition to testing the quality of sub-area dynamic OD tables obtained from the traditional static sub-area analysis and the filtering simulation step, these experiments are designed to further examine their effects on the entire dynamic OD estimation procedures. The proposed methodologies are then applied to the I-880 corridor in order to investigate their efficacy under practical settings.

In CHAPTER 5, developments are presented for several innovative Bayesian methods to efficiently utilize traffic information from small samples of probes collected over multiple days. These developments include a set of models used for estimating link travel time distributions, route-choice fractions, and destination fractions. In addition, the chapter provides a new method for estimating initial OD flows from estimated destination

fractions and observed counts, and a Monte Carlo simulation technique for estimating route-link fractions.

In CHAPTER 6, a new dynamic OD estimation models which simultaneously adjusts OD flows, route fractions, and route-link fractions along with a specially designed algorithm to solve the constrained non-convex optimization problem based on the Block Coordinate Descent method are presented. Some practical guideline on how to choose OD bounds, and weights are provided.

In CHAPTER 7, all proposed methodologies under the second framework are tested using a freeway-arterial network located in Irvine, California. The evaluations are conducted with the probe data from small sampling rates ranging from 1 to 5 percent collected over 21 days.

Finally, CHAPTER 8 summarizes the concluding remarks and findings of the dissertation. The directions for future research are also suggested.

CHAPTER 2 ASSESSMENT OF THE CURRENT STATE-OF-THE-ART

2.1 INTRODUCTION

This chapter reviews several topics related to the preparation of dynamic demand input for planning applications. Section 2.2 begins with an overview of major developments in the literature on link count based static OD estimation. Then, the extensions of these models to the dynamic case are reviewed. A special focus is on identifying the assumptions commonly made and examining how these affect the real-world implementations. The potentials and challenges in using trajectory data available from AVI systems to overcome several common theoretical and practical problems are then discussed. Section 2.3 presents various schemes, commonly employed in practice or suggested in the literature for preparing dynamic OD tables. Critical aspects associated with these schemes are also identified. Section 2.4 points out other important aspects commonly overlooked from most estimation models. Application aspects of dynamic demand inputs are emphasized. Conclusions are finally given in section 2.5.

2.2 LINK COUNT BASED OD ESTIMATION MODELS

2.2.1 Static OD estimation

Estimating or updating a static OD demand table from observed link counts is regarded as the most cost effective way in deriving the underlying travel pattern. For this reason, it has received substantial attention from many researchers. Under the static approach, traffic is assumed to be in the steady state over the entire analysis period; and vehicles, once assigned to the network, are assumed to exist simultaneously on all links on their chosen route. The static OD estimation can thus be thought of as the inverse of the static traffic assignment.

Before discussing the developments in the literature along this direction, it is instructive to first mention the *static* relationship between observed link counts and static OD flows. This relationship is commonly used in most static OD estimation models and can be expressed as follows.

$$v_l = \sum_{ij} a_l^{ij} x^{ij} \tag{2-1}$$

where

v_l The traffic count on link l during the analysis period

x^{ij} The OD flow from i to j during the analysis period

a_l^{ij} The fraction of x^{ij} that contributes to v_l (assignment fraction)

Fundamentally, assignment fractions can be calculated by combining the static route-link incidence matrix with route-choice fractions. Nonetheless, since the route-choice fractions are usually not known a priori and not independent on the unknown OD flows, early studies mainly focused on developing effective static OD estimation models for linear networks; or assumed that the route-choice fractions are independent of OD flows (or traffic) and thus can be exogenously determined before entering the OD estimation process.

Based on the information and entropy theory, Van Zuylen and Willumsen (1980) developed two models to estimate the OD table that are consistent with both available traffic counts and the prior (seed) OD table. The first model, based on the information minimization, considers each traffic count on a network link as an observation unit. The estimates can be obtained by solving the following problem.

$$\min_{x} I \approx \sum_{l} \sum_{ij} \left[x^{ij} a_l^{ij} \ln\left(\frac{s_l x^{ij}}{v_l \breve{x}_{ij}} \right) \right] \tag{2-2}$$

subject to

$$v_l = \sum_{ij} a_l^{ij} x^{ij} \tag{2-3}$$

where $s_l = \sum_{ij} \breve{x}_{ij} a_l^{ij}$, and $\breve{x}_{ij}$ is the prior (seed) OD flow from i to j.

One shortcoming of this model is that it requires node and path flow continuity. This condition is rarely the case in practice due to both traffic dynamics and measurement

errors. Therefore, data pre-processing is needed before using the model. The second model, based on the entropy maximization, regards each trip as a unit of observation. The solution of this model can be obtained by solving

$$\max_{x} W \approx \sum_{ij} \left[x^{ij} \ln\left(\frac{x^{ij}}{q_{ij}} \right) \right] \tag{2-4}$$

subject to

$$v_l = \sum_{ij} a_l^{ij} x^{ij} \tag{2-5}$$

where $q_{ij} = \dfrac{\breve{x}_{ij}}{\sum_{ij} \breve{x}_{ij}}$.

The main shortcoming of this model is twofold. First, the model assumes that the total number of trips is constant. This is a very restrictive assumption given that an outdated seed OD table is often used in practice. Second, the estimates are highly sensitive to the magnitude of the prior OD table. These two models lead to the trip distribution models of the gravity type. Similar models have also been proposed by different authors such as Wilson (1970); Snickars and Weibull, (1977); and Nguyen (1983). Nguyen (1984) provided a comprehensive review of several models under this approach.

Cascetta (1984) developed a Generalized Least Squares (GLS) model for estimating a static OD table. Under this approach, the solution can be obtained by solving the following quadratic program.

$$\min \begin{bmatrix} \breve{\mathbf{x}} - \mathbf{x} \\ \breve{\mathbf{v}} - \breve{\mathbf{a}}\mathbf{x} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{U}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}^{-1} \end{bmatrix} \cdot \begin{bmatrix} \breve{\mathbf{x}} - \mathbf{x} \\ \breve{\mathbf{v}} - \breve{\mathbf{a}}\mathbf{x} \end{bmatrix} \tag{2-6}$$

where

$\mathbf{U}$ = The variance-covariance matrix of the error term of the OD flows

$\mathbf{W}$ = The variance-covariance matrix of the error term of the link counts

Under the assumptions that the seed OD table is unbiased and that the errors from assignment matrix are negligible, the GLS estimator is the best linear unbiased estimator (BLUE). One advantage of this approach is that distributional assumptions for the random error terms of link counts and OD flows are not required. However, if the error terms can be considered to be distributed according to the multivariate normal, the GLS estimator coincides with the Maximum likelihood (ML) estimator or Bayesian estimator with multivariate normal assumption (Maher, 1983). The solution obtained is then the minimum variance estimator among all unbiased ones. Cascetta and Nguyen (1988) discussed various methods to calculate the variance-covariance matrices of the error terms. Note that although the OD flows should be constrained by the non-negativity condition, the author only developed the expression for the case where such constraints are inactive. Bell (1991) overcame this problem by proposing an algorithm which can be used in solving the problem with active non-negativity constraints.

The maximum likelihood (ML) approach (see, for example, Spiess, 1987) maximizes the likelihood of observing sampled OD table and traffic counts conditional on the underlying OD demand. For simplicity, these two sets of observations are often assumed to be independent, leading to the following form of the likelihood function:

$$L(\breve{\mathbf{x}}, \breve{\mathbf{v}} \mid \mathbf{x}) = L(\breve{\mathbf{x}} \mid \mathbf{x}).L(\breve{\mathbf{v}} \mid \mathbf{x}) \tag{2-7}$$

The ML estimator can then be obtained by maximizing the likelihood function or more conveniently its natural logarithm.

$$\mathbf{x} = \arg\max_{\mathbf{x} \in \mathbf{S}} \left[\ln L(\breve{\mathbf{x}} \mid \mathbf{x}) + \ln L(\breve{\mathbf{v}} \mid \mathbf{x}) \right] \tag{2-8}$$

where $\mathbf{S}$ = The set of feasible OD flows.

The specification of the objective function depends on the distributional assumption made on $\breve{\mathbf{x}}$ and $\breve{\mathbf{v}}$; or equivalently on their residual term. For example, if the observed traffic counts are assumed to be independent Poisson variates (Van Zuylen and Branston, 1982), then the natural logarithm of the corresponding likelihood becomes

$$\ln L(\breve{\mathbf{v}} \mid \mathbf{x}) = \sum_l \left[\breve{v}_l \ln \breve{\mathbf{a}}(\mathbf{x})_l - \breve{\mathbf{a}}(\mathbf{x})_l \right] + \text{constant} \tag{2-9}$$

Alternatively, if the multivariate normal is assumed (see, Maher, 1983; Cascetta, 1984), then

$$\ln L(\breve{\mathbf{v}} \mid \mathbf{x}) = -\frac{1}{2}(\breve{\mathbf{v}} - \breve{\mathbf{a}}(\mathbf{x}))' \mathbf{W}^{-1} (\breve{\mathbf{v}} - \breve{\mathbf{a}}(\mathbf{x})) + \text{constant} \tag{2-10}$$

For the sampled OD table, if the sample is obtained by simple random sampling of n_i out of x_i vehicles at each origin zone i (sampling fraction $\alpha_i = n_i / x_i$) with the sample $\breve{\mathbf{x}}$ assumed to have a multinomial distribution, then

$$\ln L(\breve{\mathbf{x}} \mid \mathbf{x}) = \sum_{ij} n_{ij} \ln(\alpha_i x_{ij}) + \text{constant} \tag{2-11}$$

with the feasible space $\mathbf{S}$: $\sum_j x_{ij} = \frac{n_i}{\alpha_i}$ and $x_{ij} \geq 0$.

On the other hand, if the Poisson distribution is assumed, then

$$\ln L(\breve{\mathbf{x}} \mid \mathbf{x}) = \sum_{ij} \left[n_{ij} \ln(\alpha_i x_{ij}) - \alpha_i x_{ij} \right] + \text{constant} \tag{2-12}$$

The advantage of the GLS, ML, and Bayesian estimators over the models based on the information and entropy theory is that these models allow one to place some measure of reliability on different seed OD flows and observed link counts. A comprehensive review of these statistical approaches can be found in Cascetta and Nguyen (1988).

The literature reports discussed thus far are confined to the methods for estimating static OD tables assuming that the assignment matrix is independent of the unknown OD flows and/or can be determined prior to the estimation process. In reality, route-fractions are dependent on the unknown underlying demand. As the flow rate on a given route increases, the cost of using the route (e.g., travel time, delay) also increases until the point where drivers change to an alternative route with the lower cost. To account for this interdependency, Nguyen (1977) developed a variable demand user equilibrium

assignment model by incorporating traffic counts as side constraints. Similar models were also proposed by Turnquist and Gur (1979); LeBlance and Farhangian (1982); and Sheffi (1985). Using game theory, Fisk (1988) extended the entropy model of Van Zuylen and Willumsen (1980) by introducing the user-equilibrium condition as constraints. Fisk (1989) showed that if the observed traffic counts are in user equilibrium, his extended entropy model has the same solution as the combined trip distribution and assignment model (see, for example, Erlander et al., 1979; Fisk and Boyce, 1983).

Spiess (1990) developed a gradient based approach for updating an OD table to reproduce traffic counts by iteratively calculating directions based on the gradient of the objective function. Yang et al. (1992, 1994) extended the GLS model of Cascetta (1984) by formulating a non-convex bi-level program. The GLS is at the upper level, and the equilibrium assignment is at the lower level. A heuristic algorithm which iteratively solves the upper and lower level was also proposed. This algorithm has been shown to converge but does not guarantee that the constraints are met. Similar methods have been proposed in Florian and Chen (1993); Chen (1994); and Sherali et al. (1994).

Lo et al. (1996) and Lo and Chan (2003) provided a statistical model which can simultaneously update the OD table and assignment matrix. However, necessary constraints to guarantee the physical meaning of the final updates were not included. Similarly, Yang et al. (2001) proposed a model which simultaneously updates OD tables and the travel-cost coefficient in the stochastic user equilibrium. More recently, Doblas and Benitez (2005) developed an efficient algorithm based on the Augmented Lagrangian Function (ALM) to update an OD table while preserving the prior structure of the seed OD table.

Despite the presence of various approaches for estimating or updating static OD table from link counts, it should be noted that most of these models require a reliable prior OD table to obtain reliable results. Also, the complexity of these models is dependent to a great deal on the assumptions made on the assignment matrix.

2.2.2 Dynamic OD estimation

As the applications of the dynamic traffic assignment (DTA) have received increasing attention from transportation researchers and practitioners, there have also been a considerable number of studies in the literature devoted to the estimation of dynamic OD tables from observed time-dependent traffic counts. Fundamentally, these studies focus on expanding the capability of existing static OD estimation models to sufficiently describe traffic dynamic and user behavior within the network (Zhou, 2004).

Again, before discussing the developments along this direction, it is helpful to first mention the general form of the *dynamic* relationship used in mapping between dynamic OD flows and observed time-dependent link counts. This relationship can be expressed as follows.

$$v_{lh} = \sum_{t=1}^{h} \sum_{ij} a_{lh}^{ijt} x^{ijt} \tag{2-13}$$

or equivalently,

$$v_{lh} = \sum_{t=1}^{h} \sum_{k \in K} m_{lh}^{kt} f^{kt} \tag{2-14}$$

where

v_{lh} The traffic count on link l observed during interval h

x^{ijt} The OD flow from i to j departing during interval t

f^{kt} The flow on route k departing during interval t

a_{lh}^{ijt} The fraction of x^{ijt} that contributes to v_{lh}, known as *assignment fraction*

m_{lh}^{kt} The fraction of f^{kt} that contributes to v_{lh}, known as *route-link fraction*

Compared to the static case, the assignment fractions include two additional dimensions: t - departure time interval, and h - observation time interval. Also, the route-link fractions, the dynamic counterpart of the zero-one static route-link fractions, can now take on any value in the range of zero and one.

Similar to the static case, all existing models can be generally categorized into two groups: non-assignment based and assignment based. The non-assignment based approach focuses on estimating dynamic OD tables for freeway networks of a linear configuration or turning movements for intersections. For freeway networks, if there is no route-choice involved, the estimation problem essentially reduces to that of estimating turning fractions from a time-series of traffic counts, and the remaining challenges pertain mostly to the modeling of traffic movement. Early studies under this approach (Cremer and Keller, 1981, 1984, 1987; Cremer, 1983; Nihan and Davis, 1987, 1989) assumed that vehicles' travel times through the junction or network is constant and is either small in relation to the chosen time interval or equal to some fixed number of time

intervals. Then, least squares, Kalman Filtering, recursive formula, or constrained optimization can be used to solve the system of equations. An insightful review of studies in this category can be found in Bell (1991a).

To account for travel time variations, Bell (1991b) developed two methods based on the constrained recursive least squares (CRLS) for estimating dynamic OD flows when the distribution of travel times spans a number of intervals. The first method assumes that travel times are geometrically distributed. This assumption is known to be appropriate only for short segments and unable to capture the build-up and dispersal of queues. Therefore, the method can only be used for uncongested intersections and small networks. The second method does not make any specific assumption regarding travel time distributions but rather require the knowledge of the shortest and longest lags for vehicles to arrive at each exit. However, there are substantially more parameters to be estimated. Chang and Wu (1994) used the macroscopic traffic characteristics to construct a set of dynamic equations to establish the interrelations between OD flows and link counts for a congested network. This method requires density measurements and section speed conversion, which are impractical since most of the current loop detectors do not provide such data. They also assumed that the speed of vehicles entering the freeway during the same interval is distributed in a small range over a number of intervals. Wu and Chang (1996) extend this model to use both the time series of link flows and screenline flows. In the study, travel times are assumed to follow the normal distribution. More recently, Lin and Chang (2007) presented a general analytical expression to incorporate travel time distributions into the formulation. That is, if the travel times of drivers from i to j departing during interval t follow a certain distribution with mean

μ_{ijt} and variance σ_{ijt}^2, then the fraction of vehicles that take m intervals to arrive at j, denoted by ρ_{ijt}^m, can be expressed as a cumulative density function.

$$\hat{\rho}_{ijt}^m = \int_{m.t_0}^{(m+1).t_0} f_{ijt}(x)dx \tag{2-15}$$

where t_0 is the length of one unit time interval, and $f_{ijt}(x)$ is the density function of travel time of vehicles traveling from i to j departing during interval t. Only a simulation study with the assumption of normal travel times was illustrated in the study, however.

The assignment-based approach, intended for estimating dynamic OD tables for a general network, assumes the existence of an accurate DTA model and available prior OD tables. Under this approach, Okutani (1987) was the first to use the Kalman Filter to estimate dynamic OD tables. Ashok and Ben-Akiva (1993) modified this model by defining the deviations of the OD flows from the prior estimates as state variables instead of the OD flows themselves to address the problem of the auto-regressive specification in the original work. Cascetta et al. (1993) extended the GLS model for dynamic OD estimation and provided two estimation methods, the simultaneous and sequential estimators. They also suggested three possible dynamic network loading models (DNL). The first method is to compute the route-link fractions as the equilibrium values using a DTA model although the problems from the interdependency between OD and route flows were not mentioned. The second method called a *discrete packet approach* assumes that all users departing from the same path during the same interval (packet) act

as a single user, and thus the route-link fractions only take on the value of either zero or one. In the last method called *continuous packet approach*, each packet is assumed to be continuously spread over the interval between the head and the tail, which are further assumed to remain separated by a constant time headway of one departure time interval (T). Under this method, the estimates of route-link fraction, $\hat{m}_{lh}^{kt}$, can be obtained as follows.

$$\hat{m}_{lh}^{kt} = \begin{cases} 1-\left(\dfrac{t_l^{kt}}{T}-(h-1)\right) & if\ (h-1).T < t_l^{kt} < h.T \\ 1-\hat{m}_{l(h-1)}^{kt} & if\ (h-2).T < t_l^{kt} < (h-1).T \\ 0 & Otherwise \end{cases} \tag{2-16}$$

where t_l^{kt} is the arrival time of the head of route flow, f^{kt}, at link l.

Yang et al. (1998) proposed a method to estimate dynamic OD tables based on the error back-propagation learning algorithm in neural network theory. Ashok and Ben-Akiva (2000) developed two different approaches for real-time estimation and prediction of dynamic OD flows based on the state-space model. Tavana and Mahmassani (2001) proposed a bi-level optimization model and an iterative solution framework. The upper level is the demand estimation model based on the GLS formulation, and the lower level is the simulation-based dynamic traffic assignment. Zhou (2004) extended this framework to utilize traffic counts from multiple days.

Ashok and Ben-Akiva (2002) suggested another method for estimating route-link fractions taking into account the effect of *stretching* and *squeezing* as stated below.

$$\hat{m}_{lh}^{kt} = \begin{cases} 1 & if\ (h-1).T < t_{1l}^{kt} < t_{2l}^{kt} < h.T \\ \left(h.T - t_{1l}^{kt}\right) \Big/ \left(t_{2l}^{kt} - t_{1l}^{kt}\right) & if\ (h-1).T < t_{1l}^{kt} < h.T < t_{2l}^{kt} \\ T \Big/ \left(t_{2l}^{kt} - t_{1l}^{kt}\right) & if\ t_{1l}^{kt} < (h-1).T < h.T < t_{2l}^{kt} \\ \left(t_{2l}^{kt} - (h-1).T\right) \Big/ \left(t_{2l}^{kt} - t_{1l}^{kt}\right) & if\ t_{1l}^{kt} < (h-1).T < t_{2l}^{kt} < h.T \\ 0 & Otherwise \end{cases} \quad (2\text{-}17)$$

where t_{1l}^{kt} and t_{2l}^{kt} are respectively the arrival time of the head and the tail of the route flow, f^{kt}, at link l.

Compared to the continuous packet approach proposed by Cascetta et al. (1993), this method can significantly improve the accuracy of the estimates, especially for the case where the trip durations are relatively large and/or travel time variation across successive intervals are significant. However, the use of both t_{1l}^{kt} and t_{2l}^{kt} relies on the assumption that the detailed vehicle trajectory data are available.

Ashok and Ben-Akiva (2002) also pointed out that the impact of the errors in the estimated assignment matrix to the accuracy of the OD estimates could be significant. Accordingly, they proposed two approaches to introduce the stochastic properties of the estimated assignment matrix into the problem. The first directly adds a set of equations of the error in assignment fraction and the latter adds two set of equations, one of the travel

time error and the other for route-fraction error. However, necessary constraints were absent.

To the best of our knowledge, none of the dynamic OD estimation models developed thus far was tested without any form of assumed knowledge regarding the prior OD table, route-choice fractions, and/or travel times. Due to the strong dependencies among these parameters, evaluating a model with even just partial information related to these factors can result in unreasonable justification. Regardless of the model performances reported, these models might not be adequately comprehensive for real-world applications where the mentioned knowledge rarely exists. It is thus fair to conclude that developing a complete framework for estimating dynamic OD tables, rather than solely an estimation model, has not been the primary goal of the studies currently in the literature.

2.2.3 OD estimation based on link count and AVI data

While there are many OD estimation models proposed in the literature, it is easy to expect that demand estimates, regardless of the type of the estimator, can significantly be improved if more reliable information regarding initial (seed) OD flows, traffic dynamics, and route-choice behavior are available. Fortunately, the intelligent transportation systems (ITS), which have rapidly emerged in recent years, often include the Automated Vehicle Identification technique (AVI) technologies. These technologies are capable of detecting and identifying individual vehicles equipped with the appropriate

device at various network locations, thus offering important information that can be used for improving dynamic OD demand estimates.

AVI systems can be grouped in different ways. Nanthawichit et al. (2003) categorized tag-reader equipped vehicles as a *space-based* probe, and wireless communication equipped vehicles as a *time-based* probe. Antoniou et al. (2004) classified AVI technologies along two dimensions: network coverage and vehicle coverage, as shown in Table 2-1.

Table 2-1 Classification of indicative AVI technologies by scope

		Network Coverage	
		Area-wide	Short-range
Vehicle Coverage	All vehicles	N/A	License plate recognition
	Equipped vehicles	GPS-based Cell phone tracking	Transponder detection

Source: Antoniou et al. (2004)

For GPS-based and cell phone tracking technologies, experienced travel times between any two locations can be calculated directly using the corresponding time-stamps. Also, information regarding OD and route flows as well as the usual route-sets is available. On the other hand, in the case of license plate recognition and transponder detection systems, probe travel times can only be obtained for sections between two stations. To determine OD demand and route flows, some assumptions need to be made. For example, Antoniou et al. (2004) assumed that each sensor location is sufficiently close to a particular origin or destination zone.

Although AVI technologies provide samples of vehicle trajectories that can be used to remedy several theoretical and practical problems in estimating OD tables, the use of

such data is not straightforward, especially when system dynamics are of concern due to low market penetration and sampling rates. Hellinga (1994) noted the use of an overall estimated probe market penetration to scale up sampled OD trip table when the AVI sampling is distributed randomly and evenly. However, this approach suffers from serious bias as the probe market penetration rate itself might vary greatly over time and space. Besides, many paths and OD pairs with zero observations cannot be updated.

Some studies consider the sampling rate as a decision variable in their optimization problem. For instance, Van der Zijpp (1996) proposed a constrained optimization formulation that is used to simultaneously update OD flows and identification rates of the automated license-plate survey. Asakura et al. (2000) focused on the off-line OD estimation and adopted the least squares model to jointly update OD flows and identification rates.

Dixon and Rilett (2002) provide a scheme to estimate tagged vehicle OD tables. By assuming that AVI stations are origin and destination zones, they used AVI data as the direct observation in the off-line GLS estimator and the online Kalman Filter. Eisenman and List (2004) suggested a simple method to calculate the overall AVI sampling rate for estimating the initial population OD table. The method first calculates location-dependent AVI sampling rates from the ratio between the total trips and the total probes on the corresponding link. Then, the average sampling rate is used for factoring up the entire sampled OD table. Dixon and Rilett (2005) extended their original work (Dixon and Rilett, 2002) to estimate initial OD tables for a freeway network. In the study, a time-dependent AVI sampling rate is calculated for each AVI station serving as an origin zone using the same method as Eisenman and List (2004). The initial OD tables for the

freeway network are then obtained by assigning the AVI OD tables according to some heuristic distributions.

Mishalani et al. (2002) and Antoniou et al. (2004) focused on the benefits from intersection turning fractions and link travel time measurements from sensing-based surveillance systems in estimating real-time OD flows. The Kalman Filtering model of Ashok and Ben-Akiva (1993) was employed. However, assumed historical OD tables were also used to overcome the rank deficient problem. Realizing the difficulty in estimating probe sampling rates, Zhou and Mahmassani (2006) propose a time-dependent OD estimation model that utilizes only point-to-point split fraction information from AVI data. Such information is combined with link counts via a multi-objective formulation. Although the use of point-to-point data avoids the need to estimate the sampling rates, the model does not fully utilize the information from trajectory data.

It should be noted that most studies mentioned above employ a constant expansion factor (in time and/or space) to gross up the observed OD flow. However, the expected market penetration and sampling rate that is currently available or will be available in the near future (from potential users and/or possible government-supported sampling schemes) is too low for a simple expansion technique to be suitable. The *No observation problem*, which can easily occur in the case of a dynamic OD estimation with a short interval, is an important example of some potential problems that can arise from small samples. Furthermore, while probe data can ideally be used to add in observations of the underlying route use proportions and traffic characteristics to the formulation of OD estimation models, there is a need for a systematic investigation of such possibilities.

2.3 CURRENT PRACTICE OF DYNAMIC DEMAND PREPARATION

While DTA has the potential to overcome many of the long-standing problems in the traditional static approach, transportation researchers and practitioners have also realized the fact that the theory of DTA is still relatively underdeveloped (Peeta and Ziliaskopoulos, 2001). Nevertheless, given the immediate need for dynamic analysis for many modern operation and planning applications, a simulation-based DTA has already been deployed by transportation agencies in recent years. As mentioned in subsection 2.2.2, because a complete procedure for preparing dynamic OD inputs does not currently exist, it should not be surprising that there are a variety of schemes, some of them rather ad-hoc, being implemented in the real-world applications and recommended in the academic literature. These schemes include:

- The use of a static OD table (or low resolution dynamic OD tables)
- The use of a static OD table with a uniform departure profile as a seed for dynamic OD estimation
- The use of a static OD table for regularizing dynamic OD flows
- The use of some behavioral rules to convert a static OD table to dynamic OD tables

2.3.1 The use of a static OD table (or low resolution dynamic OD tables)

This approach sets the resolution of the OD table equal to the entire analysis period or a long period (e.g. hourly) and initiates the estimation process with a static OD table cut out from a static planning model (see for example, Breiland et al., 2006). Having assumed that the demand is in the static state over some defined period, the static relationship (e.g. equation 2-1) is used as the basis for adjusting OD flows to replicate observed link counts. A variation of this approach includes holding the total demand flows fixed while heuristically adjusting the departure profile from each origin zone.

Although the methods mentioned above might appear instinctive as one can expect the seed OD flows to be adjusted with observed link counts, the quality of such a seed demand table is usually very poor. The static OD table is derived from aggregate demand models and the static traffic assignment which may yield oversaturated flow on network links and unrealistic route patterns. OD flows in this table can overflow to the point that they cannot be loaded into the traffic simulator, and large reduction factors may be needed to initiate the algorithm. Even if the algorithm eventually converges, it only yields one among myriad possible solutions which can differ significantly from the underlying trip and traffic patterns despite approximate matches the observed link counts.

On the other hand, it should be noted that for the case of static OD estimation where the assignment matrix is solely determined by the unknown route patterns, it is practically justified to input the seed OD table from a planning model directly to the static OD estimation process. As long as one is willing to make an assumption about the underlying route pattern such as in User Equilibrium (UE) and to place his/her

confidence on the quality of such seed OD table, there exists theoretical consistency between the assignment procedure in the original planning model and in the static OD estimation model through the use of a monotonic-increasing cost function (e.g. the BPR-function). It may well be argued, however, that a network may not be in UE, and neither does the BPR function even come close to representing the actual traffic behavior under smaller period of congested conditions. Thus such theoretical consistency is in fact only giving solace to the researcher who wants to forget the difficult problems and the practitioner who does not want to understand the problems, rather than providing any confidence that model would yield useful results.

2.3.2 The use of a static OD table with a uniform departure profile as a seed

The second approach commonly adopted in practice is to first convert a static OD table cut out from a planning model to a set of dynamic ones by applying a uniform pattern to all corresponding time intervals. These OD tables are then used as seeds for a dynamic OD estimation model.

This approach is subject to the same difficulties as in the first approach. For congested networks, the simulation becomes quickly congested preventing further iterations of the algorithm. Unlike the static OD approach, this approach employs a dynamic OD estimation model, so there is a much higher chance that the updated OD tables can better reproduce the traffic counts as well as temporal bottlenecks. However,

this does not necessary imply the improvement on the quality of the updated OD tables in terms of its similarity to the underlying trip patterns.

2.3.3 The use of a static OD table for regularizing dynamic OD flows

A more sophisticated approach suggested by Zhou et al. (2002) is to incorporate an additional term into the objective function which minimizes the difference between the static seed OD table and the summation of the dynamic OD tables over the study period. As such, the overall spatial distribution of the estimated dynamic OD tables is regularized by the pattern from the static planning model while the temporal patterns are adjusted with observed dynamic link counts.

In addition to the overload problem as with the previous two approaches, the main shortcoming of this approach is twofold. First, the problem is rank-deficient because the numbers of observed link counts together with those of the static OD flows are nearly always less than those of the dynamic OD flows to be estimated. Second, it is difficult to determine the proper weight to place on each term in the objective function.

2.3.4 The use of some behavioral rules

The last approach converts a static OD table obtained from a static planning model to the dynamic ones based on some behavioral rules. Li (2001) proposed a model that considers

the travelers' choice of departure time. The model, however, needs to be calibrated with disaggregated data. Friesz et al. (2001) developed the *fixed arrival time demand model* which accounts for the users' desire to arrive at their destination at a certain time. The dynamic OD tables from these models can then be used as seed OD tables for a synthetic dynamic OD estimation model.

2.4 OTHER ISSUES

2.4.1 Additional complexities in dynamic OD estimation

As mentioned in subsection 2.2.2, the assignment-based dynamic OD estimation inevitably involves the estimations of route-choice fractions and route-link fractions in addition to the OD demand flows. This is commonly done by means of an iterative algorithm. Unfortunately, most algorithms of this kind often fail to converge due to physical constraints (e.g. maximum flow) in the traffic simulation or DTA model that are not explicitly considered in the estimation model. In fact, the algorithm might be trapped with incorrect OD flows and assignment pattern because the underlying relationship between link flows and OD flows is essentially non-monotonic (e.g. the bend at the maximum flow of any flow-density relation). Moreover, the need to simultaneously model the interactions between demand and (uncalibrated) supply parameters can result in untraceable errors.

2.4.2 Application Aspects

As discussed thus far, the validity of estimated OD tables from a dynamic estimation model is based significantly on the validity of the assumptions made pertaining to route and traffic patterns. Unless the application aspects are well understood, and justified assumptions are made, estimated OD tables might be very misleading.

While this dissertation focuses on the preparation of dynamic demand inputs for short-term planning applications, this subsection briefly discusses a spectrum of transportation analysis to provide a background for understanding the motivation of the proposed frameworks. Figure 2-1 shows a graphical taxonomy in terms of traffic modeling, aggregation level, analysis time-scale, and data requirement for various kinds of transportation analysis.

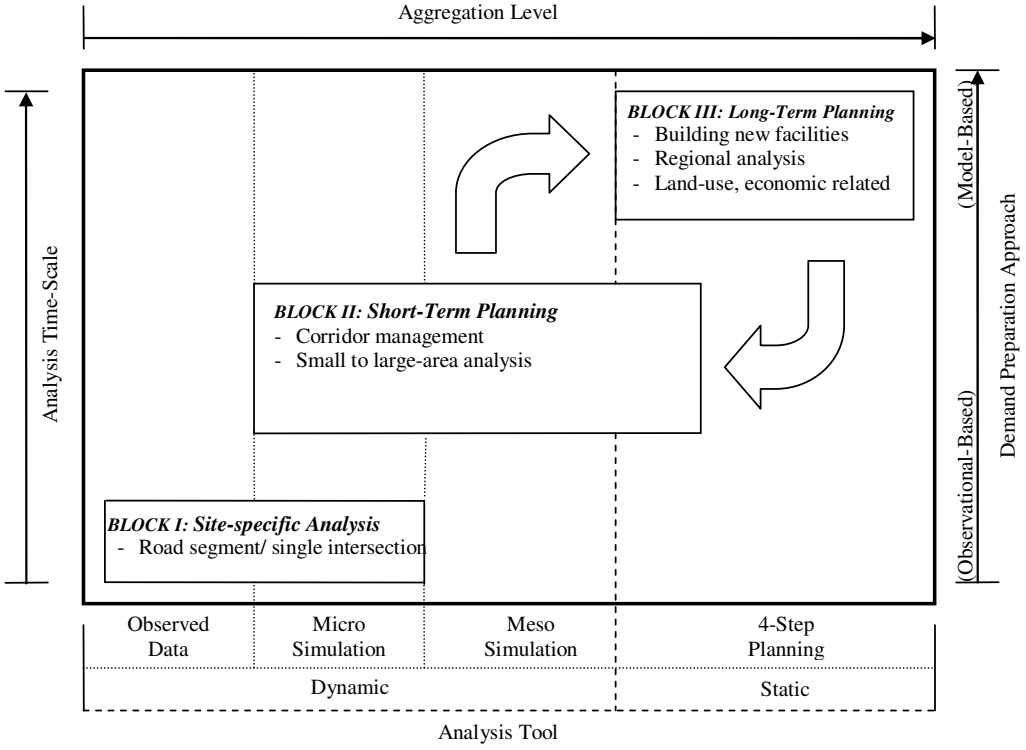


Figure 2-1 Graphical taxonomy of traffic modeling, aggregation level, analysis time-scale, and data requirement for different kinds of transport analysis

Block I: Site-specific analysis refers to the procedure of generating and evaluating small-scale facility improvement alternatives such as intersection/road segment designs, site-specific TSM (Transportation System Management) strategies, and land use development, etc. In general, analyses in this block do not require comprehensive network modeling and can be performed analytically or using microscopic traffic simulation with demand inputs that can be directly collected from the traffic surveillance systems or surveys.

Block II: Short-term planning analysis focuses on improving the productivity and efficiency of the existing facilities with minimum new construction. An example of short-term planning analysis is corridor management including operations and control

management, geometry redesign, and minor capacity expansion. Depending on the type of application, size of the network, and planning horizon, microscopic- or mesoscopic traffic simulation can be used. Since these two models are dynamic, they require dynamic demand inputs.

Block III: Long-term planning analysis refers to the traditional planning procedure in creating and evaluating new major regional transportation facility projects. This process typically involves a very large network of regional size and thus has the most aggregate analysis form.

For short term planning, it is important to consider a simulation model of the real-world network to be rooted in a planning model that is already in place in most urbanized areas. This is because a simulation model cannot exist without any relation to the socio-economic studies and forecasts done for the region.

2.5 SUMMARY

Link count based OD estimation is a complex inverse problem of updating the current (seed) OD flows from observed link counts. Under this process, unknown demand flows are mapped to observed link flows via an assignment matrix. Since this matrix is not known in practice, it needs to be estimated based on mathematical models that sufficiently describe the underlying factors establishing such a physical relation. Further problems arise due to the fact that the number of available link counts is usually less than that of the OD flows to be estimated. To obtain a unique solution, most existing models thus rely heavily on the quality of seed OD tables.

In other words, in addition to solving a complicated inverse problem, OD estimation encompasses a series of approximations under assumptions pertaining to the information contained in seed OD flows, individuals' behavior, and system dynamics. Altogether, these aspects are interrelated such that one interactively affects all others. Unfortunately, developing a complete framework for estimating dynamic OD tables, rather than solely an estimation model, has not been the primary goal of the current studies in the literature.

Because a scientifically-sound framework for preparing dynamic demand input does not currently exist, many incorrect ad-hoc schemes are commonly adopted in practice. For instance, while it is well known that a trip table from sub-area analysis cut out from a regional planning model is derived from aggregate demand models and static traffic assignment, the state of practice has been to use such a static input with ad-hoc dynamic profiles as seed for a dynamic OD estimation model. There is an obvious need for a systematic framework for proper demand preparation when a planning model is used to build a dynamic model.

The type of information availability is also important. Recent advances in technologies, such as Automatic Vehicle Identification (AVI), Global Positioning System (GPS), and cellular phone tracking, are allowing vehicles themselves to be a reliable high-fidelity data source. There are several recent studies that focus on incorporating sampled OD tables into OD estimation formulations. Most studies in this line employ constant (in time and/or space) expansion factors to gross up the observed OD tables. However, the expected market penetration that is currently available or will be available in the near future is too low for the simple expansion technique to be suitable.

Implementable sampling schemes based on participation that is voluntary or motivated-by-rewards, and redistribution of necessary add-on equipments among vehicles, can be employed in an alternating and continuing manner over space and time, with small rates. Although it may be reasonable to anticipate a much higher market penetration of vehicles with location devices such as GPS in the near future, it is highly unlikely that one can achieve a sufficient sampling rate during a particular bin of time. While probe data can ideally be used to bring observations of the underlying individuals' behavior and traffic dynamics into the formulation of OD estimation models, these possibilities have not yet been sufficiently investigated.

CHAPTER 3 SIX-STEP PLANNING MODEL

3.1 INTRODUCTION

This chapter presents a framework for preparing dynamic OD demand tables for microscopic traffic simulation. For ease of describing the work to planning practitioners who are well-versed with the Four-step planning process, this process that incorporates simulation may be considered as two additional steps of the traditional process, and hence the term "six step planning process" is used here. The scope of the framework is essentially practical. It addresses the problem of demand estimation when the prior ones are not available and a traditional demand forecasting model (planning model) is to be used in building a dynamic model. The details of the framework are provided in the next section. Section 3.3 proposes a new bounded OD estimation model which effectively combines the information from demand, traffic, and route modeling at different levels with the information from observed link flows. The model is based mainly on the iterative bi-level scheme and can easily be incorporated into the proposed framework. Section 3.4 presents a solution algorithm specially designed for solving a large-scale problem. Then, section 3.5 discusses some additional practical considerations. While a comprehensive evaluation of the proposed framework, estimation model, and solution algorithm are to be presented in the next chapter, numerical examples intended to illustrate the benefits of the proposed framework in term of modeling traffic and route

dynamics are presented in section 3.6. Finally, section 3.7 provides important concluding remarks and recommendations.

3.2 SIX-STEP PLANNING MODEL

The term *six-step planning process* evolved during a large corridor modeling exercise prepared for the California Department of Transportation (Caltrans), when it became necessary to explain to practitioners how the simulation-based analysis ties in with the existing four-step traffic demand forecasting models (Jayakrishnan et al., 2006). As the traditional four-step process has been applied in practice as four independent sequential steps, it is possible to consider further independent steps to bring dynamics into the modeling. Simulation-based or analytical DTA models are possibilities for these further steps. However, this dissertation focuses only on simulation because it is currently the only approach being implemented in real-world applications. In the six-step planning process, the 5^{th} step refers to a dynamic model built on a network without significant additional details than in the 4^{th} step (static traffic assignment), and the 6^{th} step refers to a dynamic model built on a network with significantly more supply modeling details than in the 4^{th} step. For instance, the 5^{th} step modeling could be a macroscopic or mesoscopic traffic simulation model built on the same network as in the 4^{th} step, and the 6^{th} step could be a microscopic traffic simulation model on a more detailed network.

Traditional demand forecasting (planning) models are often built for a county-wide network area (of 100s of square miles), while microscopic simulation models are

currently built for a short freeway corridor or a small freeway/arterial network of 5 to 25 square miles. These correspond to the network size for the 4th and 6th step, respectively. Ideally, the 5th step of the process can be on the same network as in the 4th step; however, it may be acceptable for computational reasons to have this analysis done on a dynamic *intermediate* model (of the size 20 to 40 square miles).

DYNASMART-P (Jayakrishnan et al., 1994) is an example of mesoscopic simulation models that can be used in the 5th step to bridge between the static and dynamic models. The model is based on the basic characteristics of the regional planning model such as length, free-flow speed, and link capacity, thus allowing for an acceptable translation from the BPR-type cost functions (applicable only for flows in longer time periods) in the planning models to a Greenshields-type speed-density function applicable for congested flows in smaller periods. This feature of the model can potentially lessen the overflow problem due to the extreme inconsistencies between the static planning and microscopic models explained in subsection 2.3.2.

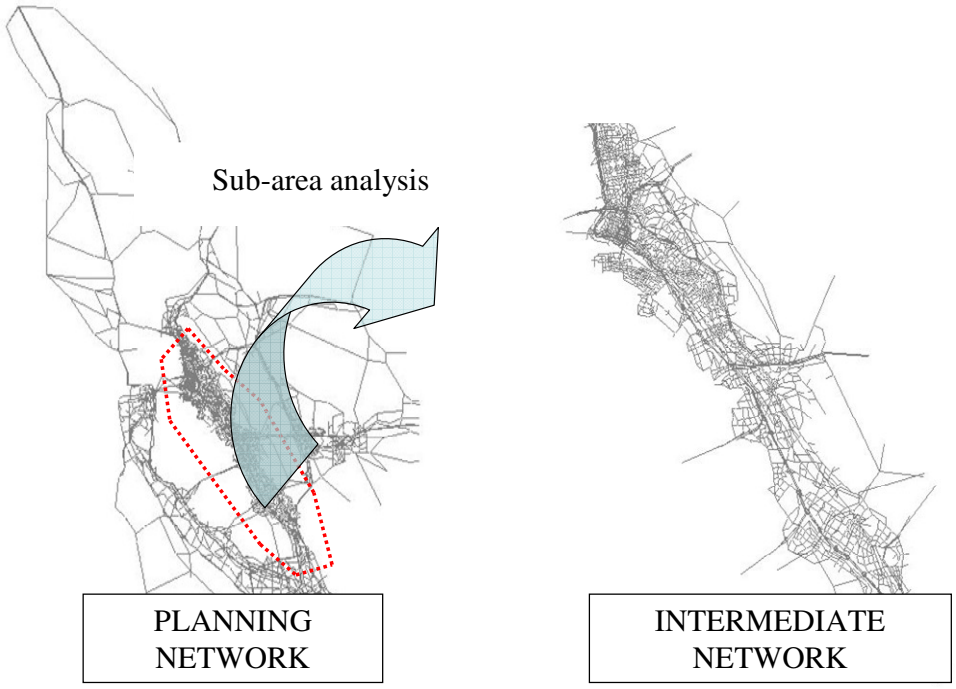


Figure 3-1 Cutting an intermediate network from a planning model

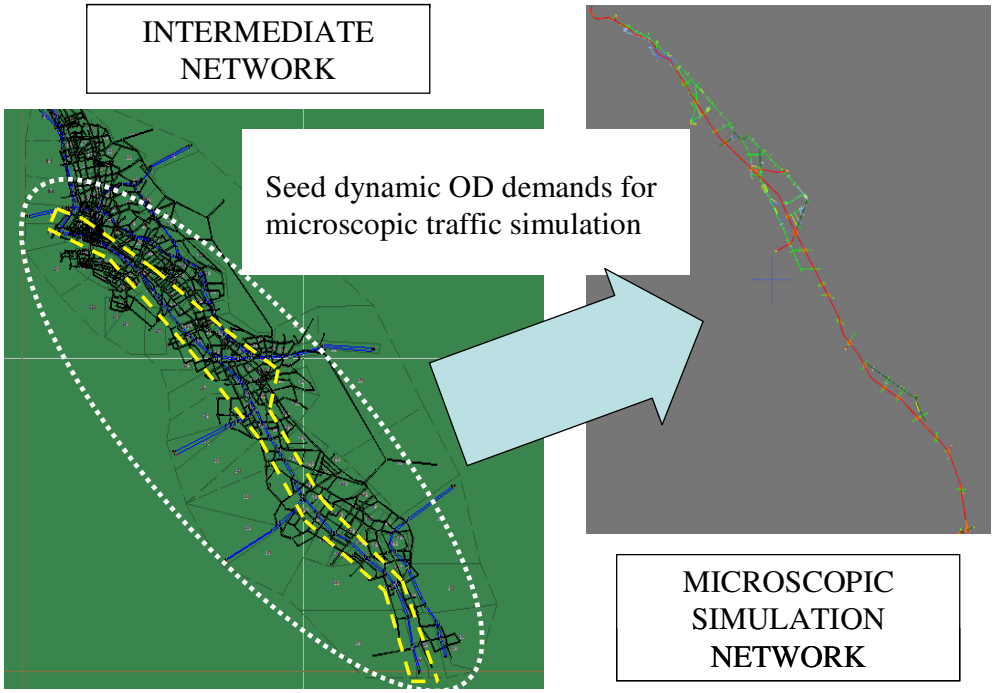


Figure 3-2 Creating a microscopic traffic simulation model

Figure 3-1 illustrates the process of cutting an intermediate dynamic network (e.g. DYNASMART-P) from a planning network (e.g. TransCAD), and Figure 3-2 shows a microscopic simulation network (e.g. Paramics) built inside such an intermediate model. Note that the intermediate simulation model provides a larger but rougher dynamic model.

