MULTICLASS ORIGIN-DESTINATION ESTIMATION USING MULTIPLE DATA TYPES

A Thesis
Presented to the Faculty of the Graduate School
of Cornell University
In Partial Fulfillment of the Requirements for the Degree of
Master of Science

by
Qing Zhao
May 2013

ABSTRACT

Estimating O-D tables for trucks is of substantial interest due to different emission characteristics, pavement damage, etc of trucks. This thesis proposes a bilevel optimization model and corresponding solution method for static multi-class O-D estimation using various types of data. Limited memory BFGS method with bounded constraints is used for solving the upper level optimization, which is used to derive O-D table entries by minimizing the sum of squared differences between observations from different data sources and the predictions of those values. A probit model is assumed in the lower-level stochastic user equilibrium problem for flow prediction. Extensive experiments have been performed on a test network with different types of link count sensors and turning movements. The tests verify the problem formulation and solution algorithm, and offer important insights into the multiclass O-D estimation process with different types of data available.

BIOGRAPHICAL SKETCH

Qing Zhao was born in 1989 in Xingjiang, China, and moved with the whole family to Shanghai when she was 3, where she finished her primary and secondary education and later earned undergraduate degree in Transportation Engineering at Tongji University. With a great love for her major, she continued her education at Cornell University as a MS-Ph.D student and earned her Master of Science degree in 2013. She will continue her program at Cornell University under the supervision of Professor Mark A. Turnquist.

致我亲爱的父亲，赵兴斌 和母亲，田静

For my father, Xingbin Zhao

and my mother, Jing Tian

ACKNOWLEDGMENTS

This work was made possible with support from Xerox Corporation through their University Affairs Committee, and also from the U.S. Department of Transportation through funding to the Region II University Transportation Research Consortium.

I wish to express here my indebtedness and gratitude to those who have helped me to bring this thesis to fruition. I am enormously grateful to my committee chair Professor Mark Turnquist, who provided intellectual and methodological guidance, as well as moral support from the moment I came to Cornell. His far-reaching vision and deep insights helped me to form the idea from the very beginning and steer my project aright. His incredible patience and rich knowledge helped me to solve the problems and make progress along the way. His high standard and detailed revision helped to bring the thesis to an end. I am extremely grateful for being the very fortunate recipient of much time and attention. The academic guidance and suggestions to the thesis from my other committee members, Professor Oliver Gao and Professor Huseyin Topaloglu, are also invaluable to this process.

I wish to offer my heartfelt thanks to Douglas Miller and Brett Musco in particular for helping with java programming. The discussions with Doug were always insightful and enjoyable. I thank Julianne Schwartz for helping with Rochester data. Many thanks to Mutiara Sondjaja for spending time with me to discuss the algorithms. Suggestions from Jiexun Xu and Victoria Wu on java programming were indispensable for the completion of the project.

I also wish to express my thanks to Marina Tharayil and other people at Xerox Corporation for their continuous support and discussions.

To my father, mother, relatives and friends, thanks for the abundance of love and always being a great source of inspiration and support. My feelings, looking back, is of immense gratitude to all these people.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AVI	Automatic Vehicle Identification
CDF	Cumulative Density Function
cls.	class
DOT	Department of Transportation
DUE	Deterministic User Equillibrium
FHWA	Federal Highway Administration
GPS	Global Positioning System
hr.	hour
LBFGS_B	Limited memory Broyden–Fletcher–Goldfarb–Shanno method with bounded constraints
OD	Origin-Destination
RMS	Root Mean Square
SAM	Stochastic Assignment Model
SCAG	Southern California Association of Governments
SF	Sioux Falls
SNL	Stochastic Network Loading
SUE	Stochastic User Equilibrium
veh.	vehicle
vol.	volume

CHAPTER 1

INTRODUCTION

One of the core concepts in analyzing performance of transportation networks is the idea that people or goods travel from an origin zone to a destination zone (geographic areas) by using paths in a connected network of directional arcs or links. As the units moving (people, vehicles, shipping containers, etc.) choose (or are routed on) paths through the network, they create flow rates on the network arcs. In general, we'll refer to the collection of trip rates as a matrix, *Q*, with elements q_{ij} representing movement from origin zone *i* to destination zone *j*. We will denote the arc (or link) flows as *x*, with the flow on a specific arc, *a*, denoted as x_a. Although these arc flows are really rates, it is common to refer to them as volumes. One of the central problems in transportation network analysis is the determination of "equilibrium" link volumes, given an input trip table and the network topology and link capacities. Equilibrium in this context is defined as a state in which no single user of the network has an incentive to switch paths between their origin and destination to make themselves better off (i.e., lower cost, reduced travel time, etc.).

One of the critical challenges in using these network analysis methods is determining the O-D flow rates. One approach to this challenge is to solve a so-called "inverse problem," where we attempt to infer the O-D table from observations taken in the network (usually link counts). Since the mid-1970's, there has been interest in estimating origin-destination matrices from link count (or other) data. For the most part, these efforts have focused on creating static O-D tables for passenger movements by automobile, using a single vehicle class. A sampling of some of the influential papers in this area includes the work of Robillard (1975), Turnquist and Gur (1979), Van Zuylen and Willumsen (1980), Carey, *et al*. (1981), Maher (1983), Cascetta (1984), Nguyen (1984), Spiess (1987), Cascetta and Nguyen (1988), Brenninger-Göthe, *et al*. (1989), Lam and Lo (1990), Yang, *et al*. (1992), Sherali, *et al*. (1994), Yang (1995) and Maher, *et al*. (2001).

In general, the rates of trip movements can be denoted as $Q(t)$, to emphasize that the rates are time-dependent. However, in this thesis, the focus will be on the static problem, where Q is treated as a collection of constant origin-destination rates of travel. This assumption is very useful for planning purposes, using averages of time-varying trip rates over defined time periods. For example, averaging is often done over a peak hour, or a morning peak period (e.g., 6 – 9 AM), a mid-day off-peak (e.g., 9 AM – 3 PM), or an average weekday.

An important aspect of this thesis is recognition that there are multiple vehicle classes using the network simultaneously. One obvious distinction is between autos and various types of trucks in a road network, but the concept of vehicle classes also transfers to other situations (multiple train classes in a railroad network; multiple aircraft types using an airport, multiple vessel types in a lock-and-dam operation on a river, etc.). These multiple vehicle classes may use different amounts of available facility capacity, have different operating characteristics, etc., so we are often interested in disaggregating the O-D table and the resulting link flows by vehicle class, *m*: q_{ij}^{m} and x_{a}^{m}.

The interest in treating multiple vehicle classes separately is growing, especially in urban areas where truck flows are of increasing interest. It has been common in transportation planning studies to assume that truck flows are a constant (and generally small) fraction of automobile flows. However, this assumption ignores the substantial differences that may exist between the origins and destinations for trucks and those for automobiles.

Estimating O-D tables for trucks is of substantial interest to transportation planners because trucks impose different levels of pavement damage than cars, they have different emission characteristics, different accident patterns, and may be the subject of different types of flow controls. However, truck O-D estimation suffers from two fundamental challenges. First, it is inherently a multi-class problem because there are several different size classes of trucks, and this makes the estimation problem more difficult. Second, the available data on truck movements is very sparse, particularly data that separates different classes of trucks. Most of the data on which O-D estimation usually depends is link counts, and these data are normally collected by loop detectors buried in the roadway. However, single-loop detectors, which are ubiquitous in practice, cannot reliably distinguish a truck from a car – they simply can count the total number of vehicles passing over the detector during a specified time interval. Because the traffic stream in most locations is heavily dominated by automobiles, these total counts have been considered usable for single-class O-D estimation with the implicit assumption that trucks follow the same pattern of movements as cars. As urban areas become more concerned about truck movements, the long-used assumption that trucks are just an “add-on” to the automobile flow has become less acceptable, but the lack of truck movement data greatly hinders the process of estimating truck O-D tables separately from automobiles.

List and Turnquist (1994, 1995) proposed a method for estimating multi-class O-D tables for truck movements, and applied the technique in the New York City metropolitan area. The procedure made use of several different types of data, including several types of vehicle classification counts that had been made for special purposes by different transportation agencies in the New York area. An extended version of this technique (List, *et al*., 2002) has been instituted as a part of the "Best Practice Model" for on-going use in the New York City area.

Another example of truck flow estimation, using a combination of link count data and regional input-output economic data for Ontario, is described by Al-Battaineh and Kaysi (2007). One of the interesting ideas represented by their work is that truck movements are connected to economic activity, so regional economic data may be of help in estimating truck flows. This builds on the general concept that link counts of traffic (truck or otherwise) are not the only useful source of data for estimating O-D tables, and we should be creating methods that can integrate a wider variety of data types.

A major difficulty with these "truck only" O-D estimation methods is that they do not account for the interactions of trucks and automobiles in congested networks. A more complete approach is to estimate separate O-D tables for several vehicle classes (automobiles and various truck size classes) simultaneously, accounting for the different operating characteristics of the various vehicle classes and their interactions in creating a flow equilibrium. However, this clearly is a more complex estimation problem than the single-class models, and depends on having some of the observed data (although not necessarily all) separated by vehicle class.

There has been very limited work on estimating O-D tables for multiple vehicle classes, but a noteworthy effort of this type is described by Wong, *et al*. (2005). They developed a formulation and solution algorithm for a five-class problem (autos, taxis, buses, light trucks, and heavy trucks) for use in Hong Kong. The method developed in this thesis differs from their work in several important ways, but their model has provided a very useful starting point for this effort.

Ha, *et al*. (2007) developed a multi-class estimation model for use in Seoul, Korea. Their method is based on minimizing squared errors between estimated and observed classification counts on network links. They compared the multi-class O-D estimation method against a single-class model using the resulting trip length distribution as a criterion, and concluded that the multi-class model performed much better, but they provide few details of the actual implementation.

Raothanachonkun, *et al*. (2005) propose a multi-class estimation model focused on automobiles, light trucks and heavy trucks as vehicle classes. Their approach has the interesting characteristic of being aimed at estimating dynamic O-D matrices, rather than static ones, and uses an assumption of stochastic equilibrium as the underlying model of flow assignment in the network. Although they sketch out the outline of the model, they provide no indication of actual computational implementation.

In a subsequent paper (Raothanachonkun, *et al*., 2006) the same authors analyze multiclass static O-D estimation using a genetic algorithm for solution. They provide illustrative results on a very small network.

The objectives of this thesis are:

1) To develop a new method for estimating static multi-class O-D tables (automobiles plus several classes of trucks), using a variety of potential data types;
2) To build that method on an underlying model of stochastic user equilibrium as the characterization of network flows, reflecting the uncertainty in how individual users choose routes through the network;

3) To consider a range of sensor types, including both traditional and innovative means of providing vehicle classification counts, and determine how the output of those sensors can be incorporated into the O-D estimation process;
4) To test the methods developed on a network with varying types and levels of data, to verify the methods and provide guidance on the relative value of different types of sensors and data.

Because much of this project is based on the ability to classify vehicles and to have sensors that can count vehicles by class, Chapter 2 focuses on vehicle classification and sensors. Chapter 3 describes the O-D estimation model formulation, and discusses the details of the solution process. Chapter 4 discusses results from a series of experiments on a test network, and Chapter 5 presents conclusions and directions for further work.

CHAPTER 2

VEHICLE CLASSIFICATION, SENSORS AND DATA TYPES

2.1 Vehicle Classification

Vehicle classification can be done in several different ways, but one of the standard classifications is that used by the Federal Highway Administration (FHWA). This classification system defines 13 vehicle classes, as listed in Table 2 - 1. Figure 2 - 1 illustrates the important truck classes. Trucks in classes 10-13 are seldom seen in urban areas, as they are mostly used for long-haul movements.

Table 2 - 1. FHWA vehicle classes.

Vehicle Class	Description
1	Motorcycles
2	Passenger cars
3	Two-axle, four-tire single unit trucks
4	Buses
5	Two-axle, six-tire single unit trucks
6	Three-axle single unit trucks
7	Four or more axle single unit trucks
8	Four or less axle single trailer trucks
9	Five-axle single trailer trucks
10	Six or more axle single trailer trucks
11	Five or less axle multi-trailer trucks
12	Six-axle multi-trailer trucks
13	Seven or more axle multi-trailer trucks

An alternative commercial vehicle classification with eight truck classes, based strictly on the gross vehicle weight, is also used quite widely. The first 6 classes are subdivisions of what are shown as types 3 and 5 in Figure 2 - 1. A "class 7" truck includes both types 6 and 7 from Figure 2 - 1, and a "class 8" truck is any tractor-trailer combination.

Type 3: Two-axle, four-tire, light trucks

Type 5: Two-axle, six-tire, medium trucks

Type 6: Three-axle single unit trucks

Type 7: Single unit trucks with > 3 axles

Type 8: Four or fewer-axle single-trailer trucks

Type 9: Five-axle single-trailer trucks

Figure 2 - 1 Truck classes defined by FHWA.

A vehicle classification scheme used in several states is based on dividing vehicles into four or five length classes. For example, a common four-class length structure is: < 26 ft., 26-39 ft., 39-65 ft. and > 65 ft. Vehicles less than 26 ft. long are assumed to be automobiles (although this grouping also would capture motorcycles and light trucks from class 3 in the FHWA scheme). The 26-39 ft. class includes single-unit trucks (FHWA types 5-7). The 39-65 ft. class includes most tractor-trailer trucks (FHWA classes 8 and 9). Vehicles over 65 ft. long are either oversize loads or trucks with multiple trailers (FHWA classes 10-13).

The analysis in this thesis is based on using the FHWA classification scheme, but the experiments in Chapter 4 use an aggregation of types into three vehicle classes: Automobiles and Light Trucks (FHWA classes 2 and 3), Medium Trucks (FHWA class 5), and Heavy Trucks (FHWA classes 6-13).

2.2 Sensors and Traffic Data

Estimation of O-D matrices was originally defined as a problem in which the only data available were traffic counts on links. The basis for this assumption was that counts from loop detectors in the pavement are the most commonly collected type of traffic data. Over the years of development of O-D estimation models, some efforts have made allowance for a few other types of data, including cordon counts, intersection turning movements, data collected at toll booths, and automatic vehicle identification (AVI) data. In the following subsections, these various types of data, and the sensors used to collect them, are discussed.

2.2.1 Loop Detectors and Link Count Data

A link count is simply the number of vehicles crossing a specific link over a defined time interval. This type of data is ubiquitous because loop detectors buried in the pavement are widely deployed as the basic data collection mechanism for traffic engineering. Figure 2 - 2 illustrates the basic set-up and technology of a simple loop detector. The loop is a few turns of wire embedded in the pavement, which forms an inductor. When a vehicle passes over the loop, the mass of metal in the vehicle changes (reduces) the inductance, and a circuit senses this as a pulse with a specific duration. If the width of the detector is d, and a vehicle (indexed by i) of length L_i

traveling at a speed u_i passes over the detector, the duration of the pulse is $\frac{L_i + d}{u_i}$. Since we can measure the duration of the pulse, if we knew either the speed or the vehicle length, we can determine the other. However, we cannot determine both speed and vehicle length simultaneously from the detector output.

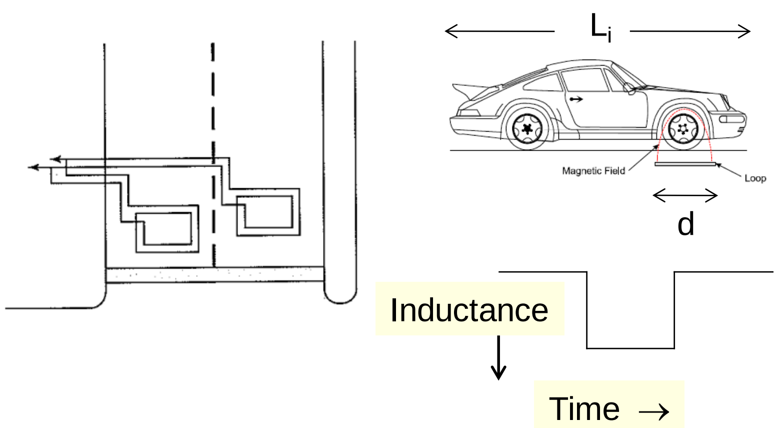

Figure 2 - 2 Simple single-loop detector implementation.

The usual output from a loop detector is a count and occupancy over some defined time period, T (often 20 sec or 30 sec). The count is simply the number of pulses recorded, or the number of vehicles counted. Occupancy is the proportion of time that there is a vehicle present over the detector. The inductance recorded at the detector over time is illustrated in Figure 2 - 3. If there are n pulses recorded over time T, then the occupancy measurement is:

$$Occ = \frac{\sum_{i=1}^{n} \frac{L_i + d}{u_i}}{T}$$ **Equation 2 - 1**

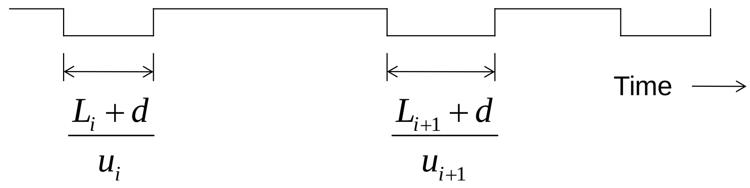

Figure 2 - 3 Loop detector recording over time.