Using a static seed OD table from the sub-area analysis within the regional planning network cut for the intermediate network with some departure temporal profile, an iterative bi-level OD estimation scheme is first employed within the intermediate model covering the final microscopic model. After the OD estimation converges, the set of time-dependent path flows entering and leaving the final microscopic simulation network is converted to a set of dynamic seed trip tables for the further dynamic OD estimation within the microscopic model. That is, the intermediate simulation is used to provide proper approximations of path and traffic dynamics. Since this scheme helps in obtaining dynamic seed OD tables which account for the system dynamics as well as supply capacities, it helps in guiding the OD estimation algorithm in the final model to a solution significantly closer to the underlying demand. Figure 3-3 provides a detailed flowchart of the six-step process.

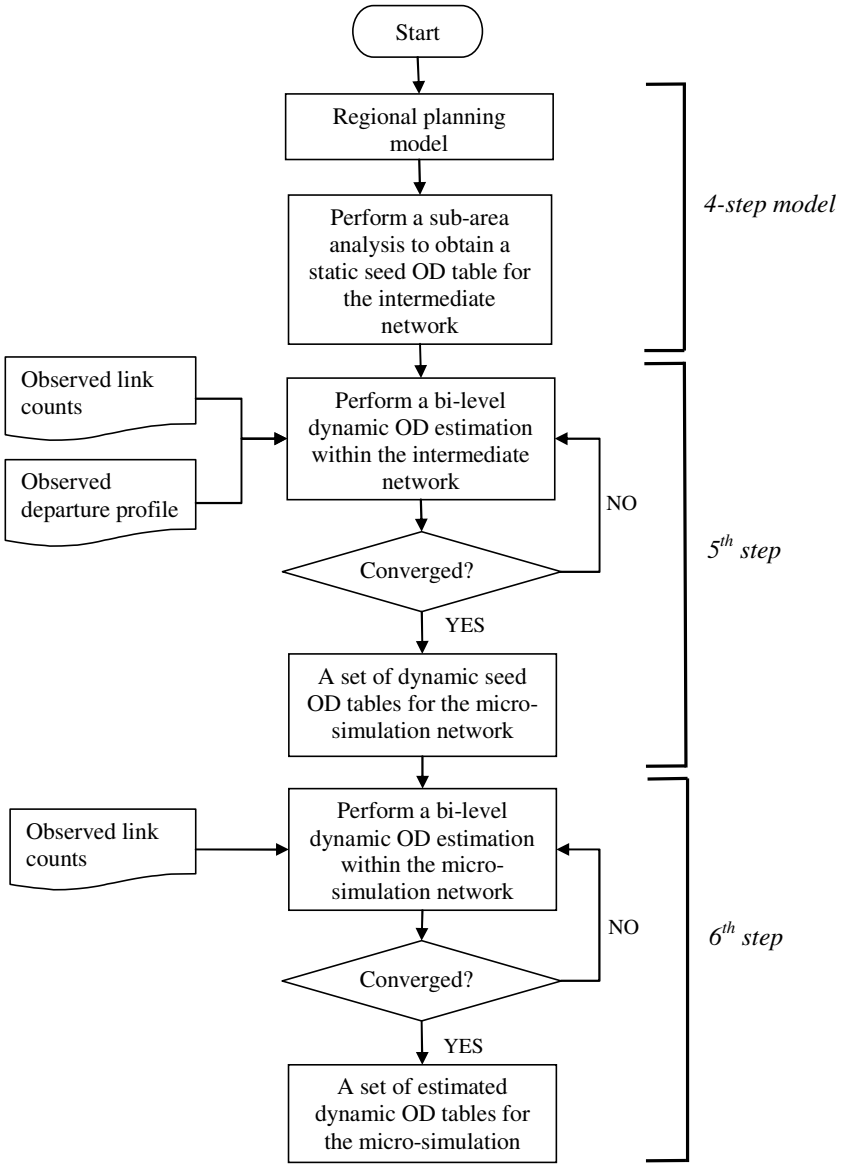


Figure 3-3 Flowchart of the six-step process modeling

It should be noted that the intermediate network surrounding the microscopic simulation model also allows for future year analyses which again require the initiation from the long-term planning level. This is another very important feature of the six-step process, since a large dynamic network surrounding the microscopic simulation is needed to capture changes in travel pattern within the regional level.

3.3 DEVELOPMENT OF A BOUNDED OD ESTIMATION MODEL

Although there are a few dynamic OD estimation models based on the iterative bi-level scheme proposed in the literature that can be incorporated into the proposed six-step process framework, a new OD estimation model which includes bound constraints is introduced in this dissertation.

Consider a network with multiple origins $i \in I$ and destinations $j \in J$ with the study period being discretized into departure time intervals $t = 1, 2, ..., T$. Denote seed OD flow between the origin i and destination j departing during departure interval t by $\breve{x}^{ijt}$. Let $\breve{v}_{lh}$ be the observed traffic count on link $l \in L_{obs}$ during observation interval h ($h = 1, 2, ..., H$). The model should seek the dynamic OD demand estimates, once assigned to the network, match with observed counts and seed OD flows as close as possible.

Hence, the objective function of the upper level formulation consists of two objectives. The first objective is to minimize the summation of the deviations between seed and estimated OD flows. The second objective is to minimize the summation of the deviations between observed and estimated link counts. Depending on the distributional assumptions made on observed link counts and seed OD tables, different functional specifications can be obtained under the Maximum Likelihood approach. In this dissertation, the weighted 2-norm distance which corresponds to multivariate normal assumption is used. It is further assumed that off-diagonal elements of the variance-covariance matrices of the error terms are zero, which is practically justified as such

information is not available from the current data source. The objective function thus coincides with the ordinary linear regression model and is similar to the model proposed by Zhou et al. (2003). Alternatively, the model can be viewed as a multi-objective optimization with fixed weights.

Unlike most dynamic OD estimation models proposed in the literature, a set of bound constraints is introduced into the optimization to ensure that each estimated OD flow falls within a specified range. There are three motivations for acting this way. First, these bound constraints help in further regularizing the estimation problem. This is a very important aspect, especially for the adopted bi-level scheme where an incorrect assignment matrix may occur from the previous simulation iteration and the algorithm might be trapped. Second, OD estimation should at least be constrained naturally by the non-negativity condition. Third, the modeler can apply his or her practical knowledge about the ranges of OD demand value. Such knowledge can be established from previously estimated OD tables or seed OD tables from the preceding simulation step within the six-step process. For example, if the OD estimation within the 5th step (say a meso-model) is already completed before the OD estimation within the 6th step then these bounds can be allowed to be closer to the seed tables. On the other hand, if the seed table is cut directly from a static OD table of a planning model, the bound may be set higher.

The optimization program can be expressed as follows.

$$\min z(x^{ijt}\ \forall i,j,t) = \frac{1}{2}\left[\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(x^{ijt}-\breve{x}^{ijt})^2}{\sigma^2_{\breve{x}^{ijt}}} + \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J}a_{lh}^{ijt}x^{ijt}-\breve{v}_{lh})^2}{\sigma^2_{\breve{v}_{lh}}}\right] \quad (3\text{-}1)$$

subject to

$$a_{lh}^{ijt} = assignment[x^{ijt}] \ \ from \ \ simulation \quad \forall i, j, t \tag{3-2}$$

$$x^{ijt\,LO} \le x^{ijt} \le x^{ijt\,UP} \qquad \forall i, j, t \tag{3-3}$$

where

x^{ijt} = The OD flow from origin i to destination j departing during time interval t

$x^{ijt\,LO}$ = The lower bound for the OD flow from origin i to destination j departing during time interval t

$x^{ijt\,UP}$ = The upper bound for the OD flow from origin i to destination j departing during time interval t

a_{lh}^{ijt} = The assignment fraction of x^{ijt} that reaches link l during observation interval h obtained from the last simulation

$\sigma_{\breve{x}^{ijt}}$ = The standard deviation of $\breve{x}^{ijt}$ or weight on the term that includes $\breve{x}^{ijt}$

$\sigma_{\breve{v}_{lh}}$ = The standard deviation of $\breve{v}_{lh}$ or weight on the term that includes $\breve{v}_{lh}$

v_{lh} = The count on link l during observation interval h

Under the six-step framework, $\breve{x}^{ijt}$ refers to the corresponding static seed OD flow from sub-area analysis combined with a dynamic fraction when performing the dynamic OD estimation within an intermediate network. It refers to the dynamic seed OD table cut from the intermediate network when performing the dynamic OD estimation at the microscopic level.

3.4 DEVELOPMENT OF A SOLUTION ALGORITHM

Since the dynamic OD estimation, especially at the regional level (say in mesoscopic simulation model), usually involves several hundred OD pairs with several departure intervals and commercial software that can handle this large amount of information is not currently available, this dissertation adopts the Frank-Wolfe (F-W) method, a memory-efficient algorithm, to solve the upper level.

For the s^{th} iteration of the F-W algorithm, the gradient of the first and second term of the objective function with respect to x^{mnd} can be calculated using equation (3-4) and equation (3-5), respectively.

$$\left.\frac{\partial z(x^{ijt}\ \forall i, j, t)}{\partial x^{mnd}}\right]_{OD\ term}^{(s)} = \frac{(x^{mnd\,(s)} - \breve{x}^{mnd})}{\sigma^2_{\breve{x}^{mnd}}} \tag{3-4}$$

$$\left.\frac{\partial z(x^{ijt}\ \forall i, j, t)}{\partial x^{mnd}}\right]_{link\ count\ term}^{(s)} = \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \frac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} a_{lh}^{ijt} x^{ijt\,(s)} - \breve{v}_{lh})}{\sigma^2_{\breve{v}_{lh}}} \cdot a_{lh}^{mnd} \tag{3-5}$$

Therefore, the gradient of the objective function with respect to x^{mnd} is

$$\left[\frac{\partial z(x^{ijt}\ \forall i,j,t)}{\partial x^{mnd}}\right]^{(s)} = \frac{(x^{mnd(s)} - \breve{x}^{mnd})}{\sigma^2_{\breve{x}^{mnd}}} + \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \frac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} a_{lh}^{ijt} x^{ijt(s)} - \breve{v}_{lh})}{\sigma^2_{\breve{v}_{lh}}} \cdot a_{lh}^{mnd} \tag{3-6}$$

For notational convenience, let

$$\upsilon_{mnd} = x^{mnd(s)} - \breve{x}^{mnd} \tag{3-7}$$

$$\vartheta_{lh} = \sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} a_{lh}^{ijt} x^{ijt(s)} - \breve{v}_{lh} \qquad \forall l,h \tag{3-8}$$

Therefore,

$$\left[\frac{\partial z(x^{ijt}\ \forall i,j,t)}{\partial x^{mnd}}\right]^{(s)} = \frac{(\upsilon_{mnd})}{\sigma^2_{\breve{x}^{mnd}}} + \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \frac{(\vartheta_{lh})}{\sigma^2_{\breve{v}_{lh}}} \cdot a_{lh}^{mnd} \tag{3-9}$$

The auxiliary OD flows $\mathbf{y}^{(s)}$ can then be easily determined by means of the following logical expressions:

$$y_{mnd}^{(s)} = \begin{cases} x^{mnd\,LO} & if \left[\frac{\partial z}{\partial x^{mnd}}\right]^{(s)} > 0 \\ x^{mnd(s)} & if \left[\frac{\partial z}{\partial x^{mnd}}\right]^{(s)} = 0 \\ x^{mnd\,UP} & if \left[\frac{\partial z}{\partial x^{mnd}}\right]^{(s)} < 0 \end{cases} \tag{3-10}$$

Correspondingly, the updated demand flows can be expressed as:

$$x^{ijt(s+1)} = x^{ijt(s)} + \lambda (y_{ijt}^{(s)} - x^{ijt(s)}) = x^{ijt(s)} + \lambda \, d_{ijt}{}^{(s)} \qquad \forall i, j, t \tag{3-11}$$

To evaluate λ, the objective function should be minimized with respect to λ. First, the objective function is written in terms of λ

$$z(\mathbf{x}+\lambda\mathbf{d}) = \frac{1}{2}\left[\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{[(x^{ijt(s)} + \lambda d_{ijt}{}^{(s)}) - \breve{x}^{ijt}]^2}{\sigma^2_{\breve{x}^{ijt}}} + \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \frac{[\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} a_{lh}^{ijt}(x^{ijt(s)} + \lambda d_{ijt}{}^{(s)}) - \breve{v}_{lh}]^2}{\sigma^2_{\breve{v}_{lh}}} \right] \tag{3-12}$$

$$= \frac{1}{2}\left[\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{[x^{ijt(s)} - \breve{x}^{ijt} + \lambda d_{ijt}{}^{(s)}]^2}{\sigma^2_{\breve{x}^{ijt}}} + \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \frac{[\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} a_{lh}^{ijt} x^{ijt(s)} - \breve{v}_{lh} + \lambda \sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} d_{ijt}^{(s)} a_{lh}^{ijt}]^2}{\sigma^2_{\breve{v}_{lh}}} \right] \tag{3-13}$$

Again, for notational convenience, let

$$\varphi_{lh} = \sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} d_{ijt}^{(s)} a_{lh}^{ijt} \tag{3-14}$$

Then,

$$z(\mathbf{x}+\lambda\mathbf{d}) = \frac{1}{2}\left[\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{[\upsilon_{ijt} + \lambda d_{ijt}{}^{(s)}]^2}{\sigma^2_{\breve{x}^{ijt}}} + \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \frac{[\vartheta_{lh} + \lambda \varphi_{lh}]^2}{\sigma^2_{\breve{v}_{lh}}} \right] \tag{3-15}$$

Taking the first derivative of the expression (3-15) with respect to λ

$$\frac{z(\mathbf{x}+\lambda\mathbf{d})}{\partial\lambda} = \sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{[v_{ijt}+\lambda d_{ijt}^{(s)}]}{\sigma^2_{\bar{x}^{ijt}}}.d_{ijt}^{(s)} + \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{[\vartheta_{lh}+\lambda\varphi_{lh}]}{\sigma^2_{\bar{v}_{lh}}}.\varphi_{lh} \tag{3-16}$$

$$= \sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{v_{ijt}d_{ijt}^{(s)}}{\sigma^2_{\bar{x}^{ijt}}} + \lambda\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(d_{ijt}^{(s)})^2}{\sigma^2_{\bar{x}^{ijt}}} + \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{\vartheta_{lh}\varphi_{lh}}{\sigma^2_{\bar{v}_{lh}}} + \lambda\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{(\varphi_{lh})^2}{\sigma^2_{\bar{v}_{lh}}} \tag{3-17}$$

Setting (3-17) equal to zero, the optimum moving size λ^* can be calculated as follows.

$$\lambda^* = -\frac{\left[\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{v_{ijt}d_{ijt}^{(s)}}{\sigma^2_{\bar{x}^{ijt}}} + \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{\vartheta_{lh}\varphi_{lh}}{\sigma^2_{\bar{v}_{lh}}}\right]}{\left[\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(d_{ijt}^{(s)})^2}{\sigma^2_{\bar{x}^{ijt}}} + \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{(\varphi_{lh})^2}{\sigma^2_{\bar{v}_{lh}}}\right]} \tag{3-18}$$

A specially designed algorithm to solve this iterative scheme is as follows.

Step 0: Initialize the main iteration counter m=0. Load seed dynamic OD tables into the traffic simulation model to obtain the initial assignment matrix.

Step 1: Set the main iteration counter $m = m + 1$.

Step 2: Substitute the assignment map from the last simulation run and solve the upper level using the F-W algorithm to update OD table as follows.

i. Set the F-W iteration counter $s = 1$.

ii. Evaluate the gradients using equation (3-9).

iii. Determine the auxiliary OD flows $\mathbf{y}^{(s)}$ using the logical expression (3-10).

iv. Calculate the optimum moving size λ^* using equation (3-18).

v. Update new OD flows $\mathbf{x}^{(s+1)}$ using equation (3-11).

vi. Set the F-W iteration counter $s = s + 1$. Check maximum number of the F-W iterations, if reached go to step vii, otherwise return to ii.

vii. Set $\mathbf{x}^{(m)}$ equal to the last F-W updates. Go to step 3

Step 3: Load the updated OD tables into the traffic simulation model to obtain the updated assignment map.

Step 4: Set the main iteration counter $m = m + 1$. Check $\left\|\mathbf{x}^{(m)} - \mathbf{x}^{(m-1)}\right\| \le \varepsilon_x$ or

$\left\|z(\mathbf{x})^{(m)} - z(\mathbf{x})^{(m-1)}\right\| \le \varepsilon_{obj}$, if satisfied go to go to step 5, otherwise go to step 2.

Step 5: End of the algorithm.

It is important to note that although the convergence criteria for the overall procedure are based on the improvement on the objective function value or the absolute difference of OD flows in two successive iterations, the convergence criteria of the F-W algorithm used in solving the upper level are based solely on a predefined maximum number of iterations. This method, suggested by Doblas and Benitez (2005), is based on the fact that the fluctuation of the assignment matrix in two successive (main) iterations causes oscillations in the value of the objective function which would make it pointless to attempt to choose the stopping criteria based on the objective function values. Note that alternative algorithms for solving the constrained optimization such as the gradient projection method can also be used but it is out of the scope of this dissertation.

3.5 PRACTICAL CONSIDERATIONS

The main hurdle in the proposed six-step framework is that the dynamic departure flow profiles are not available from the planning level. One method to overcome this problem, as suggested by Tavana and Mahamassani (2001), is to assume the uniform temporal pattern by splitting the static OD table into equal portions. Another scheme is to assume that the dynamic profile from each origin is approximately the same across its destinations so that the profile of the traffic counts from all links leaving from that origin can be applied. These schemes may, however, bias the estimation results. To lessen such impacts, it is possible to introduce relative weights into the objective function (3-1).

$$\min z(x^{ijt}\ \forall i, j, t) = \frac{1}{2}\left[\begin{array}{l} w\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(x^{ijt}-\breve{x}^{ijt})^2}{\sigma^2_{\breve{x}^{ijt}}} + \\ (1-w)\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J}x^{ijt}a_{lh}^{ijt}-\breve{v}_{lh})^2}{\sigma^2_{\breve{v}_{lh}}} \end{array}\right] \tag{3-19}$$

These weights are used to reflect the relative confidence that the modeler has on the dynamic seed OD tables and the observed link counts. Correspondingly, w should be set relatively small for the dynamic OD estimation within the 5th step. Other different schemes for determining weight values can be found in Zhou et al. (2003) and will not be further discussed here.

In most cases, standard deviations or weights for individual seed OD flows are not known, so the default values of one can be used unless the modeler has enough grounds to place various weights on different cells. Lastly, although the values set for the upper

and lower bounds depend on the modeler and the characteristics of the problem, it is usually sensible to set the maximum deviation as a function of the magnitude of the seed OD flow.

3.6 NUMERICAL EXAMPLES

Before presenting an evaluation of the proposed six-step framework and dynamic OD estimation model in the next chapter, it is helpful to first investigate the benefits from the above suggestion to use an intermediate mesoscopic-level dynamic network modeling (step 5^{th}) before attempting microscopic modeling (step 6^{th}) of a network, particularly in terms of how much better such modeling can replicate the real-world conditions. However, since the observations of the underlying trip, route, and traffic pattern sufficient for a comprehensive evaluation are not currently available for a real transportation network, a hypothetical network context with simulated ground truth data is used.

Specifically, this section focuses on an evaluation of the sub-area dynamic seed OD tables and route-link fractions from the intermediate network modeling using a mesoscopic simulation model. The relevant questions are

- How different are the sub-area OD flow profiles generated by a mesoscopic model from the static (constant) profiles?
- How different are the route-link fractions generated by a mesoscopic model from the static (constant) route-link fractions?

- Is a less-detailed dynamic model in step 5th, such as a mesoscopic model, enough to provide traffic approximations that are sufficiently close to the ground truth?

The first two questions have an intuitive answer that the temporal OD profiles and route-link fractions of the sub-area could be quite different from the constant pattern predicted by the static assignment model. However, the last question does not have an easy answer, as the mesoscopic simulation may model traffic flow dynamics differently than the ground truth microscopic simulation. The following subsections investigate these issues.

3.6.1 Hypothetical networks and simulated data

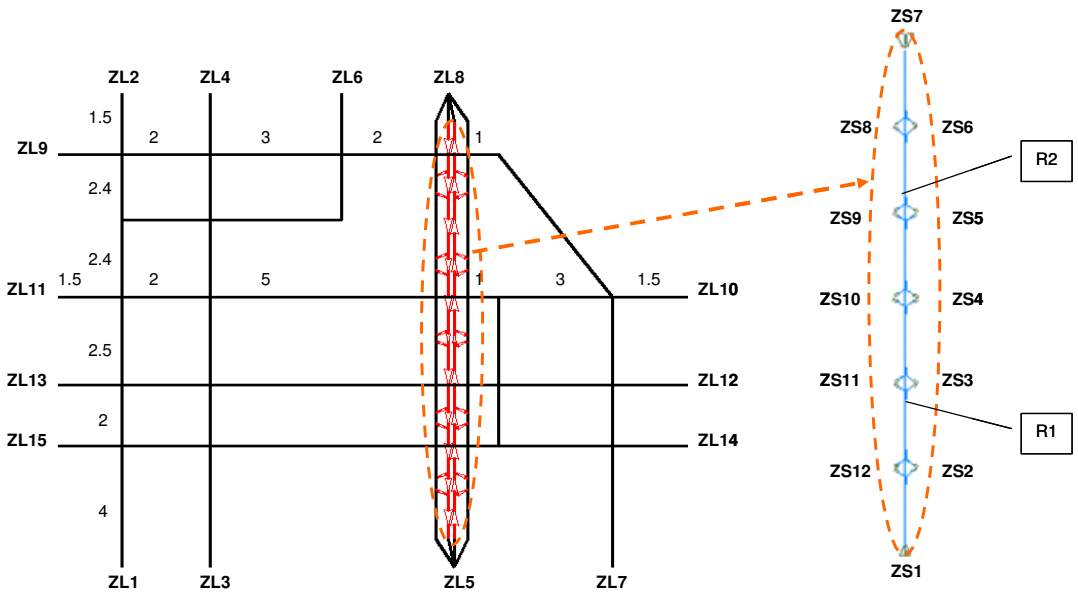


Figure 3-4 Study networks: a large freeway-arterial network (left), a freeway network to be simulated in a microscopic simulation model (right)

A large freeway-arterial network (Figure 3-4, left) is coded in a static planning model, a dynamic mesoscopic model, and also a microscopic model. TransCAD, DYNASMART-P, and Paramics are selected, respectively, as the software for the three models. Unless noted otherwise, these versions of the network are referred to as LPl, LMe, and LMi, respectively. Parameters used in these models are summarized in Table 3-1. In this section of the dissertation, the LMi microscopic simulation on the complete network is only for ground truth, akin to a real-world network traffic context. Clearly, the use of a microscopic simulation to generate the ground truth conditions would impact the evaluation results since different suits of software would perform differently. However, given the objectives of the following experiments which are to obtain a general idea about the benefits from the intermediate mesoscopic network over the static planning model and its consistency with the final microscopic model, such use can be justified.

This network includes 208 links, 15 zones, and 27 fixed-time traffic signals. Zones are assumed, without loss of generality, to be located along the network periphery. Long links (lengths shown in miles) are deliberately used in order to illustrate the effects of traffic dynamics on the temporal profile of sub-area OD flows and route-link fractions.

Table 3-1 Parameters used in the three models

	Software	Link type	Model	Parameter	
Planning model	TransCAD	Freeway	BPR function	Free-flow speed (mph)	75
				Capacity (pcphpl)	2300
		Arterial	BPRfunction	Free-flow speed (mph)	55
				Capacity (pcphpl)	900
Mesoscopic simulation	Dynasmart-P	Freeway	Dual-regime modified Greenshields	Density breakpoint (pcpmpl)	30
				Speed Intercept (mph)	92
				Minimum speed (mph)	6
				Jam density (pcpmpl)	200
				Shape term alpha	2
		Arterial	Single-regime modified Greenshields	Minimum speed (mph)	10
				Jam density (pcpmpl)	120
				Shape term alpha	1
Microscopic simulation	Paramics	Freeway	Car-following	Free-flow speed (mph)	75
				Mean headway (sec)	1
				Mean reaction time (sec)	1
				Minimum gap (ft)	6.56
				Queue speed (mph)	4.5
		Arterial	Car-following	Free-flow speed (mph)	55
				Mean headway (sec)	1
				Mean reaction time (sec)	1
				Minimum gap (ft)	6.56
				Queue speed (mph)	4.5

The network of relevance for the study is in fact a smaller sub-area network meant for microscopic traffic simulation. This network consists of 12 zones along a stretch of freeway (Figure 3-4, right) and is referred to as SMi. The following experiments compare the temporal profiles of the cut OD tables of the small network SMi and route-link fractions from within the larger LMe without performing the OD estimation, to that occurring in the ground truth simulation. Differences are expected because the LMe involves only a model of the ground truth flows based on a (mesoscopic) traffic modeling

paradigm and does not include all the detailed microscopic traffic process in the ground truth LMi.

3.6.2 Experimental designs and results

EXP 3-1 Uninterrupted freeway flow

Given the set of true 5-min dynamic OD flows from ZL5 to ZL8 and from ZL8 to ZL5 which do not share any route, the large microscopic simulation (LMi) is used to simulate the ground truth of dynamic link counts, route-link fractions, and dynamic OD flows from R1 to R2 and from R2 to R1 (see Figure 3-4, right). The summation of the ground true dynamic OD flows over all departure intervals for the large network OD pairs are regarded here as the same as the flows for that OD demand in the static planning model (LPl). Note that in practice a static OD table of a planning model is often different from the underlying one due to various sources of data and modeling errors; however, this discrepancy is neglected for a moment so that significant aspects of the dynamic traffic modeling can be highlighted. The dynamic profiles from link counts at the downstream links from ZL5 and ZL8 (as created by LMi, analogous to real-world observations) are applied to the corresponding hourly OD flow to generate a set of dynamic seed OD flows, which are then assigned in the mesoscopic simulation (LMe).

Figure 3-5 and Figure 3-6 provide a graphical comparison of the subarea dynamic seed OD flows from R1 to R2 and from R2 to R1, respectively. In the figures, dynamic

OD flows from *Micro* refers to the ground truth, from *Planning* refers to the uniform static OD flows, and from *Meso* refers to the subarea dynamic seed OD tables obtained from the mesoscopic simulation.

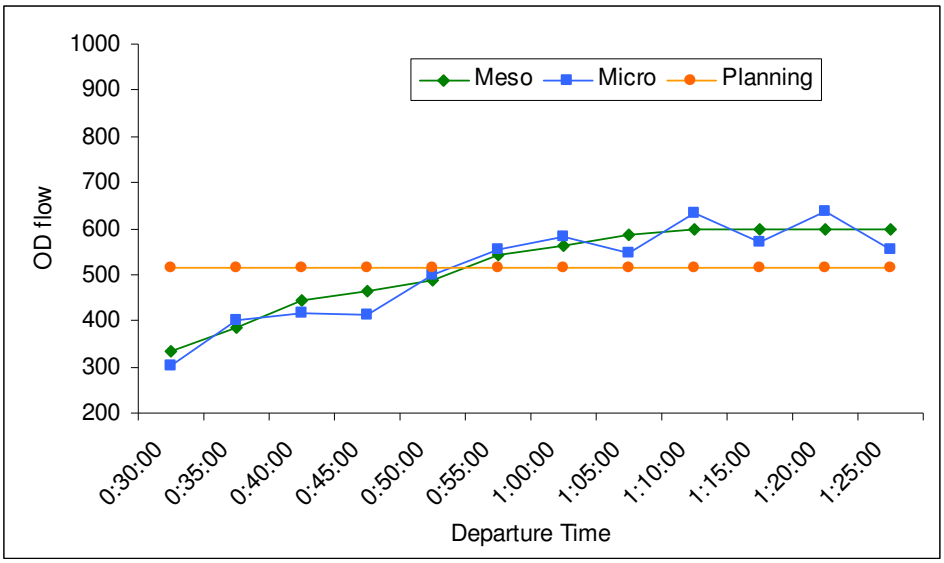


Figure 3-5 Comparison of sub-area OD flows from R1 to R2 obtained from the meso-model and the planning model to the ground truth

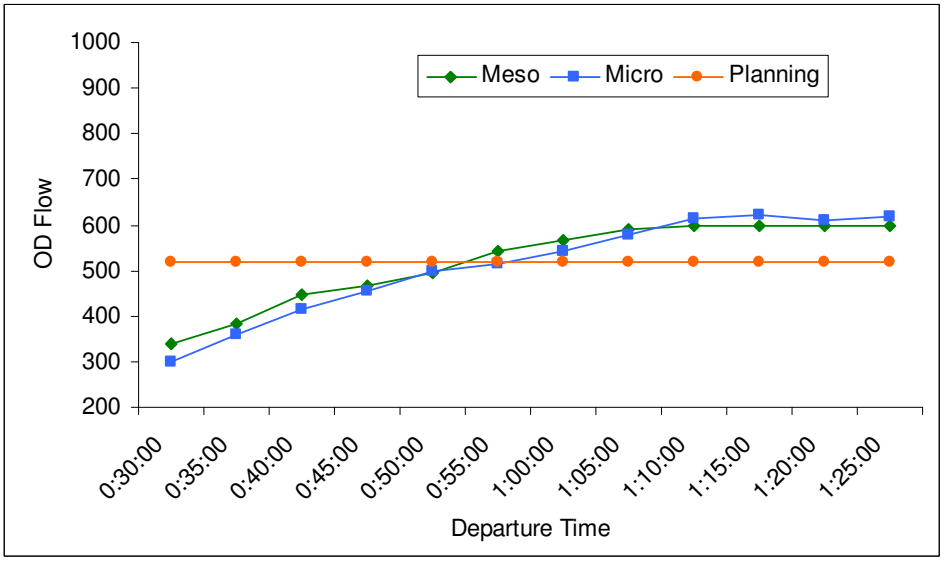


Figure 3-6 Comparison of sub-area OD flows from R2 to R1 obtained from the meso-model and the planning model to the ground truth

Table 3-2 summarizes the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) from the simulated ground truth, for the subarea dynamic OD flows and route-link fractions from the mesoscopic simulation and planning model.

Table 3-2 OD flow and route-link fraction discrepancies from EXP 3-1

	OD Errors		Route-link fraction errors	
	MAE	RMSE	MAE	RMSE
Meso-model	24.708	27.637	0.238	0.267
Planning	86.111	102.490	0.560	0.649

It can be easily seen that the mesoscopic simulation significantly improves the quality of dynamic OD flows and route-link fractions in the subarea. It is noteworthy that although the meso-model uses a simple traffic flow model without the microscopic details, it still obtains a flow pattern in the subarea that is close to the ground truth. This underscores the usefulness of developing a meso-model before cutting out a sub-area seed OD table to perform the OD estimation in a microscopic model.

EXP 3-2 General network flow (without route-choice)

To conduct the same experiment but under the complexities of traffic merging, conflicts at intersections, and signal controls, EXP 3-1 is repeated here but with non-zero OD flows applied to all OD pairs which share some common routes. The OD flow for each

zone pair is assigned to a single route to eliminate the route-choice modeling issues. Table 3-3 shows the discrepancies of the dynamic OD flows and route-link fractions within SMi sub-area from the ground truth, as compared to the results from the mesoscopic model and the planning model in terms of their MAE and RMSE.

Table 3-3 OD flow and route-link fraction discrepancies from EXP 3-2

	OD Errors		Route-link fraction errors	
	MAE	RMSE	MAE	RMSE
Meso-model	34.741	53.193	0.384	0.482
Planning	68.8	98.539	0.488	0.573

As expected, the discrepancies of the results generated from the microscopic and mesoscopic models increase in this case. This is clearly because this scenario includes several merging and bottleneck locations and signal controls, which are modeled considerably differently by the mesoscopic traffic model compared to the ground truth microscopic context. However, the mesoscopic model still provides a much better approximations of the sub-area traffic patterns and OD flows compared to the planning model.

These two experiments were intended to display the difference in the cut sub-area OD tables and route-link fractions, when one uses a planning model as compared to a mesoscopic model, and indicated that the 5^{th} step is useful even before a seed OD table is made. Note that the results were based on a *one-shot* mesoscopic simulation only and that no attempt was made to employ the iterative OD estimation within the mesoscopic model itself.

3.7 SUMMARY

This chapter introduced a systematic and practical framework for preparing dynamic demand inputs for microscopic simulation under planning applications. The framework may be considered a six-step planning process, which incorporates the traditional four-step planning process with two additional steps. The 5th step is to prepare a dynamic model with similar network modeling details as the planning network, using a computationally efficient dynamic traffic model, and the 6th step involves much greater network and traffic flow details on a sub-area cut out from the model in the 5th step. A new dynamic OD estimation model along with a solution algorithm for a large-scale problem, applicable for both 5th and 6th steps, is also presented.

For real-world implementation, the OD estimation within the 5th step (say, within a mesoscopic level) should be routinely performed for large planning models in order to keep this level of modeling close to the real-world dynamic conditions, so that desired sub-area OD tables can be cut from this model to be used as dynamic seed OD tables for OD estimation within the microscopic level. It may be argued that instead of introducing the six-step process the microscopic simulation can replace the 4th step of the traditional planning process directly. This may be true, but the computational capability of personal computers would need to take several more years to overcome. Moreover, the effort required for supply and demand calibration would not be easily arranged at a large regional network level. A more reasonable argument is then to replace the 4th step with

the macroscopic or mesoscopic simulation step, followed by the 5th step of the microscopic simulation. This is viable once the computational problems can be resolved. For this latter recommendation, the methodology presented in the 5th and the 6th steps can be directly used as the 4th and the 5th step, respectively.

It is also noteworthy that feedback loops among the steps may have to be considered, just as within the tradition four-step process, for the proper modeling consistency. However, the joint analytical models, which are possible for the four steps, cannot be easily developed unless analytical DTA models are used for the dynamic modeling steps.

The next chapter presents two additional experimental studies. These studies are designed to further investigate the efficacy of the proposed framework despite the presence of the errors from demand, route, and traffic modeling in different models. In addition, an application of the proposed framework to a real-world network, I-880 corridor, is reported.

CHAPTER 4 EVALUATION OF THE SIX-STEP MODEL

4.1 INTRODUCTION

The previous chapter described the six-step framework and the development of a new bounded dynamic OD estimation model. The benefits of a mesoscopic simulation model were illustrated as a tool for the 5th step using hypothetical networks and simulated data.

This chapter presents comprehensive experimental studies to fully evaluate the proposed methodologies and presents their application to a real network. The chapter is organized as follows. The next section describes two additional experimental studies based on the same networks presented in the previous chapter. The comparisons between the results from the proposed framework and a more traditional scheme are made. Section 4.3 applies the framework to I-880 corridor in order to illustrate the efficacies of the methodologies in real-world applications. Because the underlying OD flows are not known in this case, several performance measures are examined. Finally, important concluding remarks are provided in section 4.4.

4.2 EXPERIMENTAL STUDIES

Additional experiments are conducted in this section to evaluate the six-step framework. The experiments are based on the same settings outlined in the previous chapter. Given a

set of ground true dynamic OD flows among ZL1-ZL15 zones, the large microscopic simulation (LMi) is used to generate the ground truth of dynamic link counts, route-link fractions, and sub-area dynamic OD flows among ZS1-ZS12 zones (i.e., for SMi). However, unlike the previous experiments, the summation of ground true dynamic OD tables over all departure intervals is regarded as mean static OD table. To further account for the errors from a planning model, the mean OD flows associated with random errors, each independently drawn from a uniform distribution with 25 percent distortion, are used to construct a static OD table, analogous to the one from a static sub-area analysis from an even larger planning network.

4.2.1 Experimental Designs

Unless noted otherwise, a traditional scheme of using a set of uniformly-cut static OD tables as seed and performing the iterative OD estimation directly within the microscopic simulation model is denoted as *M1*. The proposed framework, denoted as *M2*, uses the static OD table with dynamic profile as seed, performs the iterative OD estimation first at the mesoscopic level, and eventually at the microscopic level. The outline of the experiments is given below.

EXP 4-1 Network without route-choice

The first experiment is designed to test the scenario where there is only one route per OD pair which is also known to the modeler. Correspondingly, the simulation of the ground truth in the microscopic simulation (LMi), and the traffic assignments in both planning model (LPl) and mesoscopic simulation (LMe) are based on all-or-nothing assignment.

EXP 4-2 Network with route-choice

On the other hand, the second experiment tests the scenario where there are multiple routes per OD pair and they are not known to the modeler. In generating the ground truth, drivers are assumed to choose their perceived shortest route. Random terms are set moderate, and mean route costs are updated every OD interval to reflect their general knowledge about traffic condition. The static traffic assignment in the planning model (LP1) is based on the user equilibrium (UE) condition, the concept commonly adopted by most transportation planning agencies. In the mesoscopic simulation (LMe), each vehicle is assumed to have a predetermined route randomly assigned from the two shortest paths ($k = 2$) that are calculated every OD interval. In practice, such a setting is clearly a choice of the modeler. This experiment deliberately uses a rather coarse assignment procedure so that the mesoscopic model would not make the same assumption as the ground truth but would yield certain acceptable route choice fractions.

4.2.2 Experimental results

The bi-level dynamic OD estimation model and the solution algorithm proposed in section 3.3 and section 3.4 are employed at the mesoscopic level 5 iterations each with 10 F-W sub-iterations (for M2 only), and at the microscopic simulation (SMi) 7 iterations each with 10 F-W sub-iterations. 30 minute warm-up and clearance periods with the ground true demands and route-link fractions are used to remove the effects of OD flows outside the estimation period. There are 216 detectors in the LMe, and 44 detectors in the SMi. The default values of $w = 0.5$ and individual weights of 1 are used to reflect a general case where such knowledge is not available. The limits of OD distortions are set to 30 percent from the seed OD flows from the previous simulation step.

The results of the two schemes at the microscopic level (SMi) in terms of the value of the objective function and OD RMSE from the EXP 4-1 are shown in Figure 4-1 and Figure 4-2, respectively; and those from the EXP 4-2 are shown in Figure 4-3 and Figure 4-4, respectively.

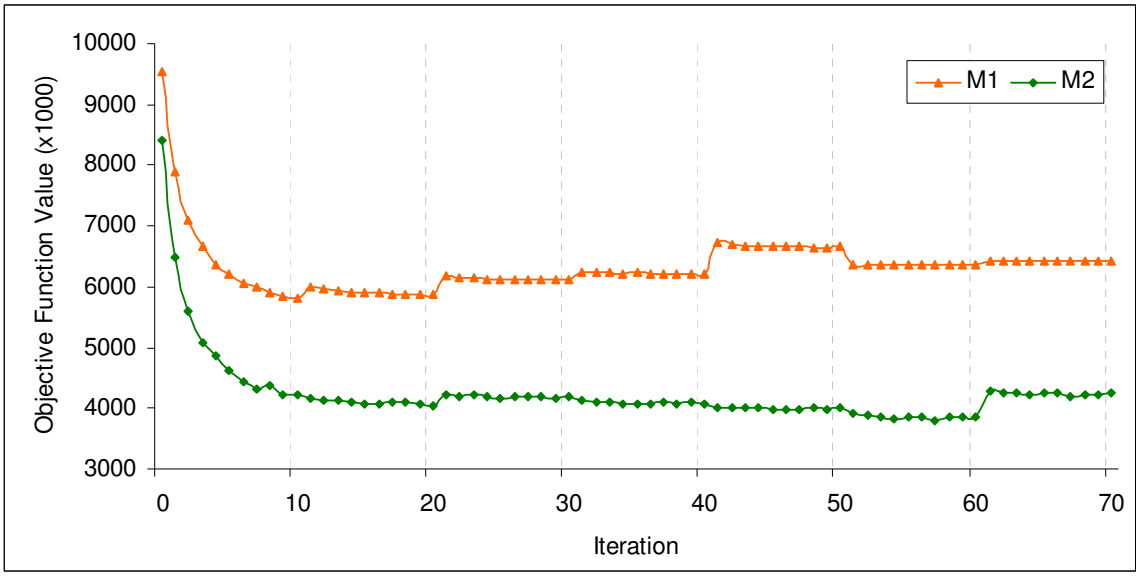


Figure 4-1 Objective function values during iterations (EXP 4-1)

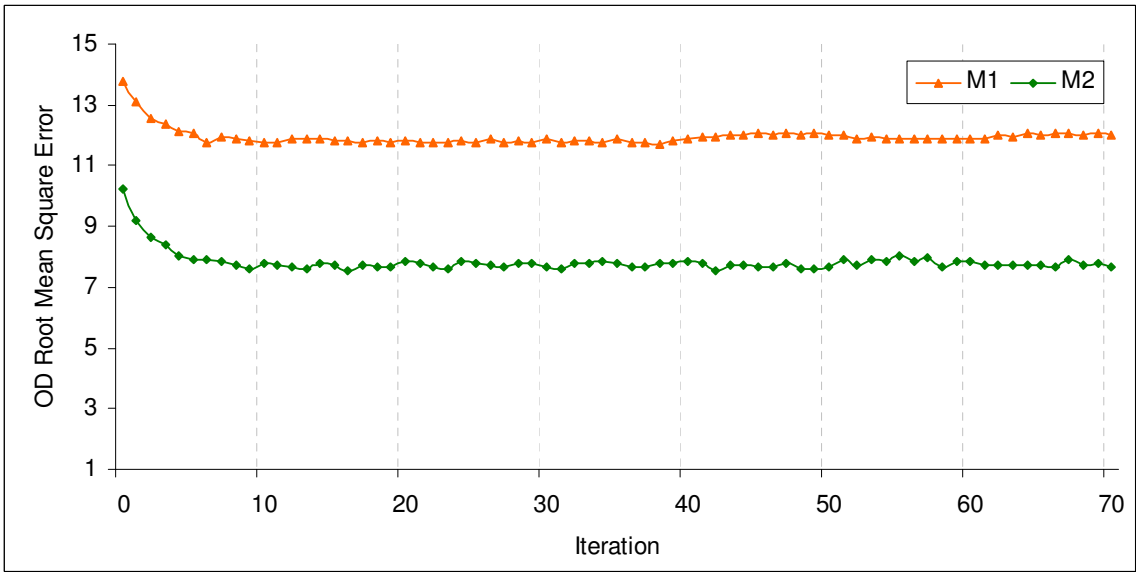


Figure 4-2 OD RMSE values during iterations (EXP 4-1)

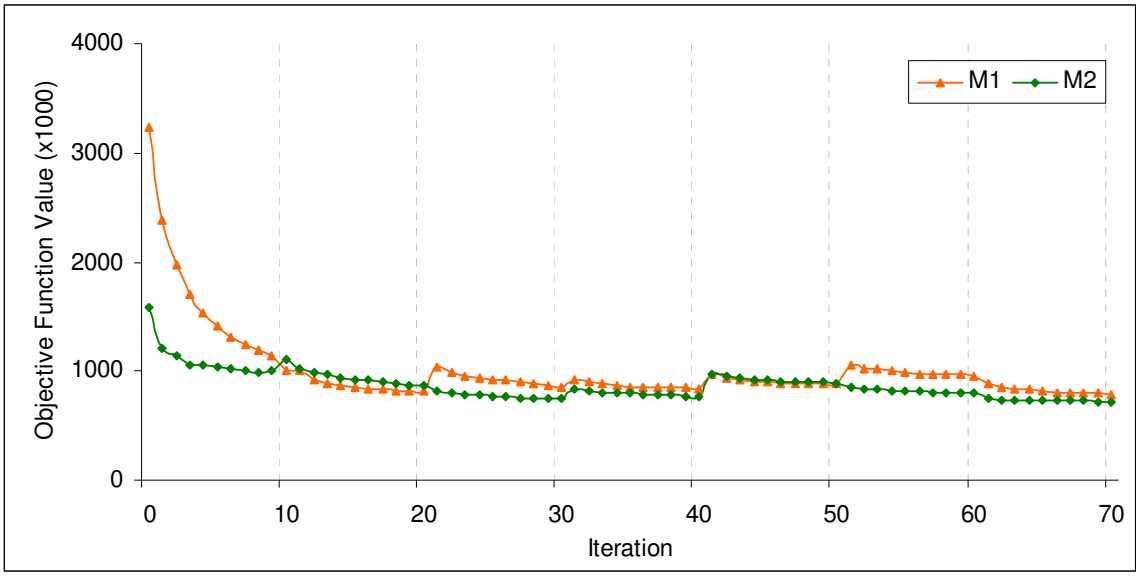


Figure 4-3 Objective function values during iterations (EXP 4-2)

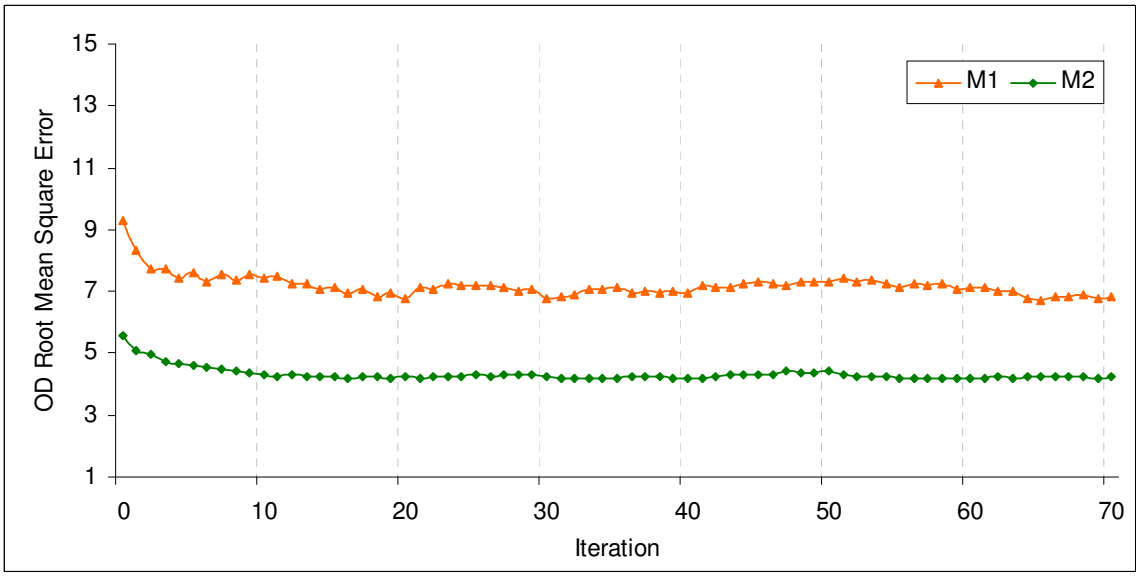


Figure 4-4 OD RMSE values during iterations (EXP 4-2)

Several conclusions can be drawn and are summarized below.

- The proposed six-step process framework (M2) provides much better seed and final estimated dynamic OD tables, in terms of both being closer to the ground truth OD flows and yielding better fit to the objective function. This is attributed

to the dynamic characteristics in the traffic and route modeling from the mesoscopic simulation level. Though the mesoscopic model is much less detailed in the traffic flow process compared to the microscopic process in the ground truth network, acceptable traffic dynamics is modeled. Similarly, even though the mesoscopic model has a rather poor route choice behavior, it still provides enough path flow dynamics.

- For both experiments, the value of the objective function reduces about 50 percent from the initial value, but the OD RMSE reduces only about 20 to 30 percent from the initial value. Moreover, as shown from the results of the EXP 4-2, even though M1 and M2 can achieve about the same objective function value, the estimated OD flows from M2 are much better than those from M1 in term of OD RMSE.
- For both experiments, the value of the objective function reduces rapidly during the first two main iterations (i.e., 20 F-W sub-iterations) and gradually afterward. In addition, at some of the F-W iterations, the value of the objective function is seen to even increase slightly after a traffic simulation run because of the fluctuation in assignment matrix between two successive iterations.
- Surprisingly, despite the fact that the EXP 4-2 involves the additional complexity from unknown route pattern, both initial and final values of the objective function and OD RMSE from the EXP 4-2 are smaller than those from the EXP 4-1**.** One reason for this could be from the fact that the traffic condition in the EXP 4-1 is more congested which greatly impacts the performance of the simulation and the OD estimation model.

- Despite the significant difference in the assignment procedures used in generating the ground truth and in preparing the seed dynamic OD table in the mesoscopic simulation, the proposed framework still provides better results when compared to the static traffic assignment. This is because a reasonable dynamic traffic assignment model can generate a more reasonable route pattern.

Lastly, Figure 4-5 and Figure 4-6 provide a graphical comparison of the demand profiles of two selected OD pairs, respectively.

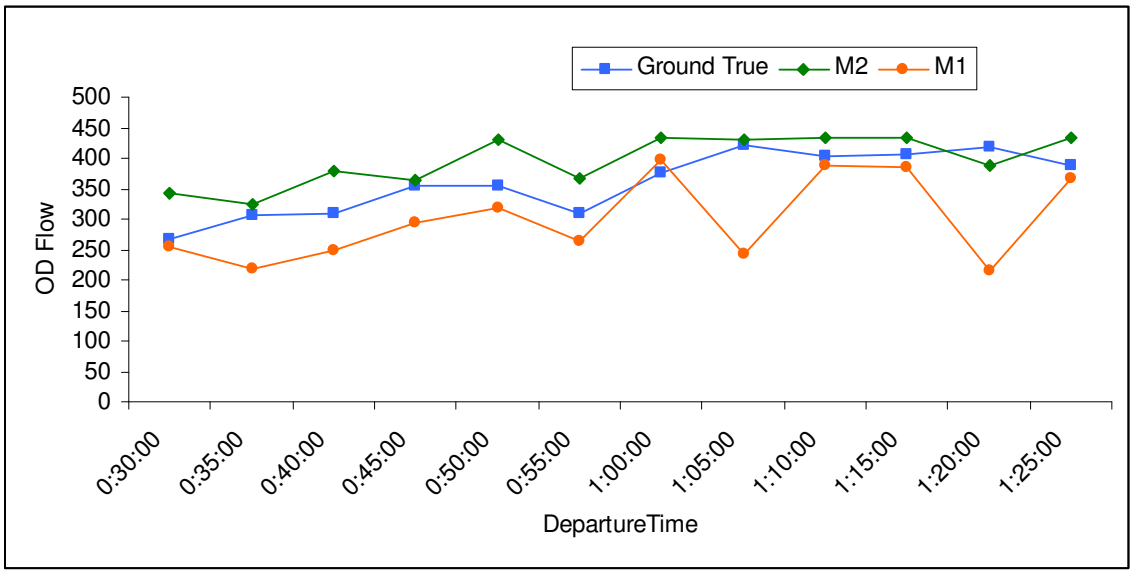


Figure 4-5 Comparison of estimated OD flows for OD pair ZS1 to ZS7 (EXP 4-2)

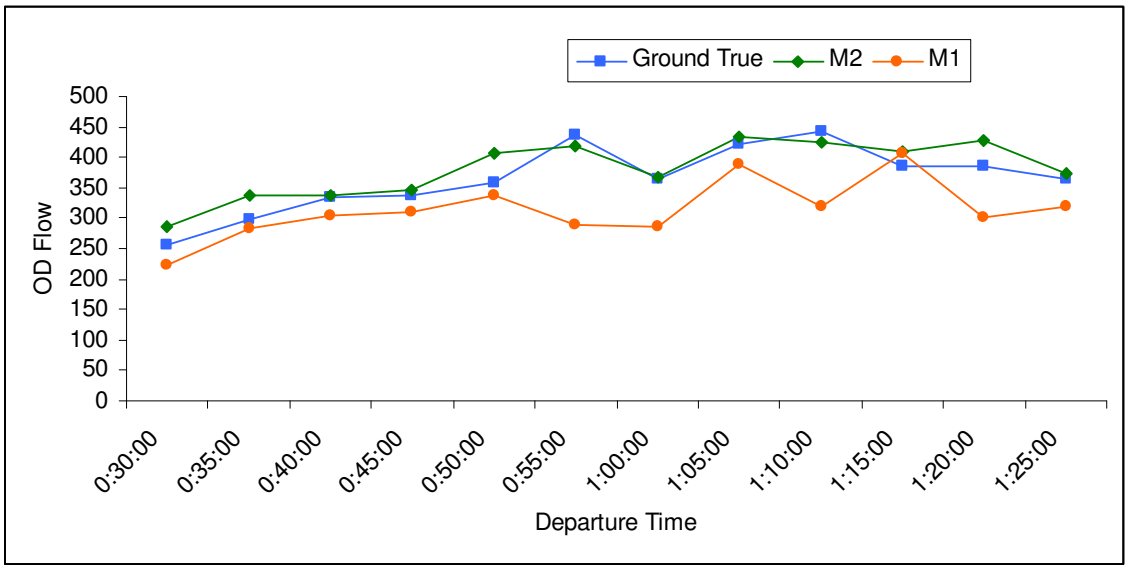


Figure 4-6 Comparison of estimated OD flows for OD pair ZS7 to ZS1 (EXP 4-2)

4.3 APPLICATION OF THE SIX-STEP PROCESS TO I-880 CORRIDOR

The efficacies of the proposed framework, dynamic OD estimation model, and solution algorithm have thus far been illustrated using several experimental studies. Even though these studies have served to illustrate various specific characteristics, they are based on hypothetical networks and simulated data, and many of the difficulties encountered in practice are not seen in the application to these networks. Therefore, the application of the proposed methodologies to the preparation of dynamic OD tables for a real-world network (the I-880 corridor) is considered in this section. Similarly, additional schemes commonly implemented in practice are also performed for comparison purposes. Unlike

the experimental studies examined previously, however, neither the underlying OD flows nor route-link fractions are known in these examples.

4.3.1 Applying the six-step process to I-880 corridor

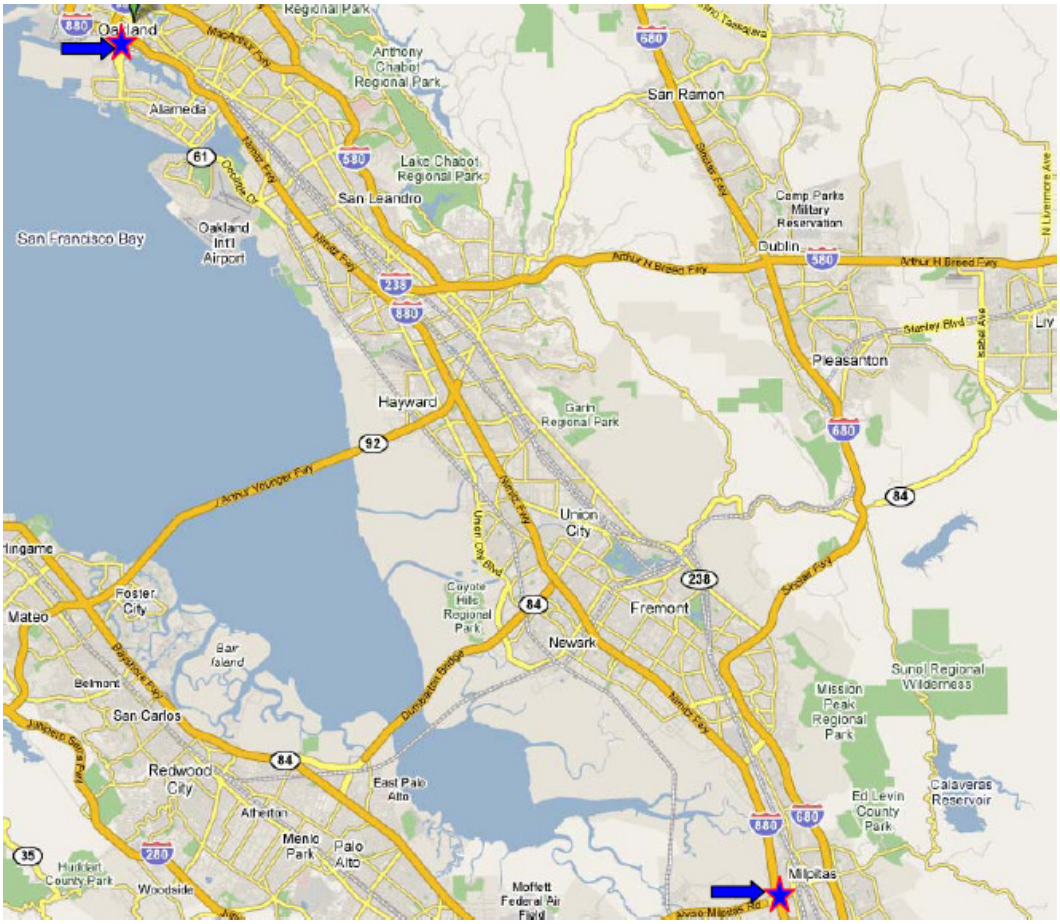


Figure 4-7 I-880 corridor (source: http://maps.google.com/)

The study site includes a 37-mile section of I-880 freeway from Grand Ave. in Oakland to SR-237 interchange in Milpitas, the adjacent highways and surface streets, located east of San Francisco (Figure 4-7). Carrying an average annual daily traffic (AADT) of

268000 vehicles, this section of I-880 is heavily congested during both AM and PM peaks and ranks among the most travelled urban freeways in the United States (USDOT, 2008).

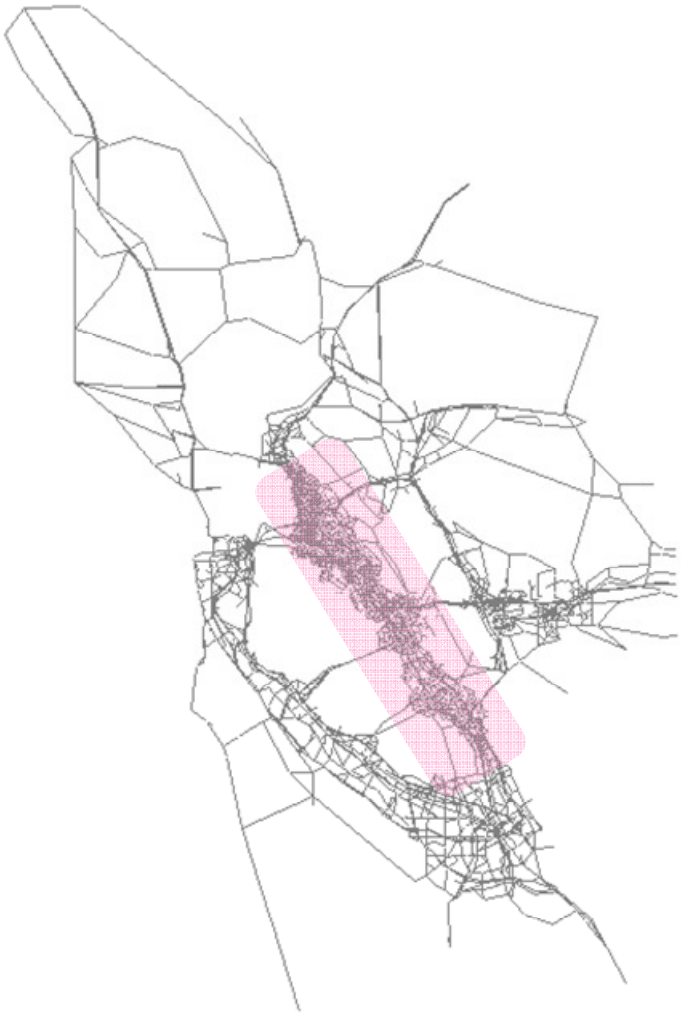


Figure 4-8 ACCMA travel demand model

Figure 4-9 Cut network for sub-area analysis

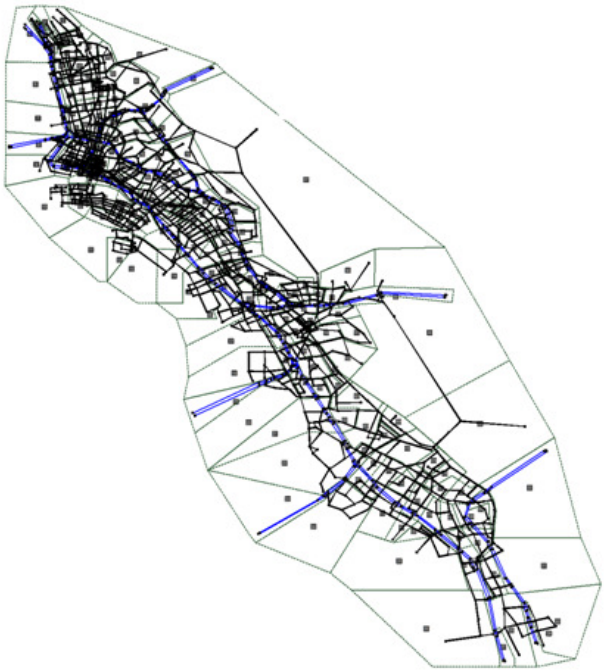


Figure 4-10 Dynamic model of the cut network coded in DYNASMART-P

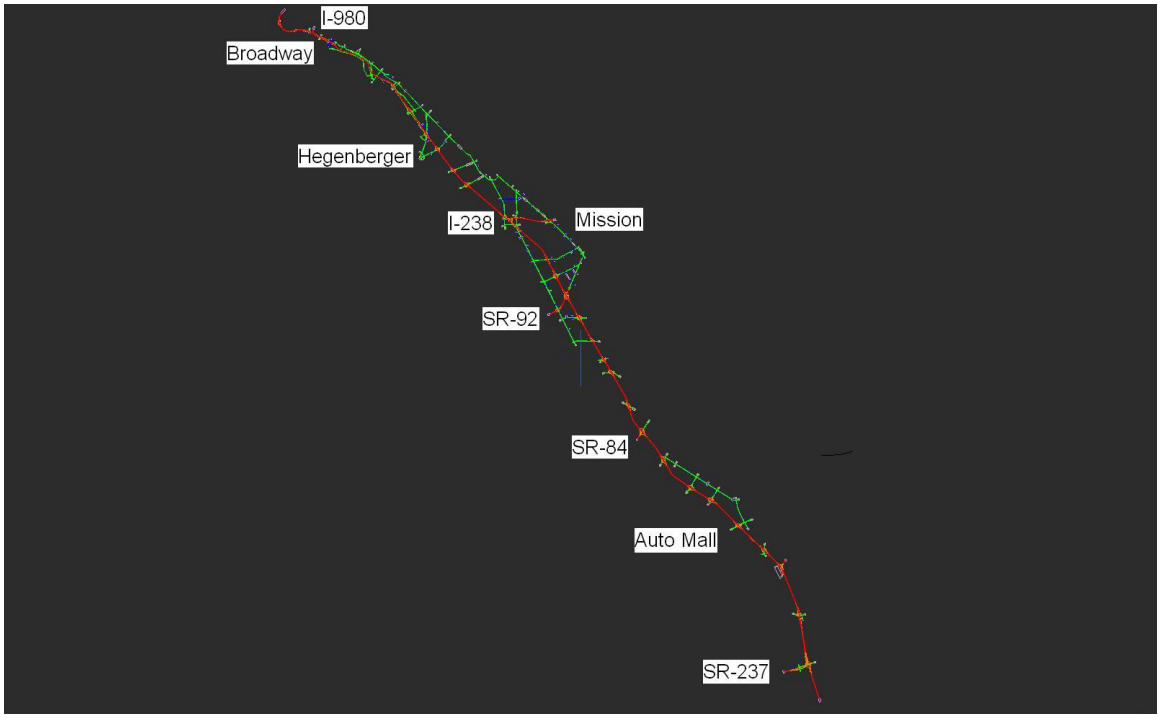


Figure 4-11 Final microscopic simulation network coded in Paramics

The entire study period, 6:30AM-9:30AM, is discretized into twelve 15-minute intervals. 15-minute counts on March 1st, 2nd, and 8th, 2005 are obtained from PeMS (Performance Measurement System), and their averages are regarded as observed mainline traffic counts from a typical day since there is no single day having data from all detectors.

The steps taken for applying the six-step process are provided as follows. Within the original ACCMA (Alameda County Congestion Management Agency) travel demand model (Figure 4-8), AM peak static traffic assignment based on the user equilibrium condition and subarea analysis for a network of intermediate size (Figure 4-9) covering the final microscopic simulation network (Figure 4-11) are first performed in TransCAD.

The output from this step is a static OD table for original zones inside the subarea and new zones along the boundary (external zones).

The subtracted area is then converted to a dynamic network in a mesoscopic traffic simulation model, DYNASMART-P (Figure 4-10). This network consists of 2174 nodes, 5887 links, and 189 zones. Since DYNASMART-P generates vehicles directly on *generation links* rather than centroids, a lookup table mapping between zones and generation links is needed. Using the static seed OD table with departure profiles derived from observed link counts from each origin zone, the bi-level OD estimation is initiated within the mesoscopic level. In the simulation, default values calibrated with real networks are used for the traffic flow parameters for each link type. Each vehicle is assigned to the best shortest route, determined every 3 minutes and updated every 30 seconds, to reduce the computational burden while taking into account the current traffic condition. In real-implementation of the six-step framework, the setting of the assignment procedure in the mesoscopic simulation is a choice of the modeler, as with the one in the traditional planning model. Warm-up and clearance periods of 30 minutes are used to approximate background traffic. Because the seed OD tables from the method mentioned above lead to many oversaturated routes which in turn can result in an unrealistic simulation, a uniform reduction factor of 0.3 is used.

The OD estimation converges after 10 iterations. At this point, the path flows entering and exiting the final microscopic simulation network are converted into seed dynamic OD flows for further OD estimation at the microscopic level. The microscopic simulation network, coded in Paramics, consists of 168 zones, 157 actuated signal intersections, 25 fixed-time signal intersections, and 143 metered lanes at 56 metering

locations. The total of 137 mainline loop detector stations are modeled according to the detection and control data provided by the cities and Caltrans. For the purpose of this study, a previously calibrated network (CCIT, 2006) is used. The issues related to the model calibration are important (see, for example, Chu et al., 2004) but out of the scope of this dissertation.

In this study, the OD estimation period of interest at the microscopic level is set as 6:30AM-7:30AM, with additional 30-minute warm-up and clearance periods. Each driver is assumed to use his/her perceived shortest route with moderate randomness, although the network is nearly linear. The default values of w=0.5 and individual weights of 1 are used. The maximum distortions (bound constraints) are set to 20 percent from the initial estimates from the mesoscopic level. However, if the initial OD flow is close to zero, the maximum bound is set to 15.

Figure 4-12 and Figure 4-13 show, respectively, the results of the OD estimation at this microscopic level in terms of the value of the objective function and traffic count RMSE during the first 6 iterations, each with 10 F-W sub-iterations.

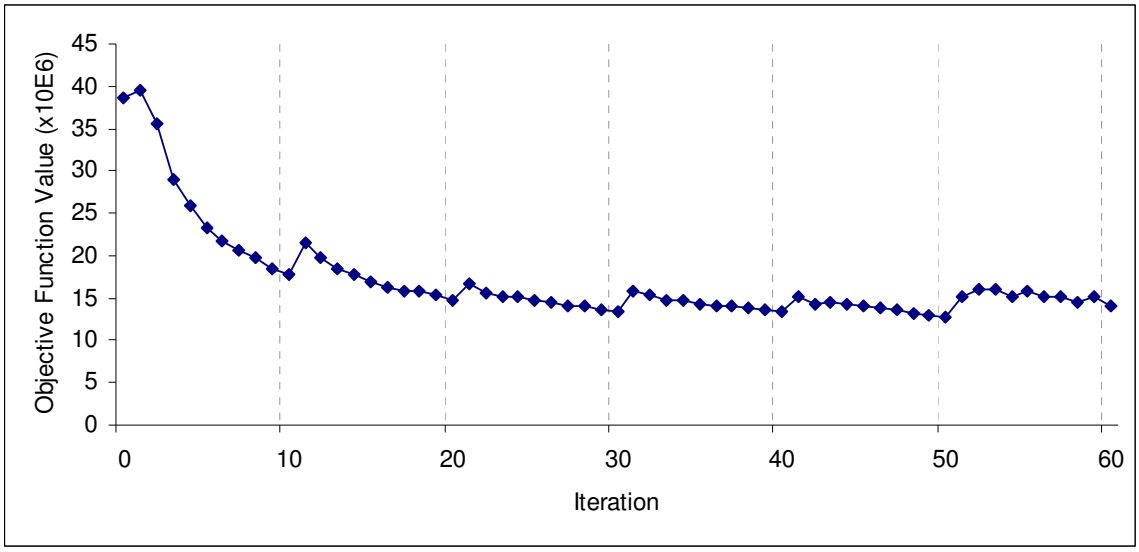


Figure 4-12 Objective function values during iterations (six-step process)

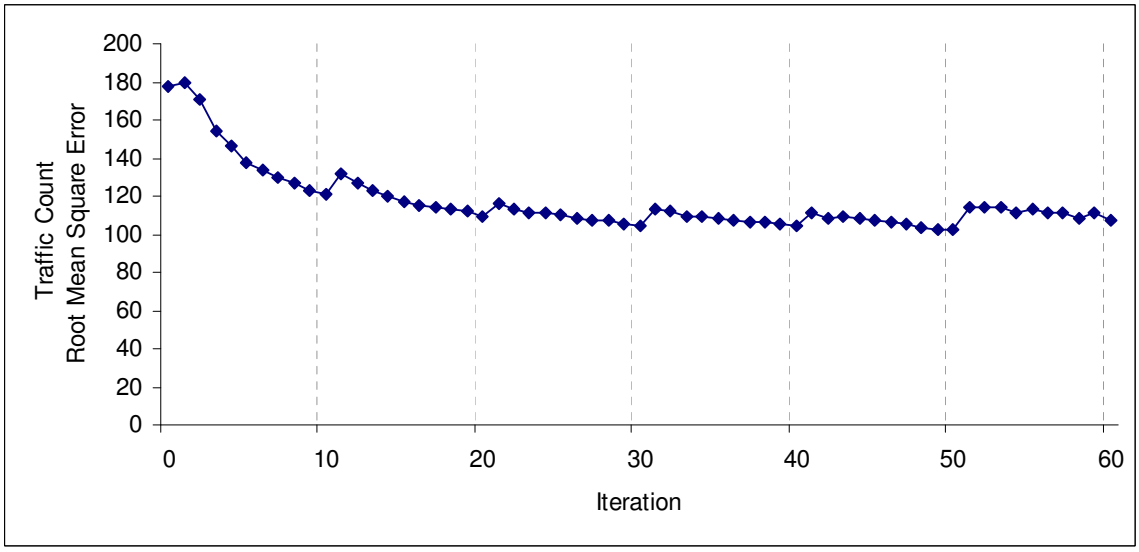


Figure 4-13 Traffic count RMSE values during iterations (six-step process)

After the 60th iteration, the values of the objective function and the traffic count RMSE reduce about 63 percent and 36 percent from their initial value, respectively. The improvements in these two measurements occur rapidly during early iterations and gradually afterward. Note that besides the fluctuations after each traffic simulation run

due to the change in assignment matrix which are expected, both sequences are generally convergent sequences.

4.3.2 Applying other schemes

To gain insight on the benefits from the above application of the six-step framework to the I-880 corridor, additional estimation schemes representing the current practice are performed in this subsection to enable the comparison.

Figure 4-14 Sub-area analysis within the ACCMA model cut for the final microscopic model

Figure 4-14 illustrates the sub-area analysis performed within the original ACCMA model for the final microscopic simulation model. The output from this procedure is a

static OD table which is then uniformly cut and regarded as seed dynamic OD tables in the following schemes.

SCHEME 1 100 percent loading

The first scheme initiates the OD estimation with 100 percent seed OD tables. Similarly, the default values of w =0.5 and individual weights of 1 are used. The maximum distortions are set to 20 percent.

As expected, the network rapidly becomes heavily congested in every simulation run due to an unrealistic gridlock caused by excessively high OD flows. The situation observed is that an entire freeway section is blocked because of a large number of vehicles stop and waiting to exit to particular locations (destinations) (see Figure 4-15).

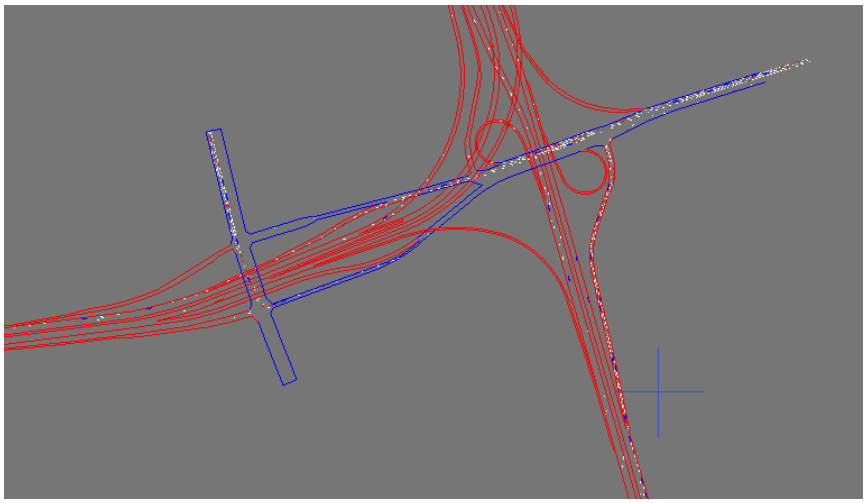


Figure 4-15 An example of unrealistic congestion

Because of the gridlock, the immediate downstream section cannot receive any flow and the length of the stopped queue keeps growing infinitely. Figure 4-6 shows a screenshot of the simulation 20 minutes later than when the bottleneck in Figure 4-15 starts to form.

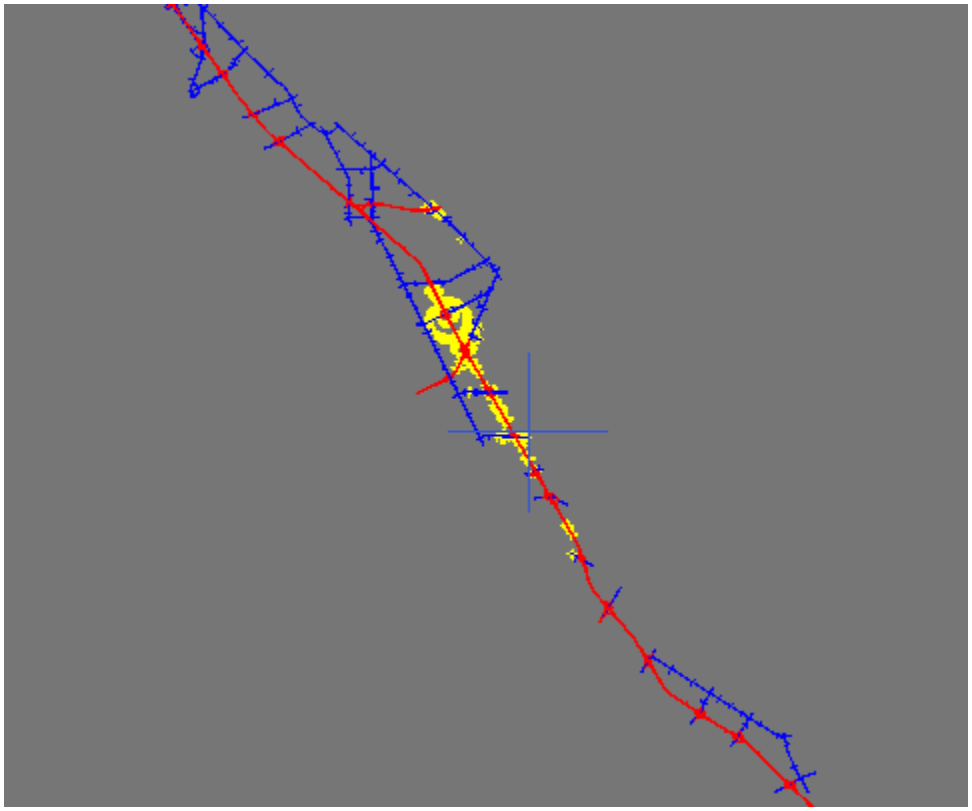


Figure 4-16 Growing queue from a gridlock

Figure 4-17 and Figure 4-18 show the results of the OD estimation in terms of the value of the objective function and RMSE of the estimated traffic counts, respectively.

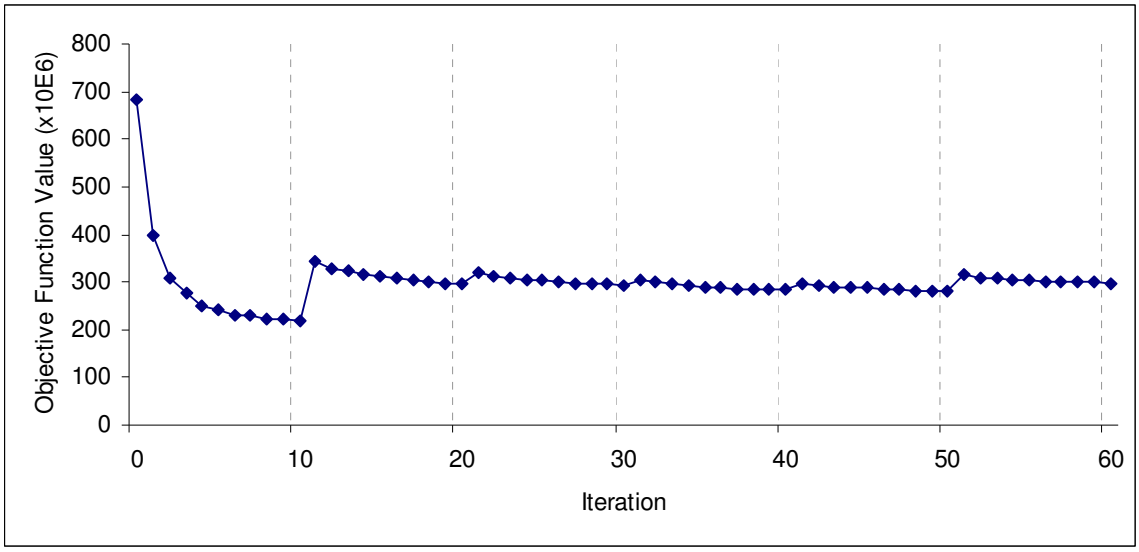


Figure 4-17 Objective function values during iterations (SCHEME 1)

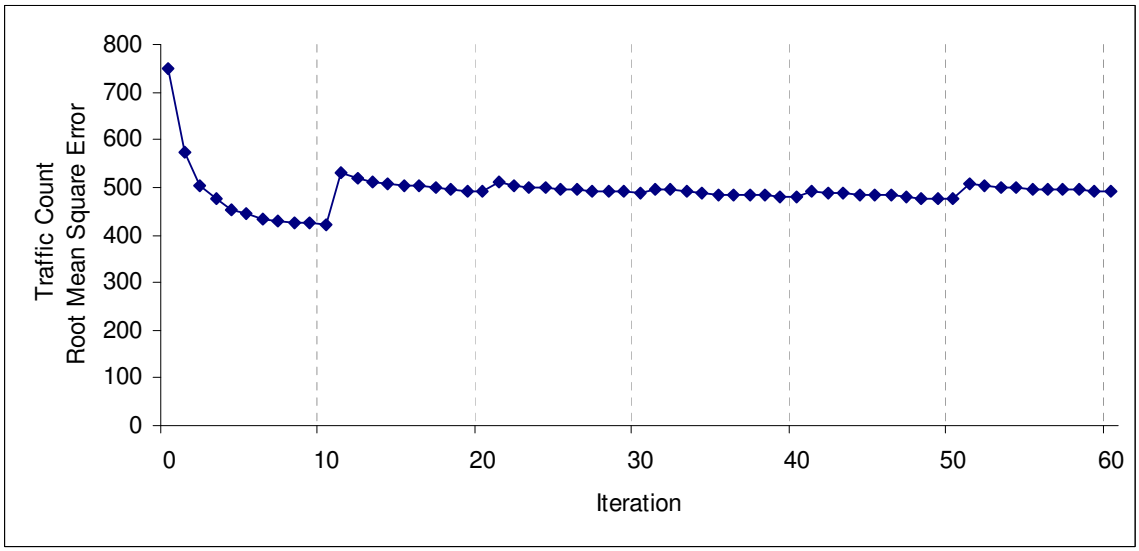


Figure 4-18 Traffic count RMSE values during iterations (SCHEME 1)

The following conclusions can be drawn.

- Excessively high flows in seed OD tables are the result of unrealistic route patterns and oversaturated link flows from the traditional static model. These demand flows cause gridlock in the microscopic traffic simulation which in turn

results in a rapid change in the assignment matrix. Usually, the change is such that the downstream links would receive more flows from different OD pairs not passing the gridlock while the upstream links tend to become increasingly congested as the simulation is unable to send vehicles from the adjusted OD tables to detector stations as predicted by the previous assignment matrix.

- The huge increases in the value of the objective function and traffic count RMSE at the 11^{th} iteration (i.e., after the 2^{nd} simulation run) are an indicator of the gridlock and sudden change in the assignment matrix between two consecutive iterations.
- After the 11^{th} iteration, the algorithm appears to be trapped with some incorrect assignment matrix and OD tables (i.e., a local solution). In practice, in order to enable further adjustment (i.e., go to another solution), it would require the modeler to manually reduce some OD flows to clear up the gridlock.
- Compared to the results from the proposed six-step process, this scheme provides much worse seed and final estimated dynamic OD tables based on the objective function value and traffic count RMSE.

SCHEME 2 70 percent loading

To prevent gridlock from occurring in the simulation, it is intuitive to uniformly reduce the magnitude of the entire seed OD tables. The second scheme thus initiates the OD estimation algorithm with 70 percent seed OD tables. The same set of weights and

restrictions are used. That is, w=0.5, individual weights = 1, and the maximum distortions = 20 percent.

The first simulation iteration runs without any gridlock. However, all simulation iterations afterward include the same gridlock location. Figure 4-19 and Figure 4-20 show the results of the OD estimation in terms of the value of the objective function and traffic count RMSE, respectively.

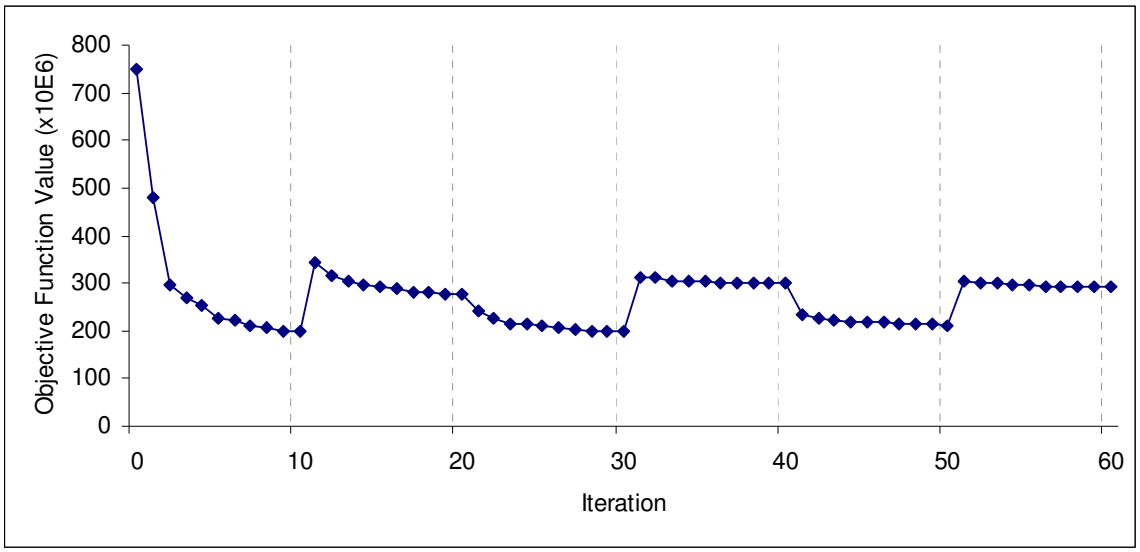


Figure 4-19 Objective function values during iterations (SCHEME 2)

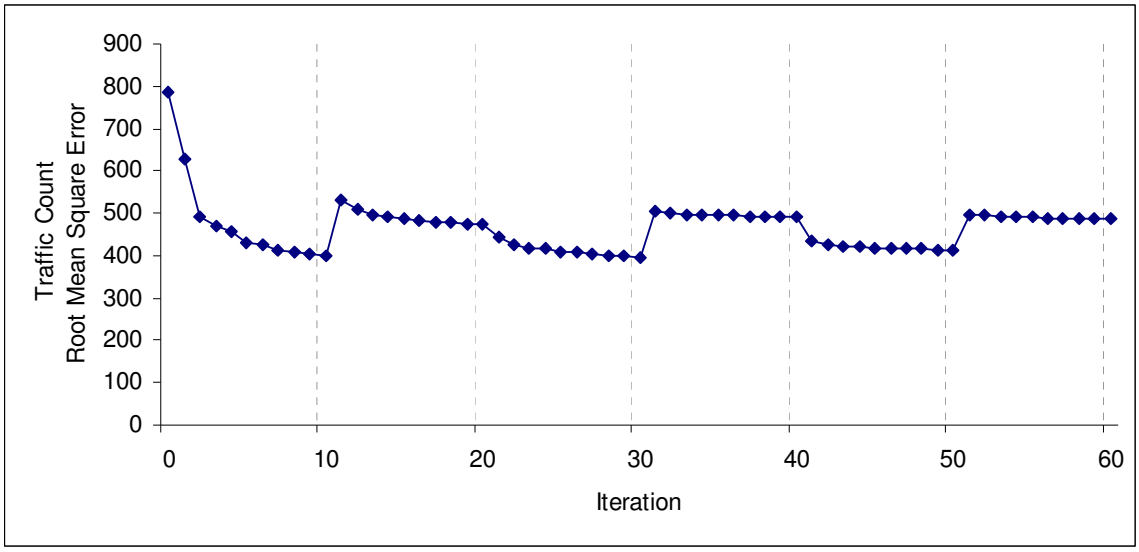


Figure 4-20 Traffic count RMSE values during iterations (SCHEME 2)

Several conclusions are drawn from the results and summarized below.

- Within the first main iteration (i.e., 10 sub-iterations), the algorithm converges very fast and yields approximately the same values of the objective function and traffic count RMSE as in the last scheme, although it initiates with the OD tables which are 30 percent lower in magnitude.
- The huge increases in the value of the objective function and traffic count RMSE indicating the sudden change in the assignment matrix at the 11^{th} iteration (i.e., after the 2^{nd} simulation run) are also presented. However, in this case, both measurements improve until the 30^{th} iteration and fluctuate afterward due to the fluctuation of congestion in simulation runs.
- Again, compared to the results from the proposed six-step process, this scheme provides much worse seed and final estimated dynamic OD tables based on the objective function value and traffic count RMSE.

It should be noted from the comparisons that seed OD tables significantly affect how the OD estimation algorithm converges. Because of the non-monotonic relation between OD and link flows and capacity restrictions in traffic simulation, the OD estimation algorithm cannot effectively update assignment matrix and OD flows if their values in the previous iteration deviate greatly from a reasonable solution.

4.3.3 Other performance measures

The discussion and comparisons presented in the previous subsections are made based mainly on the improvements in the objective function value and traffic count RMSE. Despite the indicated benefits of the six-step process, these performance measures are indirect one, and it is helpful to further examine and compare the OD tables with respect to the following supplemental performance measures.

Zonal dynamic departure flows and arrival flows

Although the underlying dynamic OD tables are not known, the underlying dynamic departure and arrival flows of those zones whose all attached links contain observed traffic counts can be determined. It is thus possible to make the comparisons between the observed and estimated row and column totals of the OD tables.

Based on the network configuration and data availability, 32 out of the total of 168 zones can be evaluated based on these two measures. However, since it is cumbersome to separately examine all 32 zones, only four samples are used in the profile comparisons. Figure 4-21 and Figure 4-22 compare the dynamic departure profiles as generated by the six-step framework and the scheme 1 (as explained in the last subsection) with the underlying one from zone 95 and zone 127, respectively. As indicated in both figures, the six-step framework yields a departure profile that is much closer to the underlying profile than is the one from the scheme 1.

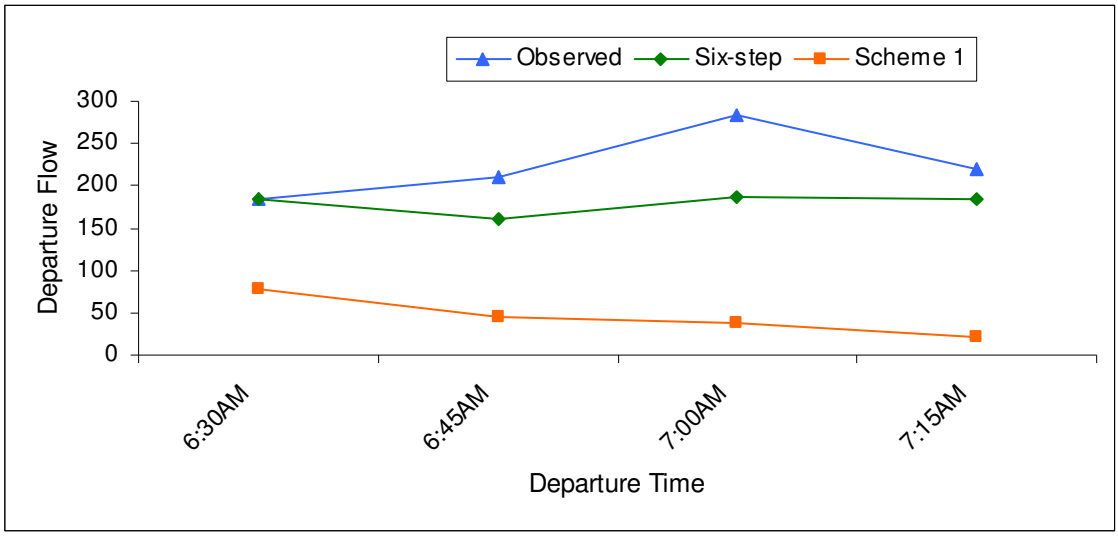


Figure 4-21 Dynamic departure flows from zone 95

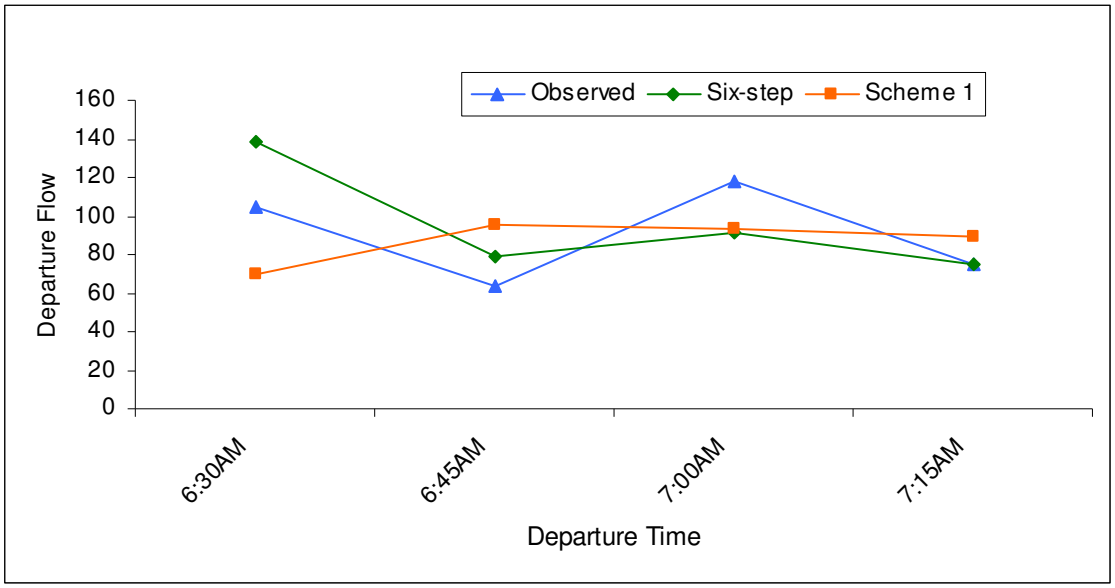


Figure 4-22 Dynamic departure flows from zone 127

The same superiority of the six-step framework is also indicated in Figure 4-23 and Figure 4-24 which show the arrival profiles as generated by both schemes of zone 8 and zone 166, respectively.