If an average vehicle length, *L*, is assumed, then occupancy is directly related to the traffic density (vehicles/mile), and average speed is tied to density and counted volume, so the usual use of the detector measurements is to report traffic volume, *q*:

$$q = \frac{n}{T}$$ **Equation 2 - 2**

and the average speed of that traffic, *u*:

$$u = \frac{q\left(L + d\right)}{Occ}$$ **Equation 2 - 3**

However, since an important aspect of vehicle classification is sensing vehicle length, the *a priori* assumption of the average vehicle length in the traffic stream makes the use of single-loop detectors for vehicle classification highly problematic. A few authors have tried to create ways around this limitation. Sun and Ritchie (1999) and Jeng and Ritchie (2008) had some success using more detailed waveform output from the detectors. However, most detectors deployed in the field are not capable of producing that waveform output, so implementation of their ideas would require some extensive retro-fitting. Wang and Nihan (2003) created a two-bin classification model ("cars" and "larger") by exploiting an ability to estimate average speed separately from the loop detector output. Equation 2-3 can be rearranged to solve for average vehicle length if average speed is known:

$$L = \frac{u(Occ)}{q} - d$$ **Equation 2 - 4**

If there are only two length classes, with known average lengths L_1 and L_2, and a proportion of the vehicles are in class 1, then the overall average length must be $L = \lambda L_1 + (1-\lambda)L_2$. If we combine this result with Equation 2-4, we can use the single-loop detector output to estimate λ. This, in turn, allows the total volume *q* to be broken down into two class counts. However, this only works when there are two classes, and if we assume that the average lengths in each class are known.

Zhang, *et al*. (2006) created a series of neural network models to attempt classification of vehicles into length groupings based on a series of single-loop detector measurements. They define a neural network for each of four vehicle length classes (< 26 ft., 26-39 ft., 39-65 ft. and > 65 ft.). Each network takes as input a set of 18 measurements (volume and lane occupancy over nine previous 20-sec. intervals, or three minutes worth of data), plus a time stamp. Each network produces an estimate of how many vehicles in a particular length class passed over the detector during that 3-minute span. They found that they could estimate the volume in the first bin (< 26 ft.; i.e., passenger cars and light trucks) reasonably accurately, but that the estimated volumes in the three bins of larger vehicles (medium and heavy trucks) had quite substantial errors. Thus, their approach seems potentially capable of separating cars from larger vehicles (i.e., two classification categories), but does not seem to be an effective way of obtaining any more detailed distinctions of truck volumes in different size classes.

Dual-loop detectors are a much more reliable means of obtaining length-based vehicle classification counts, although the deployment of dual-loop detectors is more limited. A dual-loop detector setup across a three-lane freeway section is illustrated in Figure 2 - 4. As a vehicle passes over a dual-loop detector, it creates two pulses, one from each loop. The beginning and end of each pulse can be labeled as a time (t_1 through t_4), as illustrated in Figure 2 - 5.

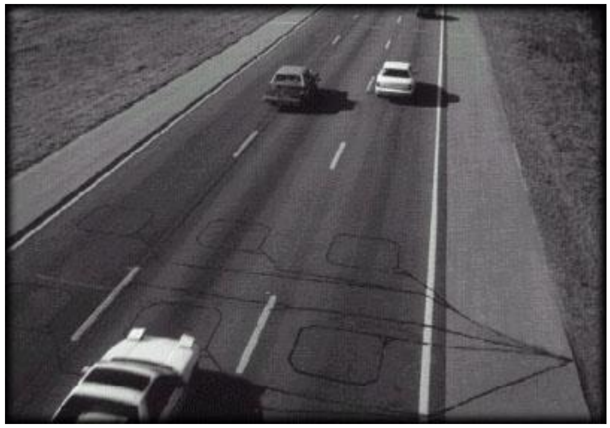

Figure 2 - 4. Dual-loop detector installation.

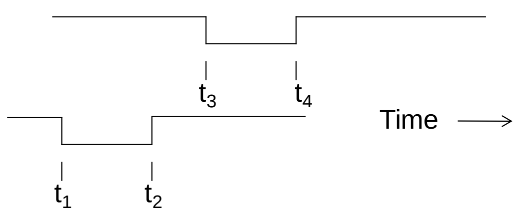

Figure 2 - 5. Dual-loop detector inductance pulses as a vehicle passes.

Then if the known distance between the two loops is D, the speed of vehicle i passing over the detector is:

$$u_i = \frac{D}{t_3 - t_1} \qquad \textbf{Equation 2 - 5}$$

Given this speed, the vehicle length can be estimated based on the average duration of the two pulses:

$$L_i = u_i \left[\frac{(t_2 - t_1) + (t_4 - t_3)}{2} \right] - d \qquad \textbf{Equation 2 - 6}$$

Vehicles of different lengths can then be grouped into some set of classes and counted. The length-based classification scheme (four vehicle length classes: < 26 ft., 26-39 ft., 39-65 ft. and > 65 ft.) described earlier may be acceptable for rural freeways, but in urban areas it is likely that few multi-trailer combinations will be observed, and it would be useful to separate FHWA type 5 trucks (two-axle, six-tire) from larger single-unit trucks of FHWA types 6 and 7. However, it is difficult to separate those single-unit classes using length alone.

In relatively uncongested conditions, dual-loop detectors can provide useful data on vehicle length, but the error rate on classifications is often quite high. Nihan, *et al*. (2002) reported on experiments that showed approximately 30-40% of trucks are misclassified, using the four-bin scheme. The most common errors are placing bin 2 trucks in bin 3, and placing bin 3 trucks in bin 4.

In congested conditions, when vehicles may often stop over a detector, the quality of the data degrades considerably. Wei (2011) has recently developed some methods for length-based vehicle classification using loop detector data under stop-and-go traffic conditions, but this remains a troublesome issue, of particular concern with loop detectors on urban streets.

2.2.2 Cordon Counts

Cordon counts identify some geographic area of interest and count all vehicles crossing the "cordon" enclosing that area (entering, leaving or both). These counts are usually done for specific purposes, and are often performed manually over a relatively short period. They are not normal ongoing data collection methods. However, when such counts are available, they may offer useful observations of total origins or total destinations within some defined area. This area may or may not correspond to one or more zones defined in the O-D table, however.

2.2.3 Turning Movement Data

Turning movements at intersections relate inbound and outbound flows on different legs (or approaches) to an intersection. Using the simple intersection in Figure 2 - 6 as an example, turning movements for all vehicles entering on approach 1 (I_1), would specify the proportions of those vehicles that turned right, went straight through, and turned left.

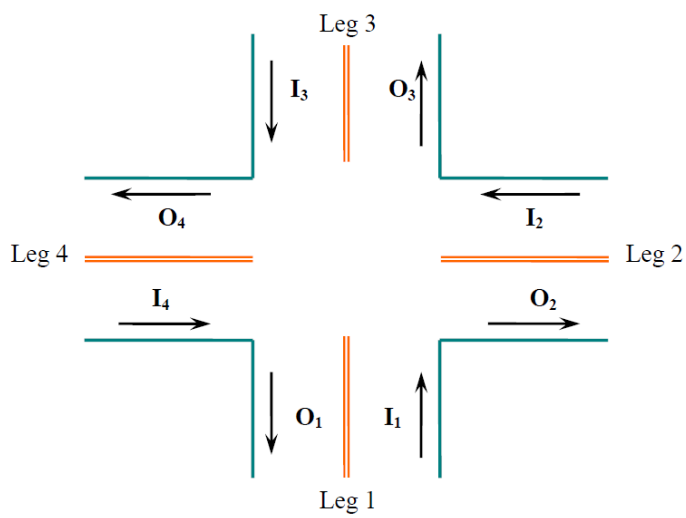

Figure 2 - 6 Simple four-leg intersection with inbound and outbound counts.

If data are collected on all movements at a given intersection with *N* approaches, the inbound and outbound counts can be related as shown in Equation 2-7 (assuming no U-turns):

$$\begin{bmatrix} O_1 \\ O_2 \\ .. \\ O_N \end{bmatrix} = \begin{bmatrix} 0 & \alpha_{21} & .. & \alpha_{N1} \\ \alpha_{12} & 0 & .. & \alpha_{N2} \\ .. & .. & .. & .. \\ \alpha_{1N} & \alpha_{2N} & .. & 0 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ .. \\ I_N \end{bmatrix} \qquad \textbf{Equation 2 - 7}$$

where α_{ij} is the proportion of traffic entering on leg *i* that exits on leg *j*.

Turning movements are not usually collected on an on-going basis as normal traffic data because in most cases, collecting the data requires manual observers (either to record data in the field or to watch video offline). However, these counts are important for optimal signal timing and thus may be collected periodically at a given intersection as part of a project to improve signal timing and intersection operation. For normal traffic engineering use, the total counts of vehicles making various turns are the important data, and it is not common to collect counts for vehicles in different classifications separately.

The potential value of turning movements for O-D estimation is that they specify a sequence of links used by a vehicle, rather than just counts on individual links. This provides an element of path information. In most previous O-D estimation efforts, the underlying assumption governing vehicle flows on the network is deterministic user equilibrium, and path flows are not uniquely identified by the equilibrium conditions, so this type of partial path information has not been of great interest. However, if the underlying flow pattern is assumed to be a stochastic user equilibrium, there is a set of identifiable path flows associated with the equilibrium, and this connection makes turning movements of significant potential value for the O-D estimation process.

A multiclass O-D estimation process also creates interest in having the turning movements separated by vehicle class. If the turning movement data are collected manually, this is certainly possible even though not commonly done. However, because manual data collection is expensive and not done frequently at any given intersection, there is also interest in automated methods of collecting turning movement data with vehicle classification that could be in place on a continuing basis.

Video data collection and processing is capable of making turning movement counts. Miovision (www.miovision.com), for example, markets a portable video data collection system designed for short term deployment at an intersection, as a substitute for manual data collection. The video processing software associated with this system is asserted to be able to provide vehicle classification data (at least at the level of separating medium trucks and heavy trucks from buses and automobile traffic, although the in-field assessments of count accuracy (see, for example, Swann (2010) and Schneider (2011)) have assessed only total counts. The accuracy of this system for total vehicle counts is quite high (generally with no more than 1-2% error), but no validation checks on classification counts have been identified in our analysis.

Further development of accurate and effective video collection and processing for turning movement data by vehicle class appears likely to be a useful endeavor, particularly when combined with a multiclass O-D estimation capability.

2.2.4 Toll Booth and Other Forms of Partial Path Data

Partial path data is sometimes available from special surveys at specific facilities. For example, truck drivers in line at a toll booth may be questioned for “last stop” and “next stop” information. This is done occasionally at locations like the toll bridges crossing the Hudson River into New York City, but collection of data in this way is labor intensive, expensive and prone to substantial reporting errors. Another example of such data is collected for U.S. import statistics. Counts of containers entering a specific port and destined for somewhere in the U.S. are collected by origin country. However, the specific destination inside the U.S. is not reported.

In urban networks this type of data can also be collected by video, using license plate matching software. That is, a specific truck may be observed at several points within a network at different times. By matching the data collected by different cameras, partial path information for the truck can be inferred. We may not know the original origin (or the final destination) of the truck, but we can infer some information about their path through the network, and whether stops within the network (which define destinations and origins of “linked” trips) are likely to have been made.

2.2.5 Automatic Vehicle Identification (AVI) Data

An increasing number of trucks have GPS receivers and many are capable of reporting their location to some central dispatching center on a regular basis. This type of data is not normally available outside the company that operates the trucks, but it is potentially a very rich data source. There has also been interest in using this type of data from automobiles (an increasing proportion of which are equipped with GPS capability) to estimate O-D information for passenger vehicles.

2.2.6 Non-Traffic Data and Use of *a priori* Trip Tables

One of the strategies used in O-D estimation since the 1970's has been to use a "target" trip table within the process. This stems from the fact that link volumes by themselves are generally insufficient to uniquely identify the trip table from which they arose (i.e., the inverse problem does not have a unique solution). One way to deal with this non-uniqueness is to postulate a trip table and then use the link volumes to find the trip table that is "closest" to this *a priori* table, and still consistent with the observed link data. Thus, the observed link data is used to adjust this *a priori* solution, rather than to dictate a solution on its own.

However, this approach depends on having some *a priori* O-D estimate available to use as a "target." In some cases, there may be an old O-D estimate that is to be updated with new data, and the target trip table approach can work quite well. Because urban areas in the U.S. have been charged with making metropolitan transportation plans for some time, there is often an automobile trip table available from a previous iteration of the planning process. However, this is more troublesome for trucks because most transportation plans have not been too concerned with truck flows in the past.

For truck flows, another approach that may be useful is to use socio-economic and land use data to estimate truck trip ends (total origins and total destinations, by zone). For example, the Southern California Association of Governments (SCAG) has developed estimates of trip rates related to employment in seven different industries as well as to numbers of households in an area. Although their estimates were created for use in the Los Angeles / Long Beach area, such estimates may also be useful in other areas. These rates, which separate light, medium and heavy trucks, are shown in Table 2-2.

Table 2 - 2 Truck trip rates estimated based on employment and numbers of households in traffic zones.

SCAG—INTERNAL TRIP RATES*

	Outbound				Inbound			
	Lt	Med	Hvy	Subtotal	Lt	Med	Hvy	Subtotal
Households	0.0390	0.0087	0.0023	0.0500	0.0390	0.0087	0.0023	0.0500
AMC	0.0513	0.0836	0.0569	0.1919	0.0513	0.0836	0.0569	0.1919
Retail	0.0605	0.0962	0.0359	0.1925	0.0605	0.0962	0.0359	0.1925
Government	0.0080	0.0022	0.0430	0.0533	0.0080	0.0022	0.0430	0.0533
Manufacturing	0.0353	0.0575	0.0391	0.1319	0.0353	0.0575	0.0391	0.1319
Transportation	0.2043	0.0457	0.1578	0.4078	0.2043	0.0457	0.1578	0.4078
Wholesale	0.0393	0.0650	0.0633	0.1677	0.0393	0.0650	0.0633	0.1677
Service	0.0091	0.0141	0.0030	0.0262	0.0091	0.0141	0.0030	0.0262

Notes:

SCAG = Southern California Association of Governments

Lt = light trucks (8,501-14,000 lb GVW) = FHWA class 3

Med = medium trucks (14,001-33,000 lb GVW) = FHWA class 5

Hvy= heavy trucks (>33,000 GVW) = FWWA classes 6 and higher

AMC = Agriculture/Mining/Construction

*Household trip rates are daily trips per household; all other trip rates are daily trips per employee

To use these trip rates, it is necessary to estimate employment and numbers of households within defined origin-destination zones. Then the trip rates can be used to construct “row totals” and “column totals” for the O-D matrices of truck classes.

2.3 Conclusions

Traffic data is available from a variety of sensor types, and the range of possibilities is increasing as better (and less expensive) video processing becomes available, and as GPS-based vehicle location technology becomes more widely deployed. Most of the O-D estimation methods that have been developed since the 1970’s are focused on using link counts because that data (often obtained via loop counters in the pavement) is most widely available. Most deployed loop counters are single-loop installations that cannot reliably distinguish vehicle classes, so the O-D estimation methods have primarily been single-class models.

In this thesis, the primary objective is to explore formulation and solution methods for a multi-class O-D estimation process. This offers the opportunity to better understand the flow patterns of different truck size classes, as well as the differences between truck O-D patterns and automobile O-D patterns. To support such a process, better vehicle classification data is vital, and this chapter has discussed some of the possible technologies for obtaining such data. The following chapter describes a model formulation for O-D estimation that can use various types of data.

CHAPTER 3

MODEL FORMULATION AND SOLUTION

3.1 Model Formulation

The problem of determining an O-D matrix from observed traffic flow data is generally underdetermined because there are typically many more unknowns (O-D matrix entries) than there are observations. This means that there are many different O-D tables that could produce the same set of observed link volumes. As a result, one aspect of O-D matrix estimation is defining what is to be considered the “best” O-D matrix from among the candidates.

A second important aspect of the problem is that observed data, whether it be traffic counts on links, classification counts from video data, turning movements, etc., always contain errors. As a result, the data may be internally inconsistent and the model for estimating an O-D matrix must be robust in the face of “noisy” or inconsistent data. An effective way of incorporating that robustness is to minimize the squared errors or deviations from the observed values, but to allow estimates that do not match observed values exactly.

A third critical aspect of the O-D estimation problem is the assumption of how a link flow pattern on the network is constructed from the O-D flows. One possible assumption is that there is a set of constant, volume-independent, path choice proportions for each O-D pair that can be determined *a priori* and used in the inverse model. This is called a proportional assignment assumption. It is useful in uncongested networks, and was widely used in the O-D estimation methods developed in the 1980’s.

A second common assumption is that link flows in the network are determined via a deterministic user equilibrium process. The equilibrium concept is important in congested networks. However, it makes the O-D estimation process much more difficult because an important parameter of the process, the route-choice proportion, depends on the O-D matrix being estimated. Early efforts to incorporate equilibrium effects were discussed by Turnquist and Gur (1979) and Nguyen (1984). Deterministic user equilibrium (DUE) is a widely used mechanism for predicting flows in congested networks, and this makes it an attractive option for O-D estimation as well. Initially, it was difficult to accommodate both the concept of DUE (which assumes perfect knowledge of travel impedances by drivers) and the fact that link data are invariably at least somewhat inconsistent with that assumption, but Sherali, *et al*. (1994) constructed an elegant approach to deal with that problem when there is a single vehicle class present. The multiclass model developed by Wong, *et al*. (2005) also assumes a DUE process for the traffic flows in the network.