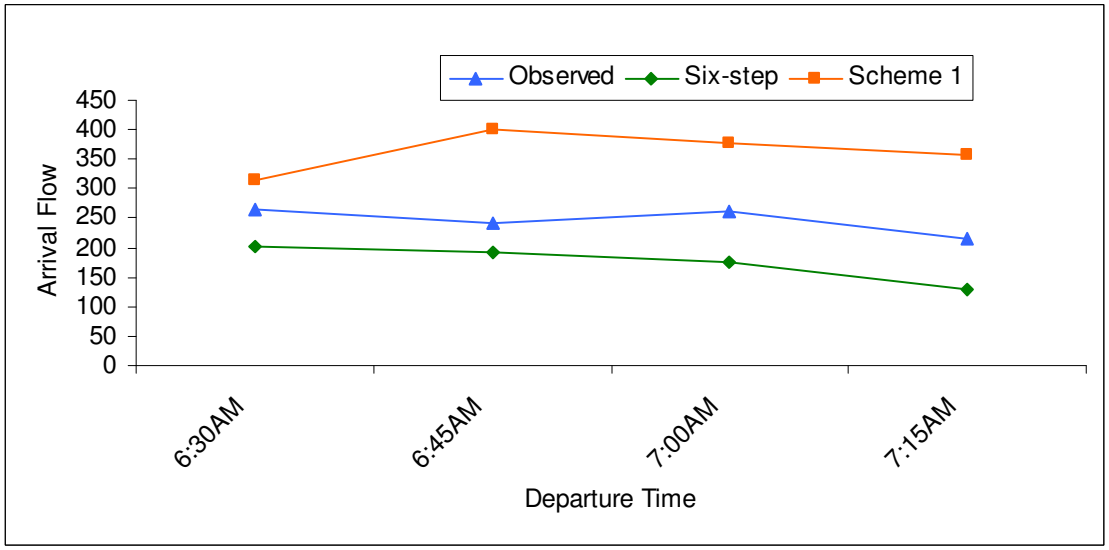


Figure 4-23 Dynamic arrival flows from zone 8

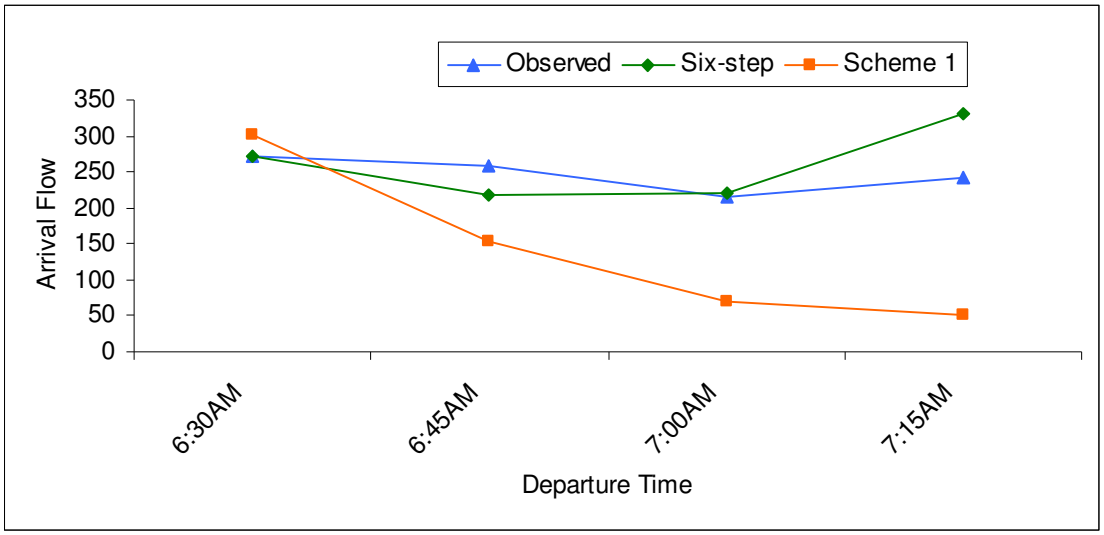


Figure 4-24 Dynamic arrival flows from zone 166

Since it may not be sufficient to make any general conclusions only from the above comparisons and one might argue that the results from these four samples are not representative, RMSE of the estimated departure and arrival flows of all 32 zones during each interval is calculated and shown in Figure 4-25. Several conclusions can be made. First, the six-step framework provides a much better match to the observed departure and arrival flows during each interval. Second, the RMSE appears to be smaller during the first interval when the traffic is not congested and becomes larger as the network gets congested. Lastly, while the RMSE from the six-step framework during the last three intervals seems to be stable, the RMSE from the scheme 1 increases continuously. Again, this can be considered as the effects of the unrealistic gridlock explained in the previous subsections.

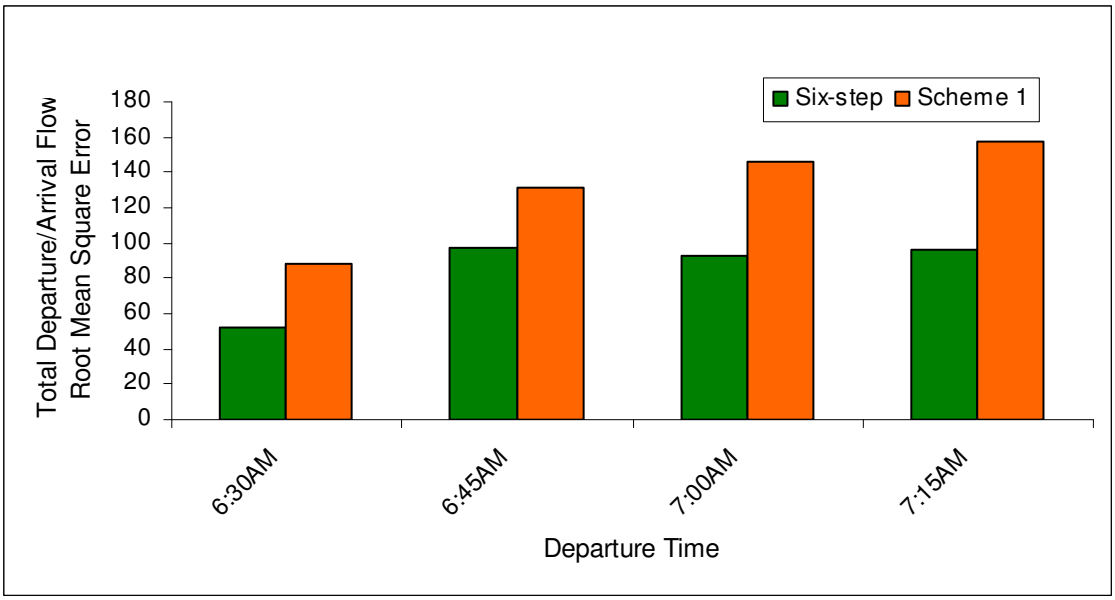


Figure 4-25 Total departure/arrival flow RMSE

Ratio between the total released vehicles and the total demand

The ratio between the total released vehicles and the total demand is the fraction of total vehicles in the OD tables that is successfully loaded into the simulation. For the network being considered which is nearly linear, a lower ratio indicates higher chance of having excessively high demand flows. Given that the comparison is made among the cases with the same total number of vehicles, this ratio can thus be also used to roughly measure the quality of OD tables.

The six-step framework and scheme 1, which result in approximately the same number of total demand, yield a ratio of 0.97 and 0.81, respectively. That is, while almost all vehicles in the OD tables from the six-step framework can successfully be loaded into the simulation, only 81 percent of the OD flows as generated from scheme 1 can be

loaded. This can be because of two reasons. First, the set of OD tables from the six-step framework is more reasonable. Second, the mesoscopic simulation model used in the 5th step provides more paths (i.e., seed OD tables that are less sparse) than the traditional static model. For this case, the first generates 15852 dynamic routes while the latter generates 14353 dynamic routes.

Demand profile of neighboring zones

Because zones in the microscopic simulation model are located at an intersection approach or a freeway ramp, neighboring zones are often very close to each other. It is thus reasonable to expect that demand profiles of zones within the same neighbor (say, to or from another particular zone) should not be very different from each other even though they are not necessary similar.

Figure 4-26 shows the location of zones used in the following comparisons. Note that zone 30 and zone 31 carry eastbound and westbound traffic of SR84, respectively; and that zone 91 and zone 92 are attached to a different approach link of the same intersection located in a residential-commercial area.

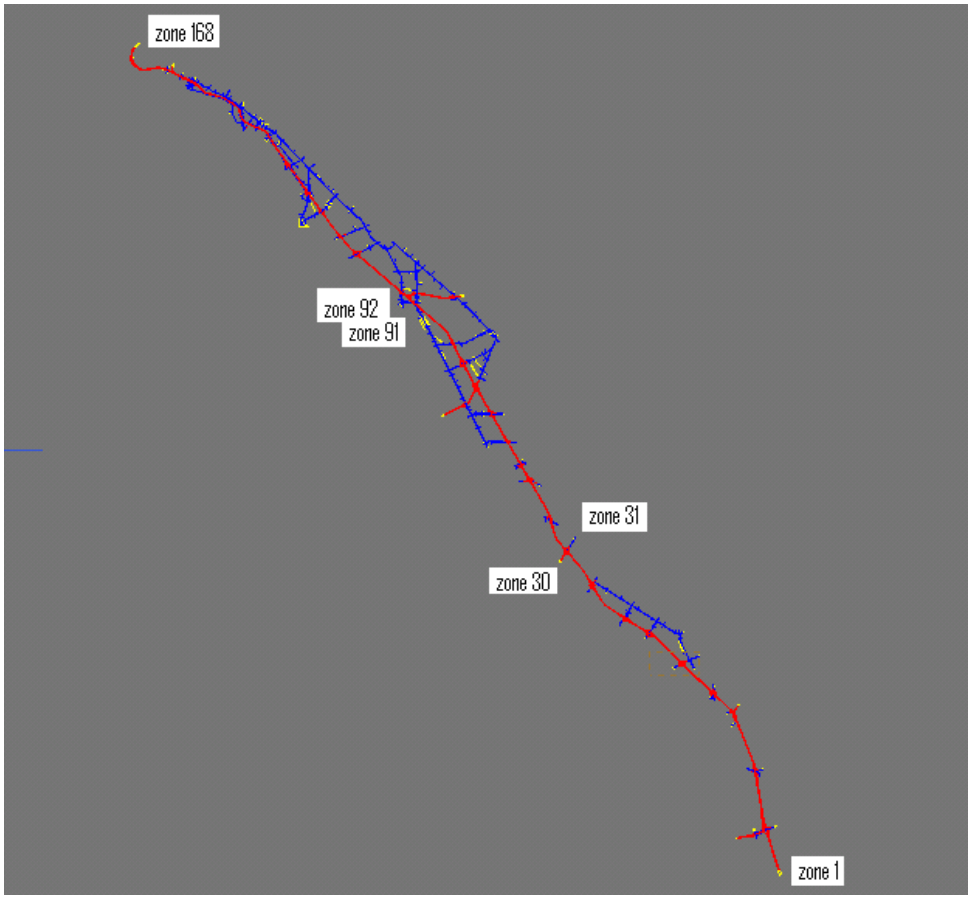


Figure 4-26 Location of the zones for comparison

Figure 4-27 and Figure 4-28 show the demand profiles from both schemes between zone 30 to zone 1 and zone 31 to zone 1, respectively. As shown, the demand profiles of these two OD pairs as generated by the six-step framework are similar in terms of both magnitude and shape, but the profiles generated from the scheme 1 are very different.

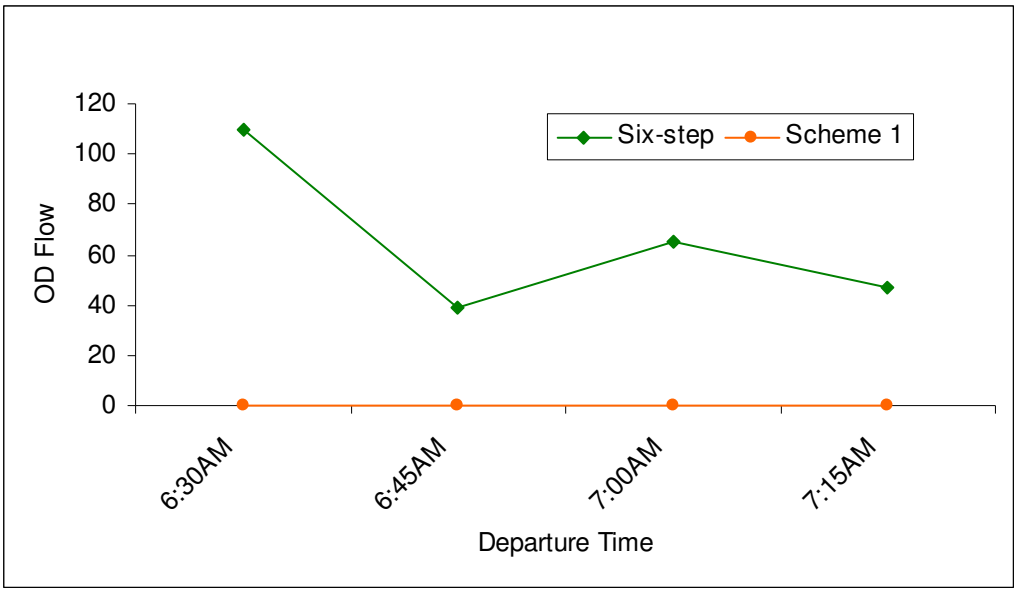


Figure 4-27 Estimated OD flows from zone 30 to zone 1

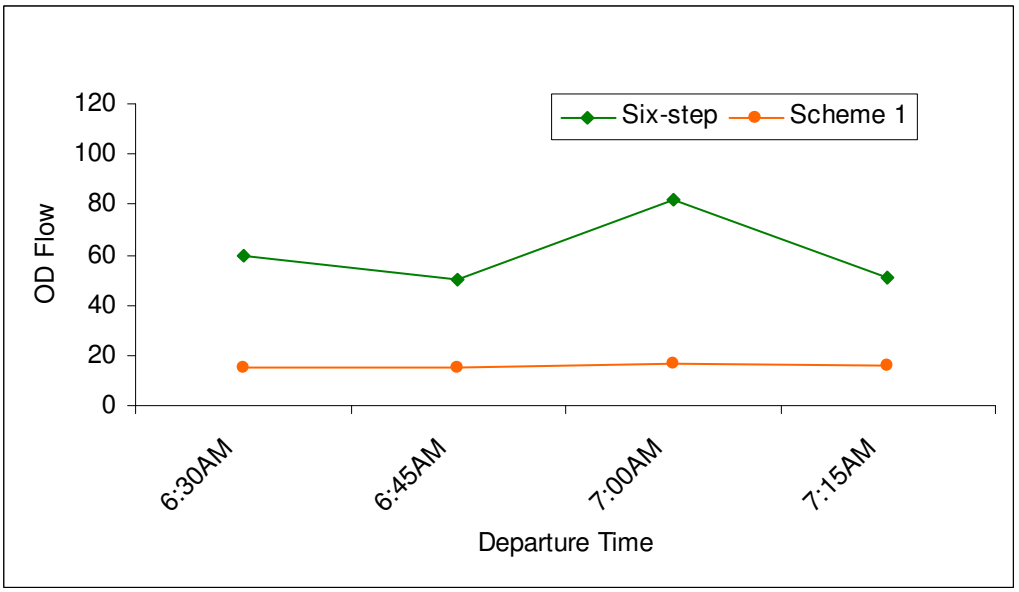


Figure 4-28 Estimated OD flows from zone 31 to zone 1

The same conclusion can also be drawn by comparing Figure 4-29 and Figure 4-30 which show the demand profiles from both schemes between zone 91 to zone 168 and zone 92 to zone 168, respectively.

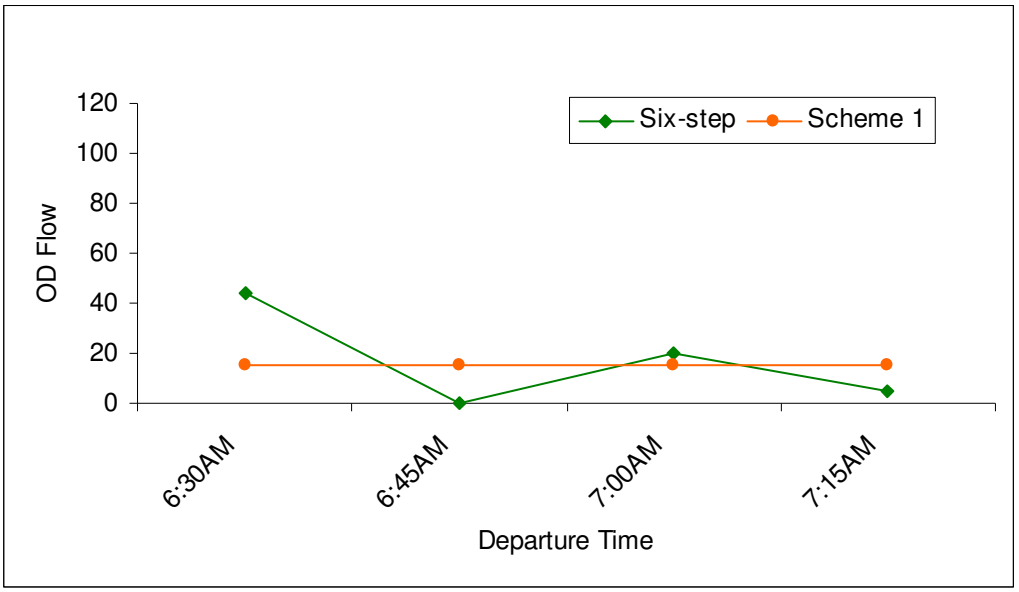


Figure 4-29 Estimated OD flows from zone 91 to zone 168

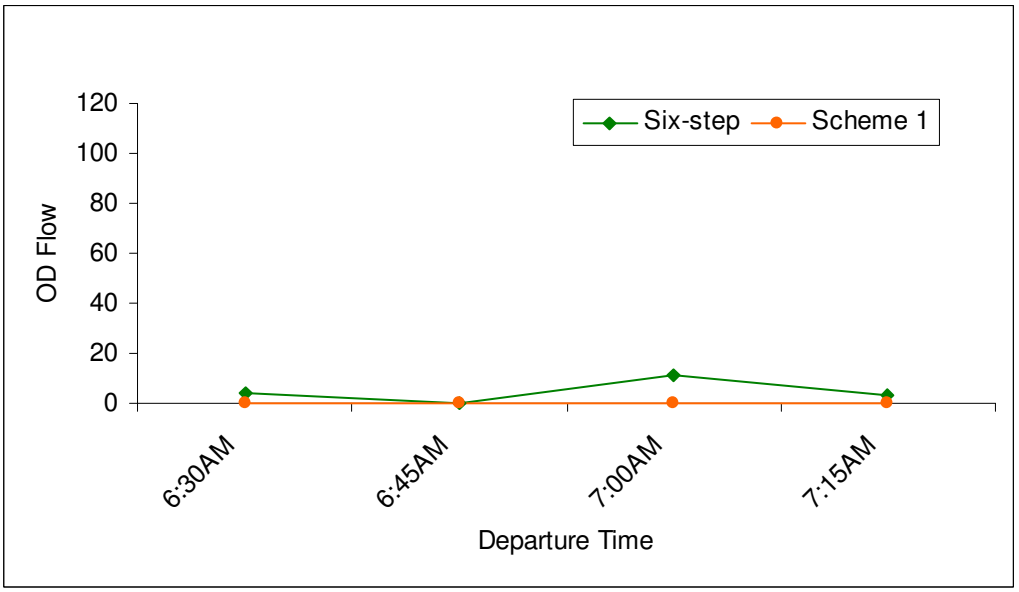


Figure 4-30 Estimated OD flows from zone 92 to zone 168

4.4 SUMMARY

This chapter evaluated the proposed six-step framework, the dynamic OD estimation model, and solution algorithm presented in the previous chapter using experimental studies as well as a real world application. Particularly, the experimental studies using hypothetical networks and simulated data were designed to illustrate various specific characteristics of the proposed methodologies while the application to I-880 corridor was intended to exemplify their efficacy under practical settings.

Based on the improvements in the objective function value and OD RMSE, the experimental studies indicated the superior abilities of the six-step framework in yielding better seed and final OD estimates. This is mainly attributed to the 5th step of the framework implemented with a mesoscopic traffic simulation model. Although the model contains much less details in traffic flow modeling compared to the microscopic simulation and the assumed route choice behavior is rather coarse, it provides more reasonable traffic and route patterns compared to the traditional static model.

The main challenges in preparing dynamic demand tables for the I-880 corridor are due to the fact that the network is heavily congested. It was found that the seed dynamic OD tables cut out directly from the static regional planning model showed excessively high demand for many OD pairs and could not be loaded into the microscopic simulation model. This is because the static traffic assignment allows oversaturated link flows. On the other hand, the 5th step of the proposed framework provided the seed dynamic OD tables, which led to a much better convergent sequence. The traffic count RMSE,

departure and arrival profiles, demand profiles, and vehicle release ratio also indicated the superiority of the estimates from the proposed framework.

The F-W algorithm suggested for solving the model's upper level showed a fairly fast convergent sequence to a reasonable point before the next simulation update in all cases and could be applied to large networks successfully. However, as with other iterative bi-level algorithms, the entire algorithm might be trapped with a local solution. In fact, the quality of the solution obtained depends largely on the quality of seed OD tables. Fluctuations in the optimization can also occur due to the change in the assignment matrix between two successive iterations. Because of the non-monotonic relationship between path and link flows and the capacity restrictions in the traffic simulation, the OD estimation algorithm cannot effectively update the assignment matrix and OD flows if their values in the previous iteration deviate greatly from a reasonable solution.

Lastly, as shown from all cases, the improvement in seed dynamic OD tables provided by the 5^{th} step of the proposed framework as compared to the ones cut directly from the static planning model is greater or at least equal to the total improvement from the OD estimation within the final microscopic simulation level. This finding should be underlined and used as the motivation for future related research to pay more attention to improving the quality of seed OD flows rather than the estimation model alone.

CHAPTER 5 DYNAMIC DEMAND PREPARATION WITH SMALL SAMPLES OF VEHICLE TRAJECTORY DATA

5.1 INTRODUCTION

The previous two chapters focused on a framework used for preparing dynamic demand inputs for microscopic simulation. The framework involves several levels of trip making and traffic modeling from the traditional static planning, as well as mesoscopic and microscopic simulation models. These schemes are necessary due to the lack of direct observations. It is not surprising that the validity of the estimates from this framework depends significantly on the validity of such models, and thus the use of these results should be justified according to the application being considered, and only with a clear understanding of the assumptions made.

Despite the fact that these modeling tools are important for various applications related to demand forecasting as well as the design and/or evaluation of alternative transportation projects, there is only little motivation for using these models in demand estimation if vehicle trajectory data are available. In fact, since such data provide not only the information regarding the OD demand pattern but also the information about the traffic and route patterns, they should ultimately be used to recalibrate both supply and demand modeling, such as those used in the previous two chapters.

The current and future challenges in using vehicle trajectory data are essentially from the fact that they are only available with small sampling rates that are both unknown

and random over time and space. Fortunately, several possible sampling schemes based on self-voluntary, motivated-by-rewards, and redistribution of necessary add-on equipments can be employed in a continuing manner over multiple days with relatively low cost. Therefore, the second framework proposed in this dissertation seeks efficient methods in using these data and combining them with the traditional data from loop detectors for the demand estimation. The specific problem of concern can be stated as follows.

Consider a transport network with several origins $i \in I$ and destinations $j \in J$. Each OD pair ij consists of multiple routes $k \in K$. The study period is discretized into departure time intervals $t = 1, 2, ..., T$. Traffic counts $\breve{v}_{lh}$ and loop occupancies $\breve{c}_{lh}$ are available on links $l \in L_{obs}$ during each observation interval $h = 1, 2, ..., H$. Assume that multiple small trajectory datasets are available from random samplings over several days, not necessary successive but including the target day m. Denote the number of observed trajectories between OD pair ij during departure time interval t on day d by $\tilde{x}_d^{ijt}$ and on route k by $\tilde{f}_d^{kt}$. The objective is to estimate dynamic OD tables for the study period on the target day using vehicle trajectories, loop occupancies, and counts without assuming the existence of prior knowledge of initial OD tables or assignment matrix or using a simulation model. Observed trajectories are assumed to have been processed and their trip ends have been correctly identified. The problem of trip end identification is important but is not within the scope of this dissertation. Du and Aultman-Hall (2007) propose a method for identifying trip ends for GPS data which can achieve a 94 percent correct identification rate.

The framework developed in this research differs radically from the previous studies in the literature. It features the use of traffic data from multiple sources to derive necessary elements in a probabilistic distribution form and the use of a stochastic estimation model to further update these initial estimates with observed traffic counts. The key contributions of the new formulations include.

- The application of Bayesian formulations to take full advantage of multiple trajectory datasets to derive the distributions of initial OD tables, link travel time distributions, and route-choice fractions. Depending on the availability of trajectory data, parameters from different hierarchies in the formulations can be flexibly used in the estimation models (optimization procedure, Kalman filtering, etc.) to ensure a high level of model observability.
- The development of an Occupancy-based Dynamic Network Loading (O-DNL) model to replace the use of a simulation model. Under this approach, a Monte Carlo simulation is set to simulate route flows by moving vehicles through each link based on the estimated travel time distribution, which in turn depends on the interval (and traffic condition) when they arrive at the link. Since the O-DNL is based on field data and is exogenous from the estimation model, the improper OD adjustment caused by the non-monotonic relationship between traffic counts and OD flows can be avoided.
- The development of a new bounded dynamic OD estimation model which simultaneously adjusts OD flows, route-link fractions, and route-choice fractions. Such adjustment is performed based on fully-estimated error variance-covariance

matrices, which reflect relative confidence in the information derived from different sources.

- The proposed formulations can be used for both off-line and on-line applications. The off-line mode refers to a training stage where data from multi-days are used to infer the distributions for different elements. The online-mode refers to an updating stage for distribution parameters without further need to handle the large dataset.

The proposed framework can be thought of as a two-step procedure: *Bayesian data analysis step* and *stochastic dynamic OD estimation step* as shown in Figure 5-1.

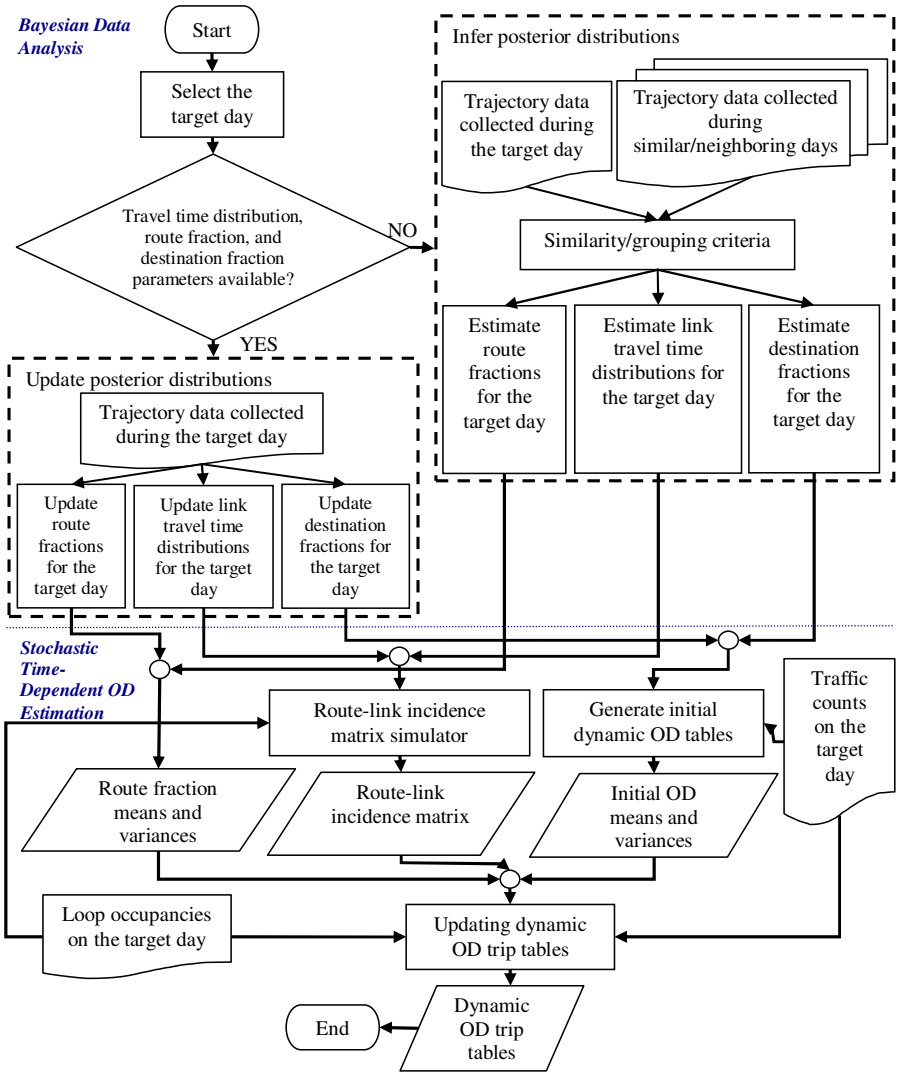


Figure 5-1 Flowchart of the framework for dynamic OD estimation with probe data

5.2 CHAPTER OUTLINE

This chapter describes the developments of several Bayesian models used in the first step of the proposed framework, and the development of new procedures for estimating initial route-link fractions and seed dynamic OD tables. The discussion on a new dynamic OD estimation with stochastic assignment matrix is postponed to the next chapter.

The rest of this chapter is organized as follows. Section 5.3 first provides some background on the current studies related to travel time estimation. Then, the development of a hierarchical Bayesian mixture model for estimating time-dependent link travel time distribution is presented. Sections 5.4 and 5.5 provide additional empirical Bayesian (EB) models used in estimating route fractions and destination fraction, respectively. Section 5.6 reports a new Monte Carlo simulation set up for estimating initial route-link fractions from estimated travel time distributions. In section 5.7, a new procedure for constructing initial (seed) OD tables is presented. Conclusions are reported in the last section.

5.3 ESTIMATION OF LINK TRAVEL TIME DISTRIBUTIONS

Travel time is one of the most important pieces of information for transportation planners, traffic operators, as well as road users. In the transportation planning context, travel time plays an important role in the studies of route choice, OD demand estimation, and other trip making behaviors. Likewise, in most advanced transportation management and information systems (ATMIS) applications, accurate estimates of travel time are required for dynamic route guidance and traveler information systems.

Unfortunately, most traffic surveillance systems are based on point detectors such as single and double inductance loops, and thus direct measurements of section travel time are not commonly available. Although these systems collect traffic counts and occupancies to estimate point speeds, such estimates are prone to significant errors. With

recent advances in vehicle probe technologies, such as Automatic Vehicle Identification (AVI), Global Positioning System (GPS), and cellular phone tracking, vehicle probes are expected to be a valuable real-time traffic data source. Many areas in the United States with toll roads have already used their existing automatic toll collection systems to collect travel time and trajectory data (Dion and Rakha, 2003). Examples of these systems include the TranStar in Houston, the Transmit in the New York/New Jersey metropolitan area, and the Fastrak in the San Francisco bay area. Some universities and government agencies have started distributing GPS devices for traffic data collection and travel survey (see, for example, McNally et al., 2003; Brickal and Bhat, 2006). In addition, the increasing popularity of electronic devices such as car navigation systems and advanced cellular phones with GPS modules is giving more road users the capability to provide and share traffic data.

Although vehicle probes provide high fidelity traffic information, implementation of such data has so far been difficult and limited, mainly due to low market penetration and low sampling rate. Previous studies have focused on the effects of small probe samples on the estimation of mean travel time and shown that the sample mean may not approach population mean as a result of correlation among samples (Hellinga and Fu, 1998; Sen et al., 1997; Oh and Jayakrishnan, 2002). There has also been research on minimum sampling rate and appropriate report frequency to guarantee reliable estimation of mean travel time (see, for example, Smith et al., 2003; Carter, 2000; Cheu et al., 2003). Yanying and Mike (2002) use a fuzzy logic model to classify driving patterns from a probe speed profile which is then used to adjust the estimated travel time.

The word *probe* conventionally refers to a floating vehicle driven by an instructed driver intended to collect information about the average traffic condition along a roadway. In recent years, the word *probe* is also used to refer to an equipped vehicle driven by a general road user which can potentially provide travel time and/or trajectory data. Data from the latter also provides extra information regarding travel time variations due to different driving behaviors, speeds across lanes, mixtures of vehicles, etc. It provides great opportunities to study travel time distributions on roadway sections. Nonetheless, there has been no attempt to use such extra information offered by probes.

This section of the dissertation develops a flexible statistical model that utilizes probe data collected successively over multiple days with low and unknown sampling rates to estimate freeway section travel time distributions. Unlike the traditional method of using historical data which applies the data of the same time of the same day of week or the neighboring day and/or last interval of the same day to an analysis (Chen et al., 2004; Rice and Van Zwet, 2004), the proposed method is able to use the historical data regardless of when the data were collected. The key concept is that geometric characteristics of each freeway section, such as section length, lane width, alignment, pavement condition, speed limit, sign-posting and connected sections, etc., are assumed to result in a unique effect which makes travel times on the section have similar distributions under similar traffic conditions. Such similarity is then used as an essential key to extract information from data collected during different intervals via a hierarchical Bayesian formulation. For each distinct interval, the corresponding travel time distribution is then modeled as a mixture of two normal components intended to capture the effects of other random factors, such as variation of traffic across lanes, mixture of

vehicle types, and differences in driving patterns across all drivers, which could not be observed by probe vehicles. As outputs of our model, each section has dynamic posterior travel time distributions (thus, also means and variances) that correspond to different intervals. The detailed explanation of this model is provided in the following subsections.

5.3.1 Traffic grouping

Critical factors that affect the distributions of freeway section travel times can be grouped into the following four categories:

1. Geometric design: length, lane width, alignment, pavement condition, speed limit, posted signs, connected sections, etc.
2. Traffic condition: flow, speed, and density.
3. Driving pattern mix: lane selection, desired speed, aggressiveness, awareness, etc.
4. Vehicle mix: proportions of trucks, buses, commercial vehicles, commuter vehicles, etc.

Factors in the first two categories are usually observed at some level in practice, but the mixtures of driving patterns and vehicle types are not observable or not usually observed. As the factors in the first category are time-invariant, and the model should be applied independently to each section, the effects from all factors in this category are trivially taken into account. It is then reasonable to expect that for each section, similar traffic conditions result in similar travel time distributions. Further, for each of the

intervals with similar traffic condition, the distribution (mixture) of driving patterns and vehicle types are the primary factors affecting the exact distributional shape.

Link density (or simply the number of vehicles on a section) is usually considered the most informative traffic parameter due to its monotonically non-increasing relationship with the overall speed. Moreover, as observed as early as four decades back (Gafarian et al., 1971; Herman et al., 1972), link density significantly influences the distributional shape of travel times. The similarity in link densities, therefore, can be used as the indicator for grouping traffic conditions. However, because loop detectors do not provide section densities, detector occupancies are used as the criteria for grouping the traffic condition. In attempts to ensure the homogeneity within a group as much as possible, a short time interval (say, 5 minutes) and occupancies from both upstream and downstream detectors are used. In this dissertation, eight occupancy ranges; 0.00-0.10, 0.10-0.15, 0.15-0.20, 0.20-0.25, 0.25-0.30, 0.30-0.35, 0.35-0.40, and > 0.40 are defined at each loop detector. Thus, for each freeway section, there are 64 (8×8) traffic condition groups.

5.3.2 The hierarchical Bayesian mixture model

The requirement that the prior be known is often regarded as a big hurdle for Bayesian analyses and is perhaps the reason why Bayesian analyses are not often used in transportation studies. Empirical Bayesian (EB) and hierarchical Bayesian (HB) models used in this dissertation, however, provide alternative approaches by incorporating the

structural advantages of Bayesian approach while requiring less rigid specifications of prior knowledge. They require only that the prior be in a certain family of distributions, indexed by hyper-parameters that are estimated from the available data. A brief introduction to the general concepts of the Bayesian mixture modeling and hierarchical modeling is given in Appendix A.

Given a traffic condition group, the exact section travel time distribution during a distinct interval is primarily determined by the mixture of driving patterns and mixture of different vehicle types. However, these factors are not easily observable. Even if they can be observed, the small sample sizes and the dynamic nature of traffic make the inference of such factors infeasible. Therefore, it is prudent to simultaneously model such factors by regarding each travel time datum as coming from one of two normal distributions. The first component refers to the *faster component* and the second to the *slower component*. Interpretations of these two components are difficult to define in a strict sense, but generally the first refers to aggressive drivers, drivers on faster lanes, fast vehicles, or simply any combination that result in relatively higher speed, and the second component refers to the opposite. The detailed explanation of the model formulation is given as follows.

The experienced travel time datum from probe z (z = 1, 2, …, Z_{lh}) on link l during distinct interval h when the link operates under the traffic condition group g, $\tilde{\tau}_{zh}^{(gl)}$, is assumed to have come from the faster component whose distribution is $\text{Normal}(\mu_{1h}^{(gl)}, \sigma_1^{(gl)})$ with probability $r_h^{(gl)}$ and from the slower component whose distribution is $\text{Normal}(\mu_{2h}^{(gl)}, \sigma_2^{(gl)})$ with probability $1 - r_h^{(gl)}$. The means of the faster component $\mu_{1h}^{(gl)}$ and the slower component $\mu_{2h}^{(gl)}$ during this time interval h are allowed

to vary from the other time intervals having the same traffic condition group g to reflect the interval-specific random effects. Namely, each of these means is further assumed to have been drawn from common distributions (of link l and traffic group g), $\mu_{1h}^{(gl)} \sim \text{Normal}(\lambda_1^{(gl)}, \alpha_1^{(gl)})$, and $\mu_{2h}^{(gl)} \sim \text{Normal}(\lambda_2^{(gl)}, \alpha_2^{(gl)})$, respectively to take advantage of the similarity among different intervals. Likewise, the interval-specific mixing proportion, $r_h^{(gl)}$, which acts as a hyper-parameter in the Bernoulli distribution of the unobservable component indicator variable $\xi_{zh}^{(gl)}$, is allowed to vary across different intervals to capture the variation of the proportions of the faster and slower components. Note that the interval-specific variances can actually vary across different intervals as well, but to reduce the number of parameters and the computational burden, it is assumed that these variances are the same across intervals (within the same traffic condition group). The structure of the hierarchical mixture model can be specified below.

$$\tilde{\tau}_{zh}^{(gl)} \mid \mu_{1h}^{(gl)}, \mu_{2h}^{(gl)}, \xi_{zh}^{(gl)}, \phi \sim \xi_{zh}^{(gl)} \text{Normal}(\mu_{1h}^{(gl)}, \sigma_1^{(gl)}) + (1 - \xi_{zh}^{(gl)}) \text{Normal}(\mu_{2h}^{(gl)}, \sigma_2^{(gl)}) \qquad (5\text{-}1)$$

$$\mu_{1h}^{(gl)} \mid \xi, \phi \sim \text{Normal}(\lambda_1^{(gl)}, \alpha_1^{(gl)})$$

$$\mu_{2h}^{(gl)} \mid \xi, \phi \sim \text{Normal}(\lambda_2^{(gl)}, \alpha_2^{(gl)})$$

$$\xi_{zh}^{(gl)} \mid \phi \sim \text{Bernoulli}(r_h^{(gl)})$$

where $\phi = \left(\lambda_1^{(gl)}, \lambda_2^{(gl)}, \sigma_1^{(gl)}, \sigma_2^{(gl)}, \alpha_1^{(gl)}, \alpha_2^{(gl)}, r_h^{(gl)}\right)$, and $\xi_{zh}^{(gl)}$ is equal to 1 or 0 if the travel time of probe z is from the faster component or the slower component, respectively. $h = 1, 2, ..., H_g$, and $z = 1, 2, ..., Z_{hl}$.

To avoid a parameter identification problem where travel time datum from the faster component is confused with the one from the slower component with negative travel time, it is useful to define a new parameter, $\mu_{sh}^{(gl)}$, such that $\mu_{2h}^{(gl)}$ is decomposed into $\mu_{1h}^{(gl)} + \mu_{sh}^{(gl)}$. The distribution of $\mu_{sh}^{(gl)}$ is also normal but now with the parameter space limited to positive.

$$\tilde{\tau}_{zh}^{(gl)} \mid \mu_{1h}^{(gl)}, \mu_{sh}^{(gl)}, \xi_{zh}^{(gl)}, \phi \sim \text{Normal}(\mu_{1h}^{(gl)} + (1-\xi_{zh}^{(gl)})\mu_{sh}^{(gl)}, \xi_{zh}^{(gl)}\sigma_1^{(gl)} + (1-\xi_{zh}^{(gl)})\sigma_s^{(gl)}) \quad (5\text{-}2)$$

$$\mu_{1h}^{(gl)} \mid \xi, \phi \sim \text{Normal}(\lambda_1^{(gl)}, \alpha_1^{(gl)})$$

$$\mu_{sh}^{(gl)} \mid \xi, \phi \sim \text{Normal}(\lambda_s^{(gl)}, \alpha_s^{(gl)}) \text{ (truncated at 0)}$$

$$\xi_{zh}^{(gl)} \mid \phi \sim \text{Bernoulli}(r_h^{(gl)})$$

where $\phi = \left(\lambda_1^{(gl)}, \lambda_s^{(gl)}, \sigma_1^{(gl)}, \sigma_s^{(gl)}, \alpha_1^{(gl)}, \alpha_s^{(gl)}, r_h^{(gl)}\right)$. Prior distributions must be given to all parameters in the highest hierarchies. A flat (diffuse) prior is assigned to reflect the lack of knowledge of parameter values with a relatively wide parameter space to ensure that prior distributions cover the parameter space of the likelihood as follows.

$$r_h^{(gl)} \sim \text{Uniform}(0.01, 0.99) \quad (5\text{-}3)$$

$$\sigma_1^{(gl)}, \sigma_s^{(gl)}, \alpha_1^{(gl)}, \alpha_s^{(gl)} \sim \text{Uniform}(0, \upsilon)$$

$$\lambda_1^{(gl)} \sim \text{Uniform}(0, \beta_1)$$

$$\lambda_s^{(gl)} \sim \text{Uniform}(0, \beta_s)$$

The range of $(0.01, 0.99)$ is used in the uniform prior distribution for the mixing proportion instead of $(0.00, 1.00)$ to prevent the situation where all data are thought of as belonging to one component.

To choose values for υ, β_1, and β_s, data from each interval with the same traffic condition group are crudely clustered via a k-means clustering algorithm (with k =2). Mean and standard deviation for each component of each interval are then calculated. The maximum standard deviation, maximum mean of the first component, and maximum difference between the means of the two components, are doubled in magnitude and used for υ, β_1, and β_s, respectively.

The model yields an estimated section travel time distribution during each distinct interval h, as explained by $\mu_{1h}^{(gl)}$, $\mu_{2h}^{(gl)}$, $\sigma_1^{(gl)}$, $\sigma_2^{(gl)}$, and $r_h^{(gl)}$. In addition, the parameters $\lambda_1^{(gl)}$, $\lambda_s^{(gl)}$, $\sigma_1^{(gl)}$, $\sigma_2^{(gl)}$, and $\bar{r}$ (average mixing proportion) can be used to construct another travel time distribution curve to be considered as the common section travel time distribution for all intervals within the same traffic condition group g. Figure 5-2 summarizes the format of each single section outputs.

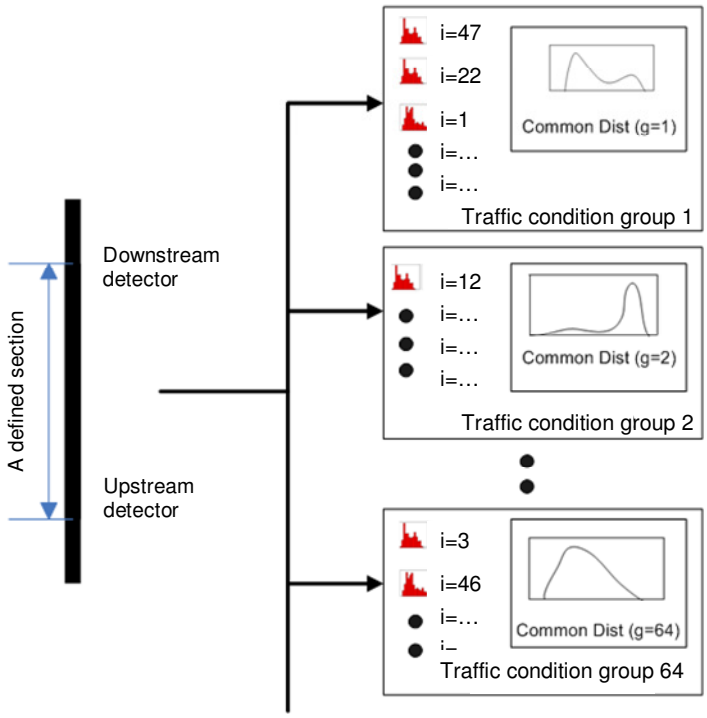


Figure 5-2 Diagram of the model outputs for each freeway section

In addition to travel time distributions, the mean travel time of each specific time interval h and of each traffic condition group g can be calculated from the equations (5-4) and (5-5), respectively.

$$\text{Mean travel time on link } l \text{ during the interval } h = r_h^{(gl)}\mu_{1h}^{(gl)} + (1-r_h^{(gl)})\mu_{2h}^{(gl)} \qquad (5\text{-}4)$$

$$\text{Mean travel time on link } l \text{ under traffic condition group } g = \bar{r}\lambda_1^{(gl)} + (1-\bar{r})\lambda_2^{(gl)} \qquad (5\text{-}5)$$

Note that even though the proposed mixture model is based on standard distributions, the inference on posterior distributions of model parameters is analytically intractable (McLachlan and Peel, 2000; Gelman et al., 2003). Fortunately, with new computational technologies, it has been recently possible to estimate the mixture model

via Markov Chain Monte Carlo (MCMC) algorithms. The Gibbs sampler, a particular MCMC algorithm, was first applied to a mixture model in Diebolt and Robert (1994). This is an efficient algorithm widely used for generating a sequence of sample draws from an unknown joint posterior distribution. In each iteration of the algorithm, a sample of each parameter is alternately drawn from the conditional posterior distribution, given the most recent draws of other parameters. If the sequence is long enough, it can be used to estimate the joint distribution. In this dissertation, the Gibbs sampler technique is carried out using WinBUGS (Lunn et al., 2000).

5.3.3 Model evaluation

Study site

Though probe data are available from many recent field studies, 100 percent samples of the kind required to validate the proposed model are not available. Thus, a microscopic traffic simulation model, Paramics, is used. As shown in Figure 5-3, the study network is a 6-mile stretch of I-405 freeway in Orange County, California. The network and 15-minute OD demands for the morning peak from 6:30 A.M. to 9:00 A.M. of a typical weekday have been calibrated using observed data on May 22, 2001 (Chu et al., 2004).

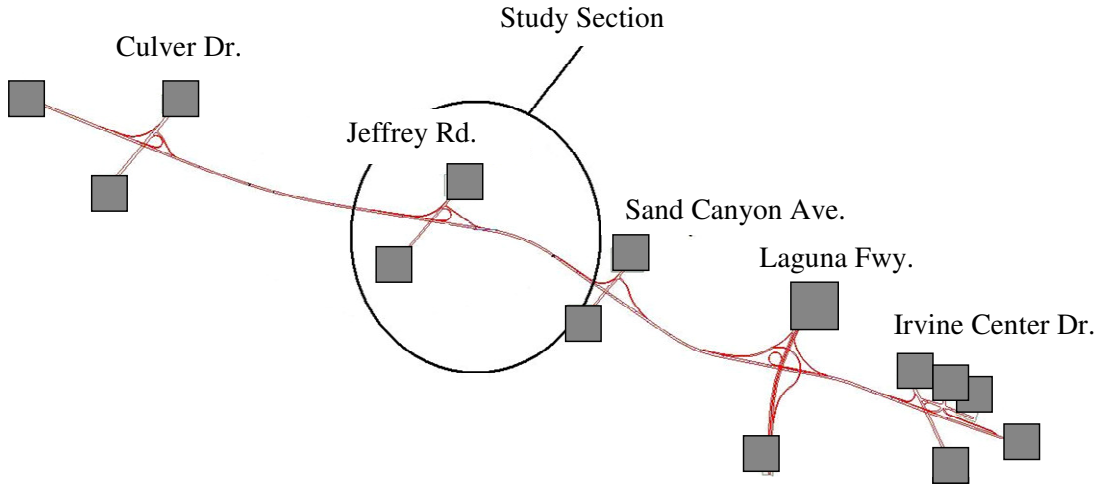


Figure 5-3 A 6-mile section of I-405 in Irvine

Traffic simulation

The simulation study aims to investigate how the model is fitted to the initial dataset from a setting where a small number of probe travel times can be collected/accessed during 7:00 A.M to 9:30 A.M. each day for 20 days (i.e., 2 hours 30 minutes each day, a total of 50 hours of data collection). To reflect variations in OD demand and small probe market penetration, each OD pair demand is allowed to vary from day to day within 15 percent of its calibrated OD demand value, and only 3 percent of all vehicles in the network are sampled as probes.

Depending on the time of arrival at each section, each travel time data is attached with the arrival time interval and the 5-minute detector occupancies from upstream and

downstream detectors indicating the traffic condition group of the data point, as defined in the subsection 5.3.1.

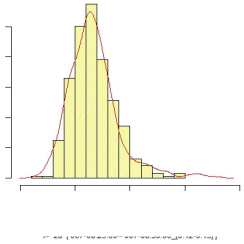
Inference on the posterior distributions and results

An R (R Development Core Team, 2005) code is written to calculate necessary statistics and successively call WinBUGS to run the Gibbs sampler. Each WinBUGS call carries out inferences on posterior distributions of parameters that explain section travel time distributions, which correspond to different distinct intervals, under one traffic condition group. Therefore, for each section s, WinBUGS is called ng_s times (ng_s = the number of traffic condition groups experienced by section s in the initial dataset). This study sets the number of Markov chains to two, each of which includes 3000 simulations, and the posterior distributions of model parameters are taken from the last 1500 simulations of each chain.

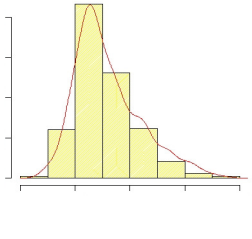
The results and analyses of one selected section in the extensive network, from the Sand Canyon on-ramp (post-mile 3.04) to the Jeffrey on-ramp (post-mile 4.03), are presented. During the twenty morning peaks, this section experienced 43 (out of 64) traffic condition groups. Figure 5-4 shows the travel time distributions of some specific intervals under some experienced traffic condition groups. In the figure, there are eight rows corresponding to eight different groups: 10, 19, 27, 31, 35, 37, 38, and 39, respectively, chosen to be displayed so that there are plots for a broad range of traffic from light to heavy. In each row, three examples of travel time distributions compared to

the histogram of the corresponding 100 percent probe sample are shown. The label below each subfigure indicates the corresponding date, time interval, and detector occupancies at upstream and downstream.

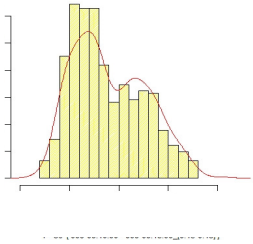
(a) Traffic Condition Group 10: OccUp(0.10-0.15) – OccDown(0.10-0.15)



Day 7-8:25-8:30, (0.12)-(0.13)

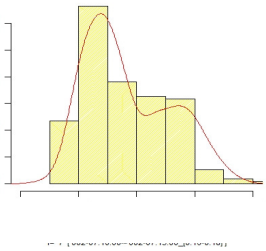


Day 1-8:25-8:30, (0.14)-(0.11)

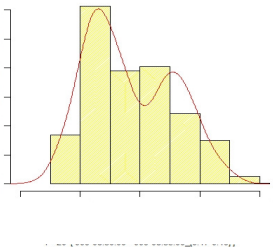


Day 9-9:10-9:15, (0.15)-(0.13)

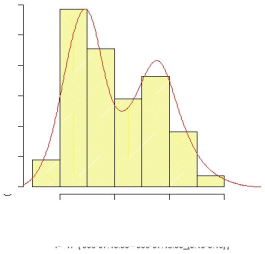
(b) Traffic Condition Group 19: OccUp(0.15-0.20) – OccDown(0.15-0.20)



Day 2-7:10-7:15, (0.16)-(0.18)

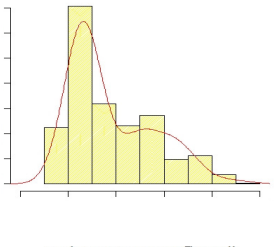


Day 9-8:30-8:35, (0.17)-(0.19)

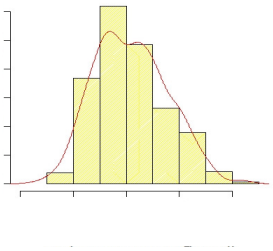


Day 6-7:10-7:15, (0.19)-(0.16)

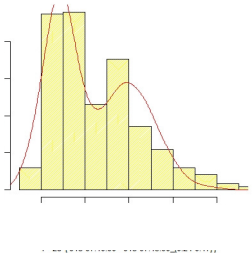
(c) Traffic Condition Group 27: OccUp(0.20-0.25) – OccDown(0.15-0.20)



Day 12-7:00-7:05, (0.23)-(0.15)

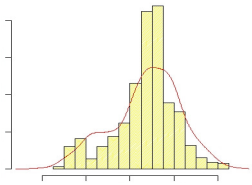


Day 14-7:25-7:30, (0.24)-(0.16)

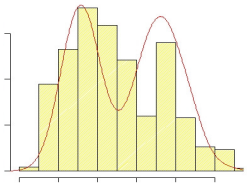


Day 13-7:10-7:15, (0.24)-(0.17)

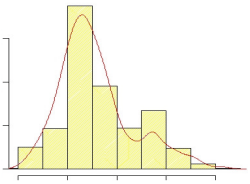
(d) Traffic Condition Group 31: OccUp(0.20-0.25) – OccDown(0.35-0.40)



Day 2-8:05-8:10, (0.22)-(0.38)

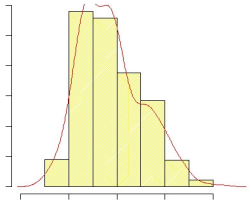


Day 2-7:45-7:50, (0.24)-(0.38)

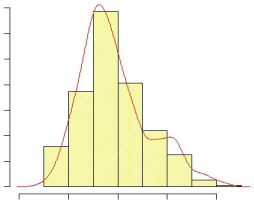


Day 11-7:30-7:35, (0.25)-(0.36)

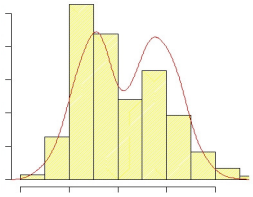
(e) Traffic Condition Group 35: OccUp(0.25-0.30) – OccDown(0.15-0.20)



Day 3-7:25-7:30, (0.26)-(0.17)

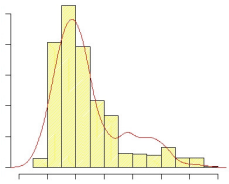


Day 13-9:15-9:20, (0.28)-(0.15)

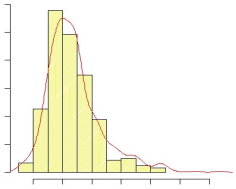


Day 8-7:05-7:10, (0.28)-(0.16)

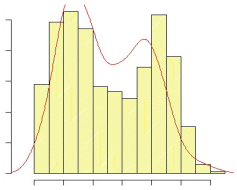
(f) Traffic Condition Group 37: OccUp(0.25-0.30) – OccDown(0.25-0.30)



Day 4-7:30-7:35, (0.25)-(0.26)

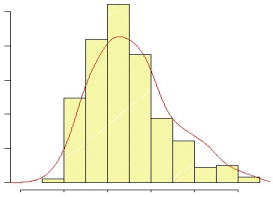


Day 15-7:20-7:25, (0.28)-(0.26)

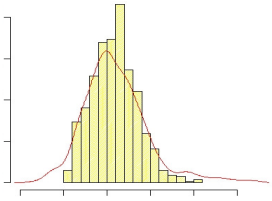


Day 17-7:05-7:10, (0.28)-(0.28)

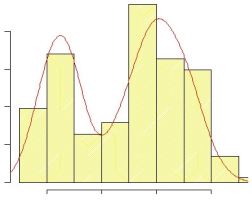
(g) Traffic Condition Group 38: OccUp(0.25-0.30) – OccDown(0.30-0.35)



Day 7-7:30-7:35, (0.28)-(0.30)

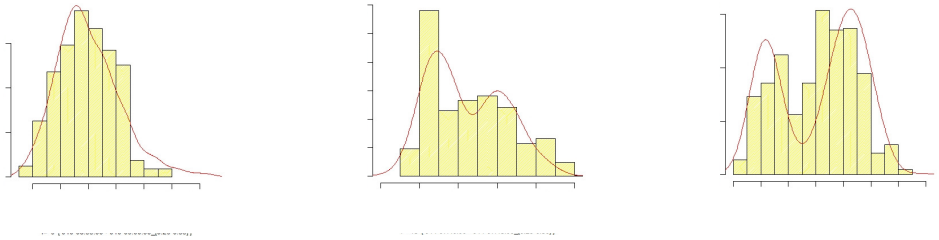


Day 20-8:35-8:40, (0.29)-(0.32)



Day 12-7:40-7:45, (0.29)-(0.33)

(h) Traffic Condition Group 39: OccUp(0.25-0.30) – OccDown(0.35-0.40)



Day 10-8:55-9:00, (0.26)-(0.35) Day 14-7:40-7:45, (0.29)-(0.36) Day 8-8:05-8:10, (0.29)-(0.38)

Figure 5-4 Examples of posterior travel time distributions on section 405n3.31ml-405n4.03ml under various traffic condition groups

It can be seen that the proposed model can fit different distributions quite well. Two main conclusions are made as follows:

1. Within the same traffic condition group, each distinct interval has its own travel time distribution that may be different from others in terms of shape, mean, mixing proportion, and data range. This can be thought of as the effect of different mixtures of driving patterns and vehicle types.
2. The travel time distribution on a section tends to be skewed-unimodal under light traffic, while under medium and heavy traffic conditions, the travel time distribution can be either skewed-unimodal or multimodal.

Furthermore, it is possible to plot another set of travel time distributions for different traffic condition groups based on the parameters in the second hierarchy. Each of these curves is the posterior distribution corresponding to a particular combination of upstream and downstream detector occupancy ranges which can be used in two ways -- to crudely predict travel time distribution on a section based solely on detectors' occupancy ranges or to combine with new travel time data in estimating travel time distribution for a new

interval. Instead of plotting all 43 curves, only 4 curves are shown in Figure 5-5 for traffic condition groups 10, 35, 38, and 39.

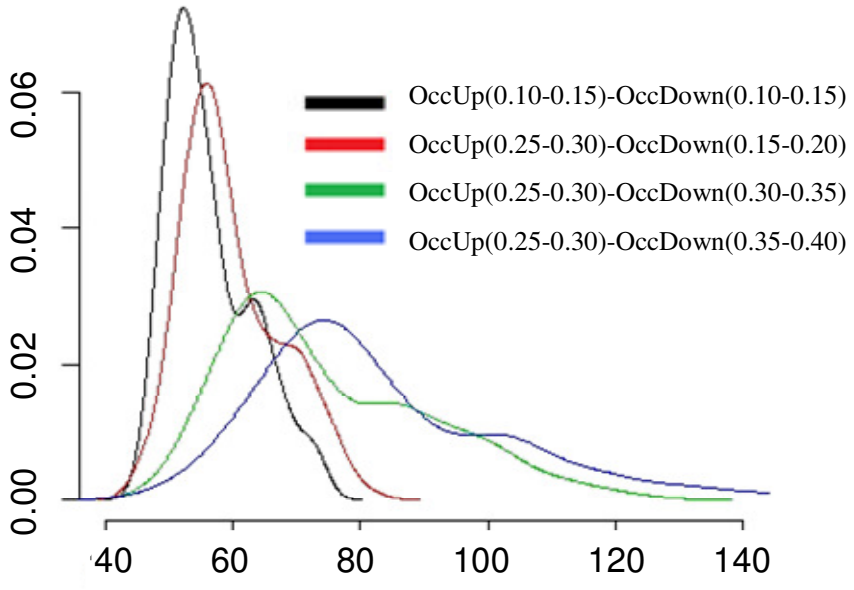


Figure 5-5 Examples of posterior travel time distributions on section 405n3.31ml-405n4.03ml for different traffic condition groups

Lastly, Figure 5-6 shows an example of the cumulative probability plots which can be used in calculating the total travel time of all drivers on a section during specific time interval and/or accessing the proportion of all drivers whose experienced travel time is greater than some specific threshold. In the figure, the cumulative plots during 7:40 A.M.-7:45 A.M. on day 12th, from 100 percent sample, 3 percent sample, and the proposed model are marked by "-", "s", and "+" respectively. It can be seen that the model can give a close approximation to the true line.

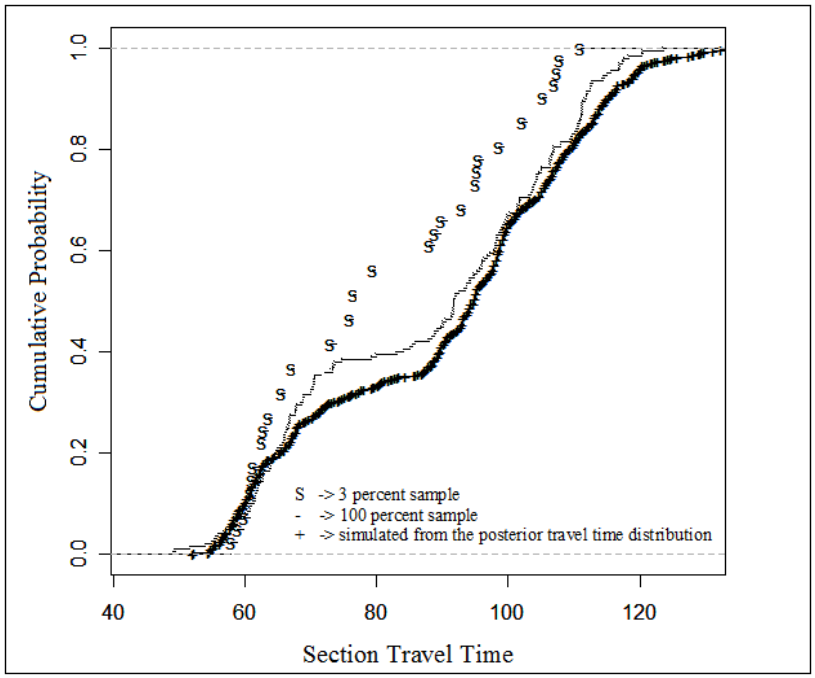


Figure 5-6 Example of cumulative probability (7:40 A.M.-7:45 A.M. - day 12th)

Numerical Examples

Table 5-1 demonstrates the evaluation of model performance for estimating the average travel time for each 5-minute interval during 7:00 A.M. to 8:30 A.M of the first day. Such estimation (column c3) can be done using equation (5-4) or directly calculating the mean value from a large number of draws from the posterior travel time distribution.

Table 5-1 Comparison of true mean speeds with simple averages from 3 percent probe samples, and estimated means from the proposed model

Time Period (i)	Date-Time	Traffic Condition Group (g)	(c1) 100 percent probe		(c2) 3 percent probe		(c3) Estimated	
			Mean	S.d.	Mean	S.d.	Mean	S.d.
1	(1)-7:00:00-7:05:00	18	56.57	6.58	54.29	3.04	56.13	6.20
2	(1)-7:05:00-7:10:00	19	58.85	8.72	58.49	9.41	59.84	8.52
3	(1)-7:10:00-7:15:00	19	58.62	8.17	60.49	8.26	59.89	7.59
4	(1)-7:15:00-7:20:00	19	57.17	6.84	58.34	7.26	58.25	6.98
5	(1)-7:20:00-7:25:00	19	59.00	7.14	58.58	6.65	58.56	6.97
6	(1)-7:25:00-7:30:00	27	58.46	6.19	60.69	9.50	57.70	6.64
7	(1)-7:30:00-7:35:00	19	60.49	9.12	57.35	5.14	60.44	8.02
8	(1)-7:35:00-7:40:00	27	59.63	7.91	59.91	9.48	60.90	8.92
9	(1)-7:40:00-7:45:00	30	85.39	27.40	66.16	12.70	76.69	20.90
10	(1)-7:45:00-7:50:00	30	67.37	12.29	77.19	22.79	66.29	14.24
11	(1)-7:50:00-7:55:00	30	62.19	9.42	62.07	7.92	62.36	10.68
12	(1)-7:55:00-8:00:00	30	99.50	16.80	96.63	15.72	95.20	15.53
13	(1)-8:00:00-8:05:00	37	99.46	20.51	94.24	25.59	95.31	17.19
14	(1)-8:05:00-8:10:00	37	68.53	29.90	70.90	15.33	71.03	15.05
15	(1)-8:10:00-8:15:00	37	64.87	26.84	63.38	13.08	63.94	12.14
16	(1)-8:15:00-8:20:00	38	64.87	31.06	78.46	19.52	79.75	16.96
17	(1)-8:20:00-8:25:00	38	88.35	26.64	84.28	16.70	85.15	12.95
18	(1)-8:25:00-8:30:00	38	70.10	24.67	66.82	10.38	67.86	10.29
Mean Absolute Relative Error (MARE)					0.0497		0.0281	
Root Relative Square Error (RRSE)					0.0781		0.0411	
Maximum Relative Error (MRE)					0.2251		0.1018	

Based on MARE, RRSE, and MRE, the proposed model performs about 44 percent, 47 percent, and 55 percent respectively better than the simple average method (column c2). Such improvements are more obvious especially for time intervals with medium and heavy traffic conditions (see time intervals 9, 10 for example). This is mainly attributed to the model's ability to capture travel time variations.

Travel time prediction can be done under different levels of data availability. Here, it is helpful to evaluate the model performance for travel time prediction under the scenario where no probe data are available on new time intervals but the exact occupancies from upstream and downstream detectors are known (say, from prediction or last time step). A new day is simulated, and the average travel time for each 5-minute interval from 8:00 to 9:00 A.M. is calculated. The prediction is then done based on the posterior travel time distributions corresponding to the occupancy combinations (i.e. from upstream and downstream detector) closest to those in the new day. Table 5-2 shows the

results and evaluation of the model. Based on all three indices, the model gives very encouraging results.

Table 5-2 Comparison of true mean speeds with predicted means from the proposed model using parameters from the look up table

Time Period (i)	Date-Time	Traffic Condition Group (g)	(c1) 100 percent probe Mean	(c2) Predicted Mean
1	(New)-8:00:00-8:05:00	27	57.90	61.23
2	(New)-8:05:00-8:10:00	27	58.46	57.21
3	(New)-8:10:00-8:15:00	30	60.96	57.31
4	(New)-8:15:00-8:20:00	27	59.63	60.79
5	(New)-8:20:00-8:25:00	30	67.37	70.62
6	(New)-8:25:00-8:30:00	30	61.14	62.18
7	(New)-8:30:00-8:35:00	30	62.19	61.43
8	(New)-8:35:00-8:40:00	31	89.28	91.67
9	(New)-8:40:00-8:45:00	30	96.84	98.33
10	(New)-8:45:00-8:50:00	30	85.39	78.06
11	(New)-8:50:00-8:55:00	31	89.02	92.33
12	(New)-8:55:00-9:00:00	30	99.45	95.35
Mean Absolute Relative Error (MARE)				0.0368
Root Relative Square Error (RRSE)				0.0431
Maximum Relative Error (MRE)				0.0858

The hierarchical Bayesian mixture model for estimating section travel time distributions presented in this section shows very promising results. Three major contributions are made as follows.

- Based on the finite mixture formulation, the proposed model can successfully fit various distributional shapes of section travel times under different traffic conditions. The model is able to capture both skewedness and multimodal shapes found in travel time data distributions.
- An occupancy-based criterion for identifying similarities in travel times is developed, so that information from probes across different time intervals can be maximally used. Through hierarchical Bayesian approach, the model can be used with a series of small samples without imposing an exchangeability assumption of travel time data across different intervals but rather assuming that component

means of travel times during different intervals, under similar traffic condition, are drawn from common distributions.

- The proposed model is implementable in a practical setting, and the required probe sample sizes are also not impractical, considering the number of vehicles with location devices such as GPS expected in the near-future.

All travel time distributions of experienced traffic conditions can be stored in a look up table. For online estimation and prediction, a proper set of parameters in the table can be chosen as the prior to combine with the likelihood from new data. In offline mode, this table can be updated (simultaneously with the estimation for new travel time distributions for new intervals) as more data become available.

It is important to note that some other observable factors affecting travel time distributions not mentioned in the section, such as weather conditions (e.g., rain, snow, fog, light), and proportions of trucks in different time intervals of day, can be easily considered. That is, besides link occupancies, such additional constraints can be used for grouping data when estimating model parameters. Also, differences are expected in the model performance when applied to the real world data which tend to be more stochastic (thus greater variance) than data generated from a microscopic traffic simulation model. However, the ability of the model to fit various distributional shapes should remain largely valid, albeit perhaps with slightly higher sample rates than used in this study.

5.4 ESTIMATION OF ROUTE FRACTIONS

With sampled trajectories, several behavioral route-choice models have been recently developed in the literature. For instance, Cascetta et al. (2002), Ben-Akiva and Bierlaire (2003) adopt the random utility concept to model individuals' decision on choosing routes. This approach can take into account topological, level of service, and user socio-economic attributes. However, it requires expensive individual data to estimate the model parameters, and its transferability can be questionable.

This section of the dissertation presents a Bayesian multinomial-Dirichlet model, which utilizes solely observed trajectories, to estimate route fractions. The key idea is that whether or not the underlying route pattern is in equilibrium, it is reasonable to expect a certain recursive route pattern from day-to-day. The method combines the multinomial likelihood from small trajectory samples during the target period with the Dirichlet prior whose parameters are empirically derived from the data from multiple days. This way, extreme multinomial estimates will be moderated toward the overall mean, with the magnitude of adjustment based on their variance. The choice of the Dirichlet prior made here is only for mathematical convenience and should not be regarded as an overly restrictive assumption.

Consider the OD flow from origin i to destination j departing during interval t. It is assumed that all routes k observed during the interval over multiple days (say, same day of week, or neighboring days) are enumerated in the exhaustive route set, K^{ijt}. This assumption, although seemingly restrictive, is to a great extent supported by an empirical finding in Cascetta et al. (2002) that only few routes are actually perceived by users.

Moreover, if the data collection is deployed continuously over time, the chance of observing all commonly used routes becomes higher. An overview of other route-set generation algorithms based on the *k*-shortest path, labeling, and simulation methods can be found in Bekhor et al. (2001).

5.4.1 The empirical Bayesian Multinomial-Dirichlet model

Denote the vector of route-choice fractions for all routes belonging to OD pair *ij* departing during interval t on the target day m by $\mathbf{q}_m^{ijt} = (q_m^{1t},..., q_m^{K^{ijt}t})$. The space of these K^{ijt} - dimensional multinomials is a $(K^{ijt}-1)$-simplex, and the Dirichlet prior given over this simplex is

$$\text{Dirichlet}(\mathbf{q}_m^{ijt} \mid \alpha_1^{ijt}, \alpha_2^{ijt}, ..., \alpha_{K^{ijt}}^{ijt}) = \frac{\Gamma(\alpha_0^{ijt})}{\prod_{k=1}^{K^{ijt}} \Gamma(\alpha_k^{ijt})} \prod_{k=1}^{K^{ijt}} (q_m^{kt})^{\alpha_k^{ijt}-1} \tag{5-6}$$

where $\alpha_k^{ijt}, \forall k$ = parameters of the Dirichlet prior, α_0^{ijt} = a concentration parameter = $\sum_{k=1}^{K^{ijt}} \alpha_k^{ijt}$, $q_m^{1t} > 0, .., q_m^{K^{ijt}t} > 0$, and $\sum_{k=1}^{K^{ijt}} q_m^{kt} = 1$.

As the Dirichlet prior is a conjugate prior for the multinomial distribution, the posterior distribution of $\mathbf{q}_m^{ijt}$ is also Dirichlet (over the simplex of the multinomials)

$$\text{Dirichlet}(\mathbf{q}_m^{ijt} \mid \alpha_1^{ijt}, .., \alpha_{K^{ijt}}^{ijt}, \tilde{f}_m^{1t}, ..., \tilde{f}_m^{K^{ijt}t}, \tilde{x}_m^{ijt}) = \frac{\Gamma(\omega_{m0}^{ijt})}{\prod_{k=1}^{K^{ijt}} \Gamma(\omega_{mk}^{ijt})} \prod_{k=1}^{K^{ijt}} (q_m^{kt})^{\omega_{mk}^{ijt}-1} \tag{5-7}$$

where $\tilde{f}_m^{kt}$ = observed trajectory flow on route k during interval t on the target day m, $\tilde{x}_m^{ijt}$ = observed trajectory flow between origin i and destination j during interval t on the target day m, $\omega_{mk}^{ijt} = \alpha_k^{ijt} + \tilde{f}_m^{kt}$, and $\omega_{m0}^{ijt} = \sum_{k=1}^{K^{ijt}} \alpha_k^{ijt} + \tilde{x}_m^{ijt}$.

It follows that the posterior mean (the Bayesian estimator minimizing the mean square error loss function), variance, and covariance of the route choice fractions can be easily calculated as follows.

$$\hat{q}_m^{kt} = \omega_{mk}^{ijt} / \omega_{m0}^{ijt} \tag{5-8}$$

$$\text{VAR}(\hat{q}_m^{kt}) = \sigma_{\hat{q}_m^{kt}}^2 = \left[\omega_{mk}^{ijt}.(\omega_{m0}^{ijt} - \omega_{mk}^{ijt})\right] / \left[(\omega_{m0}^{ijt})^2.(\omega_{m0}^{ijt}+1)\right] \tag{5-9}$$

$$\text{COV}(\hat{q}_m^{kt}, \hat{q}_m^{lt}) = \left[-\omega_{mk}^{ijt}.\omega_{ml}^{ijt}\right] / \left[(\omega_{m0}^{ijt})^2.(\omega_{m0}^{ijt}+1)\right] \tag{5-10}$$

The parameters of the Dirichlet prior α_k^{ijt}'s can be estimated by maximizing the log-likelihood function of the marginal multinomial-dirichlet distribution based empirically on a set of sampled route fractions $\boldsymbol{\theta}_d^{ijt}$'s from multiple days, $D_{ijt}^{route} = \{\boldsymbol{\theta}_1^{ijt}, \boldsymbol{\theta}_2^{ijt}, ..., \boldsymbol{\theta}_m^{ijt}, ...\}$. Several efficient methods, such as the gradient ascent, fixed point iteration, and Newton-Raphson methods, for numerically maximizing such objective function are well documented and thus will not be discussed here.

5.4.2 Numerical example

Consider a simple network with one OD pair, as shown in Figure 5-7. In the network, there are five routes from the origin zone to the destination zone.

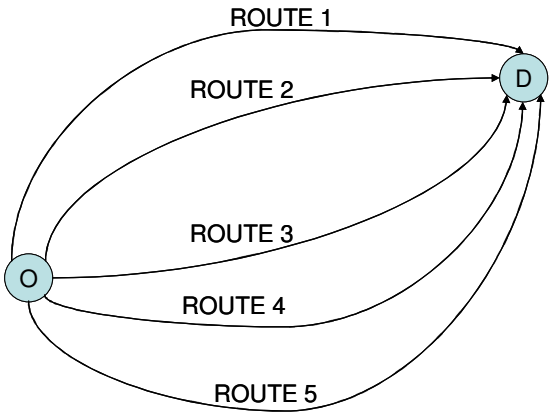


Figure 5-7 A simple network with one OD pair and five routes

In order to test the proposed method with a set of small trajectory samples collected from multiple days, simulated ground truth of route fractions during a particular interval for 21 days is generated (see Table 5-3). For a moment, let us ignore the underlying behavior used in generating these fractions and focus solely on how the proposed method can improve the estimates when compared to the raw estimates using only one-day data. More elaborate simulation studies will be presented in CHAPTER 7.

Table 5-3 Ground true route fractions for 21 days

	Route 1	Route 2	Route 3	Route 4	Route 5
Day 1	0.714	0.286	0.000	0.000	0.000
Day 2	0.556	0.222	0.111	0.111	0.000
Day 3	0.500	0.250	0.000	0.250	0.000
Day 4	0.857	0.143	0.000	0.000	0.000
Day 5	0.857	0.000	0.000	0.000	0.143
Day 6	0.500	0.375	0.125	0.000	0.000
Day 7	0.714	0.000	0.143	0.143	0.000
Day 8	0.714	0.143	0.000	0.143	0.000
Day 9	0.667	0.000	0.000	0.333	0.000
Day 10	0.429	0.000	0.429	0.143	0.000
Day 11	0.750	0.000	0.000	0.250	0.000
Day 12	0.500	0.125	0.125	0.125	0.125
Day 13	0.750	0.125	0.125	0.000	0.000
Day 14	0.625	0.250	0.000	0.125	0.000
Day 15	0.714	0.000	0.000	0.286	0.000
Day 16	0.778	0.000	0.222	0.000	0.000
Day 17	0.857	0.000	0.000	0.143	0.000
Day 18	0.857	0.143	0.000	0.000	0.000
Day 19	0.875	0.125	0.000	0.000	0.000
Day 20	0.500	0.167	0.167	0.000	0.167
Day 21	0.556	0.222	0.111	0.111	0.000

As shown in the table, these simulated underlying fractions fluctuate from day-to-day but show roughly similar patterns. Moreover, some routes have very small shares, such that they are not being used every day.

Assume that a total of 100 vehicles depart from the origin zone during this interval each day, from which it is possible to sample 5 vehicles. Table 5-4 shows a set of sampled route flows during the 21 days, based on a random sampling with 5 percent sampling rate. It is noteworthy that the knowledge of sampling rate is not required in applying the proposed model.

Table 5-4 Sampled route flows for 21 days

	Route 1	Route 2	Route 3	Route 4	Route 5
Day 1	3	2	0	0	0
Day 2	3	1	1	0	0
Day 3	1	3	0	1	0
Day 4	5	0	0	0	0
Day 5	5	0	0	0	0
Day 6	3	2	0	0	0
Day 7	2	0	1	2	0
Day 8	4	1	0	0	0
Day 9	2	0	0	3	0
Day 10	2	0	3	0	0
Day 11	4	0	0	1	0
Day 12	4	0	1	0	0
Day 13	4	0	1	0	0
Day 14	3	0	0	2	0
Day 15	4	0	0	1	0
Day 16	3	0	2	0	0
Day 17	5	0	0	0	0
Day 18	5	0	0	0	0
Day 19	4	1	0	0	0
Day 20	3	0	1	0	1
Day 21	1	3	0	1	0

Table 5-5 shows raw estimated fractions based only on the observation on each day. It can be seen that these estimates deviate greatly from the underlying ones. Moreover, there are many cells with the estimate of zero value.

Table 5-5 Raw-estimated route fractions for 21 days

	Route 1	Route 2	Route 3	Route 4	Route 5
Day 1	0.600	0.400	0.000	0.000	0.000
Day 2	0.600	0.200	0.200	0.000	0.000
Day 3	0.200	0.600	0.000	0.200	0.000
Day 4	1.000	0.000	0.000	0.000	0.000
Day 5	1.000	0.000	0.000	0.000	0.000
Day 6	0.600	0.400	0.000	0.000	0.000
Day 7	0.400	0.000	0.200	0.400	0.000
Day 8	0.800	0.200	0.000	0.000	0.000
Day 9	0.400	0.000	0.000	0.600	0.000
Day 10	0.400	0.000	0.600	0.000	0.000
Day 11	0.800	0.000	0.000	0.200	0.000
Day 12	0.800	0.000	0.200	0.000	0.000
Day 13	0.800	0.000	0.200	0.000	0.000
Day 14	0.600	0.000	0.000	0.400	0.000
Day 15	0.800	0.000	0.000	0.200	0.000
Day 16	0.600	0.000	0.400	0.000	0.000
Day 17	1.000	0.000	0.000	0.000	0.000
Day 18	1.000	0.000	0.000	0.000	0.000
Day 19	0.800	0.200	0.000	0.000	0.000
Day 20	0.600	0.000	0.200	0.000	0.200
Day 21	0.200	0.600	0.000	0.200	0.000

Assume now that it is desired to estimate route fractions on the 21st day. The proposed model can be applied as follows. The set of sampled fractions $\boldsymbol{\theta}_d^{ijt}$'s from 21 days (Table 5-5) is first used to estimate the parameters of the Dirichlet prior α_k^{ijt}'s. Then, the posterior mean fractions for the target day can be calculated using equation (5-8).

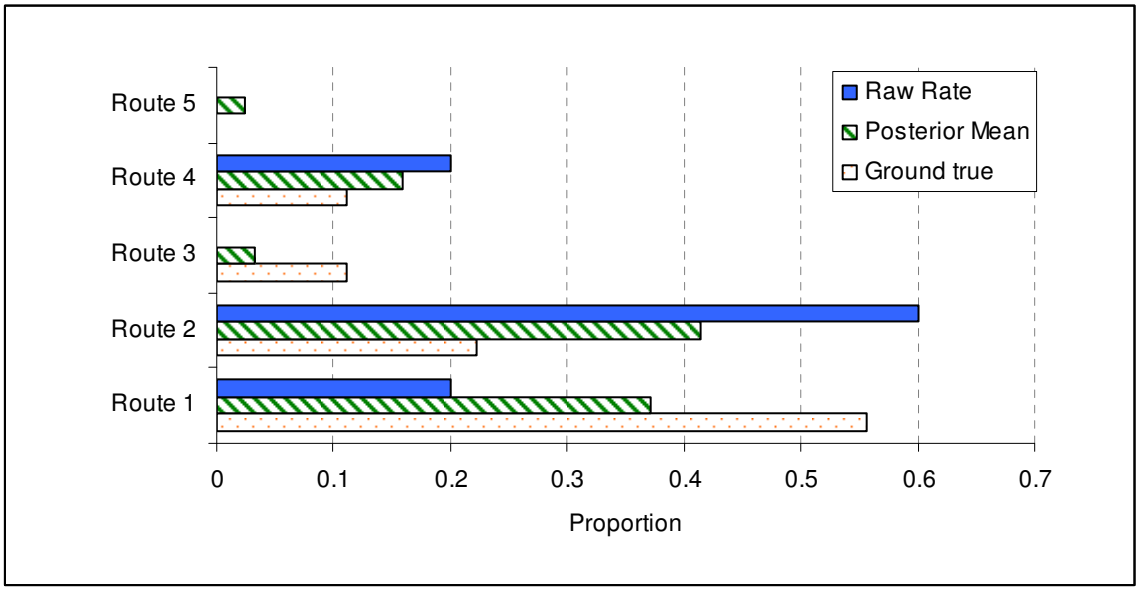


Figure 5-8 Comparison of ground true route fractions with posterior (mean) route fractions from the proposed model, and raw route fractions based on the data from the target day

Figure 5-8 compares the ground true route fractions on the 21st day with the corresponding posterior mean fractions from the proposed model, and the raw estimates based on one-day data. From the comparison, the benefits from the proposed model can be easily seen. First, the posterior mean fractions on routes 1, 2, and 4 are much closer to the ground truth when compared to the raw estimates which seem to be extreme due to the small sample. Second, although no vehicle using route 3 is observed on the target day and the corresponding raw estimate is zero, the proposed model yields a small estimate on this route.

On the other hand, while the underlying fraction on route 5 is zero, the model results in a small fraction of 0.023. This is due to the Dirichlet prior derived from the data from multiple days, some of which include samples on the route. An obvious question is then how such behavior might affect further OD estimation steps. To answer this question, it is helpful to plot the coefficients of variation (CV) of the estimates, calculated

based on equation (5-9). Figure 5-9 indicates that the route with the highest CV value is route 5 followed by route 3; and that route 1, 2, and 4 have relatively much lower CV values. These statistics correspond to the comparison presented above. Namely, low CV values indicate high consistency between the prior and the likelihood from the data, and vice versa. Therefore, further OD estimation model steps should incorporate some measure of variation from the posterior distribution.

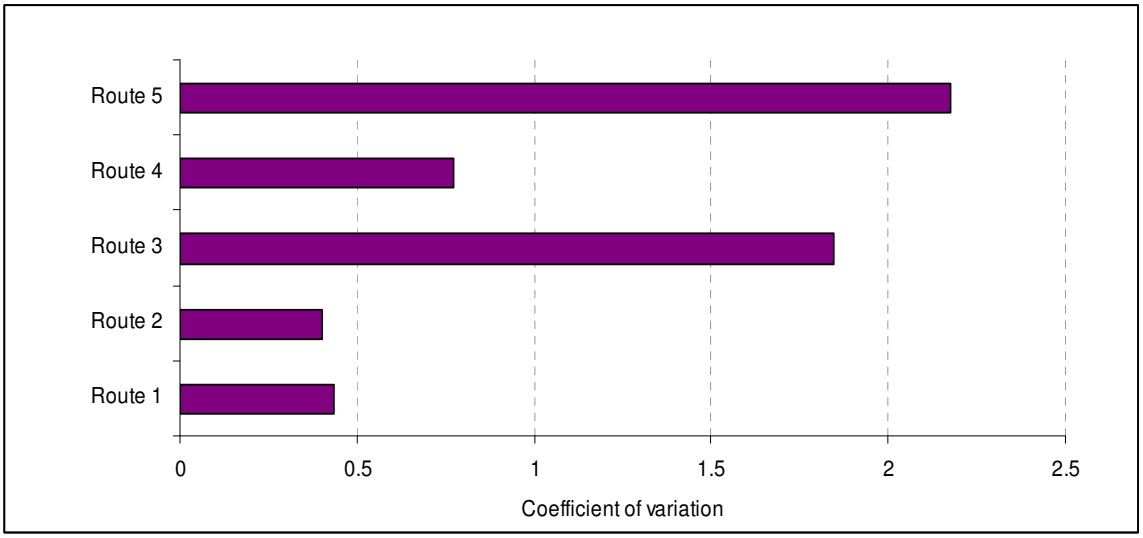


Figure 5-9 Coefficient of variation of the estimated route fractions

5.5 ESTIMATION OF DESTINATION FRACTIONS

Similarly, it is natural to expect that, for a vehicle departing from zone i during time interval t on the target day m, the probability that zone j ($j = 1, 2, ..., J$) is the destination, p_m^{ijt}, relates to the probabilities from different days, p_d^{ijt}'s.

Let $\mathbf{p}_m^{it} = (p_m^{i1t}, ..., p_m^{iJt})$ be a vector of destination fractions of vehicles departing from origin zone i during interval t on the target day m (over the J - dimensional multinomials). By assigning over the simplex the Dirichlet prior

$$\text{Dirichlet}(\mathbf{p}_m^{it} \mid \beta_1^{it}, \beta_2^{it}, ..., \beta_J^{it}) = \frac{\Gamma(\beta_0^{it})}{\prod_{j=1}^{J} \Gamma(\beta_j^{it})} \prod_{j=1}^{J} (p_m^{ijt})^{\beta_j^{it}-1} \tag{5-11}$$

where $\beta_j^{it}, \forall j$ = parameters of the Dirichlet prior, β_0^{it} = a concentration parameter = $\sum_{j=1}^{J} \beta_j^{it}$, $p_m^{i1t} > 0, .., p_m^{iJt} > 0$, $\sum_{j=1}^{J} p_m^{ijt} = 1$.

The Dirichlet posterior for $\mathbf{p}_m^{it}$ is

$$\text{Dirichlet}(\mathbf{p}_m^{it} \mid \beta_1^{it}, .., \beta_J^{it}, \tilde{x}_m^{i1t}, .., \tilde{x}_m^{iJt}, \tilde{o}_m^{it}) = \frac{\Gamma(\eta_{m0}^{it})}{\prod_{j=1}^{J} \Gamma(\eta_{mj}^{it})} \prod_{j=1}^{J} (p_m^{ijt})^{\eta_{mj}^{it}-1} \tag{5-12}$$

where $\tilde{x}_m^{ijt}$ = observed trajectory flow between origin i and destination j during interval t on the target day m, $\tilde{o}_m^{it}$ = observed trajectory flow departing from zone i during time interval t on the target day m, $\eta_{mj}^{it} = \beta_j^{it} + \tilde{x}_m^{ijt}$, and $\eta_{m0}^{it} = \sum_{j=1}^{J} \beta_j^{it} + \tilde{o}_m^{it}$.

The posterior mean, variance, and covariance of the destination fractions can be computed as follows:

$$\hat{p}_m^{ijt} = \eta_{mj}^{it} / \eta_{m0}^{it} \tag{5-13}$$

$$\text{VAR}(\hat{p}_m^{ijt}) = \sigma_{\hat{p}_m^{ijt}}^2 = \left[\eta_{mj}^{it}.(\eta_{m0}^{it} - \eta_{mj}^{it})\right] / \left[(\eta_{m0}^{it})^2.(\eta_{m0}^{it} + 1)\right] \tag{5-14}$$

$$\mathrm{COV}(\hat{p}_m^{ijt},\ \hat{p}_m^{ilt}) = \left[-\eta_{mj}^{it}.\eta_{ml}^{it}\right] \Big/ \left[(\eta_{m0}^{it})^2.(\eta_{m0}^{it}+1)\right] \tag{5-15}$$

To estimate the parameters of the dirichlet prior α_j^{it}'s empirically, a set of sampled destination fractions $\boldsymbol{\theta}_d^{it}$'s from multiple days, $D_{it}^{dest} = \left\{\boldsymbol{\theta}_1^{it}, \boldsymbol{\theta}_2^{it}, ..., \boldsymbol{\theta}_m^{it}, ...\right\}$ is utilized in the maximum likelihood estimation. A numerical example is omitted here as the method provided in this section is fundamentally the same as the one presented in the previous section. A comprehensive simulation study which includes the procedure provided here will be presented in CHAPTER 7.