A third possibility for the mechanism of determining link flows from O-D volumes is to assume a stochastic user equilibrium (SUE). The SUE assumption allows for errors in perception on the part of drivers as they make route choice decisions, and thus offers a richer view of equilibrium flows in the network. Daganzo (1983) showed that the equilibrium conditions for an SUE solution can be written as the following set of nonlinear equations:

$$x_a^m - \sum_{ij} q_{ij}^m p_{ij}^{am}\left(Q,c\right) = 0, \qquad a = 1,\ldots,n; \quad m = 1,\ldots,M \qquad \textbf{Equation 3 - 1}$$

where x_a^m is the class m volume on link a, and $p_{ij}^{am}\left(Q,c\right)$ represents the link utilization coefficients (i.e., the fraction of the class m O-D volume for pair ij that appears on link a). These link utilization coefficients are functions of both the vector of costs on the links of the network (c), and of the O-D volumes themselves (Q).

If we have observed link volumes (on some, but not necessarily all, links), the conditions in Equation 3-1 provide a way of connecting that data to the estimates of O-D volumes. However, doing this in a computationally practical way is non-trivial. Several authors have approached this issue (for a single vehicle class) using a bi-level programming formulation, where the upper-level problem adjusts the q_{ij} values, and the lower-level problem finds link flows, given the q_{ij}'s. Maher, *et al*. (2001) developed a bi-level programming approach for a logit-based SUE formulation and one vehicle class.

The logit model of route choice through the network (which produces the $p_{ij}^{am}\left(Q,c\right)$ coefficients) is attractive for its computational simplicity, but it imposes some relatively unrealistic assumptions on the route choice probabilities, as a result of basic assumptions in the choice structure on which the model is built. One of its obvious weaknesses as a route choice model is the assumption that the characteristics of routes are independent of one another, even when those routes may overlap for a large fraction of their length. Cascetta, *et al*. (1996) proposed a method for relaxing the route independence assumption, but it creates quite a complicated O-D estimation process.

An alternative approach is to assume a probit model of route choice. The underlying assumption of the probit model (that uncertain route characteristics, or driver perceptions of those characteristics, are approximately Normal) is intuitively appealing, but the computations required for assigning traffic to a network using probit-based methods are significantly more complex than the computations for logit-based assignment. This has inhibited more widespread use of the probit model for SUE predictions. However, we have decided that its conceptual advantages outweigh its computational issues, and the model developed here is built using a probit-based SUE assumption.

The SUE conditions represented in Equation 3-1 can be incorporated into the O-D estimation process in two different ways. One way is to incorporate the conditions directly as constraints into the optimization for determining the O-D elements. Maher, *et al*. (2001) refer to this approach as “equilibrium programming” (following terminology defined by Garcia and Zangwill, 1981). A second way is to formulate a bi-level model, where the upper-level problem adjusts the q_{ij} values, and the lower-level problem finds link flows, given the q_{ij}’s. If a logit-based SUE model is used, the partial derivatives of the link utilization coefficients can be constructed in a reasonably straightforward way, and an equilibrium programming approach can be effective. However, this is much more difficult with a probit-based model and the bi-level programming approach seems likely to be more effective. The bi-level approach has been adopted here.

Under ideal conditions, the link count data will separate each of the defined vehicle classes and there will be counts available for all network links. In practice, however, this is unlikely to be the case. The observed values for the link volumes may correspond to aggregations of the vehicle classes defined, and there may not be observations for all links. We will index the available observations by *l*, and the set of all observed link counts as $\breve{X}$. The count for observation *l* will be denoted $\tilde{x}_l$, and include vehicles from a set of classes, M_l. The relevant link for observation *l* will be denoted a_l. Then the observed link count data and the observations’ connection to the O-D tables for individual vehicle classes under the SUE conditions, can be written as:

$$\tilde{x}_l = \sum_{m \in M_l} \sum_{ij} p_{ij}^{a_l m}\left(Q, c\right) q_{ij}^m \qquad \forall l \qquad \textbf{Equation 3 - 2}$$

The general form of constraints representing non-link count data (for an observation indexed by *l*) can be expressed as:

$$\sum_{m \in M_l} \sum_{ij \in \mathrm{N}_l} u_{ij}^{ml} q_{ij}^{m} = B_l \qquad \textbf{Equation 3 - 3}$$

where B_l is an observed value and the u_{ij}^{ml} values are known coefficients. Equation 3-3 states that some linear combination of elements in the estimated O-D tables (for some subset of vehicle classes denoted M_l and a subset of the O-D pairs *ij* in the set N_l) should sum to an observed value. This is a very general form, because the set N_l may have any combination of *ij* pairs in it, and the set M_l may contain one or more vehicle classes.

The observed values, regardless of type, must be assumed to contain errors. Thus, there is likely to be no solution that satisfies all the constraints specified by Equation 3-2 and Equation 3-3 exactly. We allow violations of the constraints 3-2 and 3-3, but with penalties in the form of a squared-error term for each constraint. This leads to a formulation of the problem as follows:

$$\underset{q_{ij}^m \geq 0}{Min}\ f = \sum_{l \in \breve{B}} \eta_l \left(B_l - \sum_{m \in M_l} \sum_{ij \in \mathrm{N}_l} u_{ij}^{ml} q_{ij}^{m} \right)^2 + \sum_{l \in \breve{X}} \tau_l \left(\tilde{x}_l - \sum_{m \in M_l} \sum_{ij} p_{ij}^{a_l m} \left(Q, c\right) q_{ij}^{m} \right)^2 \qquad \textbf{Equation 3 - 4}$$

The sets $\breve{X}$ and $\breve{B}$ define the observed values ($\tilde{x}_l$ and B_l) for the link counts and non-link data, respectively. The incorporation of the weighting constants η_l and τ_l allows us to control the degree to which emphasis is placed on various observations within the observed data.

One of the primary difficulties in solving the optimization problem in Equation 3-4 is that the p_{ij}^{am} values are determined as part of the assignment of the O-D trips, *Q*, for all the vehicle classes to the network to achieve a stochastic equilibrium flow pattern. That SUE assignment problem is itself an optimization, and that problem needs to be solved in order to have the parameters to evaluate the function in Equation 3-4. This illustrates the bi-level character of the overall problem. Evaluation of the objective function for the “upper-level” problem Equation 3-4 requires solution of the “lower-level” optimization (the SUE flow problem). The following section describes a method for accomplishing the solution of the minimization expressed in Equation 3-4.

3.2 A Solution Method

The general flow of information in the solution approach is that a trial solution for the upper-level problem (a set of q_{ij}^{m} values) is passed to the lower-level problem. Assignment of these trips via an SUE algorithm results in link and path flows, and a set of p_{ij}^{am} values that are consistent with the input q_{ij}^{m} values. These p_{ij}^{am} values are passed back to the upper-level problem and treated as constants while the q_{ij}^{m} values are adjusted to create a new trial solution to the upper-level problem. If the upper-level problem has converged (i.e., the q_{ij}^{m} values are not changing), a solution to Equation 3-4 has been reached. However, if the new trial q_{ij}^{m} values are different from the previous values, they are passed back to the lower-level problem, where new p_{ij}^{am} values are created.

For solution of the lower-level problem, a probit-based SUE algorithm, based on the work of Maher and Hughes (1997) and Connors, *et al*. (2007), is used. The SUE calculations are iterative, and depend on a core method for doing stochastic network loading (SNL). This can be accomplished using the Stochastic Assignment Model (SAM) algorithm first described by Maher (1992).

In a more formal sense, the algorithm for solution of Equation 3-4 is as follows:

1) Set $n \leftarrow 0$ (an iteration counter). Based on free-flow conditions for travel times and costs, use SAM to do a SNL for each vehicle class. The result of this loading is a set

of p_{ij}^{am} values. Denote the collection of values as P_0. For some types of data, the associated u_{ij}^{ml} values also depend on P_0. Denote the entire collection of u_{ij}^{ml} values as U_0.

2) Using the values of p_{ij}^{am} from P_0 and u_{ij}^{ml} from U_0 (as constants), solve the problem:

$$\underset{q_{ij}^m \geq 0}{Min} \quad \sum_{l \in \breve{B}} \eta_l \left(B_l - \sum_{m \in M_l} \sum_{ij \in \mathrm{N}_l} u_{ij}^{ml} q_{ij}^m \right)^2 + \sum_{l \in \breve{X}} \tau_l \left(\tilde{x}_l - \sum_{m \in M_l} \sum_{ij} p_{ij}^{a_l m} q_{ij}^m \right)^2$$

Denote this solution as Q_0.

3) Using the O-D tables Q_n for the various vehicle classes, do the SUE calculations to get a feasible solution for the network link volumes, denoted X_n, the link utilization values, P_n, and the associated data coefficients U_n. The collection (Q_n, X_n, P_n, U_n) will be termed a *consistent* solution. (Note that when n = 0, the P_n and U_n values computed in this step replace the initial values estimated in step 1.)

4) Using the values of p_{ij}^{am} from P_n and u_{ij}^{ml} from U_n (as constants), solve the problem:

$$\underset{y_{ij}^m \geq 0}{Min} \quad \sum_{l \in \breve{B}} \eta_l \left(B_l - \sum_{m \in M_l} \sum_{ij \in \mathrm{N}_l} u_{ij}^{ml} y_{ij}^m \right)^2 + \sum_{l \in \breve{X}} \tau_l \left(\tilde{x}_l - \sum_{m \in M_l} \sum_{ij} p_{ij}^{a_l m} y_{ij}^m \right)^2$$

Denote this solution as Y_n.

5) Using the O-D tables Y_n for the various vehicle classes, do the SUE calculations to get a feasible solution for the network link volumes, denoted V_n. This will be termed a *trial* solution.

6) Linearize the assignment map between the two feasible solutions so that we can express potential trip tables as a function of a step size γ, where $0 \le \gamma \le 1$:

$$Q(\gamma) = Q_n + \gamma\left(Y_n - Q_n\right)$$

7) Perform a one-dimensional search on γ to find:

$$\underset{0\le\gamma\le1}{Min} \quad \sum_{l\in\breve{B}} \eta_l \left(B_l - \sum_{m\in M_l} \sum_{ij\in \mathrm{N}_l} u_{ij}^{ml} q_{ij}^{m}(\gamma) \right)^2 + \sum_{l\in\breve{X}} \tau_l \left[\tilde{x}_l - p_{ij}^{a_l m} q_{ij}^{m}(\gamma) \right]^2$$

The values of u_{ij}^{ml} and $p_{ij}^{a_l m}$ are the values associated with the solution Q_n. Denote the optimal value of γ as γ^*.

8) If the difference $Q(\gamma^*) - Q_n$ is small enough, the solution has converged, and we can stop, using the last consistent solution as the final solution. If the process is not converged, make $n \leftarrow n + 1$, and $Q_n = Q(\gamma^*)$. Go to step 3.

Further discussion of the steps in this algorithm is facilitated by an example, so we'll consider a small network described in the following sub-section.

3.3 An Example Network

Figure 3 - 1 shows a small network with nine nodes and 12 two-way links. The values alongside the links are lengths, in miles. For analysis purposes, each of the two-way links is replaced by a pair of directional links, so the network has 24 directed links.

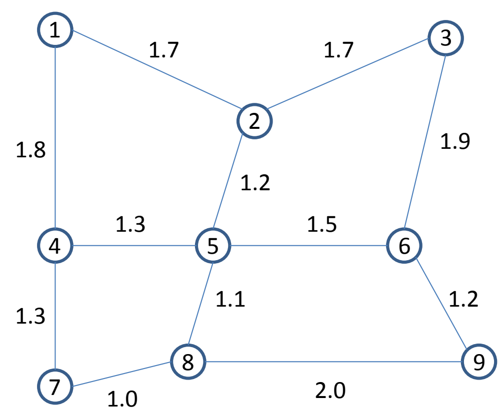

Figure 3 - 1. Example network.

Each link has a delay function to relate travel time to flow volume. Some commonly used delay function forms are the FHWA function, shown in Equation 3-5 and the Akçelik function (Akçelik , 1991) shown in Equation 3-6. Both functions are defined for all non-negative flows. Either form could be used in this example, but we will choose to assume the Akçelik functions. In this form of delay function, when flow exceeds capacity (i.e., $q_a > 1$), the delay increases rapidly, representing transient queuing on the link. For this example, we will assume capacity U_a = 1200 veh/hour on all links, a free-flow speed of 40 mph, and B = 0.4. Figure 3 - 2 shows travel time for a link that is 1.0 miles in length, as a function of traffic flow on that link, using these parameters.

$$t_a(x_a) = t_{0a}\left[1+\alpha(q_a)^{\beta}\right] \qquad \textbf{Equation 3 - 5}$$

$$t_a(x_a) = t_{0a} + \left\{0.25T\left[(q_a-1)+\sqrt{(q_a-1)^2+\frac{8B}{K_aT}q_a}\right]\right\} \qquad \textbf{Equation 3 - 6}$$

where: t_{0a} = free-flow time for link *a* (e.g., length divided by the speed limit)

T = period under analysis (in this example, 1 hour)

x_a = flow on link *a* (e.g., veh/hour)

K_a = capacity of link *a* (in same units as flow variable)

q_a = x_a / U_a (saturation level)

B = parameter that varies by link type

α, β = parameters (often $\alpha = 0.15$ and $\beta = 4$).

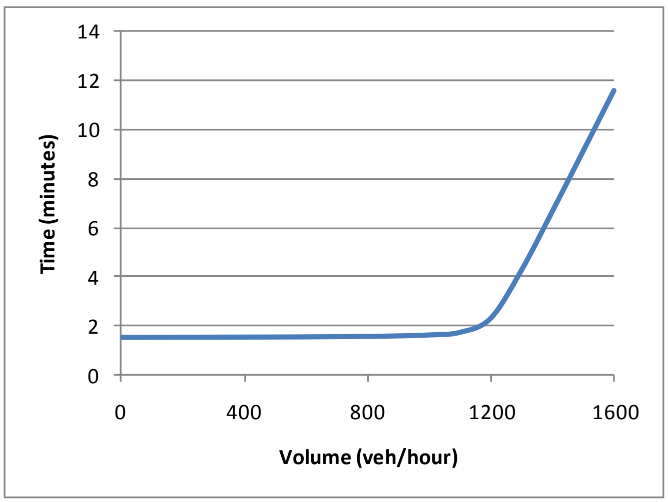

Figure 3 - 2. Travel time as a function of volume for a 1.0-mile link.

For this example, we will assume three vehicle classes – automobiles and two truck classes. The first truck class is two-axle, six-tire medium trucks (Type 5 shown in Figure 2 - 1). The second class includes all heavier trucks (Types 6-9 in Figure 2 - 1). Light trucks (Type 3 in Figure 2 - 1) are included with the automobiles. For the purposes of the multiclass network flow assignment, we define a *standardized flow*, z_a, as a weighted combination of the flows of the three classes x_a^m:

$$z_a = x_a^1 + 1.8 x_a^2 + 2.4 x_a^3 \qquad \textbf{Equation 3 - 7}$$

The standardized flow converts the truck flows into passenger car equivalents. Passenger cars (class 1) have a weight of 1. A medium truck (class 2) is assumed to be equivalent (in terms of contribution to volume-related congestion) to 1.8 passenger cars, and a heavy truck (class 3) is equivalent to 2.4 passenger cars.

The link delay functions for this example are written in terms of the standardized flow z_a, where L_a is the length of the link in miles and the resulting travel time is expressed in minutes:

$$t_a\left(z_a\right)=1.5\,L_a+\left\{15\left[\left(\frac{z_a}{1200}-1\right)+\sqrt{\left(\frac{z_a}{1200}-1\right)^2+\frac{3.2}{1200}\left(\frac{z_a}{1200}\right)}\right]\right\} \quad \textbf{Equation 3 - 8}$$

We will assume that there are four O-D pairs in this network: 1-9, 3-7, 7-3 and 9-1. Thus, the core problem is to estimate three O-D tables (one for each vehicle class), each of which contains four elements.

As drivers decide on paths from origin to destination, they are assumed to be attempting to minimize a cost that depends on both distance and travel time. The cost functions for each link (for the three vehicle classes) are:

$$c_a^1(z_a)=0.5L_a+0.33t_a(z_a) \quad \textbf{Equation 3 - 9}$$

$$c_a^2(z_a)=L_a+0.33t_a(z_a) \quad \textbf{Equation 3 - 10}$$

$$c_a^3(z_a)=1.5L_a+0.5t_a(z_a) \quad \textbf{Equation 3 - 11}$$

The terms in Equation 3-9 through 3-11 that relate to the length of the link, L_a, do not depend on the flow volume, but as flow, z_a, increases on a link, $t_a(z_a)$ increases (using Equation 3-8), and thus the cost of using that link increases to all vehicle classes, but at rates that are class-dependent.