Note that both empirical Bayesian models proposed in the previous section and this section can alternatively be formulated hierarchically by assigning a distribution with another set of hyper-parameters (in a lower hierarchy) to the vector $\boldsymbol{\alpha}_k^{ijt}$ (or $\boldsymbol{\beta}_j^{it}$). However, due to a large dimension of multinomials, the empirical Bayesian approach presented above which provides approximate solutions to the hierarchical model, is recommended.

5.6 ESTIMATION OF ROUTE-LINK FRACTIONS

5.6.1 Decomposition of assignment matrix

Under the bi-level OD estimation scheme such as the one used in the six-step process, an assignment matrix is iteratively updated based on a traffic simulation run with the current

updated OD tables. The need for such estimation is mainly due to the lack of adequate observations regarding the physical relations mapping OD flows and link flows. On the other hand, with observations of vehicle trajectory, it is possible to predetermine an assignment matrix before entering the OD estimation process.

The relationships between a link count and dynamic OD flows or dynamic route flows were given in CHAPTER 2. However, they are reproduced here for the reader's convenience. A link count observed on link l during observation interval h, denoted by v_{lh}, can be expressed as the summation of the fractions of different OD flows or route flows as in equation (5-16) and (5-17), respectively.

$$v_{lh} = \sum_{t=1}^{h} \sum_{ij} a_{lh}^{ijt} x^{ijt} \tag{5-16}$$

$$v_{lh} = \sum_{t=1}^{h} \sum_{k \in K} m_{lh}^{kt} f^{kt} \tag{5-17}$$

where x^{ijt} is the OD flow from i to j departing during interval t. f^{kt} is the route flow on route k departing during interval t. a_{lh}^{ijt} (assignment fraction) and m_{lh}^{kt} (route-link fraction) are, respectively, the fractions of x^{ijt} and f^{kt} contributing to v_{lh}.

Let q^{kt} denote the fraction of x^{ijt} using route k, then

$$f^{kt} = x^{ijt}\, q^{kt} \tag{5-18}$$

$$v_{lh} = \sum_{t=1}^{h} \sum_{ij} x^{ijt} \sum_{k \in K_{ij}} m_{lh}^{kt} q^{kt} \tag{5-19}$$

By comparing equation (5-16) and equation (5-19), Cascetta et al. (1993) shows that

$$a_{lh}^{ijt} = \sum_{k \in K_{ij}} m_{lh}^{kt} q^{kt} \tag{5-20}$$

As shown in equation (5-20), an assignment fraction a_{lh}^{ijt} can be decomposed into the summation of the products between route-link fraction m_{lh}^{kt} and route-choice fraction q_m^{kt}. Since the estimates of the latter can be obtained independently using the model described in section 5.4, this section focuses solely on how to estimate the route-link fractions $\hat{m}_{lh}^{kt}$'s based on the estimated link travel time distributions from section 5.3.

5.6.2 The Monte Carlo simulation approach

Using a Monte Carlo simulation, the proposed method generates a large number of vehicles moving through each route and calculates the fraction of vehicles arriving at each link l during each observation interval h. Each vehicle can be assumed to move over each link with a travel time drawn independently from the link travel time distribution that corresponds to the interval and occupancy group when it arrives at that link. Nonetheless, to take into account the correlations of the same drivers driving on

different links, it is assumed that each vehicle tends to have the travel time on each link drawn from the same region of the corresponding distribution.

Figure 5-10 illustrates an example of one vehicle transferring to the next link. It is assumed that the corresponding travel time distribution on each link is equally divided into three regions. Let RD and RU be an index of the region from which a travel time datum is drawn at downstream link and upstream link, respectively. In the figure, the vehicle which has the travel time on link l from region 2 has a high probability of $p(RD=2 \mid RU=2)$ to have a travel time on the link $l+1$ drawn also from region 2, and smaller probabilities of $p(RD=1 \mid RD=2)$ and $p(RD=3 \mid RU=2)$ to have the travel time drawn from region 1 and region 3, respectively.

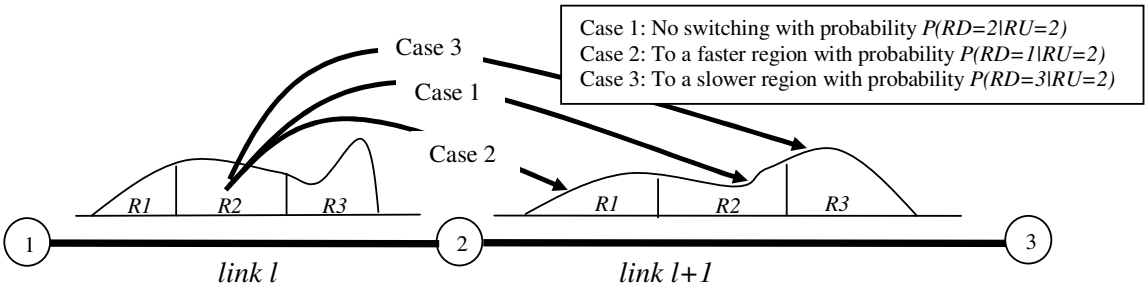


Figure 5-10 An example of a vehicle transferring to the next link

The detailed procedure for the Monte Carlo simulation is as follows:

Step 0: Select route k, departure interval t, and N_p the number of vehicles to simulate per simulation run. Determine the number of simulation run N_s and set the simulation index s to zero.

Step 1: Determine the region thresholds for travel time distributions of all related links l ($l = 1, 2, ...L_{kt}$) and intervals h ($t \le h \le H$).

Step 2: If $s < N_s$, set the iteration index $s = s + 1$ and go to step 3. Otherwise, go to step 7.

Step 3: Set all simulated counts $\vec{v}_{lh}$ and the vehicle index p to zero.

Step 4: If $p < N_p$, set the vehicle index $p = p + 1$ and go to step 5. Otherwise, go to step 6.

Step 5: Uniformly draw a departure time from the departure interval t. Move the vehicle over the first link with a travel time drawn from the corresponding distribution, assign the initial region correspondingly, and record the arrival time at the second link (departure time + link travel time). For each of the rest of the links, draw a region of the distribution from which a travel time will be drawn based on the predetermined probabilities conditioned on the previous region of the upstream link, draw a travel time from the region on the corresponding distribution, move vehicle to the end of the link, and record the arrival time at the next link. Go back to step 4.

Step 6: For each count location l during each time interval h, set the traffic counts $\vec{v}_{lh}(s)$ to the sum of all arrival records within that interval. Calculate the route-link fractions $\hat{m}_{lh}^{kt}(s)$'s using equation $\hat{m}_{lh}^{kt}(s) = \vec{v}_{lh}(s) / N_p$. Go back to step 2.

Step 7: Calculate the statistics such as mean and variance for each route-link fraction $\hat{m}_{lh}^{kt}$ using $\hat{m}_{lh}^{kt}(s)'s$ ($s = 1, 2, .., N_s$).

To determine the conditional probabilities of each region, geometric characteristics such as freeway ramps and turning links, which determine the relative region of a vehicle's travel time at the downstream link, should be taken into account. A

comprehensive simulation study which includes the procedure provided herein will be presented in CHAPTER 7.

5.7 ESTIMATION OF SEED OD TABLES

Section 5.5 presents an empirical Bayesian model for estimating destination fractions of each origin zone during each interval. In this section, new methods used for constructing dynamic seed OD tables and the variance-covariance matrix from the estimated destination fractions and observed link counts (on the target day) are presented. Let

$\hat{x}_m^{ijt}$ The seed (initial) OD flow from origin i to destination j departing during time interval t on the target day m.

$\sigma_{\hat{x}_m^{ijt}}$ The standard deviation of $\hat{x}_m^{ijt}$.

$\mathrm{COV}(\hat{x}_m^{ijt}, \hat{x}_m^{ilt})$ The covariance between $\hat{x}_m^{ijt}$ and $\hat{x}_m^{ilt}$.

$\breve{o}_m^{it}$ The total departure flow from zone i during time interval t on the target day m obtained from detectors or traffic counts.

5.7.1 Intersection and freeway networks

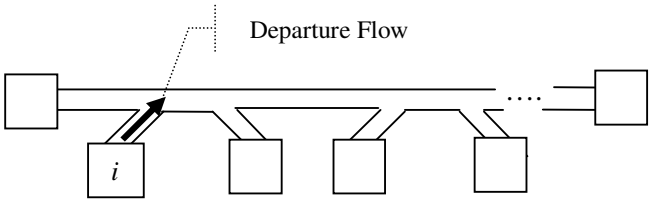


Figure 5-11 Typical zone setting for a freeway network

For intersection and freeway networks with zones located at the boundary (see, for example, Figure 5-11), the initial OD flows and variance-covariance matrix can be calculated as follows

$$\hat{x}_m^{ijt} = \breve{o}_m^{it} . \hat{p}_m^{ijt} \tag{5-21}$$

$$\sigma^2_{\hat{x}_m^{ijt}} = \left(\breve{o}_m^{it}\right)^2 . \left[\eta_{mj}^{it} . (\eta_{m0}^{it} - \eta_{mj}^{it})\right] \Big/ \left[(\eta_{m0}^{it})^2 . (\eta_{m0}^{it} + 1)\right] \tag{5-22}$$

$$\mathrm{COV}(\hat{x}_m^{ijt}, \hat{x}_m^{ilt}) = \left(\breve{o}_m^{it}\right)^2 . \left[-\eta_{mj}^{it} . \eta_{ml}^{it}\right] \Big/ \left[(\eta_{m0}^{it})^2 . (\eta_{m0}^{it} + 1)\right] \tag{5-23}$$

$$\mathrm{COV}(\hat{x}_m^{ijt}, \hat{x}_m^{ljt}) = 0 \tag{5-24}$$

$$\mathrm{COV}(\hat{x}_m^{ijt}, \hat{x}_m^{ijl}) = 0 \tag{5-25}$$

Note that the above expressions assume that $\breve{o}_m^{it}$'s are error-free. However, if the errors from traffic counts are to be taken into account, the expressions above can be easily modified.

5.7.2 General networks

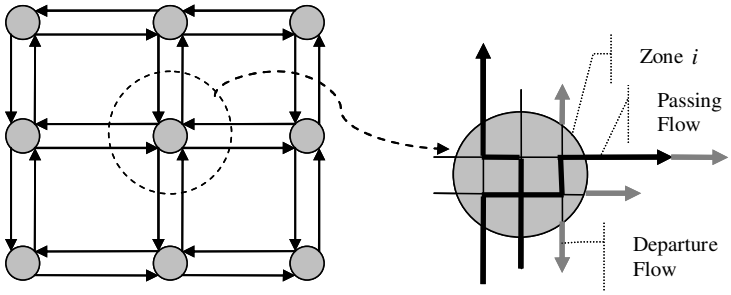


Figure 5-12 Typical zone setting for a general network

For general networks (see, for example, Figure 5-12), on the other hand, total departure flows from origin zones are not readily available. That is, total outflows from detectors or traffic counts encompass both passing and departure flows. To estimate the total departure flow, it is natural to estimate departure fraction using observed trajectories and then to apply that estimate to the total observed outflow. As with other models discussed thus far, a similar Bayesian model is proposed here to overcome the small sample problem.

Denote the total observed outflow trajectories from zone i during time period t on day d by $\tilde{u}_d^{it}$ and departure flow by $\tilde{o}_d^{it}$. It is assumed that $\tilde{o}_d^{it}$ values are independent binomial data with a sample size of $\tilde{u}_d^{it}$, and day-specific departure probability of $p_d^{it(Dep)}$. Further, $\text{logit}\left(p_d^{it(Dep)}\right)$ values are assumed to be normally distributed with parameters $(\mu^{it(Dep)}, \sigma^{it(Dep)})$. That is, day-specific departure probabilities $p_d^{it(Dep)}$ are assumed to be samples drawn from a common population distribution to reflect day-to-day variations

due to the stochasticity in both traffic dynamics and trip making patterns. The structure of the hierarchical model can be specified below and shown in Figure 5-13.

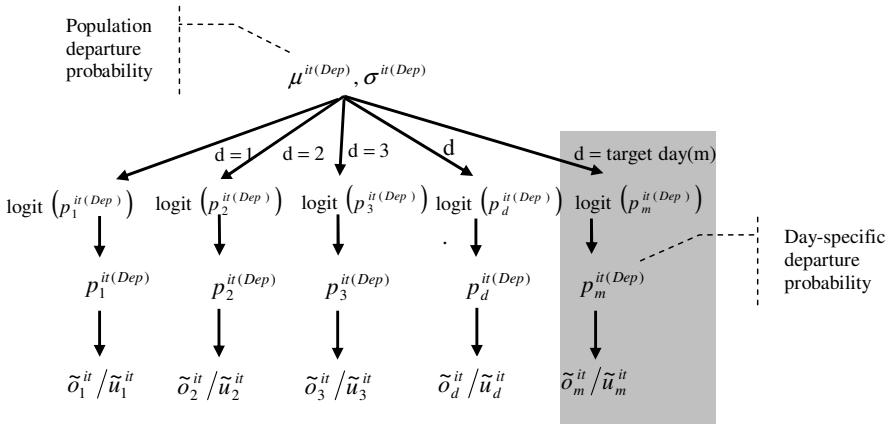


Figure 5-13 Hierarchical model for estimating departure fraction for zone i during interval t

$$\tilde{o}_d^{it} \mid \tilde{u}_d^{it}, p_d^{it(Dep)}, \phi \sim \text{Binomial}(\tilde{u}_d^{it}, p_d^{it(Dep)}) \tag{5-26}$$

$$\text{logit}(p_d^{it(Dep)}) \mid \phi \sim \text{Normal}\left(\mu^{it(Dep)}, \sigma^{it(Dep)}\right) \tag{5-27}$$

$$\mu^{it(Dep)} \sim \text{normal}(0, 1000) \tag{5-28}$$

$$\sigma^{it(Dep)} \sim \text{gamma}(0.001, 0.001) \tag{5-29}$$

where $\phi = \left(\mu^{it(Dep)}, \sigma^{it(Dep)}\right)$

The relations (5-28) and (5-29) are just flat (diffused) priors to reflect a lack of knowledge about the hyper-parameters.

The Gibbs sampler can be used to infer the posterior distributions of $p_d^{it(Dep)}$'s, $\mu^{it(Dep)}$, and $\sigma^{it(Dep)}$. The posterior means $\hat{p}_m^{it(Dep)}$'s and the corresponding standard

deviations $\sigma_{\hat{p}_m^{it(Dep)}}$'s of the day-specific departure fractions can be directly calculated by simulations.

For simplicity, it is further assumed that there is no correlation between destination fractions and departure fractions. It follows that

$$\hat{x}_m^{ijt} = \breve{u}_m^{it} . \hat{p}_m^{it(Dep)} . \hat{p}_m^{ijt} \tag{5-30}$$

$$\sigma^2_{\hat{x}_m^{ijt}} = \left(\breve{u}_m^{it}\right)^2 . \left[\sigma^2_{\hat{p}_m^{it(Dep)}} . \sigma^2_{\hat{p}_m^{ijt}} + (\hat{p}_m^{it(Dep)})^2 . \sigma^2_{\hat{p}_m^{ijt}} + (\hat{p}_m^{ijt})^2 . \sigma^2_{\hat{p}_m^{it(Dep)}}\right] \tag{5-31}$$

$$\mathrm{COV}(\hat{x}_m^{ijt}, \hat{x}_m^{ilt}) = \left(\breve{u}_m^{it}\right)^2 . \left[\sigma^2_{\hat{p}_m^{it(Dep)}} + \mathrm{COV}(\hat{p}_m^{ijt}, \hat{p}_m^{ilt})\right] \tag{5-32}$$

Note that $p_d^{it(Dep)}$ values can alternatively be assumed to be samples drawn from a beta distribution, but it was found that the logistic-normal distribution provides more accurate results, especially in the case of missing data on the target day.

5.8 SUMMARY

Though probe data offer high fidelity trip making and traffic information, the use of such data for estimating the distributions of the time-dependent OD flows and assignment matrix is not straightforward. This is because possible sampling schemes in practice can only be implemented with low sampling rates that also fluctuate over time and space. Provided that the sampling of vehicle trajectories can be performed continuously over multiple days, the empirical and hierarchical Bayesian models presented in this chapter

can be used to derive necessary elements required for estimating dynamic OD flows. These elements include time-dependent travel time distributions, route fractions, destination fractions, and departure fractions. Different similarity criteria for each element have been suggested, so that the information across different time intervals can be effectively used. As such, it is not necessary to impose the exchangeability assumption of the data across different intervals. Rather, it is assumed that each parameter explaining the distribution of interest in each interval is a sample drawn from another distribution that is common for all time intervals within the same group (see Figure 5-14).

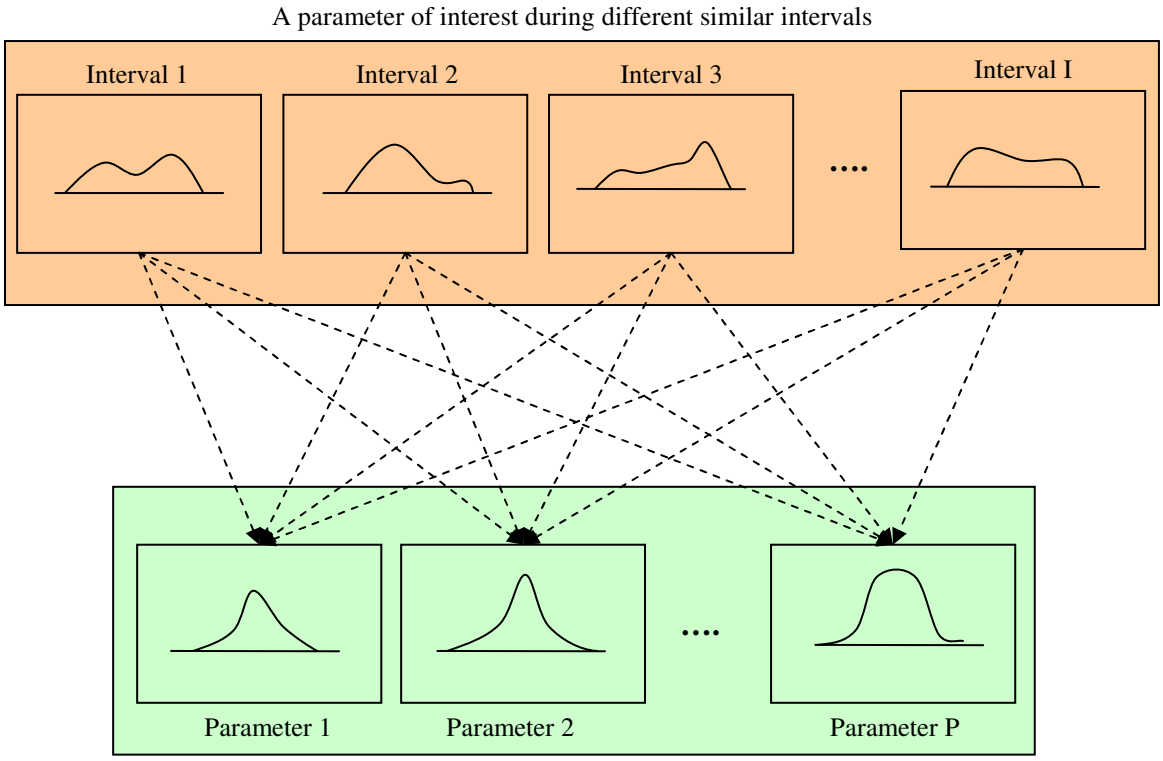


Figure 5-14 Concept of the hierarchical and empirical Bayesian models

In addition to the gain in the estimation accuracy, these models feature multi-level posterior parameter distributions which can be used in the further OD estimation step.

Common parameters, which are typically associated with a greater variance, can be used in the case of missing data. This is a very good model feature given that there are many OD pairs, routes, and links that contain no observation during the interval of interest. This chapter also presented a new method, based on a Monte Carlo simulation, to estimate the underlying route-link fractions. Under this method, vehicles move over each link according to the estimated travel time distribution, which depends on the time and traffic condition when they arrive at the link.

CHAPTER 6 DYNAMIC OD DEMAND ESTIMATION WITH STOCHASTIC ASSIGNMENT MATRIX

6.1 INTRODUCTION

The previous chapter addressed several challenges in utilizing small probe samples collected over multiple days to estimate the initial OD flows, route-choice fractions, and route-link fractions for each interval within the study period. On the one hand, the data from probe vehicles are considered to be direct high-fidelity observations of trip making and traffic patterns, but on the other hand, they are only samples from different subpopulations. It is thus desirable to further update these estimates with link count observations during the study period, which provide (indirect) information regarding the prevailing population route and OD flows. This chapter aims to explore the possibility in this direction.

Section 6.2 presents a new OD estimation model, which simultaneously updates the initial estimates of OD table and assignment matrix from observed link counts. The adjustment is done based on the relative confidence placed on each measurement or initial estimate through the use of estimated error variance-covariance matrices from the first step of the framework. A set of natural constraints is introduced to ensure a meaningful solution, and bound constraints are used as an additional regularization mechanism. Section 6.3 presents an efficient solution algorithm, particularly in term of the amount of the memory required for solving a large scale problem. Based on the Block Coordinate Descent method, the problem is decomposed into sub-problems, which are

then solved iteratively. To solve the sub-problems involving both bound and functional constraints, a combined use of the Augmented Lagrangian Function and the Frank-Wolfe method is presented. Section 6.4 discusses some practical considerations regarding the initial estimates, weights, and bound constraints. Concluding remarks are then given in the last section.

6.2 OD ESTIMATION WITH STOCHASTIC ASSIGNMENT MATRIX

Recall the relationship between a link count and OD flows. If the underlying assignment fractions a_{lh}^{ijt} are known with certainty, expression (5-16) can be rewritten to include the link count measurement error as

$$\breve{v}_{lh} = \sum_{t=1}^{h} \sum_{ij} a_{lh}^{ijt} x^{ijt} + \pi_{lh} \tag{6-1}$$

where $\breve{v}_{lh}$ is an erroneous count on link l during interval h, x^{ijt} is the OD flow from i to j departing during interval t, and π_{lh} is a random error due to the link count measurement.

Nevertheless, the underlying assignment fractions are unknown in practice, and their estimates $\hat{a}_{lh}^{ijt}$, which are also prone to error, can be expressed as

$$\hat{a}_{lh}^{ijt} = a_{lh}^{ijt} + \delta_{lh}^{ijt} \tag{6-2}$$

where δ_{lh}^{ijt} is a random error due to the travel time measurement, traffic loading, and route-choice modeling.

By substituting equation (6-2) to equation (6-1), it follows that

$$\breve{v}_{lh} = \sum_{t=1}^{h} \sum_{ij} \hat{a}_{lh}^{ijt} x^{ijt} - \sum_{t=1}^{h} \sum_{ij} \delta_{lh}^{ijt} x^{ijt} + \pi_{lh} \tag{6-3}$$

$$\breve{v}_{lh} = \sum_{t=1}^{h} \sum_{ij} \hat{a}_{lh}^{ijt} x^{ijt} + \dot{\pi}_{lh} \tag{6-4}$$

where $\dot{\pi}_{lh} = \pi_{lh} - \sum_{t=1}^{h} \sum_{ij} \delta_{lh}^{ijt} x^{ijt}$

As pointed out by Ashok and Ben-Akiva (2002), since $\hat{a}_{lh}^{ijt}$ and $\dot{\pi}_{lh}$ are correlated, the dynamic OD estimation with erroneous assignment fractions becomes a standard *error-in-variables* problem; and the application of the standard least squares method to the relation (6-4) may yield biased and inconsistent estimates. Accordingly, they present two methods to handle the stochastic properties of the assignment fractions. The first method is to include an additional term from the relationship (6-2) to the GLS-based formulation, leading to an estimation model which adjusts both OD flows and assignment matrix. This method can be viewed as an extension of a static OD estimation model proposed by Lo et al. (1999), which is based on the maximum likelihood with sampled OD flows, assignment fractions, and link counts assumed to have a multivariate normal (MVN) distribution. Because an assignment matrix is usually very large, and more

importantly the information regarding the variance-covariance matrix of the error term in expression (6-2) is often not available, this approach is not practical. Instead, the second method incorporates two additional error terms related to travel times and route-choice fractions. These models provide an appealing way to account for the errors in the estimated assignment fractions. Nonetheless, obtaining the samples or initial estimates of these parameters and the corresponding variance-covariance matrices remains a challenging issue in these studies. Since these parameters are unbounded in the formulations, the natural constraints may be infringed upon after the update, leading to fruitless estimates of both OD flows and assignment fractions.

A new dynamic OD estimation model proposed in this chapter handles the stochastic elements in an assignment matrix by incorporating two additional error terms according to the decomposition of assignment matrix discussed in subsection 5.6.1 into the objective function. The first term is of the route-choice fractions, and the second term is of the route-link fractions. Under the Maximum Likelihood approach, different specifications of these two terms can be obtained, depending on the distributional assumptions made on the observations. This dissertation adopts the 2-norm distance between the parameters being estimated and their initial estimates or observations, weighted by the inverses of variance-covariance matrix of the corresponding error terms. As such, the model can be viewed as based on the multivariate normal assumption or the GLS method which does not require any specific distribution. A set of constraints is also introduced to the optimization in order to guarantee the following conditions. First, all route fractions for each OD pair during each departure interval must sum to one, and each fraction can only take on a value from the range of zero and one. Second, for a given

route, departure interval, and observation interval, a route-link fraction bounded within the range of zero and one cannot be greater than the fraction on any of its upstream links. Third, OD flows are non-negative. Further, thanks to the information regarding the parameter spaces from the posterior distributions available from the Bayesian data analysis step, it is possible to apply the notion of general upper and lower bounds to guarantee a reasonable range for each parameter. This way, the bound constraints are used both to guarantee the natural requirements and as a regularization technique to overcome the rank-deficient problem should the total number of available information is less than the rank of the parameters being estimated (Aster et al., 2004).

For notational convenience, drop the target day index m. Let

IJ_{obs} The total number of OD pairs with available initial (seed) estimates

L_{obs} The total number of links with available observed counts

K^{ijt} The total number of routes from origin i to destination j during departure period t

L_{kt} The total number of links on route k with vehicles departing during departure period t.

H The total number of observation intervals.

$\hat{\mathbf{x}}_t$, $\mathbf{x}_t$ The (IJ_{obs})-vector of initial (seed) OD flows departing during the interval t, and of its corresponding OD flows to be estimated, respectively.

$\breve{\mathbf{v}}_h$, $\mathbf{v}_h$ The (L_{obs})-vector of observed link counts during the interval h, and of its corresponding link counts by assigning the OD flows using the current assignment fractions, respectively.

$\hat{\mathbf{q}}_{ijt}$, $\mathbf{q}_{ijt}$ The (K^{ijt})-vector of initial route-choice fractions specific to OD pair ij departing during period t, and of its corresponding route-choice fractions to be estimated, respectively.

$\hat{\mathbf{m}}_{kt}$, $\mathbf{m}_{kt}$ The ($L_{kt}.(H-t+1)$)-vector of initial route-link fractions of route k departing during the interval t, and of its corresponding route-link fractions to be estimated, respectively.

The optimization program can be expressed as

$$\min z(\mathbf{x}_t \ \forall t, \mathbf{q}_{ijt} \ \forall i,j,t, \mathbf{m}_{kt} \ \forall k,t \ ) = \begin{bmatrix} \sum_{t=1}^{T} (\mathbf{x}_t - \hat{\mathbf{x}}_t)^T \mathbf{W}_{\mathbf{x}_t}^{-1} (\mathbf{x}_t - \hat{\mathbf{x}}_t) + \\ \sum_{h=1}^{H} (\mathbf{v}_h - \breve{\mathbf{v}}_h)^T \mathbf{W}_{\mathbf{v}_h}^{-1} (\mathbf{v}_h - \breve{\mathbf{v}}_h) + \\ \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{t=1}^{T} (\mathbf{q}_{ijt} - \hat{\mathbf{q}}_{ijt})^T \mathbf{W}_{\mathbf{q}_{ijt}}^{-1} (\mathbf{q}_{ijt} - \hat{\mathbf{q}}_{ijt}) + \\ \sum_{k=1}^{K} \sum_{t=1}^{T} (\mathbf{m}_{kt} - \hat{\mathbf{m}}_{kt})^{\mathbf{T}} \mathbf{W}_{\mathbf{m}_{kt}}^{-1} (\mathbf{m}_{kt} - \hat{\mathbf{m}}_{kt}) \end{bmatrix} \quad (6\text{-}5)$$

subject to

$$\sum_{\forall k \in K^{ijt}} \mathbf{q}_{ijt}[k] = 1 \qquad \forall i, j, t \quad (6\text{-}6)$$

$$\mathbf{q}_{ijt}^{LO} \le \mathbf{q}_{ijt} \le \mathbf{q}_{ijt}^{UP} \qquad \forall i, j, t \quad (6\text{-}7)$$

$$\mathbf{m}_{kt}^{LO} \le \mathbf{m}_{kt} \le \mathbf{m}_{kt}^{UP} \qquad \forall k, t \quad (6\text{-}8)$$

$$\mathbf{x}_t^{LO} \le \mathbf{x}_t \le \mathbf{x}_t^{UP} \qquad \forall t \quad (6\text{-}9)$$

where $\mathbf{W}_{\mathbf{x}_t}, \mathbf{W}_{\mathbf{v}_h}, \mathbf{W}_{\mathbf{q}_{ijt}}, \mathbf{W}_{\mathbf{m}_{kt}}$ is the variance-covariance matrix of the error term corresponding to the indexed OD flows, link counts, route-choice fractions, and route-link fractions, respectively.

With the proposed Bayesian data analysis step presented in the previous chapter, the estimated matrices $\hat{\mathbf{W}}_{\mathbf{x}_t}$, $\hat{\mathbf{W}}_{\mathbf{q}_{ijt}}$, and $\hat{\mathbf{W}}_{\mathbf{m}_{kt}}$ are readily available. The elements of $\hat{\mathbf{W}}_{\mathbf{v}_h}$ are commonly determined from link counts observed from the field over multiple days or by assuming that each link count is independent and that the standard deviation is a certain fraction of the observed value.

It is noteworthy that the formulation above assumes the independency entirely across different sets of random variables $\mathbf{x}$, $\mathbf{q}$, and $\mathbf{m}$; and modestly within each set. The problem is, however, still very difficult to solve because it is highly non-convex and involves a very large number of parameters. Two simplifications are therefore considered. The first simplification is to hold the estimated route-link fractions fixed. This is practically justified under our framework as the variances of route-link fractions are generally much smaller than those of route-choice fractions. Moreover, proper bounds for each route-link fraction are difficult to determine. The second simplification is to assume further independency across all individual parameters, which is commonly done in several previous studies under the ordinary least squares (OLS) scheme. This way, the problem reduces to that of simultaneously adjusting OD flows and route-choice fractions. The optimization program can now be conveniently expressed in the scalar form as follows.

$$\min z(x^{ijt}\ \forall i,j,t,\ q_{kt}\ \forall k,t) = \left[\begin{array}{l} \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(x^{ijt}-\hat{x}^{ijt})^2}{\sigma^2_{\hat{x}^{ijt}}} + \\ \frac{1}{2}\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J}x^{ijt}\sum_{k\in K^{ijt}}\hat{m}_{lh}^{kt}q_{kt}-\bar{v}_{lh})^2}{\sigma^2_{\bar{v}_{lh}}} + \\ \frac{1}{2}\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{(q_{kt}-\hat{q}_{kt})^2}{\sigma^2_{\hat{q}^{kt}}} \end{array}\right] \tag{6-10}$$

subject to

$$\sum_{k\in K^{ijt}} q_{kt} = 1 \qquad \forall K^{ijt} \tag{6-11}$$

$$q_{kt}^{LO} \le q_{kt} \le q_{kt}^{UP} \qquad \forall k,t \tag{6-12}$$

$$x^{ijt\,LO} \le x^{ijt} \le x^{ijt\,UP} \qquad \forall i,j,t \tag{6-13}$$

6.3 SOLUTION ALGORITHM

The formulation presented in the previous section can take into account the randomness in an assignment matrix. This, however, comes at the cost of introducing a large number of extra decision variables to the optimization problem. Moreover, because of the multiplicative relationship between OD flows and route-choice fractions, the objective function (6-10) is non-convex. As a result, usual numerical algorithms such as Newton-

Raphson and quasi-Newton methods may not be suitable for seeking the optimal solution, especially for a large-scale network.

The algorithm presented in this section is therefore developed based on the Block Coordinated Descent or nonlinear Gauss-Seidel method. Under this method, the optimization is iteratively performed with respect to each of the block coordinate vectors, taken in a cyclic order. For the problem being considered, it is convenient to decompose it into two sub-problems. Denote the main iteration counter by m. These two sub-problems are specified below.

Problem 1

$$\min \left(z_1(\mathbf{x}) + z_2(\mathbf{x}, \mathbf{q}^{(m-1)}) + z_3(\mathbf{q}^{(m-1)})\right) \tag{6-14}$$

subject to

$$\mathbf{x}_t^{LO} \le \mathbf{x}_t \le \mathbf{x}_t^{UP} \qquad \forall t \tag{6-15}$$

Problem 2

$$\min \left(z_1(\mathbf{x}^{(m)}) + z_2(\mathbf{x}^{(m)}, \mathbf{q}) + z_3(\mathbf{q})\right) \tag{6-16}$$

subject to

$$\sum_{\forall k \in K^{ijt}} \mathbf{q}_{ijt}[k] = 1 \qquad \forall i, j, t \tag{6-17}$$

$$\mathbf{q}_{ijt}^{LO} \le \mathbf{q}_{ijt} \le \mathbf{q}_{ijt}^{UP} \qquad \forall i, j, t \tag{6-18}$$

The main algorithm used for optimizing this problem is outlined as follows.

Step 0: Initialize the main iteration counter $m = 0$. Set $\mathbf{x}^{(0)} = \hat{\mathbf{x}}$ and $\mathbf{q}^{(0)} = \hat{\mathbf{q}}$.

Step 1: Set the main iteration counter $m = m + 1$.

Step 2: Solve problem 1 to obtain $\mathbf{x}^{(m)}$

Step 3: Solve problem 2 to obtain $\mathbf{q}^{(m)}$

Step 4: Check the convergence criteria. If met, stop. Otherwise, go to step 1.

Namely, a sequence of $\mathbf{x}^{(m)}$ and $\mathbf{q}^{(m)}$ iteratively minimizes the objective function (6-10) while holding the values of the variables in the other set fixed. Since the objective function is differentiable with respect to both $\mathbf{x}$ and $\mathbf{q}$ with the Hessian matrices being positive definite, each sub-problem is strictly convex and the objective function value is strictly decreasing along the direction generated by the sub-problems (see, Bertsekas, 1999; Lo et al., 1999).

In the decomposition, problem 1 is a bounded dynamic OD estimation, similar to the one presented in CHAPTER 3. Thus, the presented algorithm based on the F-W algorithm can be modified to solve this bounded problem. On the other hand, problem 2 to update route-choice fractions is more complicated as it involves the functional constraints that the summations of route fractions must equal to one in addition to the bound ones.

Let us first consider solving problem 1. For notational convenience, sub-iteration index s is used in the following discussion to differentiate the F-W sub-iterations from

the main iterations, m. Recall from problems (6-10) and (6-14), the objective function of problem 1 can be expressed as follows.

$$\min P_1(x^{ijt}\ \forall i,j,t \mid {q_{kt}}^{(m-1)}\ \forall k,t) = \left[\begin{array}{l} \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(x^{ijt}-\hat{x}^{ijt})^2}{\sigma^2_{\hat{x}^{ijt}}} + \\ \frac{1}{2}\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J}x^{ijt}\sum_{k\in K^{ijt}}\hat{m}^{kt}_{lh}{q_{kt}}^{(m-1)}-\breve{v}_{lh})^2}{\sigma^2_{\breve{v}_{lh}}} + \\ \frac{1}{2}\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{({q_{kt}}^{(m-1)}-\hat{q}_{kt})^2}{\sigma^2_{\hat{q}^{kt}}} \end{array}\right] \quad (6\text{-}19)$$

Since the last term in (6-19) is constant and does not affect the optimization, it can be omitted from further optimization process. Problem 1 can then be rewritten as

$$\min P_1(x^{ijt}\ \forall i,j,t \mid {q_{kt}}^{(m-1)}\ \forall k,t) = \left[\begin{array}{l} \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(x^{ijt}-\hat{x}^{ijt})^2}{\sigma^2_{\hat{x}^{ijt}}} + \\ \frac{1}{2}\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J}x^{ijt}\sum_{k\in K^{ijt}}\hat{m}^{kt}_{lh}{q_{kt}}^{(m-1)}-\breve{v}_{lh})^2}{\sigma^2_{\breve{v}_{lh}}} \end{array}\right] \quad (6\text{-}20)$$

subject to

$${x^{ijt}}^{LO} \le x^{ijt} \le {x^{ijt}}^{UP} \quad \forall i,j,t \quad (6\text{-}21)$$

At the s^{th} iteration of the F-W algorithm, the gradients of the first and second term in the objective function with respect to x^{mnd} can be calculated as follows.

$$\left.\frac{\partial P_1(\mathbf{x}\mid\mathbf{q}^{(m-1)})}{\partial x^{mnd}}\right]^{(s)}_{OD\ term} = \frac{(x^{mnd\,(s)} - \hat{x}^{mnd})}{\sigma^2_{\hat{x}^{mnd}}} \tag{6-22}$$

$$\left.\frac{\partial P_1(\mathbf{x}\mid\mathbf{q}^{(m-1)})}{\partial x^{mnd}}\right]^{(s)}_{link\ count\ term} = \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \frac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} x^{ijt\,(s)} \sum_{k\in K^{ijt}} \hat{m}_{lh}^{kt} q_{kt}^{\ (m-1)} - \breve{v}_{lh})}{\sigma^2_{\breve{v}_{lh}}} \sum_{k\in K^{mnd}} \hat{m}_{lh}^{kd} q_{kd}^{\ (m-1)} \tag{6-23}$$

Let

$$\upsilon_{mnd} = x^{mnd\,(s)} - \hat{x}^{mnd} \tag{6-24}$$

$$\vartheta_{lh} = \sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} x^{ijt\,(s)} \sum_{k\in K^{ijt}} \hat{m}_{lh}^{kt} q_{kt}^{\ (m-1)} - \breve{v}_{lh} \qquad \forall l, h \tag{6-25}$$

The gradient of the objective function (6-20) with respect to x^{mnd} is thus

$$\frac{\partial P_1(\mathbf{x}\mid\mathbf{q}^{(m-1)})}{\partial x^{mnd}} = \frac{\upsilon_{mnd}}{\sigma^2_{\hat{x}^{mnd}}} + \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \frac{(\vartheta_{lh})}{\sigma^2_{\breve{v}_{lh}}} \sum_{k\in K^{mnd}} \hat{m}_{lh}^{kd} q_{kd}^{\ (m-1)} \tag{6-26}$$

The auxiliary OD flow $\mathbf{y}^{(s)}$ can be determined by means of the following logical expressions:

$$y_{mnd}^{(s)} = \begin{cases} x^{mnd\,LO} & if \left[\dfrac{\partial P_1(\mathbf{x}\mid\mathbf{q}^{(m-1)})}{\partial x^{mnd}}\right]^{(s)} > 0 \\ x^{mnd\,(s)} & if \left[\dfrac{\partial P_1(\mathbf{x}\mid\mathbf{q}^{(m-1)})}{\partial x^{mnd}}\right]^{(s)} = 0 \\ x^{mnd\,UP} & if \left[\dfrac{\partial P_1(\mathbf{x}\mid\mathbf{q}^{(m-1)})}{\partial x^{mnd}}\right]^{(s)} < 0 \end{cases} \tag{6-27}$$

Then, the OD flows can be updated as

$$x^{ijt(s+1)} = x^{ijt(s)} + \lambda(y_{ijt}^{(s)} - x^{ijt(s)}) = x^{ijt(s)} + \lambda\, d_{ijt}^{(s)} \qquad \forall i, j, t \tag{6-28}$$

Write the objective function in term of λ:

$$\mathrm{P}_1(\mathbf{x} + \lambda\mathbf{d} \mid \mathbf{q}^{(m-1)}) = \left[\begin{array}{l} \dfrac{1}{2}\displaystyle\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J} \dfrac{((x^{ijt(s)} + \lambda d_{ijt}^{(s)}) - \hat{x}^{ijt})^2}{\sigma_{\hat{x}^{ijt}}^2} + \\ \dfrac{1}{2}\displaystyle\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \dfrac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J}(x^{ijt(s)} + \lambda d_{ijt}^{(s)}) \sum_{k\in K^{ijt}} \hat{m}_{lh}^{kt} q_{kt}^{(m-1)} - \bar{v}_{lh})^2}{\sigma_{\bar{v}_{lh}}^2} \end{array} \right] \tag{6-29}$$

$$= \left[\begin{array}{l} \dfrac{1}{2}\displaystyle\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J} \dfrac{(x^{ijt(s)} - \hat{x}^{ijt} + \lambda d_{ijt}^{(s)})^2}{\sigma_{\hat{x}^{ijt}}^2} + \\ \dfrac{1}{2}\displaystyle\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \dfrac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} x^{ijt(s)} \sum_{k\in K^{ijt}} \hat{m}_{lh}^{kt} q_{kt}^{(m-1)} - \bar{v}_{lh} + \lambda \sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} d_{ijt}^{(s)} \sum_{k\in K^{ijt}} \hat{m}_{lh}^{kt} q_{kt}^{(m-1)})^2}{\sigma_{\bar{v}_{lh}}^2} \end{array} \right] \tag{6-30}$$

Let

$$\varphi_{lh} = \sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} d_{ijt}^{(s)} \sum_{k\in K^{ijt}} \hat{m}_{lh}^{kt} q_{kt}^{(m-1)} \tag{6-31}$$

Then,

$$\mathrm{P}_1(\mathbf{x} + \lambda\mathbf{d} \mid \mathbf{q}^{(m-1)}) = \left[\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(\upsilon_{ijt} + \lambda d_{ijt}^{(s)})^2}{\sigma_{\hat{x}^{ijt}}^2} + \frac{1}{2}\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \frac{(\vartheta_{lh} + \lambda\varphi_{lh})^2}{\sigma_{\bar{v}_{lh}}^2} \right] \tag{6-32}$$

To evaluate λ, the expression (6-32) should be minimized with respect to λ.

$$\frac{\partial P_1(\mathbf{x}+\lambda\mathbf{d}\,|\,\mathbf{q}^{(m-1)})}{\partial\lambda}=\left[\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(\upsilon_{ijt}+\lambda d_{ijt}^{(s)})}{\sigma_{\hat{x}^{ijt}}^2}d_{ijt}^{(s)}+\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{(\vartheta_{lh}+\lambda\varphi_{lh})}{\sigma_{\bar{v}_{lh}}^2}\varphi_{lh}\right] \tag{6-33}$$

$$=\left[\begin{array}{l}\displaystyle\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\upsilon_{ijt}d_{ijt}^{(s)}}{\sigma_{\hat{x}^{ijt}}^2}+\lambda\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(d_{ijt}^{(s)})^2}{\sigma_{\hat{x}^{ijt}}^2}+\\ \displaystyle\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{\vartheta_{lh}\varphi_{lh}}{\sigma_{\bar{v}_{lh}}^2}+\lambda\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{(\varphi_{lh})^2}{\sigma_{\bar{v}_{lh}}^2}\end{array}\right] \tag{6-34}$$

Setting the expression (6-34) to zero, the optimal moving size is

$$\lambda^*=\frac{\left[-\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{\upsilon_{ijt}d_{ijt}^{(s)}}{\sigma_{\hat{x}^{ijt}}^2}-\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{\vartheta_{lh}\varphi_{lh}}{\sigma_{\bar{v}_{lh}}^2}\right]}{\left[\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(d_{ijt}^{(s)})^2}{\sigma_{\hat{x}^{ijt}}^2}+\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{(\varphi_{lh})^2}{\sigma_{\bar{v}_{lh}}^2}\right]} \tag{6-35}$$

Before the entire algorithmic procedure can be given, let us next consider solving problem 2. Recall from problems (6-10) and (6-16), the objective function of problem 2 can be expressed as follows.

$$\min P_2(q_{kt}\ \forall\ k,t \mid x^{ijt(m)} \forall i,j,t) = \left[\begin{array}{l} \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(x^{ijt(m)}-\hat{x}^{ijt})^2}{\sigma^2_{\hat{x}^{ijt}}} + \\ \frac{1}{2}\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J}x^{ijt(m)}\sum_{k\in K^{ijt}}\hat{m}^{kt}_{lh}q_{kt}-\breve{v}_{lh})^2}{\sigma^2_{\breve{v}_{lh}}} + \\ \frac{1}{2}\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{(q_{kt}-\hat{q}_{kt})^2}{\sigma^2_{\hat{q}^{kt}}} \end{array} \right] \quad (6\text{-}36)$$

Again, since the first term in (6-36) is constant and does not affect the optimization, it can be ignored from the problem. Then,

$$\min P_2(q_{kt}\ \forall\ k,t \mid x^{ijt(m)} \forall i,j,t) = \left[\begin{array}{l} \frac{1}{2}\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J}x^{ijt(m)}\sum_{k\in K^{ijt}}\hat{m}^{kt}_{lh}q_{kt}-\breve{v}_{lh})^2}{\sigma^2_{\breve{v}_{lh}}} + \\ \frac{1}{2}\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{(q_{kt}-\hat{q}_{kt})^2}{\sigma^2_{\hat{q}^{kt}}} \end{array} \right] \quad (6\text{-}37)$$

subject to

$$\sum_{k\in K^{ijt}} q_{kt} = 1 \qquad \forall\ K^{ijt} \quad (6\text{-}38)$$

$$q^{LO}_{kt} \le q_{kt} \le q^{UP}_{kt} \qquad \forall\ k,t \quad (6\text{-}39)$$

The Augmented Lagrangian Function (ALM) or Method of Multipliers is chosen to optimize this problem. This method tends to find the solution that meets the Karush-Kuhn-Tucker conditions. Moreover, with linear constraints, the method results in $\nabla^2 L$

that is independent of the multipliers, and the contour does not alter in shape from stage-to-stage. Hence, the objective function does not suffer from distortion, which helps in preventing the ill-conditioning in successive problems. An overview of the ALM is provided in Appendix B.