3.4 Stochastic Network Loading (Algorithm Step 1)

The basis of equilibrium in networks is that drivers of individual vehicles are choosing paths through the network to minimize their own costs. In deterministic equilibrium, each driver is assumed to have complete and accurate information about the costs of all links as the basis for a path choice. In stochastic equilibrium, the perceptions of travel time (or cost) are assumed to contain errors. These errors are modeled as random variables ς_a^m, and the (random) perceived costs of using link *a* are then:

$$\tilde{c}_a^1(z_a) = 0.5L_a + 0.33t_a(z_a) + \varsigma_a^1 \quad \textbf{Equation 3 - 12}$$

$$\tilde{c}_a^2(z_a) = L_a + 0.33t_a(z_a) + \varsigma_a^2 \quad \textbf{Equation 3 - 13}$$

$$\tilde{c}_a^3(z_a) = 1.5L_a + 0.5t_a(z_a) + \varsigma_a^3 \quad \textbf{Equation 3 - 14}$$

Along the links of a path from origin to destination, the costs add, so we can also consider the cost along a path, *p*, as a random variable:

$$\tilde{g}_p^m = g_p^m + \psi_p^m \quad \textbf{Equation 3 - 15}$$

where $g_p^m = \sum_{a \in p} c_a^m\left(z_a\right)$ is the deterministic portion of the path cost and $\psi_p^m = \sum_{a \in p} \varsigma_a^m$ is the random portion. It is usual to assume that the link random variables have zero mean, so ψ_p^m also has zero mean, and the expected path cost is g_p^m.

If we consider two alternative paths, p_1 and p_2, for a given origin and destination, the probability of a driver choosing path p_1 is the probability that path p_1 is perceived as having lower cost:

$$\Pr\left[\tilde{g}_{p_1}^m \le \tilde{g}_{p_2}^m\right] = \Pr\left[g_{p_1}^m + \psi_{p_1}^m \le g_{p_2}^m + \psi_{p_2}^m\right] \quad \textbf{Equation 3 - 16}$$

Re-arranging Equation 3-16 to put the random variables all on one side of the inequality, we can write it as:

$$\Pr\left[\tilde{g}_{p_1}^m \le \tilde{g}_{p_2}^m\right] = \Pr\left[\psi_{p_1}^m - \psi_{p_2}^m \le g_{p_2}^m - g_{p_1}^m\right] \qquad \textbf{Equation 3 - 17}$$

Equation 3-17 is written in the form of the cumulative distribution function (CDF) of the random variable $\psi_{p_1}^m - \psi_{p_2}^m$: i.e., the probability that the random variable takes on a value no greater than $g_{p_2}^m - g_{p_1}^m$. Thus, the path choice probability depends directly on the distributional assumption made for the error terms. There are two common choices: the Gumbel distribution, which leads to the logit model for the path choice probabilities, and the Normal distribution, which leads to expressing Equation 3-17 as a cumulative Normal distribution (the probit model).

Expanding Equation 3-16 and 3-17 to the case where there are more than two alternative paths means that we are concerned with identifying the smallest of a set of random variables, and if the probit model (i.e., Normally distributed errors) is assumed, this means the smallest of a set of Normal random variables. The central result for effective computations is Clark's Approximation (Clark, 1961), which says that if Y_i and Y_j are Normally distributed random variables with means μ_i and μ_j, variances σ_i^2 and σ_j^2 and covariance v_{ij}, then the random variable T defined to be the minimum of Y_i and Y_j is itself approximately Normal with mean μ_T and variance σ_T^2, where:

$$\mu_T = \pi_i \mu_i + \pi_j \mu_j - a\phi(\alpha) \qquad \textbf{Equation 3 - 18}$$

$$\sigma_T^2 = \left(\mu_i^2 + \sigma_i^2\right)\pi_i + \left(\mu_j^2 + \sigma_j^2\right)\pi_j - \left(\mu_i + \mu_j\right)a\phi(\alpha) - \mu_T^2 \qquad \textbf{Equation 3 - 19}$$

In Equation 3-18 and 3-19, the quantities a and α are computed from:

$$a^2 = \sigma_i^2 + \sigma_j^2 - 2v_{ij} \qquad \textbf{Equation 3 - 20}$$

$$\alpha = \left(\mu_i - \mu_j\right)/a \qquad \textbf{Equation 3 - 21}$$

$\phi(.)$ is the standard Normal density function. The probability that it is Y_i that is smaller is π_i, given by:

$$\pi_i = \Phi(-\alpha) \qquad \textbf{Equation 3 - 22}$$

$\Phi(.)$ is the standard Normal CDF. As a result of Equation 3-22, we also have:

$$\pi_j = 1 - \pi_i = \Phi(\alpha) \qquad \textbf{Equation 3 - 23}$$

Finally, if Y_i and Y_j were correlated with some third random variable Y_k, then the covariance of T with Y_k is:

$$v_{Tk} = \pi_i v_{ik} + \pi_j v_{jk} \qquad \textbf{Equation 3 - 24}$$

Equations 3-18 through 3-23 create a straightforward way of obtaining the mean and variance of the minimum of two Normal random variables. The fact that the minimum is also approximately Normal, combined with Equation 3-24, further allows this result to be applied iteratively across an arbitrary set of Normal random variables to obtain the minimum of the entire set of *n* random variables. This makes the result of immense value in doing probit-based stochastic network flow assignment.

This is the idea exploited by Maher (1992) in constructing what he termed a Stochastic Assignment Model (SAM). Given a set of random link costs, $\tilde{c}_a$, assumed Normally distributed, SAM operates from an origin node and moves through a network, performing two basic operations – *scanning* and *merging*. Outbound links from a given node are scanned to construct the distributions of cost to reach the ends of those links. Sets of inbound links to a given node are merged to construct the distribution of cost required to reach that node. Once a node has been merged, the links outbound from it are eligible for scanning. When all inbound links to a node have been scanned, that node is eligible to be merged. Maher (1992) developed a very efficient algorithm for performing the required computations, based on Clark's (1961) results.

In concept, SAM calculates a set of probabilities that can be used to load all trips originating at the origin node (and terminating at all destinations) in a single pass. However, when doing so, it is quite easy to encounter a situation called "lock-up" in the algorithm, where no node is eligible to be merged and the algorithm cannot complete. This is not a problem in acyclic networks, but most transportation applications operate in networks where cycles are possible. It is possible to avoid the potential for lock-up by limiting the operations (and flows) to a set of "efficient" links, but doing so in a careful way forces the SAM algorithm to operate on a single O-D pair at a time, making it much more computationally intensive. However, for the current purposes, the benefits of using the concept of link efficiency seem to outweigh the additional computational burden, so this has been incorporated.

The concept of efficient links for an origin-destination pair originated with Dial (1971), who used the idea in his logit-based SNL algorithm. For a given origin-destination pair, crossing an efficient link moves one further from the origin and closer to the destination. Links that are not on the shortest path from the origin to the destination can be efficient, as long as they are on paths that are generally moving in the right direction. However, the concept precludes cycling in the network because somewhere along a cyclic sub-path one must be moving "backwards" – i.e., getting closer to the origin and further from the destination. Thus, if flows are limited to paths made up of only efficient links there can be no cycles in the paths. For a single O-D pair, this allows the SAM algorithm to avoid the potential problem of lock-up.

The idea of defining and using only efficient links can be illustrated using our simple example network. We'll define the costs for medium trucks (vehicle class 2) at a set of initial conditions when $z_a = 0$ for all links. Equation 3-8 is used to compute travel times (minutes) for all links, $t_a(0)$. Then Equation 3-10 is used to compute the costs for medium trucks: $c_a^2(0)$. These costs are shown along the links in Figure 3 - 3. We consider the O-D pair 1-9, and Figure 3 - 3 also includes a set of node labels, C_O , representing the cost of the minimum-cost path from the origin (node 1) to that node. The minimum-path *tree* is indicated with the heavier directed links.

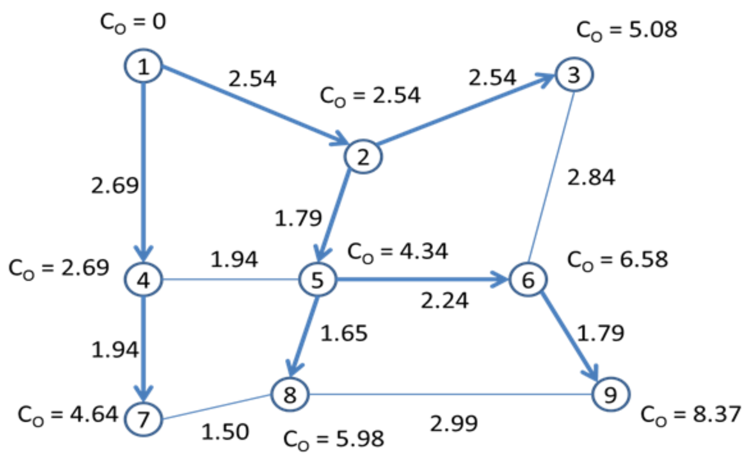

Figure 3 - 3. Example network with link costs and minimum path tree from node 1.

We can also construct the minimum-cost path from each node to the destination (node 9), and use the label C_D to indicate the cost from each node, as shown in Figure 3 - 4. The destination-focused minimum-path tree is also indicated in Figure 3 - 4, using the heavier directed arcs.

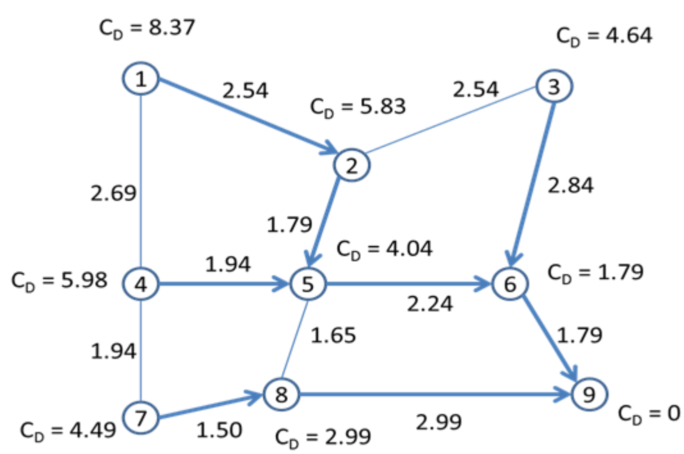

Figure 3 - 4. Example network with link costs and minimum path tree to node 9.

Figure 3 - 5 shows the combination of C_O and C_D labels at the nodes. The concept of an efficient link (for a specific O-D pair) is that traversing that link takes one further from the origin (C_O is increasing from the beginning node of the link to the ending node), and closer to the destination (C_D is decreasing from the beginning node of the link to the ending node). For example, we can see that traversing link 1→2 satisfies both criteria, and thus link 1→2 must be efficient for O-D pair 1-9. Figure 3 - 5 shows the collection of all efficient links for O-D pair 1-9.

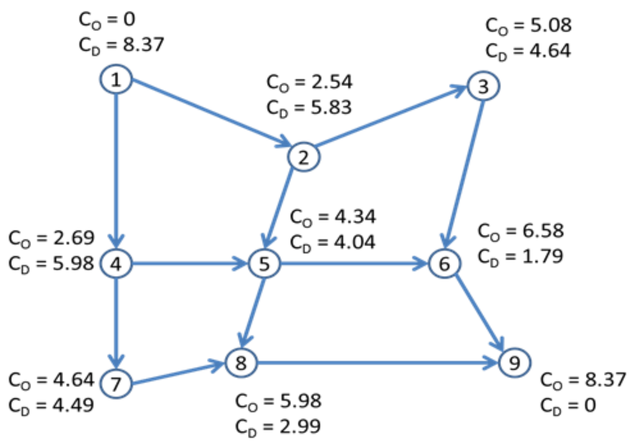

Figure 3 - 5 Example network with efficient links for O-D pair 1- 9.

Using only efficient links for the assignment of flow from 1-9 results in an acyclic network and avoids any potential problems of lock-up in the SAM algorithm. The set of efficient links also defines implicitly a set of possible paths between node 1 and node 9, but we have not had to explicitly enumerate those paths. The SAM algorithm, however, will produce probability values for use of links by each vehicle class on this O-D pair (the p_{ij}^{am} values noted earlier), as well as enough information to reconstruct probabilities of path use which will be important for constructing some constraints of the form described in Equation 3-3. The results of the SAM calculation of link probabilities for O-D pair 1-9 and vehicle class 2 are shown in Figure 3 - 6.

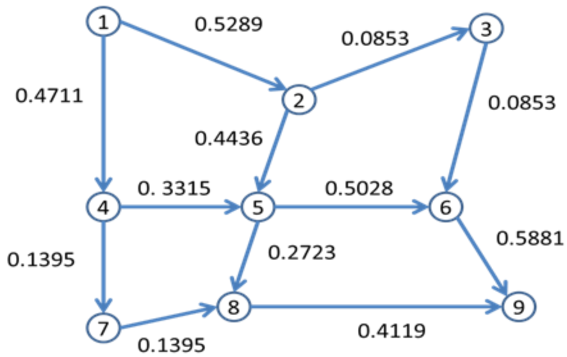

Figure 3 - 6 Link use probabilities for vehicle class 2 (medium truck) and O-D pair 1- 9.

If we interpret these probabilities as proportions of a total flow of vehicles from origin to destination, we have conservation of flow at the nodes, and the product $p_{ij}^{am} q_{ij}^{m}$ represents part of the flow volume for link *a*. For example, if the O-D volume of class 2 trucks from 1 to 9 is 500, the expected volumes (rounded to integers) on the links of the network contributed by that vehicle class and O-D pair, are as shown in Figure 3 - 7.

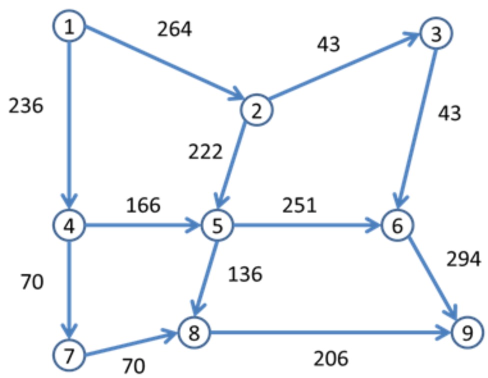

Figure 3 - 7 Expected flow volumes for vehicle class 2(medium truck) and O-D pair 1- 9, if the O-D flow is 500.

Of course, the total flow within each vehicle class on any network link may have contributions from many O-D pairs, and the observed counts on any link may include multiple vehicle classes, so in general, we have the relationship shown earlier as Equation 3-2.

The link utilization probabilities also implicitly define probabilities for use of different paths between the origin and destination. Consider again the situation shown in Figure 3 - 6. If the probability of a truck using link 1-2 is 0.5289, that value must be the sum of probabilities for all paths that contain link 1-2. At node 2, that probability is split between paths that continue to node 3 and those that go to node 5. Thus, we can say that the probability of the partial path 1-2-3 is 0.0853, and the probability of the partial path 1-2-5 is 0.4436. At node 5, the partial path 1-2-5 merges with the partial path 1-4-5, and an assumption is necessary about how vehicles entering node 5 from different directions exit that node. This information may be provided by explicit turning probabilities (which could be calibrated from video surveillance of an intersection, for example). In the absence of such specific information, we might make a Markov assumption that each inbound partial path splits in the same proportions outbound. That is, since the ratio of the outbound probabilities at node 5 is $\frac{0.5028}{0.2723}$, each inbound partial path will divide in that proportion. Thus, the probability of the partial path 1-2-5-6 will be $0.4436\left(\frac{0.5028}{0.5028+0.2723}\right)=0.2878$, and the probability of the partial path 1-2-5-8 will be $0.4436\left(\frac{0.2723}{0.5028+0.2723}\right)=0.1559$.

We continue splitting partial paths and computing probabilities until we reach the destination, at which point we have complete origin-destination paths identified. In this small example, there are six potential paths between node 1 and node 9, with usage probabilities computed as follows:

Path	Probability
1-2-3-6-9	0.0853
1-2-5-6-9	0.2878
1-2-5-8-9	0.1559
1-4-5-6-9	0.2150
1-4-5-8-9	0.1165
1-4-7-8-9	0.1395

It can easily be verified that these path probabilities sum to 1.0. We also note that the path (1-2-5-6-9) that was computed to be the shortest path (see Figure 3 - 3) has the highest probability of use. In this stochastic flow assignment, shorter paths have a higher probability of use, but less direct paths (like 1-2-3-6-9, for example) have non-zero probability of use. That is, not all traffic is assumed to use the single shortest path.

Being able to compute expected path probabilities for a given vehicle class and O-D pair allows us to take advantage of potential observations of vehicles at multiple locations. For example, suppose some number of class 2 trucks were observed on the sequence of links 2-5 and 5-8. Those trucks could be traveling from node 1 to node 9, using path 1-2-5-8-9, or they could be part of some other O-D pair (in this case, from 3 to 7, using path 3-2-5-8-7). The relative probabilities for the various paths give us useful information about how to factor such observations into our overall analysis.