Following Doblas and Benitez (2005), the proposed method includes the equality constraints in the Lagrangian function but directly handles the bound constraints using the F-W method. This yields a very memory-efficient approach suitable for a large-scale problem. For notational convenience, sub-iterations of the ALM are denoted by z and sub-iterations of the F-W method by s. The Lagrangian function can be expressed as

$$\min \mathrm{L}_2(\mathbf{q} \mid \mathbf{x}^{(m)}, \boldsymbol{\zeta}^{(z)}) = \left[\begin{array}{l} \dfrac{1}{2}\displaystyle\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \dfrac{\left(\displaystyle\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} x^{ijt(m)} \sum_{k\in K^{ijt}} \hat{m}_{lh}^{kt} q_{kt} - \breve{v}_{lh}\right)^2}{\sigma_{\breve{v}_{lh}}^2} + \\ \dfrac{1}{2}\displaystyle\sum_{k=1}^{K}\sum_{t=1}^{T} \dfrac{(q_{kt} - \hat{q}_{kt})^2}{\sigma_{\hat{q}^{kt}}^2} + \\ C\displaystyle\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\left[\left(\sum_{k\in K^{ijt}} q_{kt} - 1 + \zeta_{ijt}^{(z)}\right)^2 - \left(\zeta_{ijt}^{(z)}\right)^2\right] \end{array} \right] \tag{6-40}$$

subject to

$$q_{kt}^{LO} \le q_{kt} \le q_{kt}^{UP} \qquad \forall\, k, t \tag{6-41}$$

where C is a scale factor which might vary for different constraints but remain constant during the optimization process. ζ_{ijt} is a multiplier specific to all routes between i to j departing during period t.

At the s^{th} iteration of the F-W algorithm, the gradient of each term in (6-40) with respect to q_{rd} can be expressed as follows.

$$\left.\frac{\partial L_2(\mathbf{q} \mid \mathbf{x}^{(m)}, \boldsymbol{\zeta}^{(z)})}{\partial q_{rd}}\right]^{(s)}_{link\ count\ term} = \sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \frac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} x^{ijt(m)} \sum_{k\in K^{ijt}} \hat{m}_{lh}^{kt} q_{kt}^{(s)} - \breve{v}_{lh})}{\sigma^2_{\breve{v}_{lh}}} x^{<rd>(m)} \hat{m}_{lh}^{rd} \quad (6\text{-}42)$$

$$\left.\frac{\partial L_2(\mathbf{q} \mid \mathbf{x}^{(m)}, \boldsymbol{\zeta}^{(z)})}{\partial q_{rd}}\right]^{(s)}_{path\ fraction\ term} = \frac{(q_{rd}^{(s)} - \hat{q}_{rd})}{\sigma^2_{\hat{q}^{rd}}} \quad (6\text{-}43)$$

$$\left.\frac{\partial L_2(\mathbf{q} \mid \mathbf{x}^{(m)}, \boldsymbol{\zeta}^{(z)})}{\partial q_{rd}}\right]^{(s)}_{constr\ term} = 2C\left(\sum_{k\in K^{<rd>}} q_{kd}^{(s)} - 1 + \zeta_{<rd>}^{(z)}\right) \quad (6\text{-}44)$$

where

$x^{<rd>}$ is the demand flow to which route rd belongs

$K^{<rd>}$ is the route set to which route rd belongs

$\zeta_{<rd>}$ is the multiplier of the constraint to which route rd belongs

Let

$$\upsilon_{rd} = q_{rd}^{(s)} - \hat{q}_{rd} \quad (6\text{-}45)$$

$$\vartheta_{lh} = \sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} x^{ijt(m)} \sum_{k\in K^{ijt}} \hat{m}_{lh}^{kt} q_{kt}^{(s)} - \breve{v}_{lh} \qquad \forall l,h \quad (6\text{-}46)$$

The gradient of the objective function (6-40) with respect to q_{rd} is

$$\frac{\partial L_2(\mathbf{q} \mid \mathbf{x}^{(m)}, \boldsymbol{\zeta}^{(z)})}{\partial q_{rd}} = \sum_{h=1}^{H} \sum_{l=1}^{L_{obs}} \frac{(\vartheta_{lh})}{\sigma^2_{\bar{v}_{lh}}} x^{<rd>(m)} \hat{m}_{lh}^{rd} + \frac{(\upsilon_{rd})}{\sigma^2_{\hat{q}^{rd}}} + 2C\left(\sum_{k \in K^{<rd>}} q_{kd}^{(s)} - 1 + \zeta_{<rd>}^{(z)} \right) \quad (6\text{-}47)$$

The auxiliary direction $\mathbf{y}^{(s)}$ can then be obtained by means of the following logical expressions:

$$y_{rd}^{(s)} = \begin{cases} q_{rd}^{LO} & if \left[\dfrac{\partial L_2(\mathbf{q} \mid \mathbf{x}^{(m)}, \boldsymbol{\zeta}^{(z)})}{\partial q_{rd}} \right]^{(s)} > 0 \\ q_{rd}^{(s)} & if \left[\dfrac{\partial L_2(\mathbf{q} \mid \mathbf{x}^{(m)}, \boldsymbol{\zeta}^{(z)})}{\partial q_{rd}} \right]^{(s)} = 0 \\ q_{rd}^{UP} & if \left[\dfrac{\partial L_2(\mathbf{q} \mid \mathbf{x}^{(m)}, \boldsymbol{\zeta}^{(z)})}{\partial q_{rd}} \right]^{(s)} < 0 \end{cases} \quad (6\text{-}48)$$

Correspondingly, route fractions can be updated as follows.

$$q_{kt}^{(s+1)} = q_{kt}^{(s)} + \lambda (y_{kt}^{(s)} - q_{kt}^{(s)}) = q_{kt}^{(s)} + \lambda \, d_{kt}^{(s)} \qquad \forall k, t \quad (6\text{-}49)$$

Write the objective function in term of λ:

$$L_2(\mathbf{q} + \lambda \mathbf{d} \mid \mathbf{x}^{(m)}, \boldsymbol{\zeta}^{(z)}) = \left[\begin{array}{l} \dfrac{1}{2} \sum_{h=1}^{H} \sum_{l=1}^{L_{obs}} \dfrac{(\sum_{t=1}^{h} \sum_{i=1}^{I} \sum_{j=1}^{J} x^{ijt(m)} \sum_{k \in K^{ijt}} \hat{m}_{lh}^{kt} (q_{kt}^{(s)} + \lambda d_{kt}^{(s)}) - \bar{v}_{lh})^2}{\sigma^2_{\bar{v}_{lh}}} + \\ \dfrac{1}{2} \sum_{k=1}^{K} \sum_{t=1}^{T} \dfrac{((q_{kt}^{(s)} + \lambda d_{kt}^{(s)}) - \hat{q}_{kt})^2}{\sigma^2_{\hat{q}^{kt}}} + \\ C \sum_{t=1}^{T} \sum_{i=1}^{I} \sum_{j=1}^{J} \left[\left(\sum_{k \in K^{ijt}} (q_{kt}^{(s)} + \lambda d_{kt}^{(s)}) - 1 + \zeta_{ijt}^{(z)} \right)^2 - \left(\zeta_{ijt}^{(z)} \right)^2 \right] \end{array} \right] \quad (6\text{-}50)$$

$$= \left[\begin{array}{l} \frac{1}{2}\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \frac{(\sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} x^{ijt(m)} \sum_{k\in K^{ijt}} \hat{m}_{lh}^{kt} q_{kt}^{(s)} - \breve{v}_{lh} + \lambda \sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} x^{ijt(m)} \sum_{k\in K^{ijt}} \hat{m}_{lh}^{kt} d_{kt}^{(s)})^2}{\sigma_{\breve{v}_{lh}}^2} + \\ \frac{1}{2}\sum_{k=1}^{K}\sum_{t=1}^{T} \frac{(q_{kt}^{(s)} - \hat{q}_{kt} + \lambda d_{kt}^{(s)})^2}{\sigma_{\hat{q}^{kt}}^2} + \\ C\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\left[\left(\sum_{k\in K^{ijt}} (q_{kt}^{(s)} + \lambda d_{kt}^{(s)}) - 1 + \zeta_{ijt}^{(z)} \right)^2 - \left(\zeta_{ijt}^{(z)}\right)^2 \right] \end{array} \right] \tag{6-51}$$

Let

$$\varphi_{lh} = \sum_{t=1}^{h}\sum_{i=1}^{I}\sum_{j=1}^{J} x^{ijt(m)} \sum_{k\in K^{ijt}} \hat{m}_{lh}^{kt} d_{kt}^{(s)} \tag{6-52}$$

Then,

$$L_2(\mathbf{q} + \lambda \mathbf{d} \mid \mathbf{x}^{(m)}, \boldsymbol{\zeta}^{(z)})$$

$$= \left[\begin{array}{l} \frac{1}{2}\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}} \frac{(\vartheta_{lh} + \lambda \varphi_{lh})^2}{\sigma_{\breve{v}_{lh}}^2} + \\ \frac{1}{2}\sum_{k=1}^{K}\sum_{t=1}^{T} \frac{(\upsilon_{kt} + \lambda d_{kt}^{(s)})^2}{\sigma_{\hat{q}^{kt}}^2} + \\ C\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\left[\left(\sum_{k\in K^{ijt}} q_{kt}^{(s)} + \lambda \sum_{k\in K^{ijt}} d_{kt}^{(s)} - 1 + \zeta_{ijt}^{(z)} \right)^2 - \left(\zeta_{ijt}^{(z)}\right)^2 \right] \end{array} \right] \tag{6-53}$$

To evaluate λ, the expression (6-53) should be minimized with respect to λ.

$$\frac{L_2(\mathbf{q}+\lambda\mathbf{d}\mid\mathbf{x}^{(m)},\boldsymbol{\zeta}^{(z)})}{\partial\lambda}=\begin{bmatrix}\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{(\vartheta_{lh}+\lambda\varphi_{lh})}{\sigma^2_{\bar{v}_{lh}}}\varphi_{lh}+\\ \sum_{k=1}^{K}\sum_{t=1}^{T}\frac{(\upsilon_{kt}+\lambda d_{kt}{}^{(s)})}{\sigma^2_{\hat{q}^{kt}}}d_{kt}{}^{(s)}+\\ 2C\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\sum_{k\in K^{ijt}}q_{kt}{}^{(s)}+\lambda\sum_{k\in K^{ijt}}d_{kt}{}^{(s)}-1+\zeta_{ijt}{}^{(z)}\right)\sum_{k\in K^{ijt}}d_{kt}{}^{(s)}\end{bmatrix} \tag{6-54}$$

$$\begin{aligned}=&\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{\vartheta_{lh}\varphi_{lh}}{\sigma^2_{\bar{v}_{lh}}}+\lambda\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{\varphi_{lh}{}^2}{\sigma^2_{\bar{v}_{lh}}}+\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\upsilon_{kt}d_{kt}}{\sigma^2_{\hat{q}^{kt}}}+\lambda\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{d_{kt}{}^{(s)2}}{\sigma^2_{\hat{q}^{kt}}}+\\ &2C\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\sum_{k\in K^{ijt}}q_{kt}{}^{(s)}\sum_{k\in K^{ijt}}d_{kt}{}^{(s)}-\sum_{k\in K^{ijt}}d_{kt}{}^{(s)}+\zeta_{ijt}{}^{(z)}\sum_{k\in K^{ijt}}d_{kt}{}^{(s)}\right)+\\ &2C\lambda\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\sum_{k\in K^{ijt}}d_{kt}{}^{(s)}\right)^2\end{aligned} \tag{6-55}$$

Setting the expression (6-55) to zero, the optimal moving size is

$$\lambda^*=\frac{\begin{bmatrix}-\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{\vartheta_{lh}\varphi_{lh}}{\sigma^2_{\bar{v}_{lh}}}-\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{\upsilon_{kt}d_{kt}}{\sigma^2_{\hat{q}^{kt}}}-\\ 2C\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\sum_{k\in K^{ijt}}q_{kt}{}^{(s)}\sum_{k\in K^{ijt}}d_{kt}{}^{(s)}-\sum_{k\in K^{ijt}}d_{kt}{}^{(s)}+\zeta_{ijt}{}^{(z)}\sum_{k\in K^{ijt}}d_{kt}{}^{(s)}\right)\end{bmatrix}}{\begin{bmatrix}\sum_{h=1}^{H}\sum_{l=1}^{L_{obs}}\frac{\varphi_{lh}{}^2}{\sigma^2_{\bar{v}_{lh}}}+\sum_{k=1}^{K}\sum_{t=1}^{T}\frac{d_{kt}{}^{(s)2}}{\sigma^2_{\hat{q}^{kt}}}+2C\lambda\sum_{t=1}^{T}\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\sum_{k\in K^{ijt}}d_{kt}{}^{(s)}\right)^2\end{bmatrix}} \tag{6-56}$$

The F-W iterations are performed until the convergence criteria are met. This is the point at which the iteration of the ALM continues. The multipliers are updated using equation (6-57).

$$\zeta_{ijt}^{(z+1)} = \left(\sum_{k \in K^{ijt}} q_{kt}^{(s)} - 1 \right) + \zeta_{ijt}^{(z)} \qquad \forall i, j, t \tag{6-57}$$

The multipliers serve as a bias in the arguments of the penalty terms, and the updating rule (6-57) tend to change the bias in a way that increase the penalty on violated constraints in successive stages, thus forcing the stationary points toward feasibility.

Finally, the algorithm for solving the entire optimization problem is outlined below.

Step 0: Initialize the main iteration counter m=0. Set $\mathbf{x}^{(0)} = \hat{\mathbf{x}}$ and $\mathbf{q}^{(0)} = \hat{\mathbf{q}}$.

Step 1: Set the main iteration counter $m = m+1$.

Step 2: Solve problem 1 to obtain $\mathbf{x}^{(m)}$

i. Set the F-W iteration counter $s = 1$.

ii. Evaluate the gradients of P_1 using equation (6-26).

iii. Compute the auxiliary directions $\mathbf{y}^{(s)}$ using logical expression (6-27).

iv. Compute the optimal moving size λ^* using equation (6-35).

v. Update OD flows $\mathbf{x}^{(s+1)}$ using equation (6-28).

vi. Set the F-W iteration counter $s = s+1$. Check $\left\| \mathbf{x}^{(s)} - \mathbf{x}^{(s-1)} \right\| \le \varepsilon_x$, if satisfied go to vii, otherwise return to ii.

vii. Set $\mathbf{x}^{(m)}$ equal to the last F-W updates. Go to step 3

Step 3: Solve problem 2 to obtain $\mathbf{q}^{(m)}$.

i. Set the ALM iteration counter $z = 1$.

ii. Set the F-W iteration counter $s = 1$.

iii. Evaluate the gradients of the Lagrangian function L_2 using equation (6-47).

iv. Compute the auxiliary directions $\mathbf{y}^{(s)}$ using logical expression (6-48).

v. Compute the optimal moving size λ^* using equation (6-56).

vi. Update route fractions $\mathbf{q}^{(s+1)}$ using equation (6-49).

vii. Set the F-W iteration counter $s = s+1$. Check $\left\|\mathbf{q}^{(s)} - \mathbf{q}^{(s-1)}\right\| \le \varepsilon_q$, if satisfied go to viii, otherwise return to iii.

viii. Set the ALM iteration counter $z = z+1$. Check $\left\|\mathbf{q}^{(z)} - \mathbf{q}^{(z-1)}\right\| \le \varepsilon_q$, if satisfied go to x, otherwise return to ix.

ix. Update the multipliers using equation (6-57). Go to ii.

x. Set $\mathbf{q}^{(m)}$ equal to the last F-W updates. Go to step 4.

Step 4: Set the main iteration counter $m = m+1$. Check $\left\|\mathbf{x}^{(m)} - \mathbf{x}^{(m-1)}\right\| \le \varepsilon_x$ and

$\left\|\mathbf{q}^{(m)} - \mathbf{q}^{(m-1)}\right\| \le \varepsilon_q$ or $\left\|z(\mathbf{x},\mathbf{q})^{(m)} - z(\mathbf{x},\mathbf{q})^{(m-1)}\right\| \le \varepsilon_{obj}$, if satisfied go to step 5, otherwise go to step 2.

Step 5: End of the algorithm.

The use of the Block Coordinate Descent method presented above can be seen to make practical sense since both sub-problems can be fairly easily solved. Note that the proposed algorithm may converge to a local minimum, as with other conventional numerical methods. However, as noted by Lo et al. (1999), since both problem 1 and problem 2 are convex, the chance of obtaining the global solution is higher. Also, if the sample is large, and the model is unique and consistent, the sequence converges to the unique global solution.

6.4 PRACTICAL CONSIDERATIONS

As mentioned in the previous chapter, the Bayesian models in the first step of the proposed framework provide posterior parameters in multiple levels. This is a very important model feature, considering the fact that there is likely to be some OD pairs, routes, and links containing no observations during a particular interval of interest (on the target day). In such a situation, the corresponding parameters from the lower hierarchy (in the HB formulation) or from the empirical prior (in the EB formation), which are generally associated with a greater variance (thus reflecting less confidence on the estimate), can be used in the formulation presented in this chapter. Again, this procedure does not imply the exchangeability assumption of the data across different intervals but is a way to adopt the more vague information derived from the data in different intervals within the same similarity criteria.

The ability of the formulated optimization problem in limiting the deviations of the estimates with respect to the initial estimates should also be emphasized. This is particularly important for the cases where the problem is underdetermined and different solutions of route fractions and OD flows can reproduce the observed link flows equally well. In practice, the choice of bound settings is up to the modeler, although it is convenient to set these bounds as a function of the variance from the corresponding posterior distribution. Some other heuristic rules may also be applied. For instance, if the initial estimate is very small, bounds that constitute a larger interval can be used to include the possibility that the estimate is highly biased due to small or missing data.

6.5 SUMMARY

This chapter presented an optimization procedure, included in the second step of the proposed framework, for updating the initial estimates of OD table and assignment matrix with observed link flows during the study period on the target day. The method is based on the least squares model which does not require the samples or initial estimates to be in a specific probability distribution. Thus, the statistics from the Bayesian data analysis step can be directly used without imposing further assumptions. Since the problem may be underdetermined and multiple solutions may exist, a set of functional and bound constraints are introduced to preserve the structure of the initial estimates as well as to guarantee a meaningful solution.

Because the model involves a non-convex objective function and also a large number of decision variables, a solution algorithm based on the Block Coordinate Descent method is developed in order to expedite the search for the global optimum. A combined use of the Augmented Lagrangian method and the Frank-Wolfe algorithm, which results in a very memory-efficient scheme, is presented for solving the sub-problems.

The next chapter will conduct a comprehensive simulation study, based partly on the data observed from the field. This study is used to evaluate the methodologies for estimating the initial OD flows and assignment matrix proposed in the previous chapter and for updating these estimates proposed in this chapter.

CHAPTER 7 EVALUATION OF THE FRAMEWORK OF DYNAMIC DEMAND PREPARATION WITH VEHICLE TRAJECTORIES

7.1 INTRODUCTION

The procedures commonly used for evaluating an OD estimation model can be generally grouped into two data approaches-- field observation or synthetic data. The first approach tests the OD estimates in terms of how close they can replicate observed link counts once assigned to the network. Since the ground true OD flows are not known, it is not possible to evaluate the quality of the estimates directly. By assigning *assumed ground truth OD flows* to the network, the second approach generates a set of synthetic link counts as the evaluation inputs. The resulting OD flows from the model are then compared against the assumed ground truth.

As mentioned in subsection 2.2.2, some assumptions pertaining to the traffic assignment, travel times in the network, and/or the availability of reliable seed OD flows are usually made in model evaluation. For instance, studies using the synthetic data often create seed OD flows by adding some random distortions to the assumed ground truth OD flows. Therefore, there exist high dependencies between the model performance and the details of the evaluation framework, although this fact is often not reported. Because the proposed framework includes the modules for estimating the seed (initial) OD, route, and traffic patterns, which are also subjected to the evaluation, a new comprehensive

simulation experiment will be conducted in this chapter to achieve the following objectives.

1) To illustrate the efficacy of the proposed framework under a setting that differs drastically from the hypothesized conditions using a real-network
2) To quantify the relative benefits of the information from small trajectory samples (say, with sampling rates of less than 5 percent) collected over multiple days in estimating the initial OD, route, and traffic patterns through the use of the proposed Bayesian models
3) To quantify the relative benefits of the proposed optimization program in updating these initial estimates with the observed link flows

The rest of this chapter is organized as follows. Section 7.2 describes the experimental design including demand generation, traffic and route simulation, and data collection. The results from the experiment are presented in section 7.3. Then, section 7.4 discusses some important aspects related to the interdependencies between the testing results and the simulation settings. The last section provides the conclusions and some recommendations.

7.2 EXPERIMENTAL DESIGNS

The proposed methodologies presented in the previous two chapters are tested using a freeway-arterial network with the total length of 37 miles, located in Irvine, California

(see Figure 7-1). This network consists of 1089 OD pairs (zones shown), 198 loop detectors, and 38 actuated signal intersections.

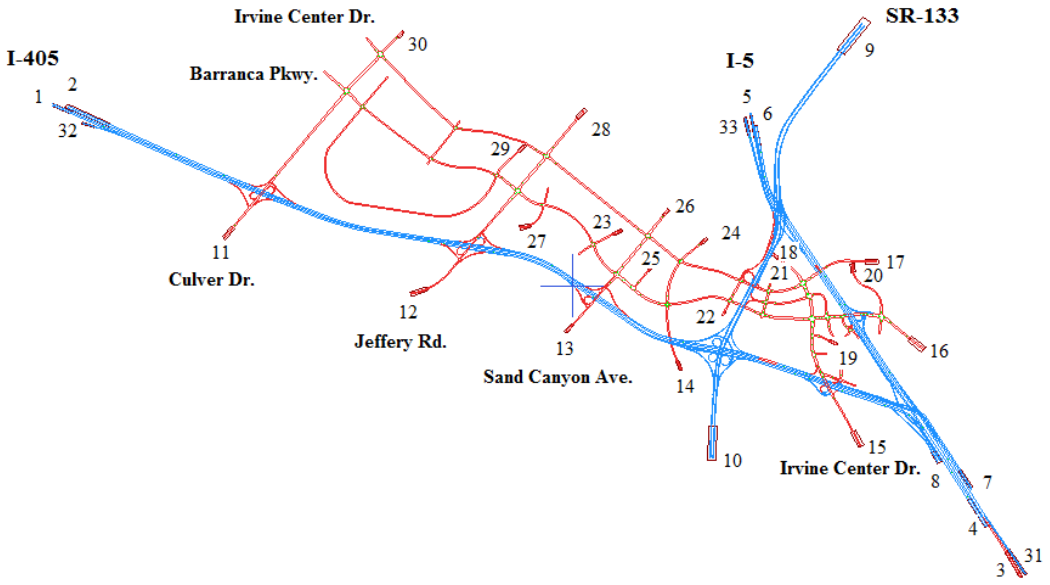


Figure 7-1 A freeway-arterial network in Irvine, California

Because the underlying dynamic OD tables, link travel time distributions, route-link fractions, and route-choice fractions required for the model validation are not available, a simulation study is necessary. A set of formerly-estimated 5-minute OD tables during 7:30AM to 9:30AM, based on the traffic data on May 22, 2001, are used as the seed to generate a series of daily dynamic OD trip tables for 21 days. This is done as follows. For each OD cell, a uniform distribution with the lower and upper limits of 20 percent from the seed value is constructed. Then, daily dynamic OD tables are independently drawn from these distributions. It should be noted that the daily demand fluctuation may be better represented by the Poisson, multivariate normal, or multinomial

distribution. However, this study deliberately uses the uniform distribution, which concentrates less around the mean, in order to evaluate the model performance among higher fluctuation using a distributional form that is radically different from the one hypothesized in the models. Mixtures of vehicle and driver types are determined also in a similar manner.

Then, a microscopic simulation, Paramics, is used to simulate the traffic on each day. In the simulations, drivers are assumed to select their perceived shortest route. To reflect their general knowledge about the traffic conditions and to generate multiple routes per OD pair, the mean route costs are updated every interval, and the random term for each individual is set moderate. During 7:30AM to 9:30AM on each day, the network serves an averaged total of 77476 vehicles, and the traffic conditions vary from the free-flow condition to heavily congested condition with the maximum travel time in the network of 42 minutes. There are about 2952 total dynamic routes. The maximum number of routes per OD pair per interval is 16, and the average number of routes per non-zero OD pair per interval is 1.94. For the validation purpose, the trajectories of all vehicles including vehicles' origin and destination, link travel times, and route-choice fractions are saved.

Next, the study simulates a random sampling scheme of collecting trajectory data with small-unknown rates varying over time and space, by drawing a random rate for each origin zone during each 5-minute interval independently from a uniform distribution with the limits of 0 and 5 percent. Without loss of generality, the 21st day is regarded as the target day for which the OD estimates are required. The sampled trajectory data from all 21 days are used in the estimation of the initial OD flows, route-link fractions, and

route-choice fractions on the target day using the models presented in CHAPTER 5. The OD estimation with stochastic route-choice fractions proposed in CHAPTER 6 is then used to further update these initial estimates based on link counts observed on the target day. Figure 7-2 summarizes the evaluation framework.

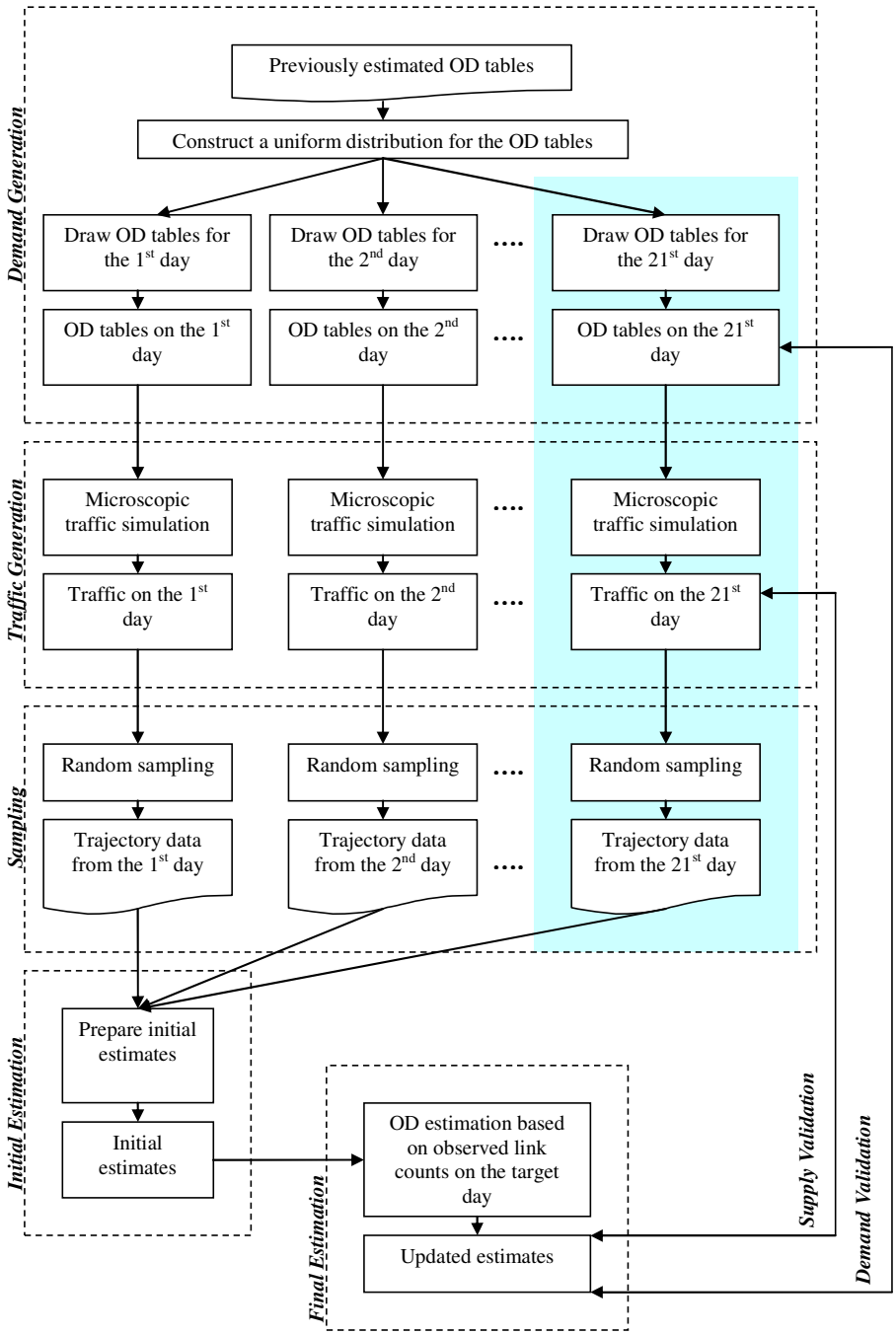


Figure 7-2 Evaluation framework

7.3 EXPERIMENTAL RESULTS

In subsection 7.3.1, the estimation results of the link travel time distributions (section 5.3) and route-link fractions (section 5.6) will be presented. The estimation results of the initial route-choice fractions (section 5.4) and OD flows (section 5.5 and section 5.7) will be presented in subsection 7.3.2, where they can also be compared with the corresponding final estimates (section 6.2 and section 6.3).

7.3.1 Travel time distributions and route-link fractions

For each freeway link, interval-specific travel time distributions are estimated provided that there are at least a few trajectories observed during that interval. Because detectors installed on the arterials usually do not provide occupancy, time-of-day is used instead as the similarity criteria in constructing the prior distributions for estimating travel time distributions on the arterials. Three Markov chains, each of which includes 3000 simulations, are employed. The posterior distributions are taken based on the last 1500 simulations from each chain. Figure 7-3, 7-4, and 7-5 show selected estimates of the travel time distributions on a three-link section on I-5, I405, and SR-133, respectively.

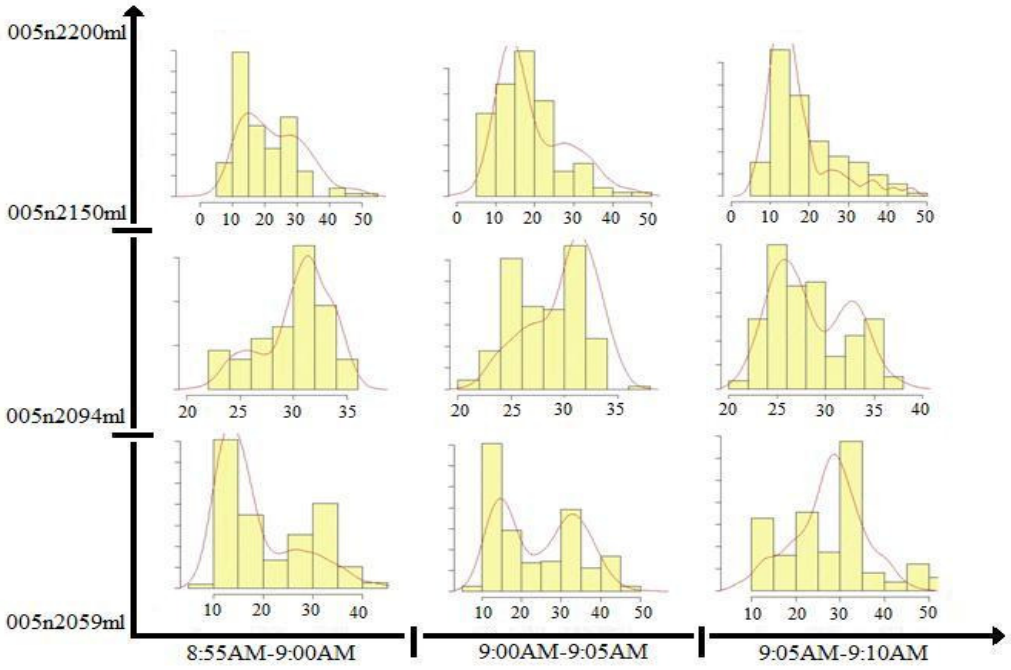


Figure 7-3 Travel time distributions on a three-link section on I-5

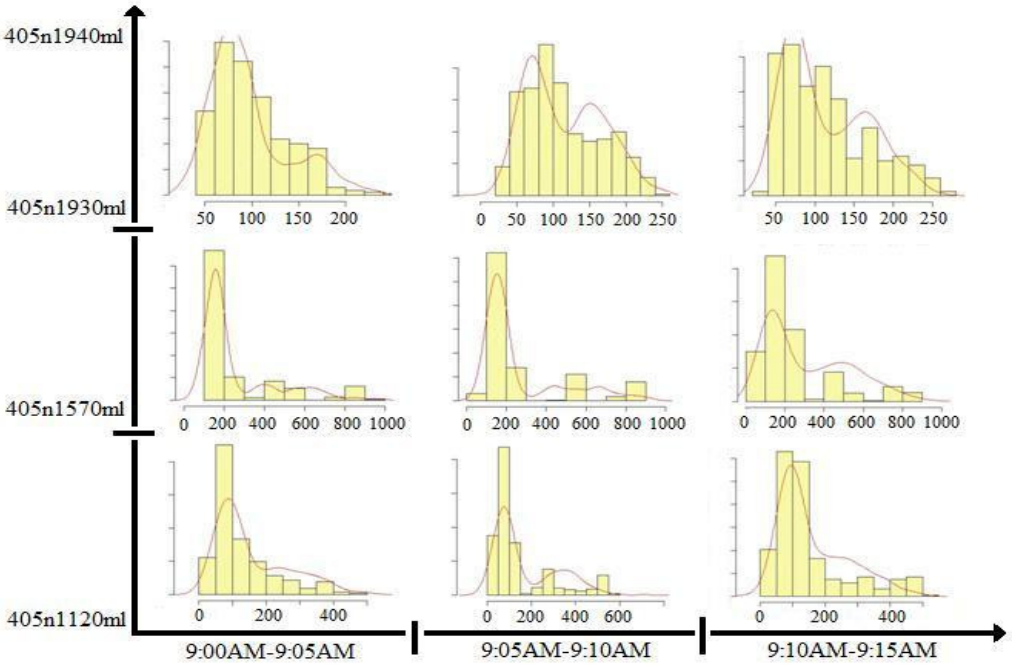


Figure 7-4 Travel time distributions on a three-link section on I-405

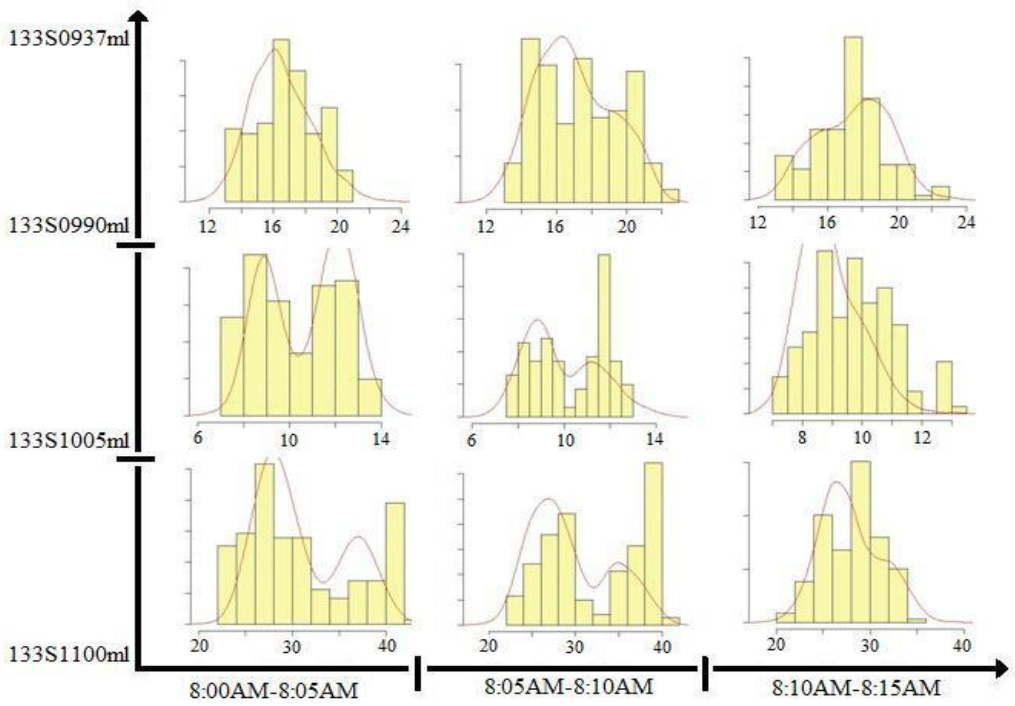


Figure 7-5 Travel time distributions on a three-link section on SR-133

It can be seen that the estimated distributions (line) approximate the underlying distributions (bar) very well. From an extensive investigation of the travel time distributions on different links during different time intervals, two main conclusions can be drawn. First, even within the same traffic condition group, each link during each distinct interval might have its travel time distribution significantly different from one another due to interval-specific factors, such as the mixtures of driving behaviors and of vehicle types. Second, link travel time distributions tend to be either skewed-unimodal or multimodal, depending on both traffic condition and mixing proportions.

It is also helpful to evaluate the performance of the proposed Bayesian mixture model in term of estimating the average travel times. Compared to the ground truth, the Mean Absolute Relative Error (MARE) and Root Mean Square Error (RMSE) of the

estimated mean travel times on all links and all intervals is 0.0263 and 0.0398, respectively. These statistics show that the proposed model outperforms the Maximum Likelihood approach (ML) using the data observed solely during each interval, which yields MARE and RMSE of 0.125 and 0.162, respectively.

The estimated travel time distributions are then used in inferring the underlying route-link fractions via the Monte Carlo simulation outlined in subsection 5.6.2. Note that for some links and intervals that contain no observation of vehicle trajectory and the interval-specific distribution is not available, the common travel time distribution of the traffic condition group corresponding to the observed occupancies is used instead. Table 7-1 compares the estimated route-link fractions based on three different approaches with the ground truth. In the table, *Simplified-DNL* refers to the discrete packet approach suggested by Cascetta et al. (1993) which moves vehicles based on mean travel times, *O-DNL I* refers to the proposed Monte Carlo simulation approach with independent link travel times, and *O-DNL II* refers to the proposed Monte Carlo simulation approach with the correlations of the same drivers on different links. Both O-DNL I and O-DNL II are taken as the average of 30 iterations, each of which simulates 1000 vehicles per route. For O-DNL II, three regions for each travel time distribution are assumed. Since the conditional probabilities for each region are not known, it is assumed that each transferring vehicle has equal probabilities of switching to another region of 0.1 each and staying in the same region of 0.8.

Table 7-1 Errors in route-link fractions

	RMSE (all data)	MARE (all data with non-zero actual value)	MAE (all data with zero actual value)
Simplified-DNL	0.163	0.333	0.071
O-DNL I	0.102	0.211	0.072
O-DNL II	0.100	0.204	0.074

It is very clear from the comparison that the proposed methods (O-DNL I and O-DNL II) yield much more accurate results than the simplified-DNL method. However, it is interesting to note that O-DNL II performs only slightly better than O-DNL I although the first can take into account the correlations of the same drivers on different links. This could be because of two main reasons. First, the assumed conditional probabilities of the travel time regions might deviate greatly from the true values, which vary according to the traffic conditions and the characteristics of link connections. Second, a significant portion of the error in both cases might indeed come from the errors in estimated travel time distributions.

7.3.2 Route-choice fractions and OD flows

The initial route-choice fractions and destination fractions are estimated using the multinomial-Dirichlet model proposed in section 5.4 and 5.5, respectively. For each origin zone and time interval that has no trajectory data, the prior parameters derived empirically from the data from similar intervals are used instead. In estimating the initial OD flows, since all zones are located along the network boundary, the estimation of departure flows is not required and the OD estimates can be determined directly using the estimated destination fractions and link flows out from the origin zones.

Within the estimation period (8:00AM to 9:00AM), there are a total of 13068 dynamic OD flows and 2167 dynamic route fractions. In order to evaluate the proposed estimation procedure used in updating these initial estimates with the observed link counts on the target day (section 6.2 and section 6.3), a warm-up and clearance period of 30 minutes are first applied to remove the effects of the OD flows outside the estimation period. During these two special periods, the underlying OD flows are assigned to the network based on the true assignment matrix, and their contributions to the link counts within the study period are eliminated accordingly. The weight associated with each route-choice fraction is calculated using equation (5-9) and the weight associated with each OD flow is calculated using equation (5-22). As with previous studies, it is assumed that the standard deviation of each link count datum is one percent of its mean. The maximum distortions (bound constraints) are set to 20 percent from the initial estimates. However, if the initial OD flow or route link fraction is close to zero, the maximum bound is set to 15 and 0.10, respectively. In practice, these restrictions are up to the modeler.

The plots comparing the ground true route-choice fractions versus the estimates from the Maximum Likelihood approach based on the data collected only on the target day (RAW), versus the estimates from the multinomial-Dirichlet model (BAY), and versus the final update using the proposed dynamic OD estimation (UPDATED) are provided in Figure 7-6, 7-7, and 7-8, respectively.

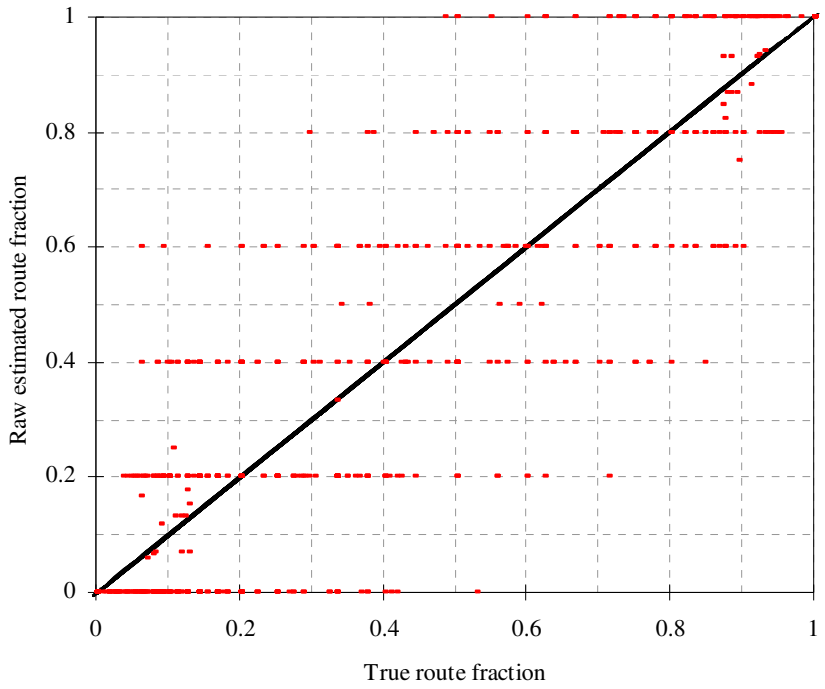


Figure 7-6 Comparison between true and raw estimated route fractions (RAW)

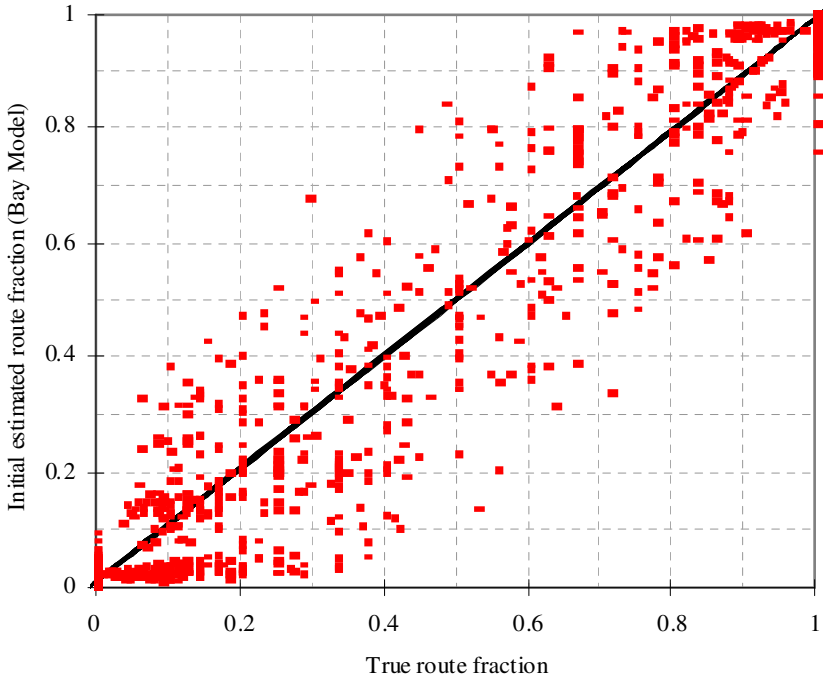


Figure 7-7 Comparison between true and initial estimated route fractions (BAY)

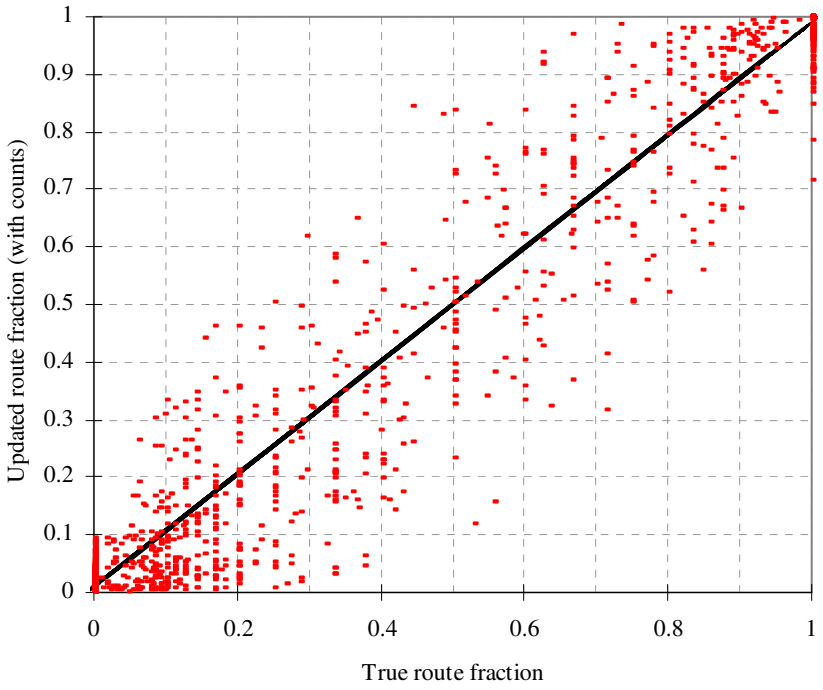


Figure 7-8 Comparison between true and updated route fractions (UPDATED)

Table 7-2 illustrates a comparison among these methods, based on RMSE, MARE, and MAE.

Table 7-2 Errors in route-choice fractions

	RMSE (all data)	MARE (all data with non-zero actual value)	MAE (all data with zero actual value)
Raw estimated route fractions	0.100	0.311	0.000
Bay estimated route fractions	0.074	0.204	0.022
Updated route fractions	0.073	0.194	0.039

Similarly, the plots for comparing the true OD flows versus the estimates from these methods are given in Figure 7-9, 7-10, and 7-11, respectively.

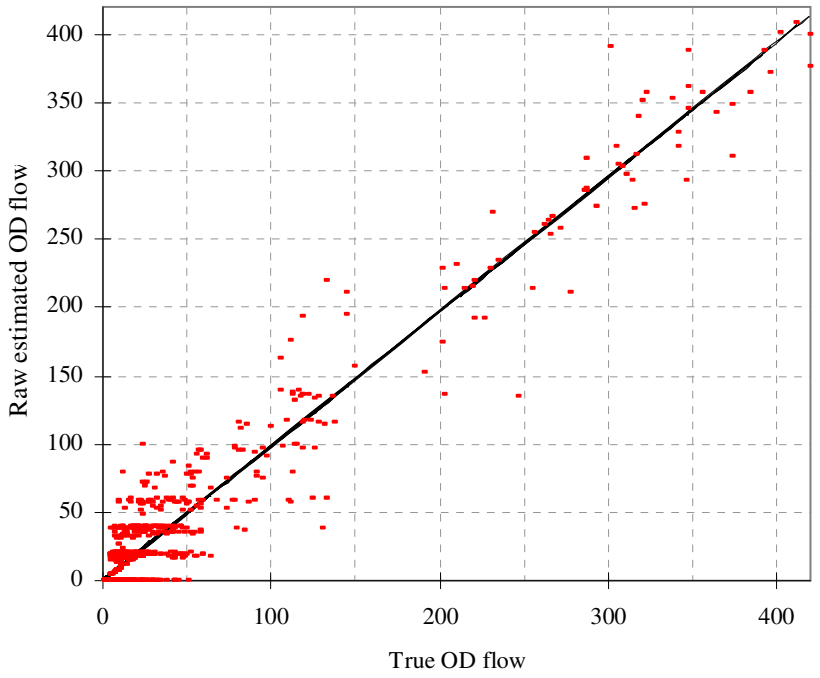


Figure 7-9 Comparison between true and raw estimated OD flows (RAW)

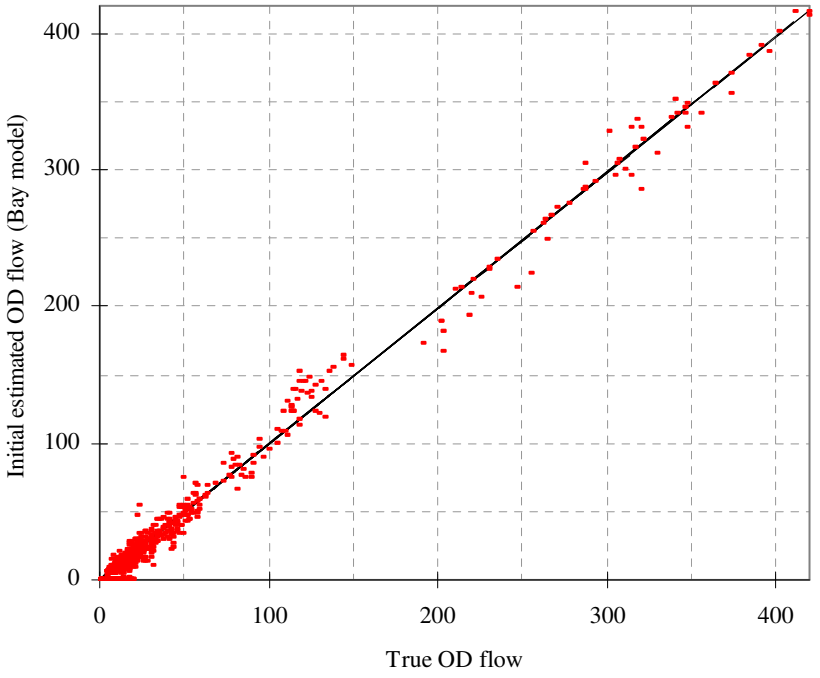


Figure 7-10 Comparison between true and initial estimated OD flows (BAY)

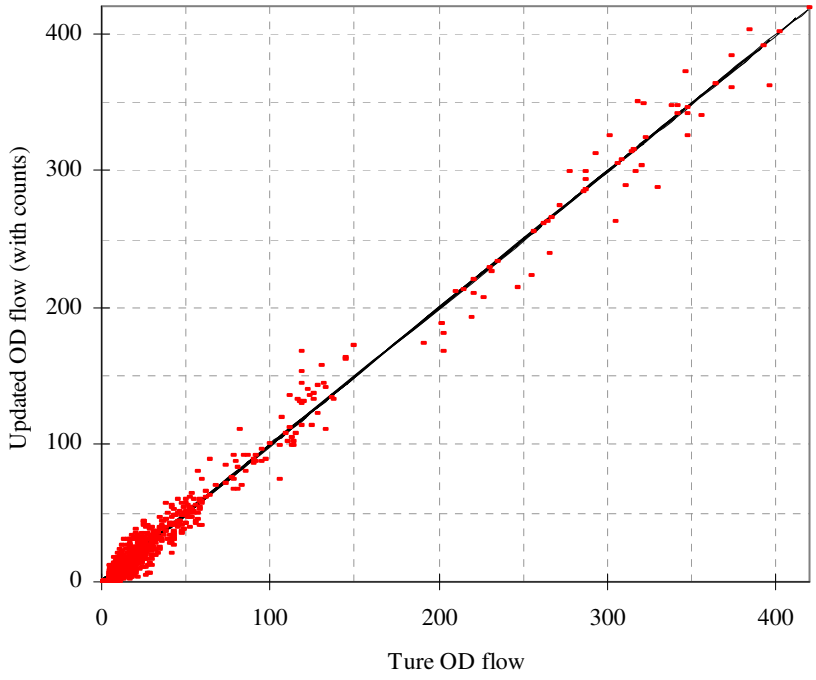


Figure 7-11 Comparison between true and updated OD flows (UPDATED)

Also, Table 7-3 illustrates a comparison among these methods, based on RMSE, MARE, and MAE.

Table 7-3 Errors in OD flows

	RMSE (all data)	MARE (all data with non-zero actual value)	MAE (all data with zero actual value)
Raw estimated OD flows	5.357	0.833	0.000
Bay estimated OD flows	2.109	0.414	0.000
Updated OD flows	1.975	0.271	0.000

Significant improvement in the estimation accuracies by the multinomial-Dirichlet model over the usual ML approach with one-day data in estimating route-choice fractions can be appreciated by comparing Figure 7-6 versus Figure 7-7 and the first two rows in

Table 7-2. The same benefit in terms of estimating the initial OD flows is also evident from the comparison of Figure 7-9 versus Figure 7-10 and the first two rows in Table 7-3. One of the model's vital features that should be pointed out is its ability to moderate extreme estimates toward the means; for instance, among all vehicles departing from origin 29 during 8:15AM to 8:20AM, the ML method with one-day data finds only 3 out of 6 destinations and 5 out of 11 used routes with non-zero flows while the multinomial-Dirichlet model captures all of them.

Although the additional improvement from the programming (6-10)-(6-13) are not obvious by comparing the Figure 7-7 versus Figure 7-8, or Figure 7-10 versus Figure 7-11; the last two rows of Table 7-2 and Table 7-3 indicates, based on the RMSE and MARE values, the marginal benefit in the estimates of route-choice fractions and the significant benefit in the estimates of OD flows, respectively. The reason for the small improvement on route-choice fractions compared to the OD flows may be that the data used in estimating the initial route-choice fractions are much smaller than that of the initial OD flows.

Lastly, a supplemental evaluation of the estimated OD tables and assignment matrix based on how well these two sets of estimates can reproduce the observed link counts is provided in Table 7-4. The table compares, based on the three error measures, the observed link counts versus the estimated values from initial OD tables (with initial assignment map), and versus the estimated values from the updated OD tables (with updated assignment map). As shown, the great improvement due to the updating procedure should be appreciated.

Table 7-4 Errors in counts

	RMSE (all data)	MARE (all data with non-zero actual value)	MAE (all data with zero actual value)
Assigning initial estimated OD flows	15.987	0.301	0.131
Assigning updated OD flows	10.142	0.168	0.197

7.4 SOME CONSIDERATIONS ABOUT THE EXPERIMENTAL DESIGN

Recall that the proposed Bayesian models infer the parameter posterior distributions based on the likelihood from the data collected during the interval of interest and the prior distributions, which are derived either hierarchically or empirically from the data observed during similar intervals. Different similarity criteria were suggested for different parameters-- the traffic condition group is used for estimating the travel time distributions, and time-of-day for estimating the destination, route-choice, and departure fractions. It is thus instructive to discuss in more details how the experiment generates the recursive patterns of the so-called similar intervals, which to a great extent affects the models' performance. The relevant aspects pertain to

- The generation of daily demand patterns
- The generation of route and traffic patterns
- The generation of trajectory samples

As mentioned, a set of uniform distributions is used in the experiment to represent the fluctuation of OD flows from day-to-day instead of the Poisson, multivariate normal or multinomial distribution, which are commonly assumed in the literature. This is due to two main reasons. First, while it is reasonable to expect some recursive demand pattern from day-to-day, there has been no empirical evidence regarding the distributional forms which may actually vary among different OD pairs and time intervals. To reflect the lack of such knowledge and to prevent unnecessary systematic biases in the evaluation, it is constructive to use the uniform distributions. Second, the use of the Poisson, multivariate normal or multinomial distribution would have led to more favorable results since these distributions are closer to the assumption made in the model such that the mean destination fractions from each origin zone over multiple days are distributed according to the Dirichlet distribution. In order to test the robustness of the entire framework, a distribution that is radically different from the one hypothesized should be used.

In the simulations, the explicit mechanism used to model the stochastic properties of route selection includes some moderate randomness into each individual's perceived route costs. However, it should be noted that this is not the only factor resulting in the fluctuation of the route fractions from day-to-day. Different demand tables used each day also result in the fluctuation of the traffic and mean route costs, which in turn lead to different route patterns. Again, regardless of how this procedure replicates the real-world situation, it creates an environment that is quite different from the hypothesized setting. In terms of the traffic modeling, it is expected that the real-world travel time data tend to be more stochastic (thus greater variance) than the data generated from the microscopic

traffic simulation model. However, the ability of the Bayesian mixture model to fit various distributional shapes should remain largely valid.

Lastly, the only sampling scheme presented in the experiment is based on random sampling conducted over multiple days from the entire population. In practice, depending on the market penetration of the tracking devices, different sampling procedures might be implemented. For example, it might be possible to obtain the trajectory data from the road users who have a tracking device. In this case, the random sampling is only possible over a subset of the population. On the other hand, if the necessary equipment are to be distributed to the sampling units in an alternating manner, it can be assumed that the random sampling is performed over the population, similar to the scheme used in the experiment. Note that the combination of these two cases is also possible.

In short, the simulation study presented in this chapter is by no means set up such that one can draw the empirical conclusions, especially regarding the relationship between the estimation accuracies and the sample rates assumed. Instead, it is designed to evaluate the robustness of the framework under the conditions that differ greatly from the assumptions in the models.

7.5 SUMMARY

A comprehensive simulation study on a real network was conducted in this chapter to test the methodologies proposed in the previous two chapters. With the simulation experiment, explicit assumptions regarding the patterns of the underlying OD flows,

traffic movements, and route-choice fractions, which may overestimate the models' performance, were avoided.

Based on the trajectory data from a random sampling over 21 days with random sampling rates of up to 5 percent and the traffic data from loop detectors, the simulation experiment indicated the applicability of the proposed framework with very promising results. In terms of the traffic estimation, thanks to the flexibility of the proposed Bayesian mixture model and the Occupancy-based Dynamic Network Loading (O-DNL) model, complex relationships between route and link flows can be efficiently captured without assuming a certain form of travel time distribution. The combined use of these two models also provides a convenient way to exogenously bring in the occupancy measurements to the OD estimation procedure, which in turn overcomes several problems from the non-monotonic relations between path flows and link counts.

The main benefits of the Bayesian multinomial-Dirichlet model for estimating the initial route-choice and destination fractions are threefold. First, the model is suitable with several possible sampling schemes where the route-set library is to be constructed from small samples in multiple days. As shown in the experiment, the route-set library created by this method is very satisfactory. Second, the model can moderate extreme estimates due to low sampling rates toward the prior means with the shrinking magnitude based on the likelihood precision. Third, the variances associated with these initial estimates are readily available and can be directly used in the proposed optimization problem. With this regard, weights placed on the historical and current data are effectively determined. Moreover, the impact from small estimates on unused routes or destinations does not appear to be significant after the updating procedure. Finally, as

indicated from the experiment, the updating procedure for OD flows and route fractions using the information from the observed link counts can further improve the quality of the initial estimates.

Future studies should include the investigation of different sampling schemes with different sampling rates. Some variations of the simulation study presented here are also possible to analyze the models' sensitivity. To enhance the performance of the Occupancy-based Dynamic Network Loading (O-DNL) model, video data are needed to obtain the empirical conclusions regarding the relationship between the link configurations and traffic conditions. Lastly, route-link fractions may enter the updating module as decision variables although it is difficult to set the constraints to ensure practical solutions.

CHAPTER 8 CONCLUSIONS AND RECOMMENDATIONS

8.1 INTRODUCTION

This chapter reports the concluding remarks of the dissertation and suggests the directions for future research. Overall conclusions and discussions are given in the next section. Section 8.3 summarizes specific conclusions from the experiments conducted with each proposed framework. The author's perspective on the contributions of the research to the state-of-art on dynamic demand preparation is then given in section 8.4. Section 8.5 discusses further extensions and directions for future research in this area.

8.2 OVERALL CONCLUSIONS

In addition to solving an inverse problem, link count based OD estimation often involves a series of approximations under some assumptions regarding the information in the prior OD flows, route-choice behavior, and traffic dynamics, which are interrelated such that one interactively affects all others. However, developing a complete framework for estimating dynamic OD flows, rather than solely an estimation model, has not been the primary goal of most current studies in the literature. It is not surprising that the lack of a

deployable framework has led to the situation where many incorrect ad-hoc schemes are commonly adopted in practice.

Recent advances in AVI, GPS, and cellular phone tracking technologies make available high-fidelity probe data which potentially provide the information about the underlying OD flows. Several recent studies focus on incorporating the sampled OD flows into OD estimation formulations. Most of these studies employ constant (over time and/or space) expansion factors on the observed OD flows although the expected market penetration that is currently available or expected to be available in the near future is too low for such a simple expansion technique to be suitable. While probe data can be used to bring in observations of the underlying route-choice fractions and traffic dynamics to the estimation formulation, these possibilities have not been sufficiently investigated in the literature.

This dissertation itemizes and examines these critical issues. Then, it proposes two new frameworks along with necessary mathematical models and solution algorithms for preparing dynamic demand inputs for planning (off-line) applications. The first framework focuses on the use of traffic simulation models and the estimation module based solely on traffic counts which are readily available in most urban areas. Under this framework, the traditional planning model is augmented with a filter traffic simulation step, which captures important spatial-temporal characteristics of route and traffic patterns within a large surrounding network, to improve the flow estimates entering and leaving the final microscopic simulation network. These inbound and outbound flows serve as seed OD flows used in a further estimation procedure. A new bounded dynamic OD estimation model which minimizes the deviations between observed and estimated

counts and the deviations between seed and estimated OD flows along with a new solution algorithm suitable for a large problem is also suggested.

The second framework incorporates additional information from multiple sets of small probe data continually collected over multiple days. There are two steps under this framework. The first step includes several innovative empirical and hierarchical Bayesian models for estimating time-dependent travel time distributions, destination fractions, and route distributions from probe data. These models provide multi-level posterior parameters and tend to moderate extreme estimates toward the overall mean with the magnitude depending on their precision. Such features overcome several critical problems due to non-uniform (over time and space) small sampling rates. The second step involves a construction of initial OD flows, an estimation of route-link fractions from estimated travel time distributions, and an updating procedure using a new estimation formulation which adjusts OD flows, route fractions, and route-link fractions simultaneously.

8.3 SPECIFIC CONCLUSIONS FROM THE EXPERIMENTAL STUDIES

To facilitate the presentation, this section summarizes specific conclusions drawn from the experiments conducted with each of the two proposed frameworks, respectively.

8.3.1 Six-step model

Several conclusions are drawn from the experiments performed and are summarized below.

- In all experiments, the 5th step can improve the quality of seed dynamic OD flows and route-link fractions in the subarea significantly although the mesoscopic simulation is based on simple traffic flow models without microscopic details. This underscores the usefulness of developing a mesoscopic model network before cutting out sub-area seed OD tables to perform further OD estimation within the microscopic level.
- Despite a significant difference between the assignment procedure used in generating the ground truth and the one in preparing seed dynamic OD tables in the mesoscopic simulation, the proposed framework provides improved results compared to the use of the static traffic assignment. This is because the mesoscopic model provides a more reasonable route pattern which takes into account traffic dynamics as well as supply capacities.
- The proposed framework provides much more accurate final OD estimates in terms of both being closer to the ground true OD flows and yielding a better fit to the objective function. In all experiments conducted, the benefit from the 5th step (i.e., the improvement in seed OD tables) is greater than or at least equal to the benefit from the OD estimation performed within the microscopic level.
- The F-W algorithm suggested for solving the model's upper level converges fairly fast to a reasonable point before the next traffic simulation update in all cases and

could be applied to large networks successfully. However, as with other iterative bi-level algorithms, the entire algorithm might be trapped with a local solution.

- Based on both experimental studies and the application to the I-880 corridor, the 5th step of the proposed framework yields seed dynamic OD tables, which lead to a much better convergent sequence.
- Fluctuations in the optimization can occur due to changes in the assignment matrix between two successive iterations. Because of the non-monotonic relation between path and link flows and the capacity restrictions in the traffic simulation, the OD estimation algorithm cannot effectively update assignment matrix and OD flows if their values in the previous iteration deviate greatly from a reasonable solution.

8.3.2 OD estimation with small trajectory samples

Similarly, several conclusions can be drawn from the experiments performed with the second framework and are summarized below:

- The Bayesian mixture model provides estimated link travel time distributions sufficiently close to the underlying one in most cases. It should be noted that even within the same traffic condition group, each link during each distinct interval might have its travel time distribution significantly different from one another due to interval-specific factors, such as the mixtures of driving behaviors and vehicle

types. In addition, link travel time distributions tend to be either skewed-unimodal or multimodal.

- The Occupancy-based Dynamic Network Loading (O-DNL) methods outperform the discrete packet approach suggested by Cascetta et al. (1993). Since route-link fractions are derived from field data and more importantly exogenously from the OD estimation model, improper OD adjustment caused by the non-monotonic relationship between traffic counts and OD flows can be avoided.
- The main benefits of the multinomial-Dirichlet model for estimating destination fractions and route-choice fractions are threefold. First, it provides much more accurate estimates compared to the maximum likelihood approach using the data observed only during the interval. Second, it is suitable with various sampling schemes that construct route sets from the observations from multiple days. Third, variance-covariance matrices of the estimates, which indicate the consistency between the prior and the likelihood of the current data, are readily available.
- In addition to the improvement in the estimation accuracy, the Bayesian models feature multi-level posterior parameter distributions that can be used in the OD estimation step. Common parameters, which are typically associated with greater variance, can be used in the case of missing data. This is an important model feature given that there are many OD pairs, routes, and links that contain no observation during the interval of interest.
- In the experiment, the proposed OD estimation formulation with stochastic assignment matrix provides marginal improvement in the estimates of route-choice fractions and significant improvement in the estimates of OD flows. The

reason for only a small improvement in route-choice fractions, compared to the OD flows, could be that the data used in estimating the initial route-choice fractions are much smaller than that of the initial OD flows.

8.4 RESEARCH CONTRIBUTIONS

The dissertation examines the current state-of-art on link count based OD estimation models proposed in the literature and investigates different common schemes currently adopted in practice. Then, it develops and evaluates two novel frameworks to be used with different data sources. The main contributions from each framework can be summarized as follows.

8.4.1 Six-step model

- The proposed six-step process offers a systematic procedure for preparing dynamic OD demand inputs for microscopic simulation under the planning context. Since the framework does not assume the availability of seed OD tables, route-choice, and/or travel time data; and the OD estimation module is based only on observed link counts; they are applicable to the current real-world applications.

- The proposed dynamic OD estimation model can incorporate weights which reflect relative reliabilities of the data from different sources and the modeling at different levels. Moreover, it includes bound constraints that further regularize the estimation.
- The proposed solution algorithm is memory efficient and can be applied to a large scale problem.

8.4.2 OD estimation with small trajectory samples

- The proposed framework offers an effective procedure for extracting dynamic OD demand and route distribution patterns from probe and traffic count data.
- The proposed Bayesian models can estimate time-dependent link travel time distributions, destination fractions, and route-choice fractions from small probe samples collected over multiple days and thus circumvent several difficulties due to small samples.
- The Occupancy-based Dynamic Network Loading (O-DNL) models proposed for constructing time-dependent route-link fractions provide a key capability to take into account traffic dynamics without placing any restrictive assumptions on the travel time distributions. More importantly, since time-dependent route-link fractions can be estimated before entering the OD estimation process, several difficulties due to the non-monotonic relationship between route and link flows can be avoided.

- The proposed dynamic OD estimation model takes into account the stochastic properties of the initial estimates of OD flows, route-choice fractions, as well as route-link fractions. Moreover, it incorporates a set of constraints used in both regularizing the problem and ensuring a meaningful solution.
- The proposed solution algorithm is memory efficient and suitable for solving a large scale problem.

8.5 FUTURE RESEARCH

Several research extensions can be considered imminent to the presented methodologies. The following improvements to each framework are suggested:

8.5.1 Six-step model

- All experiments presented in this dissertation assume that supply parameters in traffic simulation models have been calibrated and thus fixed constant over all iterations of the algorithm. In practice, the model calibration is usually performed simultaneously with the OD estimation process. This could affect the convergence of the algorithm and thus should be further investigated.
- Feedback loops among the steps may have to be considered, just as within the traditional four-step process, for proper modeling consistency. However, the joint

analytical models, which are possible for the four steps, cannot be easily developed unless analytical DTA models are used for the dynamic modeling steps.

- To improve the OD estimation module used in the framework, occupancy and speed may also be considered as supplemental variables. Since the measurements of these quantities are more informative in term of indicating traffic conditions, they can help overcome the difficulties due to the non-monotonic relationship between path and link flows.
- Other solution algorithms for solving bounded problem such as the projected gradient can also be considered.

8.5.2 OD estimation with small trajectory samples

- To improve the Bayesian mixture model, other observable factors affecting travel time distributions such as weather conditions (e.g., rain, snow, fog, and light), and proportions of trucks in different time intervals of day, etc. can be easily considered by adding more constraints when grouping travel time data. Also, the number of mixture components can be modeled as unknown parameter.
- The Bayesian multinomial-Dirichlet model for estimating OD and route distributions can be extended to include independent variables to obtain a general model which can be applied to different sites.

- The complete OD estimation model which considers route-link fractions as decision variables in the optimization should be tested, in which case, the problem should be decomposed into three sub-problems.
- Although the proposed framework includes the posterior updating procedure (see Figure 5-1) which corresponds to the case of real-time OD estimation, a numerical example to illustrate such applications is absent as the dissertation focuses mainly on the off-line case. However, such an extension can be straightforwardly performed.

APPENDIX A Bayesian Mixture Modeling Concepts

In the situations where the data measurements are taken under different conditions or from different sub-populations, standard distributions often fail to sufficiently describe the different aspects of data. To capture such complexity, mixture models constitute unknown distributional shapes as a convex combination of multiple distributions, which can be expressed as

$$f(y \mid \theta, q) = \sum_{k=1}^{K} q_k f_k(y \mid \theta_k)$$

$$\sum_{k=1}^{K} q_k = 1$$

$$0 \le q_k \le 1 \quad (k = 1, 2, ..., K)$$

The quantities q_k and $f_k(y \mid \theta_k)$ are respectively the mixing proportion and the probability density function (PDF) for the k^{th} component. In a special case where all mixture components are normal, the mixture distribution is a weighted average sampling distribution with vector q as a description of the variation in parameter vector θ across the entire population. From the Bayesian paradigm, each data point y_i is modeled conditionally on an unobserved indicator variable ξ_i, which is itself assumed to have a multinomial PDF with the proportion vector q as hyper-parameter.

The advantage of Bayesian modeling is that it allows the prior PDF, $f(\theta)$, that may come from previous studies or expert knowledge, to be combined with the data likelihood, $L(y \mid \theta)$, in calculating the posterior distribution through the relation:

$$f(\theta \mid y) \propto L(y \mid \theta).f(\theta)$$

The hierarchical Bayesian approach further allows more complex structures of the prior specification in that hyper-parameters that describe the probability distribution of any lower-hierarchy model parameter can themselves be given probabilistic distributions. The posterior distribution can be calculated through the relation:

$$f(\theta, \phi \mid y) \sim L(y \mid \theta).f(\theta \mid \phi).f(\phi)$$

The use of hierarchical Bayesian in specifying the prior is a very powerful mechanism to incorporate information from different data sets or previous studies.

APPENDIX B Augmented Lagrangian Function (ALM)

Consider the generic form of constrained optimization problem below.

$\min f(\mathbf{x})$

subject to

$g_j(\mathbf{x}) \geq 0 \qquad \forall j = 1,2,...J$

$h_k(\mathbf{x}) = 0 \qquad \forall k = 1,2,...K$

The expression for ALM (Doblas and Benitez, 2005) is the following:

$$L(\mathbf{x},\upsilon,\zeta) = f(\mathbf{x}) + C\sum_{j=1}^{J}\left\{< g_j(\mathbf{x}) + \upsilon_j >^2 - \upsilon_j{}^2\right\} + C\sum_{k=1}^{K}\left\{[h_k(\mathbf{x}) + \zeta_k]^2 - \zeta_k{}^2\right\}$$

where

$$< \psi > = \begin{cases} \psi, \; for \;\; \psi < 0 \\ 0, \; for \;\; \psi \geq 0 \end{cases}$$

C = the scale factor.

The method consists of minimizing the lagrangian function until an estimate is reached. Denote a sub-problem by s. The multipliers are updated using the following equations.

$$\upsilon_j{}^{(s+1)} = < g_j(\mathbf{x}^{(s)}) + \upsilon_j{}^{(s)} > \qquad \forall j = 1,2,...,J$$

$$\zeta_k^{(s+1)} = h_k(\mathbf{x}^{(s)}) + \zeta_k^{(s)} \qquad \forall k = 1,2,...,K$$

Parameters $\upsilon,$ and ζ serve as a bias in the arguments of the penalty terms, and the updating rules tend to change the bias in a way that increase the penalty on violated constraints in successive stages, thus forcing the stationary points toward feasibility.

REFERENCES

Antoniou, C., Ben-Akiva, M., Koutsopoulos, H., 2004. Incorporating automated vehicle identification data into origin-destination estimation. Transportation Research Record 1882, pp. 37-44.

Asakura, Y., Hato, E., Kashiwadani, M., 2000. OD matrices estimation model using AVI data and its application to the Han-Shin Expressway Network. Transportation 27, pp. 419-438.

Ashok, K., Ben-Akiva, M.E., 1993. Dynamic O-D matrix estimation and prediction for real-time traffic management systems. In Transportation and Traffic Theory, C.F. Daganzo (Ed), Elsevier, New York, pp. 465-484.

Ashok, K., Ben-Akiva, M.E., 2000. Alternative approaches for real-time estimation and prediction of time-dependent origin-destination flows. Transportation Science 34, pp. 21–36.

Ashok, K., Ben-Akiva, M.E., 2002. Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows. Transportation Science 36, pp. 184–198.

Aster, R.C., Borchers, B., Thurber, C.H., 2004. Parameter estimation and Inverse Problems. Oxford. Elsevier Inc.

Bekhor, S., Ben-Akiva, M.E., Ramming, S., 2001. Route choice: choice set generation and probabilistic choice models. Proceeding of the 4th TRISTAN Conference, Azores, Portugal.

Bell, M.G.H., 1991a. The estimation of origin-destination matrices by constrained generalized least squares. Transportation Research 25B, pp. 13-22.

Bell, M.G.H., 1991b. The real time estimation of origin-destination flows in the presence of platoon dispersion. Transportation Research Part 25B, pp. 115-125.

Ben-Akiva, M., Bierlaire, M., 2003. Discrete choice models with applications to departure time and route choice. Handbook of Transportation Science, second ed. Kluwer (Chapter 2).

Bertsekas, D., 1999. Nonlinear programming. Athena Scientific.

Breiland, C., Chu, L., Benouar, H., 2006. Operational effect of single-occupancy hybrid vehicles in high-occupancy vehicle lanes. Transportation Research Record 1959, pp. 151-158.

Brickal, S., Bhat, C. R., 2006. Comparative analysis of global positioning system-based and travel survey-based data. Transportation Research Record 1972, pp. 9-20.

Carter, M., 2000. Automated vehicle identification tags in San Antonio: Lessons learned from the metropolitan model deployment initiative. FHWA-OP-00-017. FHWA, U.S. Department of Transportation.

Cascetta E., 1984. Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator. Transportation Research 18B, pp. 289-299.

Cascetta, E., Inaudi, D., Marquis, G., 1993. Dynamic estimators of origin-destination matrices using traffic counts. Transportation Science 27, pp. 363-373.

Cascetta, E., Nguyen, S., 1988. A unified framework for estimating or updating origin/destination matrices from traffic counts. Transportation Research 18B, pp. 437-455.

Cascetta, E., Russo, E., Viola, F., Vitetta, A., 2002. A model of route perception in urban road network. Transportation Research 36B, pp. 577-592.

CCIT, 2006. Corridor management plan demonstration. Referred Website: http://www.calccit.org

Chang, G., Wu, J., 1994. Recursive estimation of time-varying O-D flows from traffic counts in freeway corridors. Transportation Research 28B, pp. 141-160.

Chen, C., Skabardonis, A., Varaiya, P., 2004. A System for Displaying Travel Times on Changeable Message Signs. Proceeding of the 83th TRB Annual Meeting, Washington, DC.

Chen, Y., 1994. Bilevel programming problems: Analysis, algorithms and applications, PhD thesis, report CRT-984, Centre de recherché sur les transports (CRT), University of Montreal, Montreal, Quebec, Canada.

Cheu, R., Chi, X., Der-Horng, L., 2003. Probe vehicle population and sample size for arterial speed estimation. Computer-Aided Civil and Infrastructure Engineering, Vol.17, No.1, pp.53–60.

Chu, L., Liu X., Oh, J., Recker, W., 2004. A Calibration Procedure for Microscopic Traffic Simulation. Proceeding of the 83th TRB Annual Meeting, Washington, DC.

Chu, L., Liu X., Recker, W., Zhang, H.M., 2004. Performance evaluating of adaptive ramp metering algorithms using microscopic traffic simulation model. Journal of Transportation Engineering 130, No. 3, pp.330-338.

Cremer, M., 1983. Determining the time dependent trip distribution in a complex intersection for traffic responsive control. Preprint of the 4th International IFAC/IFIP/IFORS Conference on Control in Transportation Systems, Germany, Baden-Baden.

Cremer, M., Keller, H., 1981. Dynamic identification of O-D flows from traffic counts at complex intersections. Proceeding of the 8th international Symposium on Transportation and Traffic Theory. University of Toronto Press, Toronto, Canada.

Cremer, M., Keller, H., 1984. A systems dynamics approach to the estimation of entry and exit O-D flows. Proceeding of the 9th international Symposium on Transportation and Traffic Theory. J. Volmuller and R. Hamerslag (Ed.). VUN Science Press, Utrecht, the Natherlands.

Cremer, M., Keller, H., 1987. A new class of dynamic methods for the identification of origin-destination flows. Transportation Research 21B, pp. 117-132.

Diebolt, J., Robert, C., 1994. Estimation of finite mixture distributions through Bayesian sampling, Journal of the Royal Statistical Society. Series B (Methodological), Vol. 56, No. 2, pp. 363-375.

Dion, F., Rakha, H., 2003. Estimating spatial travel time using automatic vehicle identification data. Proceeding of the 82nd TRB Annual Meeting, Washington, DC.

Dixon, M.P., Rilett, L.R., 2005. Population Origin–Destination estimation using automatic vehicle identification and volume data. Journal of Transportation Engineering 131, pp. 75-82.

Doblas, J., Benitez, F., 2005. An approach to estimating and updating origin-destination matrices based upon traffic counts preserving the prior structure of a survey matrix. Transportation Research Part 39B, pp. 565-591.

Du, J., Aultman-Hall, L., 2007. Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues. Transportation Research Part 41A, pp. 220–232.

Eisenman, S., List, G., 2004. Using probe data to estimate OD matrices. Proceeding of the 7th International IEEE Conference on Intelligent Transportation Systems, Washington, D.C.

Erlander, S., Nguyen, S., Stewart, N., 1979. On the calibration of the combined distribution/assignment model. Transportation Research 13B, pp. 259-267.

Fisk, C.S., 1988. On combining maximum entropy trip matrix estimation with user-optimal assignment. Transportation Research Part 22B, pp. 69-79.

Fisk, C.S., 1989. Trip matrix estimation from link traffic counts: The congested network case. Transportation Research Part 23B, pp. 331-336.

Fisk, C.S., Boyce, D.E., 1983. A note on trip matrix estimation from link traffic count data. Transportation Research 17B, pp. 245-250.

Florian, M., Chen, Y., 1993. A coordinate descent method for the bilevel OD matrix adjustment problem. Presented at the IFORS conference in Lisbon, Portugal. Available in an earlier version as Publication CRT-750 at the CRT, Universite de Montreal, Montreal, Quebec, Canada.

Gafarian, A.V., Munjal, P.K., Pahl, J., 1971. An experimental validation of two Boltzmann-type statistical models for multi-lane traffic flow. Transportation Research 5, pp. 211-224.

Gelman, A., Carlin, J., Stern, H., Rubin, D., 2003. Bayesian Data Analysis. Chapman & Hall/CRC.

Global Ecology Corporation, 2008. Global Positioning System. Referred Website: http://www.geco.us/gps.asp#

Hellinga, B., 1994. Estimating dynamic origin-destination demands from link and probe counts. PhD. Thesis, Queen's University, Ontario, Canada.

Hellinga, B., Fu, L., 1998. Assessing expected accuracy of probe vehicle travel time reports. Journal of Transportation Engineering 125, No. 6, pp. 524-530.

Herman, R., Lam, T., Prigogine, I., 1972. Kinetic theory of vehicular traffic: Comparison with data. Transportation Science 6, pp. 440-452.

Jayakrishnan, R., Kim, H-M., Jintanakul, K., Chu. L., 2006. Large corridor simulation: Model validation issues, Presented at the International Symposium on Transport Simulation, Lausanne, Switzerland.

Jayakrishnan, R., Mahamassani, H., Hu, T., 1994. An evaluation tool for advanced traffic information and management systems in urban networks. Transportation Research 2C, pp. 129-147.

LeBlance, L. J., Farhangian, K. 1982. Selection of a trip table which reproduces observed link flows. Transportation Research 16B, pp. 83–88.

Lo, H.P., Zhang, N., 1996. Estimation of an origin-destination matrix with random link choice proportions: a statistical approach. Transportation Research Part 30B, pp. 309-324.

Lo, H.P., Zhang, N., Lam, W.H.K., 1999. Decomposition algorithm for statistical estimation of OD matrix with random link choice proportions from traffic counts. Transportation Research Part 33B, pp. 369-385.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS -- a Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing, Vol. 10, pp. 325-337.

Maher, M., 1983. Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach. Transportation Research Part 17B, pp. 435-447.

McLachlan, G., Peel, D., 2000. Finite Mixture Models. John Wiley & Sons. Inc.

McNally, M.G., Marca, J.E., Rindt, C.R., Koos, A.M., 2003. TRACER: In-vehicle, GPS-based, wireless technology for traffic surveillance and management. California PATH Research Report, UCB-ITS-PRR-2003-23.

Mishalani, R., Coifman, B., Gopalakrishna, D., 2002. Evaluating real-time origin-destination flow estimation using remote sensing-based surveillance data. Proceeding of the 7th International Conference on the Applications of Advanced Technology in Transportation, ASCE, Cambridge, MA.

Nanthawichit, C., Nakatsuji, T., Suzuki, H., 2003. Application of probe vehicle data for real-time traffic state estimation and short-term travel time prediction on a freeway. Transportation Research Record 1855, pp. 49-59.

Nihan, N.L., Davis, G.A., 1987. Recursive estimation of origin-destination matrices from input/output counts. Transportation Research 21B, pp. 149-163.

Nihan, N.L., Davis, G.A., 1989. Application of prediction-error minimization and maximum likelihood to estimate intersection O-D matrices from traffic counts. Transportation Science 23, pp. 77–90.

Nguyen, S., 1977. Estimating an OD matrix from network data: A network equilibrium approach. Publication 87. CRT, University of Montereal, Montreal, Canada.

Nguyen, S., 1983. Modele de distribution spatiale tenant compte des itineraries. INFOR 21(4), pp. 270-292.

Nguyen, S., 1984. Estimating origin-destination matrices from observed flows. In Transportation Planning Models (M. Florian, ed.), Elsevier Science Publisher B. V. (North Holland), Amsterdam, pp. 363-380.

Oh, J-S., Jayakrishnan, R., 2002. Emergence of private advanced traveler information system providers and their effect on traffic network performance. Transportation Research Record 1783, pp.167-177.

Okutani, I., 1987. The Kalman filtering approach in some transportation and traffic problems. In Transportation and Traffic Theory, N.H. Garner and N.H.M. Wilson (Ed), Elsevier, New York, pp. 397-416.

Peeta, S., Ziliaskopoulos, A., 2001. Foundations of dynamic traffic assignment: The past, the present and the future. Networks and Spatial Economics, Vol. 1, pp. 233-265.

R Development Core Team, 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, Referred Website: http://www.R-project.org.

Rice, J., Van Zwet, E., 2004. A simple and effective method for predicting travel times on freeways, IEEE Transactions on Intelligent Transportation Systems, Vol.5, No. 3, pp. 200-207.

Road Bureau MLIT, 2008. The number of navigation system units in market. Referred Website: http://www.mlit.go.jp/road/ITS/

Schrank, D., Lomax, T., 2005. Urban Mobility Report. Texas Transportation Institute. Referred Website: http://mobility.tamu.edu/ums/

Sen, A., Thakuriah, P., Zhu, X.Q., Karr, A., 1997. Frequency of Probe Reports and Variance of Travel Time Estimates. Journal of Transportation Engineering 123, No. 4, pp. 290-297.

Sherali, H.D., Sivanandan, R., Hobeika, A.G., 1994. A linear programming approach for synthesizing origin-destination trip tables from link traffic volumes. Transportation Research Part 28B, pp. 357-378.

Smith, B., Zhang, H., Fontaine, M., Green, M., 2003. Cell phone probes as an ATMS tool. Center for Transportation Studies at the University of Virginia Research Report, UVACTS-15-5-79.

Sheffi, Y. 1985. Urban transportation networks: Equilibrium analysis with mathematical programming methods. Prentice-Hall, Englewood Cliffs, NJ.

Snickars, F., Weibull, J.W., 1977. A minimum information principle, theory and practice. Region. Sci. Urban Econ. 7, pp. 137-168.

Spiess, H., 1987. A maximum-likelihood model for estimating origin-destination matrices. Transportation Research 21B, pp. 395-412.

Spiess, H., 1990. A descent based approach for the OD matrix adjustment problem. Publication no. 693 at Centre de recherchesur les transports, Univeristy de Montreal, Montreal, Canada.

Tavana, H., Mahamassani, H., 2001. Estimation of dynamic origin-destination flows from sensor data using bi-level optimization method. Proceeding of the 80th TRB Annual Meeting, Washington, DC.

Turnquist, M., Gur, Y., 1979. Estimation of trip tables from observed link volumes. Transportation Research Record 730, pp. 1-6.

USDOT, 2000. Our nation's highway-2000. FHWA-PL-01-1012. Referred Website: http://www.fhwa.dot.gov/ohim/onh00/

USDOT, 2008. 2006 Highway Performance Monitoring System (HPMS). Referred Website: http://www.fhwa.dot.gov/policyinformation/tables/02.cfm

Van der Zijpp, N.J.,1996. Dynamic origin-destination matrix estimation on motorway nethwork. Ph.D. thesis, Transportation Planning and Traffic Engineering Subsection of the Faculty of Civil Engineering of Delft University of Technology.

Van Zuylen, H.J., Branston, D. M., 1982. Consistent link flow estimation from counts. Transportation Research 16B, pp. 473–476.

Van Zuylen, H.J., Willumsen, L.G., 1980. The most likely trip matrix from traffic counts. Transportation Research 14B, pp. 281-293.

Wilson, A.G., 1970. Entropy in unban and regional modeling. Methuen, Inc., New York.

Wu, J., Chang, G., 1996. Estimation of time-varying origin-destination distributions with dynamic screenline flows. Transportation Research 30B, pp. 277-290.

Yang, H., Akiyama, T., Sasaki, T., 1998. Estimation of time-varying origin-destination flows from traffic counts: A neural network approach. Mathematical and Computer Modeling, Vol. 27, pp. 323-334.

Yang, H., Iida, Y., Sasaki, T., 1994. The equilibrium-based origin-destination matrix estimation problem. Transportation Research Part 28B, pp. 23-33.

Yang, H., Meng, Q., Bell, M., 2001. Simultaneous estimation of the origin-destination matrices and travel-cost coefficient for congested networks in a stochastic user equilibrium. Transportation Science 35, pp. 107–123.

Yang, H., Sasaki, T., Iida, Y., Asakura, Y., 1992. Estimation of origin-destination matrices from link traffic counts on congested networks. Transportation Research 26B, pp. 417-434.

Yanying, L., Mike, M., 2002. Link Travel Time Estimation Using Single GPS Equipped Probe Vehicle. Proceeding of the IEEE 5th International Conference on Intelligent Transport Systems, Singapore.

Zhou, X., 2004. Dynamic origin-destination demand estimation and prediction for off-line and on-line dynamic traffic assignment operation. Ph.D. thesis. University of Maryland, College Park, USA.

Zhou, X., Mahmassani, H. S., 2006. Dynamic OD Demand Estimation Using Automatic Vehicle Identification Data. IEEE Transactions on Intelligent Transportation Systems, Vol. 7, No. 1, pp. 105- 114.

Zhou, X., Qin, X., Mahamassani, H., 2003. Dynamic origin-destination demand estimation using multi-day link traffic counts for planning applications. Transportation Research Record 1831, pp. 30-38.