3.5 Observations and Estimating O-D Tables (Algorithm Step 2)

To illustrate the use of count data to estimate an O-D table, let us suppose that traffic counts on links 1-4, 3-6, 5-6 and 5-8 are available. On link 5-6, the counts separate the vehicle classes in each direction. On links 1-4 and 5-8, the counts are from dual loop detectors and separate automobiles from trucks, but both classes of trucks are combined. Furthermore, on link 4-1, the truck count is missing. On link 3-6, counts are available in both directions, but are simply total vehicles (from a single loop counter). The 15 available observed data are as follows:

Link	Count
1-4	642 (automobiles – class 1)
1-4	48 (total trucks – classes 2 and 3)
4-1	654 (automobiles – class 1)
3-6	676 (total vehicles – all classes)
6-3	639 (total vehicles – all classes)
5-6	916 (automobiles – class 1)
5-6	34 (medium trucks – class 2)
5-6	45 (heavy trucks – class 3)
6-5	918 (automobiles – class 1)
6-5	43 (medium trucks – class 2)
6-5	67 (heavy trucks – class 3)
5-8	1004 (automobiles – class 1)
5-8	131 (total trucks – classes 2 and 3)
8-5	1006 (automobiles – class 1)
8-5	109 (total trucks – classes 2 and 3)

For this example, there are four O-D pairs for each vehicle class, or a total of 12 O-D entries to be estimated. Thus, we have more observations than unknowns, but the data may not all be consistent, so there may be no trip table that would match these observations exactly. Also, in large networks, there are normally many more O-D table elements than we have observations, so the problem is usually underspecified.

We use the SAM algorithm to compute link utilization probabilities and path probabilities for vehicle classes 1 and 3, just as for vehicle class 2 above, but the probabilities are different because the cost functions (3-9) and (3-11) for vehicle classes 1 and 3 are different from vehicle class 2. Figure 3 - 8 shows the link use probabilities for vehicle class 1 and O-D pair 1-9; Figure 3 - 9 shows the comparable values for vehicle class 3. These figures can be compared to Figure 3 - 6, which pertains to vehicle class 2. The probabilities for the other O-D pairs are also computed, but not shown explicitly in figures.

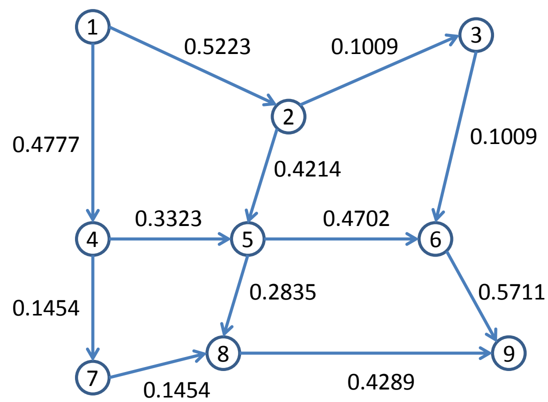

Figure 3 - 8 Link use probabilities for vehicle class 1 (automobiles) and O-D pair 1- 9.

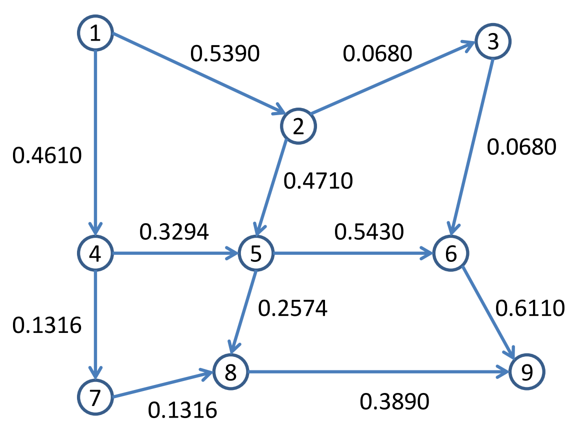

Figure 3 - 9. Link use probabilities for vehicle class 3 (heavy trucks) and O-D pair 1- 9.

The following equations relate the observed data to the unknown trip table values.

$$642 = 0.4777 q_{19}^{1} + 0.0577 q_{37}^{1} + e_1$$

$$48 = 0.4711 q_{19}^{2} + 0.0415 q_{37}^{2} + 0.4610 q_{19}^{3} + 0.0266 q_{37}^{3} + e_2$$

$$654 = 0.0648 q_{73}^{1} + 0.4805 q_{91}^{1} + e_3$$

$$676 = 0.1009 q_{19}^{1} + 0.4060 q_{37}^{1} + 0.0853 q_{19}^{2} + 0.3757 q_{37}^{2} + 0.0680 q_{19}^{3} + 0.3365 q_{37}^{3} + e_4$$

$$639 = 0.3961 q_{73}^{1} + 0.0964 q_{91}^{1} + 0.3642 q_{73}^{2} + 0.0809 q_{91}^{2} + 0.3241 q_{73}^{3} + 0.0640 q_{91}^{3} + e_5$$

$$916 = 0.4702 q_{19}^{1} + 0.2934 q_{73}^{1} + e_6$$

$$34 = 0.5028 q_{19}^{2} + 0.2786 q_{73}^{2} + e_7$$

$$45 = 0.5429 q_{19}^{3} + 0.2575 q_{73}^{3} + e_8$$

$$918 = 0.2996 q_{37}^{1} + 0.4656 q_{91}^{1} + e_9$$

$$43 = 0.2848 q_{37}^{2} + 0.4985 q_{91}^{2} + e_{10}$$

$$67 = 0.2635 q_{37}^{3} + 0.5387 q_{91}^{3} + e_{11}$$

$$1004 = 0.2835 q_{19}^{1} + 0.5530 q_{37}^{1} + e_{12}$$

$$131 = 0.2723 q_{19}^{2} + 0.6044 q_{37}^{2} + 0.2574 q_{19}^{3} + 0.6638 q_{37}^{3} + e_{13}$$

$$1006 = 0.5518 q_{73}^{1} + 0.2868 q_{91}^{1} + e_{14}$$

$$109 = 0.6041 q_{73}^{2} + 0.2748 q_{91}^{2} + 0.6645 q_{73}^{3} + 0.2589 q_{91}^{3} + e_{15}$$

Using these 15 observations, we can estimate O-D tables for the three vehicle classes by solving:

$$\underset{q_{ij}^{m} \geq 0}{Min} \quad \sum_{l=1}^{15} \left(e_l\right)^2$$

The solution to the minimization problem (rounded to the nearest integer number of trips) is:

$$q_{19}^{1} = 1199 \qquad q_{19}^{2} = 50 \qquad q_{19}^{3} = 41$$

$$q_{37}^{1} = 1200 \qquad q_{37}^{2} = 98 \qquad q_{37}^{3} = 71$$

$$q_{73}^{1} = 1200 \qquad q_{73}^{2} = 32 \qquad q_{73}^{3} = 88$$

$$q_{91}^{1} = 1199 \qquad q_{91}^{2} = 30 \qquad q_{91}^{3} = 90$$

Because the link utilization probabilities from the SAM algorithm were constructed in step 1 based on free-flow conditions, and the trip table we have just estimated, if assigned to the network, will produce link volumes that are not free-flow, the entire process is not yet converged. That is the reason for steps 3-8 in the algorithm. However, we have constructed an initial estimate of the three trip tables that can be the basis for further iterations.

In this small example, the actual trip tables that were assigned to the network and used to generate the observed flows are known, even though in field applications of the O-D estimation process that is not the case. The actual O-D tables have not been used in any way in the estimation process, but we can compare the current estimated trip tables with the actual values to judge how similar they are. The actual trip tables are as follows:

$$q_{19}^{1} = 1200 \qquad q_{19}^{2} = 50 \qquad q_{19}^{3} = 40$$

$$q_{37}^{1} = 1200 \qquad q_{37}^{2} = 100 \qquad q_{37}^{3} = 70$$

$$q_{73}^{1} = 1200 \qquad q_{73}^{2} = 30 \qquad q_{73}^{3} = 90$$

$$q_{91}^{1} = 1200 \qquad q_{91}^{2} = 30 \qquad q_{91}^{3} = 90$$

It is clear that the initial estimates are very close to the true values, even though data were not available for all network links, some of the data aggregated vehicle classes, and the portion of the algorithm representing the SUE flow calculations has not yet been invoked.

Although in this example the trip tables estimated based on link counts only are very close to the true values, the 9-node example network is also useful to illustrate the incorporation of turning count data at intersections. For this illustration, turning movement data from three of the four approaches at node 5 is used and the assumption is that those turning movements are total counts (i.e., not specific by vehicle class). The available data are summarized in Figure 3 - 10. Turning movements are not specified for the approach from node 2. This helps illustrate that the computations can be done using any available data, and that not all approaches need to be covered.

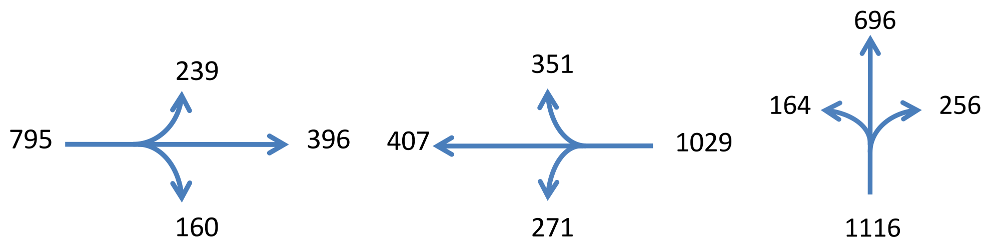

Figure 3 - 10. Available turning movement data at node 5.

A turning movement observation is equivalent to observing a vehicle on a sequence of two links. That is, the 396 vehicles that go straight through the intersection in the leftmost diagram in Figure 3 - 10 are vehicles that use the sequence of links 4-5 and then 5-6. This sequence is part of two paths for different O-D pairs:

O-D Pair	Path
1-9	1-4-5-6-9
7-3	7-4-5-6-3

As part of the SAM calculations, we have determined that the probabilities of these paths for the three vehicle classes on the respective O-D pairs are:

Table 3 - 1 Example path probabilities in 9-node network

		Probabilities		
O-D Pair	Path	Vehicle Class 1	Vehicle Class 2	Vehicle Class 3
1-9	1-4-5-6-9	0.2073	0.215	0.2234
7-3	7-4-5-6-3	0.0989	0.0845	0.0679

This allows construction of another observation constraint for the O-D table estimation:

$$396 = 0.2073 q_{19}^{1} + 0.2150 q_{19}^{2} + 0.2234 q_{19}^{3} + 0.0989 q_{73}^{1} + 0.0845 q_{73}^{2} + 0.0679 q_{73}^{3} + e_{16}$$

The observations of turning movements are not simple link counts and the coefficients in the constraint are not link probabilities, so this is a constraint of the form in Equation 3-3 and the coefficients are examples of u_{ij}^{ml} values. Each of the movements illustrated in Figure 3 - 10 can be converted into an observational constraint of the general form given in Equation 3-3:

$$239 = 0.1818\, q_{73}^{1} + 0.1785\, q_{73}^{2} + 0.1702\, q_{73}^{3} + e_{17}$$

$$160 = 0.1250\, q_{19}^{1} + 0.1165\, q_{19}^{2} + 0.1059\, q_{19}^{3} + e_{18}$$

$$351 = 0.2618\, q_{91}^{1} + 0.2867\, q_{91}^{2} + 0.3185\, q_{91}^{3} + e_{19}$$

$$407 = 0.1014\, q_{37}^{1} + 0.0864\, q_{37}^{2} + 0.0693\, q_{37}^{3} + 0.2038\, q_{91}^{1} + 0.2117\, q_{91}^{2} + 0.2202\, q_{91}^{3} + e_{20}$$

$$271 = 0.1982\, q_{37}^{1} + 0.1984\, q_{37}^{2} + 0.1943\, q_{37}^{3} + e_{21}$$

$$164 = 0.1255\, q_{91}^{1} + 0.1167\, q_{91}^{2} + 0.1058\, q_{91}^{3} + e_{22}$$

$$696 = 0.3579\, q_{73}^{1} + 0.4099\, q_{73}^{2} + 0.4750\, q_{73}^{3} + 0.1613\, q_{91}^{1} + 0.1581\, q_{91}^{2} + 0.1531\, q_{91}^{3} + e_{23}$$

$$256 = 0.1945\, q_{73}^{1} + 0.1941\, q_{73}^{2} + 0.1895\, q_{73}^{3} + e_{24}$$

When the above observations are added to the minimization problem, the resulting solution is:

$q_{19}^{1} = 1199$ $\qquad q_{19}^{2} = 48$ $\qquad q_{19}^{3} = 42$

$q_{37}^{1} = 1201$ $\qquad q_{37}^{2} = 91$ $\qquad q_{37}^{3} = 78$

$$q_{73}^{1} = 1200 \qquad q_{73}^{2} = 34 \qquad q_{73}^{3} = 85$$

$$q_{91}^{1} = 1199 \qquad q_{91}^{2} = 34 \qquad q_{91}^{3} = 86$$

This solution is very similar to the previous solution. In this example, the addition of turning movement data is not important for improving the solution, but the example does illustrate how non-link count data can be incorporated into the O-D estimation process. In subsequent experiments on a larger and more complicated network in Chapter 4, the value of turning movement data will be clearer.

Other types of partial path information may also be represented in similar constraints if the data acquisition equipment can provide it. For example, license plate matching capability could produce an observation that a certain number of vehicles were observed on both links 1-2 and 7-8 (i.e., non-sequential links). These two links will be parts of specific paths for which probabilities have been computed, and constraints of the same form as the ones above can be constructed. The key is in identifying the potential "efficient" paths within the SAM algorithm and then tying counts of vehicles observed on parts of those paths to the possible O-D pairs on which they could be traveling.

3.6 Consistent Solutions and Stochastic Equilibrium

If the O-D volumes resulting from the zero-flow estimates of link use probabilities were assigned to flow over the network, they would imply a set of link volumes that would create different link travel times and costs than those used to construct the link utilization probabilities, and thus the probabilities used so far are not consistent with the estimated O-D volumes. Step 3 in the overall algorithm described at the beginning of section 3 updates those probability estimates and resulting link volumes to be consistent with the current O-D table estimates, thus creating a collection of O-D volumes, link volumes and probabilities [denoted by (*Q, X, P, U*)] that are all internally consistent. This will be termed a *consistent solution* to the overall problem.

Finding a consistent solution depends on solving for stochastic user equilibrium (SUE) in the network -- a condition in which the link volumes, resulting travel times and costs, and estimated link utilization probabilities are all consistent with the behavioral principle that shorter (or faster or cheaper) paths for any given O-D pair should have higher probability of use than longer (or slower or more expensive) paths.

Finding the multiclass SUE solution is a process of solving the equilibrium conditions posed by Daganzo (1983) – Equation 3-1, repeated here for easy reference.

$$x_a^m - \sum_{ij} q_{ij}^m p_{ij}^{am}\left(Q,c\right) = 0, \qquad a = 1,\ldots,n; \quad m = 1,\ldots,M$$ **Equation 3-1**

Equation 3-1 may be viewed as the first-order conditions for an optimization problem, as described (for a single vehicle class) by Sheffi (1985). This optimization is:

$$\underset{x_a}{Min} = -\sum_a \int_0^{x_a} c_a(w)dw + \sum_a x_a c_a(x_a) - \sum_{ij} q_{ij} S_{ij}(c)$$ **Equation 3 - 25**

The first term in Equation 3-25 is the function that would be minimized at a deterministic user equilibrium solution. The second term is the total cost of travel in the network, and would be minimized at a system-optimal solution. The third term is the expected value of the total perceived network cost, using the Satisfaction function, $S_{ij}(c)$, defined as the expected perceived minimum cost of travel for O-D pair *ij*:

$$S_{ij}(c) = E\left[\min_k \tilde{g}_{ij}^k\right]$$ **Equation 3 - 26**

$\tilde{g}_{ij}^k$ is the perceived path cost for path *k* from origin *i* to destination *j*. Along the links of a path the costs add, so the cost along a path, *k*, is a random variable as described earlier. The expectation in Equation 3-26 is with respect to the distribution of the error terms in the perceived path costs, which depend on the link costs, so we write $S_{ij}(c)$ as a function of the set of link costs, *c*.

Maher and Hughes (1997) developed an algorithm for solving the optimization in Equation 3-25, taking advantage of the fact that the $S_{ij}(c)$ function is implicitly evaluated during the performance of the SAM algorithm. Their algorithm was originally designed for use with a single vehicle class, but the modification to include multiple vehicle classes is straightforward. Their algorithm involves iterative application of the SAM algorithm, combined with a search mechanism to average the successive solutions. Extending this process to include multiple vehicle classes doing the SAM calculations at each iteration for all vehicle classes separately, calculating a satisfaction function, $S_{ij}^m(c^m)$, function for each vehicle class, and updating link travel times for the next iteration based on the aggregate equivalent flow on each link.

Step 3 of the multiclass O-D estimation algorithm listed at the beginning of this chapter is the application of the multiclass version of the Maher-Hughes method for finding SUE solutions, using the current estimated trip tables, to produce the necessary consistent solution at a given iteration.

3.7 Solving the Nonlinear Optimization at Steps 2 and 4 of the Procedure

The construction of estimated trip tables, Q, with the link utilization probabilities, P, and coefficients, U, treated as constants, requires solution of a problem of the form:

$$\underset{q_{ij}^m \geq 0}{Min} \quad \sum_{l \in \bar{B}} \eta_l \left(B_l - \sum_{m \in M_l} \sum_{ij \in \mathrm{N}_l} u_{ij}^{ml} q_{ij}^m \right)^2 + \sum_{l \in \bar{X}} \tau_l \left(\tilde{x}_l - \sum_{m \in M_l} \sum_{ij} p_{ij}^{a_l m} q_{ij}^m \right)^2 \qquad \textbf{Equation 3 - 27}$$

This is a quadratic optimization, and for the solution to be sensible the trip table elements must be non-negative. One effective method for finding a solution to this problem is the L-BFGS-B algorithm (Byrd, *et al*. ,1995; Zhu and Nocedal, 1997; and Morales and Nocedal, 2011). The core idea on which this algorithm is based is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method for nonlinear optimization. The L-BFGS version of this algorithm uses some approximations in the storage of derivatives to create a limited-memory (L) version of the method that is effective for large problems. The further extension to handle simple box constraints on variables (the B suffix) creates an effective method for the problem faced here.

The L-BFGS-B method is one of a class of algorithms called quasi-Newton methods. These algorithms are based on a core idea that if a function $f(\vec{q})$, where $\vec{q}$ is an n x 1 vector, is twice continuously differentiable, we can construct a quadratic approximation to it at a point $\vec{q}$ by using a Taylor series expansion around a point $\vec{q}^k$:

$$f(\vec{q}) \cong f(\vec{q}^k) + \left(\vec{q} - \vec{q}^k\right) \nabla f(\vec{q}^k) + \frac{1}{2}\left(\vec{q} - \vec{q}^k\right)^T \nabla^2 f(\vec{q}^k)\left(\vec{q} - \vec{q}^k\right) \qquad \textbf{Equation 3 - 28}$$

If $\nabla^2 f(\vec{q})$ is positive definite, then the minimum of the quadratic approximation occurs at the point $\vec{q}*$ where $\nabla f(\vec{q}*) = 0$. This point is found by solving:

$$\vec{q}* = \vec{q}^k - \left[\nabla^2 f(\vec{q}^k)\right]^{-1} \nabla f(\vec{q}^k) \qquad \textbf{Equation 3 - 29}$$

Newton's Method involves using Equation 3-29 iteratively, finding the next approximation $\vec{q}^{k+1}$ to the minimum of $f(\vec{q})$ by evaluating the gradient and the Hessian (matrix of second partial derivatives) of f at a point $\vec{q}^k$, and then computing the next estimate as:

$$\vec{q}^{k+1} = \vec{q}^k - \left[\nabla^2 f(\vec{q}^k)\right]^{-1} \nabla f(\vec{q}^k) \qquad \textbf{Equation 3 - 30}$$

The process is then repeated with $\vec{q}^{k+1}$ replacing $\vec{q}^k$.

In practice, Newton's method has several shortcomings. It requires storage and repeated computation of the inverse Hessian, and if the Hessian is ill-conditioned, the algorithm can be unstable or fail to converge. The techniques known as quasi-Newton methods have been developed to overcome these problems. The general ideas underlying quasi-Newton methods are described in several texts on nonlinear optimization and solution of nonlinear equations (e.g., Dennis and Schnabel, 1996; Luenberger and Ye, 2008). The central idea is to construct and update an approximation to the inverse Hessian in a way that does not require explicit inversion and that ensures that the approximation is always positive definite. Quasi-Newton method contains a whole family of algorithms. The difference lies in how they do the approximation of Hessian. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm has proven over time to be an effective and reliable means for doing that. The L-BFGS variation stores a simplified version of the whole *n* by *n* Hessian approximation matrix, using a history of the past *r* updates of the position and the corresponding gradient. L-BFGS-B, the bounded variation of L-BFGS, uses the idea of an active set to accommodate simple constraints. At every step, the algorithm identifies fixed and free variables using a simple gradient method and runs the L-BFGS only on the free variables.

In our implementation, we used an open-source java wrapper (Kobos, 2012), which wraps from the original Fortran code to solve the optimization problem in step 2 and 4 of the O-D estimation procedure.

3.8 The Search Process at Steps 6 and 7 of the Algorithm

At the end of step 5 of the algorithm, we have two feasible solutions to the O-D estimation problem – a current solution created through steps 2 and 3, and a new trial solution created in steps 4 and 5. The trial solution should be in a direction from the current solution that produces an improvement overall, but there is an important question of how far to move in that direction. This creates a search problem that operates along the line between the current solution and the trial solution.

To implement that search process, we linearize the assignment map between the two feasible solutions (the current solution Q_n and the trial solution Y_n) so that we can express potential trip tables and associated link volumes as a function of a step size, γ , where $0 \le \gamma \le 1$:

$$Q(\gamma) = Q_n + \gamma\left(Y_n - Q_n\right) \quad \textbf{Equation 3 - 31}$$

$$X(\gamma) = X_n + \gamma\left(V_n - X_n\right) \quad \textbf{Equation 3 - 32}$$

Then we can search for the value of γ that solves:

$$\underset{0\le\gamma\le1}{Min} \quad f(\gamma)=\sum_{l\in\breve{B}}\eta_l\left(B_l-\sum_{m\in M_l}\sum_{ij\in N_l}u_{ij}^{ml}q_{ij}^{m}(\gamma)\right)^2+\sum_{l\in\breve{X}}\tau_l\left[\tilde{x}_l-x_l(\gamma)\right]^2 \qquad \textbf{Equation 3 - 33}$$

This is a one-dimensional optimization because the only variable is γ. We'll denote the optimal value of γ as γ^*.

Since $f(\gamma)$ is convex, we can solve for γ^* analytically by setting $f'(\gamma)=0$. This gives us

$$\boldsymbol{\gamma}^* = \frac{\sum_l \boldsymbol{\eta}_l(\boldsymbol{B}_l-\boldsymbol{c}_l)(\boldsymbol{d}_l-\boldsymbol{c}_l)+\sum_l \boldsymbol{\tau}_l(\widetilde{\boldsymbol{x}}_l-\boldsymbol{x}_l)(\boldsymbol{v}_l-\boldsymbol{x}_l)}{\sum_l \boldsymbol{\eta}_l(\boldsymbol{d}_l-\boldsymbol{c}_l)^2+\sum_l \boldsymbol{\tau}_l(\boldsymbol{v}_l-\boldsymbol{x}_l)^2} \qquad \textbf{Equation 3 - 34}$$

where:

$$\boldsymbol{c}_l = \sum_m \sum_{ij} \boldsymbol{u}_{Qij}^{ml} \cdot \boldsymbol{Q}_{ij}^{m} \qquad \textbf{Equation 3 - 35}$$

$$\boldsymbol{d}_l = \sum_m \sum_{ij} \boldsymbol{u}_{Yij}^{ml} \cdot \boldsymbol{Y}_{ij}^{m} \qquad \textbf{Equation 3 - 36}$$

$$\boldsymbol{x}_l = \sum_m \sum_{ij} \boldsymbol{p}_{Qij}^{ml} \cdot \boldsymbol{Q}_{ij}^{m} \qquad \textbf{Equation 3 - 37}$$

$$\boldsymbol{v}_l = \sum_m \sum_{ij} \boldsymbol{p}_{Yij}^{ml} \cdot \boldsymbol{Y}_{ij}^{m} \qquad \textbf{Equation 3 - 38}$$

$$l \in \breve{B}_l$$

$$m \in M_l$$

$$i,j \in N_l$$

These required calculations are straightforward and can be done quite easily.

3.9 Conclusions

This chapter describes a formulation of the multiclass O-D estimation problem that meets specified criteria. It is constructed to accept a variety of different types of data that relate O-D volumes to observed values. This includes traditional link count data, but also accommodates other types of data such as turning movements at intersections, partial path observations from individual vehicles, etc. The data may be either class-specific or may aggregate a subset of vehicle classes. A particular focus of this formulation is to allow better estimation of truck flows in urban networks, using various size classes of trucks, and separating them from estimation of automobile O-D patterns.

The O-D estimation process is constructed using a probit-based stochastic model for network loading and user equilibrium. Using a stochastic model of user equilibrium (SUE) allows a relaxation of the common assumption that all drivers have complete and accurate knowledge of link travel costs everywhere in the network. The stochastic model includes errors in individual drivers' perceptions of those costs. Although the probit model of SUE is more demanding computationally than the logit-based model, it has important conceptual advantages with respect to the assumption about the distribution of errors and in reflecting covariance among path alternatives that share common links.

Sections 3.3-3.8 describe in detail an algorithm for solving the O-D estimation problem formulated in section 3.2. This algorithm borrows pieces from the work of several previous authors, but it represents a significant advance in capability for estimating multiclass O-D tables.

The following chapter describes an extensive set of experiments with the algorithm on a test network. These experiments illustrate the use of the algorithm, and also contribute insight into the relation between the amount and character of available data and the quality of the resulting O-D estimates.

CHAPTER 4

TESTS AND ANALYSIS

To test the concepts and solution algorithm, a series of experiments has been developed. These experiments are designed to test performance of the solution method under varying conditions and with varying types of data. Section 4.1 introduces the test network and the method to generate observation data we used in all following tests. Section 4.2 proposes several questions we are interested in and thus designing the tests upon. Section 4.3 describes the algorithm implementation details and how we evaluate test results. Section 4.4 to section 4.10 answers the questions proposed in section 4.2 in details using test results and corresponding analysis. Section 4.11 concludes what we found.

4.1 Test Network

All the tests are performed on the “Sioux Falls” (SF) network, shown in Figure 4 - 1. It originated from a representation of part of the street network in Sioux Falls, South Dakota. This network, first constructed and used by LeBlanc, et al. (1975), has since become a “standard” test network for many types of transportation network algorithms. We have borrowed the basic structure of the network from the original version used by LeBlanc, et al. (1975), but we have created O-D tables and link characteristics that enhance the network’s usefulness as a test bed for our multiclass O-D estimation algorithm. By creating an O-D table for each vehicle class and using those tables to create assigned link volumes, we create a test bed where we can select different sets of volume data and see how well the algorithm can reproduce the known O-D tables under varying conditions of available data.

Three vehicle classes (see Table 4 - 1) and two theoretical O-D tables have been created. The first O-D table contains 4 origin and destination zones, which generates a 4 by 4 table for each vehicle class (See Table 4 - 2, Table 4 - 3, and Table 4 - 4). This set of O-D tables is relatively simple and the flow pattern of automobile and truck are quite similar. The second set of O-D tables contains 7 origin and destination zones and has considerably different flow pattern between automobile and truck (see Table 4 - 5, Table 4 - 6, and Table 4 - 7). In the 7 by 7 O-D tables, node 10 is assumed to be a zone at downtown area, where a lot of automobile trips begin and end. Node 13 is a loading area at the periphery of the city, where a relatively higher percentage of truck trips begin and end.

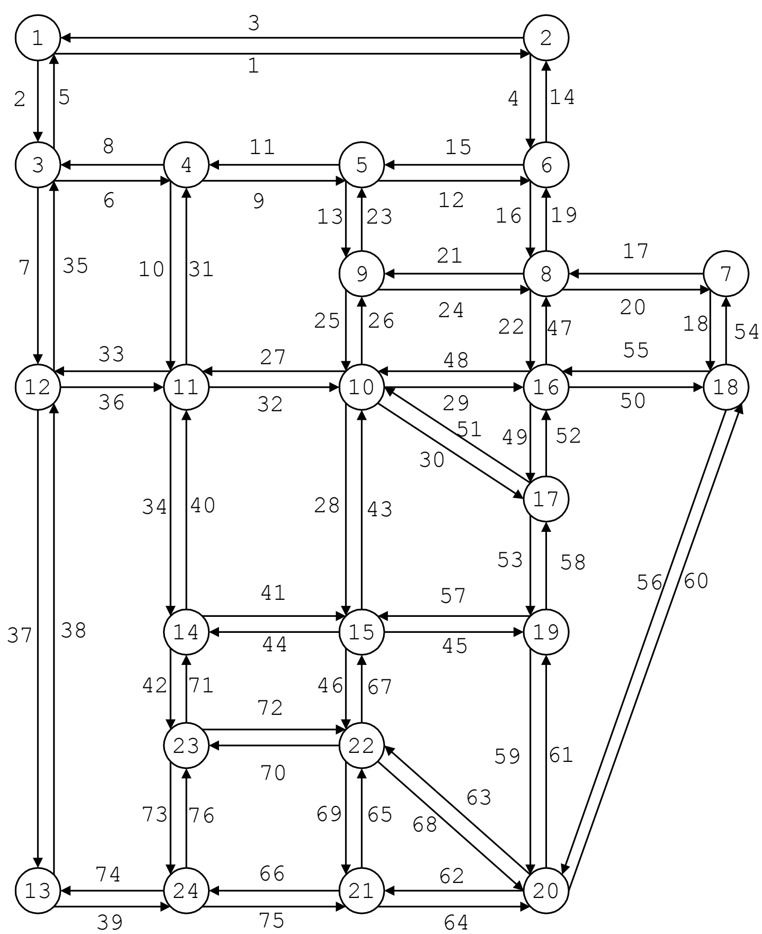

Figure 4 - 1 Sioux Falls network

Table 4 - 1: Vehicle class information in Sioux Falls network

Class number	Type	Cost of time Coefficient	Cost of distance Coefficient	Vehicle equivalents
1	Automobile	0.2	0.25	1
2	Medium truck	0.33	1	2
3	Heavy truck	0.5	1.5	3

Table 4 - 2: Four-zone O-D table for vehicle class 1 (veh/hr)

O\D	1	7	15	20
1	0	3109	839	891
7	3117	0	2462	1994
15	846	2456	0	1788
20	881	2010	1779	0

Table 4 - 3: Four-zone O-D table for vehicle class 2 (veh/hr)

O\D	1	7	15	20
1	0	215	61	57
7	214	0	163	141
15	58	161	0	123
20	64	131	110	0

Table 4 - 4: Four-zone O-D table for vehicle class 3 (veh/hr)

O\D	1	7	15	20
1	0	176	50	52
7	169	0	125	103
15	46	133	0	89
20	55	109	111	0

Table 4 - 5: Seven-zone O-D table for vehicle class 1 (veh/hr)

O\D	1	6	7	10	13	15	20
1	0	491	440	1070	532	839	359
6	491	0	306	269	224	266	264

7	442	308	0	1348	90	723	90
10	1064	267	1343	0	532	533	794
13	526	221	90	537	0	708	91
15	846	267	714	534	719	0	1069
20	355	267	88	807	89	1071	0

Table 4 - 6 : Seven-zone O-D table for vehicle class2 (veh/hr)

O\D	1	6	7	10	13	15	20
1	0	14	37	121	69	24	19
6	25	0	19	37	16	7	10
7	30	19	0	50	37	7	18
10	116	45	62	0	116	38	49
13	68	23	30	113	0	27	25
15	19	7	14	33	25	0	15
20	19	3	18	33	28	3	0

Table 4 - 7: Seven-zone O-D table for vehicle class 3 (veh/hr)

O\D	1	6	7	10	13	15	20
1	0	42	53	121	41	48	39
6	38	0	8	19	13	6	8
7	68	20	0	37	18	10	18
10	115	34	40	0	26	28	23
13	33	15	8	36	0	18	14
15	27	7	17	32	8	0	14
20	32	13	21	23	5	8	0

Through SUE calculation, these trip table produce link and path flows for each vehicle class. We can then sample link counts, turning observations at intersections, etc., and create observations that may be separated by vehicle class or aggregated in various ways. This creates a test environment where different sets of assumed data can be specified and the algorithm can be run to estimate multiclass O-D tables. Because the original "correct" O-D tables are known, the result of various experiments can be evaluated by comparing the estimated tables with the correct tables.

4.2 Questions for testing

The following are the questions investigated. All the tests are designed to answer these questions, and are discussed individually in the indicated sections.

1) If classified link count data are available on all the links in the network, how accurately can the three vehicle class O-D tables be estimated? (Section 4.4)
2) If classified link counts are available, but not all links are covered, how does the accuracy of the O-D estimates change as the proportion of links with available count data changes? (Section 4.5)
3) How does the accuracy of O-D estimation change if only some of the available link counts classify three vehicle classes, and the rest simply report total vehicles in all classes? (Section 4.6)
4) Are dual loop detectors (capable of separating automobiles from trucks, but reporting an aggregate of the two truck classes) an acceptable substitute for classified link count sensors? (Section 4.7)
5) In many cities, the available count data is limited to a relatively small fraction of the total links, and vehicle classification information is available on only a portion of the links with counts. Under these conditions, is it possible to estimate multiclass O-D tables with reasonable confidence? (Section 4.8)

6) If, in addition to link count data, turning movement observations are also available at key intersections, by how much is the accuracy of the multiclass O-D estimates improved? (Section 4.9)
7) If the observed data contain significant errors, how is the accuracy of the resulting O-D estimates degraded? (Section 4.10)

4.3 Algorithm Implementation and Evaluation of Test Results

The algorithm is coded in java based on the solution method in section 3.2. The implementation of the L-BFGS-B algorithm is adapted from an open-source java wrapper (Kobos 2012). To check whether the solution has converged, root mean square (RMS) is used as a criterion. RMS is a statistical measure of the magnitude of a varying quantity. It measures how close the two values or two sets of values are. The calculation of RMS is shown in Equation 4-1.

$$RMS = \sqrt{\frac{\sum_m \sum_{ij} \frac{(\tilde{q}_{nij}^m - \tilde{q}_{(n-1)ij}^m)^2}{0.5 \times (\tilde{q}_{nij}^m + \tilde{q}_{(n-1)ij}^m)}}{\sum_m N_m}} \qquad \textbf{Equation 4 - 1}$$

where:

$\tilde{q}_{nij}^m$ = the O-D volume estimated in n[th] iteration for vehicle class m from origin i to destination j

N_m = the number of O-D pairs for vehicle class m

The termination criteria for the algorithm is set to be maximum 200 iterations or RMS of two successive iterations of O-D volume estimates < 0.1, whichever comes first. An RMS of 0.1 means on average the squared error of an O-D pair between the two successive iterations is 1%.

From our experiments, most tests converge at about 10 to 30 iterations. If a test doesn't converge even at 60 iterations, normally it will go all the way until the preset maximum iteration. When this happens, the objective value is just bouncing back and forth around the optimum value without much improvement. Therefore setting maximum iteration to 200 should be enough. After the termination criteria are met, the code will output the last set of O-D pairs calculated from the optimization.

RMS can only ensure that the successive two optimization results are close enough. However, it doesn't necessarily mean that the estimation is close enough to the "correct" value. The estimation error is computed by the percent difference between the estimated value and true value (see Equation 4-2)

$$\boldsymbol{\varepsilon_{ij}^m = \frac{(\tilde{q}_{ij}^m - q_{ij}^m)}{q_{ij}^m} \times 100\%} \qquad \textbf{Equation 4 - 2}$$

where:

$\tilde{q}_{ij}^m$ = the estimated O-D volume for vehicle class m from origin i to destination j

q_{ij}^m = the true O-D volume for vehicle class m from origin i to destination j

ε_{ij}^m = the estimation error for O-D pair from i to j for vehicle class m

For each test run, we can derive a set of estimation errors, one for each entry of the O-D table. After sequencing all the estimation errors from the smallest to the biggest from one run of test, we found that it basically follows an 's' curve. Figure 4 - 2 is an example when running the algorithm on SF network with classified link count observations on all the links for 4 by 4 O-D tables. This pattern shows in all other tests as well.

We want to know how this solution method performs overall. We considered an O-D pair as estimated acceptably correctly if the estimation error is within $\pm$5%. Then the percentage of

O-D pairs estimated correctly for each vehicle class and overall can be used as performance measures to compare the results between different tests. Not only do we care about how many O-D pairs are estimated correctly, but also the volume captured within that range. If the corresponding O-D volumes for the O-D pairs estimated within ±5% range is small, that means a large proportion of traffic is estimated incorrectly, which is undesirable even if the method captures a high percentage of O-D pairs. Therefore, the percentage of O-D volume (classified or aggregated) that is estimated within ±5% range is also an important performance measure.

For the example shown in Figure 4 - 2, 89% of total O-D pairs fall into the range between -5% and +5%, which accounts for 96% of O-D volume. After decompose them into different vehicle classes, 91.7% of automobile, 91.7% of medium tracks, and 83.3% of heavy truck O-D pairs are estimated within the acceptable error, with corresponding volume of 96.2%, 95.9% and 91.6% respectively.

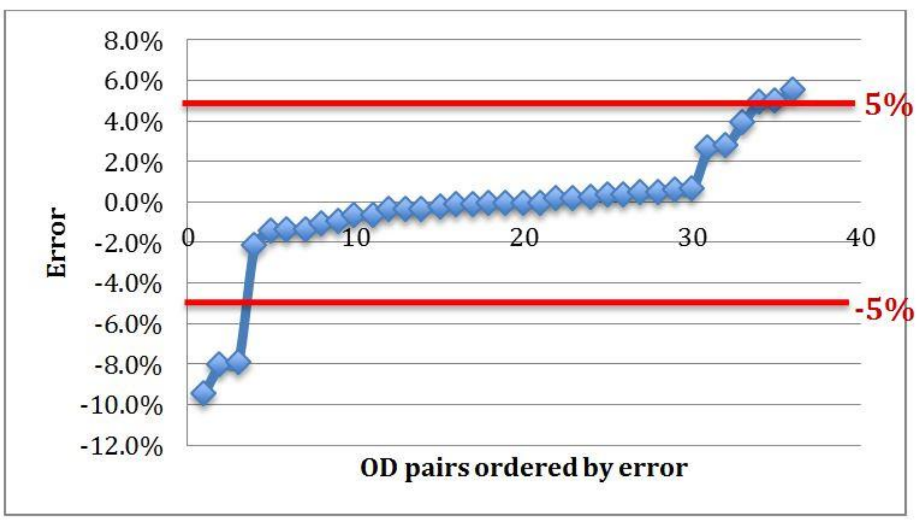

Figure 4 - 2: Sequenced error for 100% link coverage with classified vehicle counts

4.4 Q1: Estimating Multiclass O-D tables from Classified Link Counts

Suppose we have classified link counts on all the links in the network. Can the true multiclass O-D tables be estimated accurately? When implemented on the 4-zone O-D tables, the result is very positive. As we can see from Table 4 - 8, 89% of O-D pairs and 96% of volume are estimated within 5% by the solution method. The results are consistent across all vehicle classes. However, in the 7-zone O-D pair tests, the outcome is quite poor. Only 12.7% of O-D pairs are estimated within 5%, which counts about 18.8% of the total traffic volume.

Table 4 - 8 Estimation results when classified link counts with full link coverage available

	4 x 4 O-D	7 x 7 O-D
Maximum error	5.5%	234.7%
Minimum error	-9.4%	-100.0%
Number of OD pairs within 5%	88.9%	12.7%
OD volumes within 5%	96.0%	18.8%
OD pairs within 5% for vehicle class 1	91.7%	11.9%
OD pairs within 5% for vehicle class 2	91.7%	16.7%
OD pairs within 5% for vehicle class 3	83.3%	9.5%
OD volumes within 5% for vehicle class 1	96.2%	18.8%
OD volumes within 5% for vehicle class 2	95.9%	20.0%
OD volumes within 5% for vehicle class 3	91.6%	17.3%

The reason behind this phenomenon might come from the relative relationship between the number of observations and the number of unknowns. As the O-D table increases in size, the number of unknowns that need to be estimated increases rapidly. For the 7-zone case, the number of unknowns is $7 \times 6 \times 3 = 126$. In 4-zone case, the number is only $4 \times 3 \times 3 = 36$. In these tests, there are three counts available on each of 76 links, or 228 observations, but not all the observations are independent because flow must be conserved at nodes within each vehicle class. In addition, some subsets of links are in the efficient sets for the same collection of O-D pairs, which further reduces the information content of the counts.

The solution method developed here makes effective use of available link count data, and in networks with only a few origins and destinations it can produce reliable multiclass O-D estimates. However, as the number of origins and destinations grows (for a constant network structure), the link count data (even separated by vehicle class) is not sufficient for reliable multiclass O-D estimation.

4.5 Q2: Effects of Reduced Link Count Coverage

From the previous section, we see that in a case where the number of unknown O-D table entries to be estimated is much smaller than the number of available link counts (i.e., the 4-zone case), the multiclass O-D estimation process can be quite accurate. However, in practice it is unlikely that counts will be available on all links. This raises the question of how the accuracy of the O-D estimates degrades as the number of observations is reduced. A set of tests on the 4-zone network is designed to address that question.

We estimated the 4-zone O-D tables under different percentages of links that are covered by classified link count sensors, ranging from 10% to 100% in increments of 10%. Each case used 20 samples of randomly selected links except for the 100% case. For each experiment, the percentage of O-D pairs that fall into the ±5% estimation error range and the percentage of total volume represented by those O-D pairs are measured.

The results are summarized in Table 4-9 and shown graphically in Figure 4-3. For this network, reasonably accurate O-D estimates are obtained as long as the percentage of links with counts exceeds about 50%. If the percentage of links covered is below that value, the overall error of estimation increases dramatically. At a given level of link coverage, the total percentage of O-D volume estimated within ±5% error is bigger than the percentage of corresponding number of O-D pairs, which means that the solution estimates of large volume O-D pairs tend to be more accurate than estimates of small volume O-D pairs.

Table 4 - 9: Results from different level of link coverage

% of links with counts	% O-D pairs within ±5%	% O-D volume within ±5%
100	88.9%	96.0%
90	86.9%	94.1%
80	84.6%	92.5%
70	81.8%	89.9%
60	81.3%	88.2%
50	80.0%	89.0%
40	74.6%	80.0%
30	53.3%	62.1%
20	36.4%	44.0%
10	12.6%	15.1%

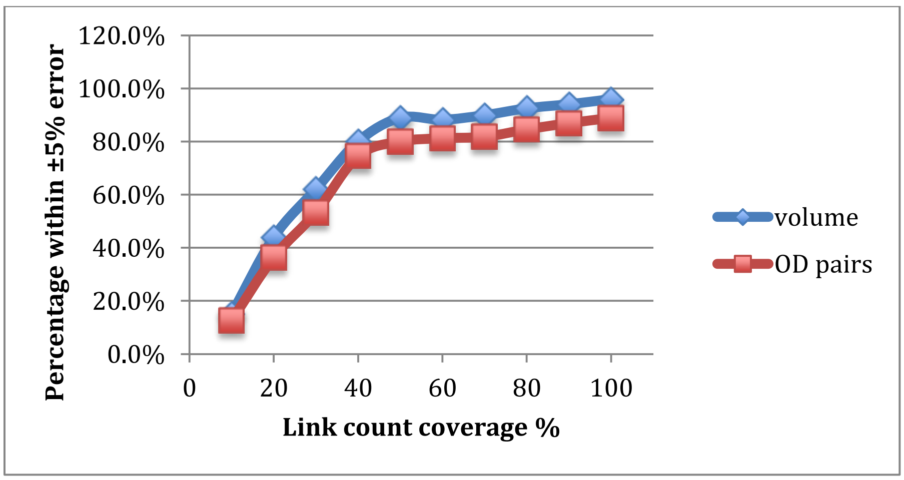

Figure 4 - 3: Percentage of O-D pairs and volume within ±5% error under different link count coverage

Table 4-10 summarizes O-D estimates (zone pair and vehicle class) that consistently have the largest errors in these experiments. Some O-D entries tend to be underestimated consistently, while others tend to be overestimated. The relative frequency column in Table 4-10 indicates the percentage of the total samples where a given O-D entry is among the three largest negative errors or the three largest positive errors.

Table 4 - 10: O-D pairs frequently estimated with large error among all samples

	Origin	Destination	Vehicle class	Volume	Relative frequency
Largest Underestimates	1	15	3	50	50.0%
	15	1	3	46	29.0%
	15	20	3	89	25.8%
	20	15	3	111	22.6%
	7	1	3	169	19.4%
Largest Overestimates	1	20	3	52	43.5%
	20	1	3	55	43.5%
	20	7	3	109	30.6%
	20	7	2	131	29.0%
	7	20	3	103	17.7%

Most of the O-D entries with largest percentage errors are in vehicle class 3 (heavy trucks), where the volumes are relatively small. Of course, on a small base volume a modest absolute error can yield a relatively large percentage error. However, these results do indicate that the truck O-D volumes, which are generally much smaller than automobile O-D volumes, are harder to estimate accurately.

It is also of interest that the O-D pairs with one end in either node 15 or node 20 appear most often. Understanding which O-D pairs and which vehicle classes are most often subject to large errors can provide insight for choosing sensor types and locations in order to measure those movements more accurately. The sensor location problem is not considered here, but is a very interesting direction for future research.

One of the main conclusions from this set of experiments is that the O-D estimation process in networks with small O-D tables is tolerant of less than full count coverage down to some critical value (in this test network, approximately 50%). If the available counts represent less than that percentage of links, the accuracy of the O-D estimates degrades rapidly. Second, the high volume O-D entries tend to be estimated more accurately than the low volume ones, and the largest percentage errors are concentrated in low volume vehicle classes, like heavy trucks. Third, these large errors may also be concentrated geographically in the network, and this implies that the locations of sensors in networks with less than full coverage is an important issue.

4.6 Q3: Effects of Losing Classification Information

In the previous two sections, we focused on O-D estimation using link counts that always include vehicle classification. However, not all sensors can classify vehicles. In particular, single loop counters that are widely deployed in practice can only provide aggregate traffic counts. Collecting vehicle classification data is generally more expensive than collecting aggregate counts. The premise of this thesis is that multiclass O-D estimation is needed to understand truck movements better, and this implies that at least some of the available data must distinguish vehicle classes. In this set of tests, we examine how the accuracy of the multiclass O-D estimates is affected when only part of the link count sensors can provide classification information. We again focus on the 4-zone case, since in the 7-zone case even with all classified counts, the link counts alone are not sufficient to estimate the O-D tables accurately.

In the first set of tests, it is assumed that link count data are available on all the links in the network. However, only a certain percentage of link count sensors can separate the three vehicle classes. The rest are single loop sensors. We checked the situation of 100%, 75%, 50%, 25%, and 0% of links having classified count data. The 100% case corresponds to the experiment done to address question 1 (in section 4.4).

The results are illustrated in Figure 4 - . As the proportion of counts that include classification information decreases, the percentage of O-D entries estimated within 5% of actual values declines almost linearly, and this relationship is particularly true for the truck classes. The automobile class is more tolerant to the loss of vehicle classification data because most of the vehicles being counted are automobiles. These figures illustrate that it is very difficult to estimate truck O-D volumes with any accuracy when a significant fraction of the sensors are reporting only total vehicles counted.

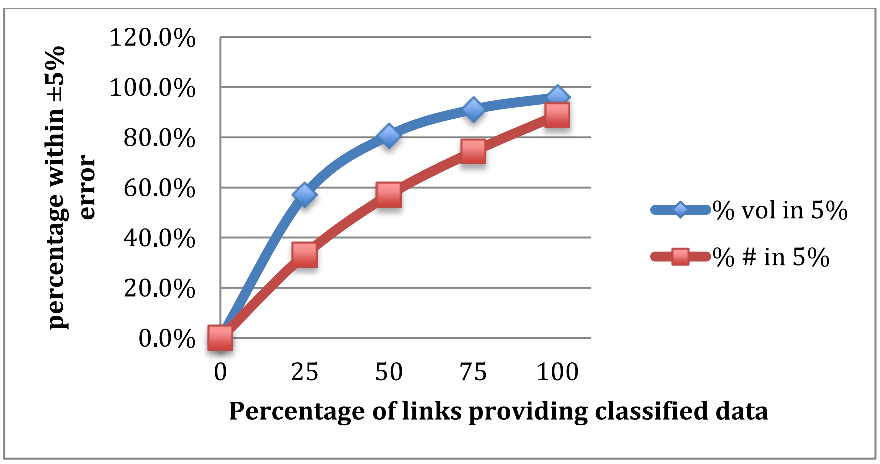

Figure 4 - 4: The percentage of O-D pairs and volume within ±5% estimation error

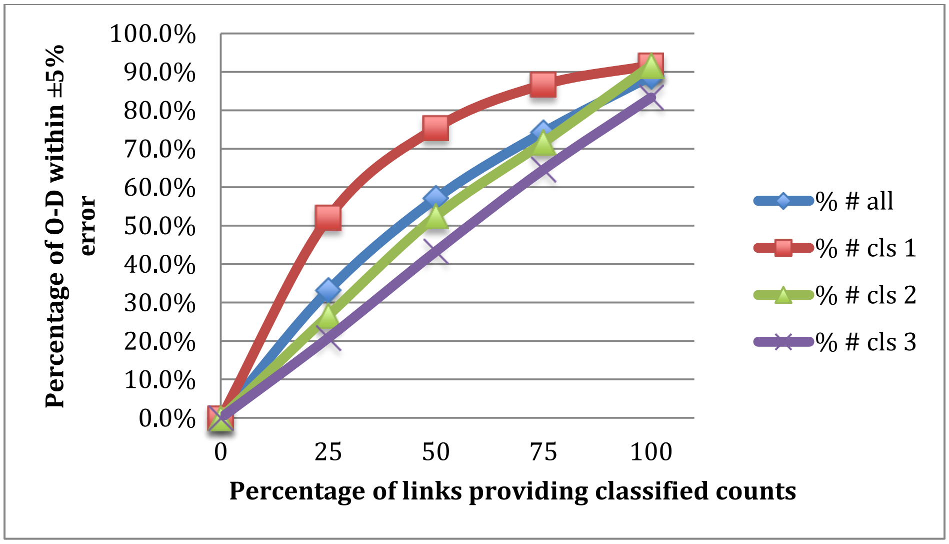

Figure 4 - 5: The percentage of O-D pairs within ±5% error for each vehicle class

A second set of tests explores the joint effect of losing some vehicle classification data and some link coverage. The tests relating to link coverage (see section 4.5) were repeated with only 50% of the available sensors able to provide classified counts. The rest are assumed to be single loop detectors that report only total vehicles. The results are summarized in Table 4 - 11, which can be compared to Table 4 - 9. When only 50% of the sensors can provide classified count information, the quality of the estimations decreases for all vehicle classes, and especially for the two truck classes. Even with 80% coverage on network links, if only half of those counts contain vehicle classification information the number of truck O-D entries that are within the ±5% error range is only about 30%.

Table 4 - 11 Estimation errors of different level of link coverage with 50 percent of sensors provide classified link counts

link coverage	OD in 5%	OD vol. in 5%	ODs cls. 1	ODs cls. 2	ODs cls. 3
100%	57.1%	80.7%	75.4%	52.5%	43.3%
90%	52.4%	75.6%	70.4%	46.3%	40.4%
80%	43.9%	72.3%	69.2%	31.7%	30.8%

70%	44.4%	68.9%	65.4%	35.8%	32.1%
60%	43.3%	64.5%	57.9%	38.3%	33.8%
50%	37.2%	59.4%	55.8%	28.8%	27.1%
40%	26.7%	46.6%	39.6%	19.2%	21.3%
30%	17.8%	33.9%	27.5%	11.3%	14.6%
20%	10.6%	18.1%	15.0%	8.3%	8.3%
10%	7.9%	7.8%	5.4%	6.7%	11.7%

These experiments reinforce the importance of having vehicle classification data as part of the traffic counts. If it is desired to do multiclass O-D estimation, it is more important to have classification data from whatever links are available than to have full coverage of the network links.

4.7 Q4: Is Dual Loop Detector Data an Acceptable Substitute?

If vehicle classification data is important, but expensive to collect, it is useful to ask if partial classification data from much cheaper dual loop detectors would be sufficient. They cannot separate all vehicle classes, but can tell an automobile from a truck.

To explore the potential for dual loop detector data, we tested a situation with 100% link coverage, where 100%, 50%, and 0% of them are dual loop detectors. The remaining counts are assumed to be total vehicle counts from single loop sensors.

The test results are organized in Table 4 - 12. Without classified link counts, the estimation of truck O-D volumes is completely ineffective. Even if all sensors can differentiate between automobile and trucks (i.e. 100links_100dual case), the estimates of truck O-D volumes are nearly all incorrect. It is preferable to have 50% classification counts than to have 100% dual loop counts. When only single loop detectors exist, even with 100% link coverage, none of the O-D entries (automobile or truck) is estimated usefully. Interestingly, even the automobile O-D estimates are degraded by substituting dual loop counts for classification counts. This is the case even though the automobile observations are the same in each case. The confusion induced in trying to sort out the truck classes from one another without actual classification data carries over into the automobile estimates as well. We conclude that using dual loop detectors to substitute for classification counts in estimating O-D tables with more than one truck class is not likely to prove useful.

Table 4 - 12: Different classification level for 100% link coverage

	ODs in 5%	OD volume in 5%	ODs in 5% for cls 1	ODs in 5% for cls 2	ODs in 5% for cls 3
100links_100clsfy	88.9%	96.0%	91.7%	91.7%	83.3%
100links_100dual	30.6%	82.7%	83.3%	0.0%	8.3%
100links_50clsfy	57.1%	80.7%	75.4%	52.5%	43.3%
100links_50dual	22.6%	70.1%	65.0%	1.7%	1.3%
100links_100single	0.0%	0.0%	0.0%	0.0%	0.0%

4.8 Q5: Effectiveness of Typical Current Data

In actual cities, not all the links are covered by sensors and most deployed sensors are single loop detectors. Even in cities noted for effective traffic management, like Rochester, NY, only 20% of links in the central city area are covered by sensors, all of which are single loop detectors. The results from the earlier sets of tests offer little hope of doing effective multiclass O-D estimation (or even single class estimation) with such sparse data, but as an explicit test case, we have postulated a situation where 20% of links have counts, and 25% of those counts are from dual loop sensors. We compare that with a case where all the counts are from single loop sensors.

The test results are summarized in Table 4 - 13. Comparing the last two rows of Table 4 – 13, we conclude that even if 25% of the sensors are upgraded to dual loop sensors, few reliable estimates will be derived, particularly for trucks. We have also tested two more extensive updates of typical current counter data – upgrading either 50% or 100% of the current link counters to provide vehicle classification data. Upgrading 50% of the counters (i.e., having classification data on 10% of total network links) to provide such data has very modest benefits. Some O-D entries are estimated more accurately, but the overall results are still quite poor. Even if all the sensors were upgraded to provide classified link counts (20links_100clsfy case – data from the set of tests in section 4.5 regarding link coverage), accuracy of the results is strongly limited by the sparse coverage of the data across the network.

Table 4 - 13: Possible classification level in real situation

	ODs in 5%	OD volume in 5%	ODs in 5% for cls 1	ODs in 5% for cls 2	ODs in 5% for cls 3
20links_100clsfy*	36.4%	44.0%	36.7%	36.3%	36.3%
20links_50clsfy	10.6%	18.1%	15.0%	8.3%	8.3%
20links_25 dual	9.2%	22.7%	19.2%	2.1%	6.3%
20links_0dual	2.8%	3.5%	3.8%	0.4%	4.2%

* Note: the notation of the cases: '20links' means 20% of total links in the network are covered by sensors. '100clsfy' means 100% of the sensors can classify three vehicle classes.

Estimating multiclass O-D tables with the level of data typically available currently is largely a futile exercise. In order to get a relative acceptable result, classified link counts are needed with preferably at least 50% of the link covered (see Table 4 - 9). This will provide at least 80% of O-D pairs within the $\pm$5% error range given all the sensors can provide classified counts. More sensors capable of providing vehicle classification information are needed, and there is also potentially an important role for data types other than link counts in order to estimate multiclass O-D tables.

4.9 Q6: Adding Turning Movements

Ever since question 1(section 4.4), we have discovered that link counts alone (even with 100% of links covered by classification counts) is not enough for 7-zone O-D table estimation. Turning movements at intersections offer one possible source of additional data. As noted in Chapter 2, most of the literature on O-D estimation has paid little attention to turning movements because the partial path information they provide is not important under deterministic user equilibrium assumptions, where path flows are not uniquely identified. However, SUE assignments do identify path probabilities, and turning movements may be of greater value. Turning movements are typically collected manually, but they can be provided by surveillance cameras and there has been some work on reconstructing them from inbound and outbound loop sensors at intersections. The tests in this section have added turning movement data to the original link count information and tested on the 7-zone O-D situation.

We ordered all the non-centroid nodes in the network in terms of their entering/ exiting traffic volume from the largest to the smallest (see Table 4 - 14) and picked out a few intersections that are likely to be important for collecting turning movements.

Table 4 - 14: Non-centroid nodes ordered by their entering traffic volume

Non-centroid node	Vehicle class 1	Vehicle class 2	Vehicle class 3
3	6100	499	488
18	4807	241	191
16	4713	146	131
12	3974	394	246
11	3772	299	227
4	3639	286	313
8	3406	180	211
19	3157	82	69
9	3042	265	240

5	2602	256	242
22	2417	121	73

We did three sets of tests, supplemented by turning movements of 2, 4, and 6 intersections respectively. We picked node 16 and 11 as the choice of two intersections, since node 3, 18 and 12 are all at the periphery of the network. For the case of 4 intersections, in addition to node 16 and 11, we added node 3 and node 22. Node 3 experiences the highest traffic volume and can cover the northern part of the network after node 16 and 11 took care of the middle part (also the downtown area) of the Sioux Falls network. Node 22 is an irregular intersection, which might get cameras installed due to some other concerns (e.g. traffic safety). Furthermore, it covers the southern part of the network. For the case of 6 intersections, node 8 and 19 are added for similar reasons. The situation for each case is summarized in Table 4 - 15.

Table 4 - 15: Case situations when turning movements are added

Case	Intersections
2 intersections added	11,16
4 intersections added	3, 11, 16, 22
6 intersections added	3, 8, 11, 16, 19, 22

Three types of turning movement data are tested. Classified turning movements can provide counts for all three vehicle classes. Dual loop turning movements are assumed to be reconstructed from dual loop counters on entry to and exit from the intersection. These counts differentiate automobile from trucks, but group both truck classes together, as with the dual loop link counts. Total vehicle turning movements give only total counts of vehicle movements. In addition, as a base case, we include the case where no turning movements are available.

We combined each type of the turning movement counts to three kinds of link counts: 100% link coverage with all 100% classified link counts; 100% link coverage with all 100% being dual loop detectors; and 80% of links covered with all assumed to be classified link counts. Each of these three cases is tested with different numbers of intersections considered. This creates a total of 36 experiments, including all combinations of the three experimental factors.

Table 4 - 16 summarizes the results for six intersections with turning movements. Parallel sets of results for four and two intersections with counts are shown in Table 4 - 17 and Table 4 - 18. Availability of turning movements obviously improve the estimation result as the information provided from observations increases. Especially when no classified information from link counts is available, classified turning movements can help to improve the truck estimation immensely.

Table 4 - 16: Estimation supplemented by different turning movement of 6 intersections

7x7 OD	6 intersections	ODs in 5%	OD volume in 5%	ODs in 5% for cls 1	ODs in 5% for cls 2	ODs in 5% for cls 3
100% link with all classified link counts	no turning	12.7%	18.8%	11.9%	16.7%	9.5%
	classified turning	35.7%	53.5%	45.2%	31.0%	31.0%
	dual turning	36.5%	54.2%	42.9%	35.7%	31.0%
	single turning	27.0%	52.8%	47.6%	16.7%	16.7%
80% link with all classified link counts	no turning	14.4%	26.7%	20.2%	12.0%	10.8%
	classified turning	35.0%	49.1%	39.6%	34.8%	30.7%
	dual turning	33.4%	54.8%	45.1%	30.1%	25.0%
	single turning	18.1%	36.8%	33.6%	11.1%	9.6%
100% link with all dual loop counts	no turning	9.5%	33.1%	23.8%	2.4%	2.4%
	classified turning	39.7%	58.3%	45.2%	45.2%	28.6%
	dual turning	12.7%	43.5%	28.6%	0.0%	9.5%
	single turning	7.1%	32.1%	21.4%	0.0%	0.0%

Table 4 - 17: Estimation supplemented by different turning movement of 4 intersections

7x7 OD	4 intersections	ODs in 5%	OD volume in 5%	ODs in 5% for cls 1	ODs in 5% for cls 2	ODs in 5% for cls 3
100% link with all classified link counts	no turning	12.7%	18.8%	11.9%	16.7%	9.5%
	classified turning	42.1%	57.1%	52.4%	42.9%	31.0%
	dual turning	41.3%	60.5%	50.0%	40.5%	33.3%
	single turning	27.0%	51.7%	47.6%	16.7%	16.7%
80% link with all classified link counts	no turning	14.4%	26.7%	20.2%	12.0%	10.8%
	classified turning	31.9%	47.1%	40.1%	28.2%	27.3%
	dual turning	30.1%	47.4%	39.8%	26.9%	23.6%
	single turning	16.6%	33.4%	30.2%	10.8%	8.8%
100% link with all dual loop counts	no turning	9.5%	33.1%	23.8%	2.4%	2.4%
	classified turning	29.4%	52.6%	42.9%	26.2%	19.0%
	dual turning	11.9%	39.6%	31.0%	2.4%	2.4%
	single turning	9.5%	31.4%	23.8%	2.4%	2.4%

Table 4 - 18: Estimation supplemented by different turning movement of 2 intersections

7x7 OD	2 intersections	ODs in 5%	OD volume in 5%	ODs in 5% for cls 1	ODs in 5% for cls 2	ODs in 5% for cls 3
100% link with all classified link counts	no turning	12.7%	18.8%	11.9%	16.7%	9.5%
	classified turning	27.8%	48.6%	38.1%	23.8%	21.4%
	dual turning	30.2%	48.5%	42.9%	23.8%	23.8%
	single turning	28.6%	52.2%	47.6%	23.8%	14.3%
80% link with all classified link counts	no turning	14.4%	26.7%	20.2%	12.0%	10.8%
	classified turning	25.5%	41.0%	35.0%	19.9%	21.7%
	dual turning	20.7%	39.0%	32.5%	14.8%	14.9%
	single turning	14.0%	27.1%	23.3%	10.4%	8.5%
100% link with all dual loop counts	no turning	9.5%	33.1%	23.8%	2.4%	2.4%
	classified turning	16.7%	34.9%	23.8%	11.9%	14.3%
	dual turning	7.9%	32.3%	23.8%	0.0%	0.0%
	single turning	7.9%	25.8%	19.0%	2.4%	2.4%

For example, before adding any turning movements, with 100% link coverage by dual loop sensors, only 2.4% of truck O-D pairs are estimated within $\pm$5% range. After adding classified turning movement counts at six intersections, the accuracy rate is raised to 45.2% for medium truck and 28.6% for heavy truck. The accuracy rate for automobiles has been doubled. Overall, the number of O-D pairs that have acceptable errors has increased from 9.5% to almost 40% with classified turning movement counts. In sum, when classification information is not available in the link count data, the classified turning movements can improve the estimation dramatically, especially for trucks. However, dual loop turning counts and total vehicle turning movements are much less effective.

Even if we had classified link counts on all the links, adding turning movements still increases the accurate rate from 12.7% to around 36%. In this case, dual loop turning movements produce a similar improvement. When the classification information is available from link counts, classified turning movements can be substituted by cheaper dual turning movements and have similar overall O-D estimation accuracy. This conclusion also holds for 80% link coverage cases.

When only total vehicle turning movements are available, the primary effect is on estimates of automobile (class 1) O-D volumes. If classification counts are available on links, addition of total turning movements at intersections improves the automobile estimation accuracy, although it provides little benefit for the truck classes.

All of the above conclusions also hold for supplementing by 4 intersections and 2 intersections. However, as the number of intersections providing turning movement is reduced, the result is less dramatic. But even so, the number of O-D pairs within the $\pm 5\%$ error range still doubled in the 2 intersection case.

Four important conclusions emerge from these experiments. First, turning movement data can be quite useful in improving the accuracy of multiclass O-D estimates. Second, classification counts at intersections and along links can be substitutes for one another, as long as classification data is available from some source. Third, for the truck classes total vehicle turning movements are not very useful, even if vehicle classification data is available from the link counts. Some distinction between automobiles and trucks is necessary, even if only from dual loop counters at the intersections. Finally, the more intersections with data the better, although even a few intersections can be useful.

4.10 Q7: Effects of Errors in Observations

All the previous tests are performed based on error-free observations. However, in reality, all the observations are subjected to a certain level of error depending on the types of sensors. We are interested in how the solution method performs under the observation errors.

Under the current technology, the error of dual loop link counts is around 4% (Yu, et al. (2010)). For classified link counts, it's around 10% (Yu, et al. (2010)). For turning movements, the current technology can provide unclassified counts at an error of 8% (Yi, et al. (2010)). We assume that both dual and classified turning movements have 10% errors. We selected 4 situations from the tests on turning movements in section 4.9 that have relative higher accuracy among all the experiments on 7-zone O-D table and introduce the error to the observations.

The error of the sensors is assumed to be the coefficient of variation of a normal distribution with the mean of true traffic volume on the link or turning path. We generated a set of observations and use the original random numbers we used in tests of section 4.9 to select links.

A summary of the results is shown in Table 4 - 19. In all four test cases, there is marked deterioration in the quality of the estimated O-D tables when the observed data contains errors of the magnitude that might normally be expected in practice. This is not unexpected, but these results emphasize two important aspects of the multiclass O-D estimation problem. First, it is important to reduce the errors in the observed data whenever possible. This may mean using better counting technology, or doing more filtering of the data before it is used. Second, in the presence of errors it is necessary to have far more observations to support O-D estimation than would be necessary in an error-free environment. This has implications for sensor location, which includes consideration of both how many sensors, and where they should be located.

Table 4 - 19: Comparison of estimation with error and no error introduced

	Error test	ODs in 5%	OD volume in 5%	ODs in 5% for cls 1	ODs in 5% for cls 2	ODs in 5% for cls 3
100% link covered by all classified link counts with dual turning in 4 intersections	no error	41.3%	60.5%	50.0%	40.5%	33.3%
	with error	11.1%	23.9%	16.7%	14.3%	2.4%
100% link covered by all dual loop counts with classified turning in 6 intersections	no error	39.7%	58.3%	45.2%	45.2%	28.6%
	with error	15.1%	7.1%	9.5%	19.0%	16.7%
80% link covered by all classified link counts with classified turning in 6 intersections	no error	35.0%	49.1%	39.6%	34.8%	30.7%
	with error	10.1%	15.7%	13.0%	7.9%	9.5%
80% link covered by all classified link counts with dual turning in 6 intersections	no error	33.4%	54.8%	45.1%	30.1%	25.0%
	with error	9.2%	14.9%	11.7%	8.3%	7.5%

4.11 Conclusions

In this chapter, a series of experiments have been carried out on a test network to address several questions regarding the estimation of multiclass O-D tables. The results of these experiments lead to 10 important conclusions.

1) In situations where the O-D tables are small relative to the set of link observations available, multiclass O-D tables can be estimated reliably if the observations are separated by vehicle class.
2) In this set of experiments, the O-D estimation process in networks with small O-D tables is tolerant of less than full count coverage down to some critical value (in this test network, approximately 50%). If the available counts represent less than that percentage of links, the accuracy of the O-D estimates degrades rapidly.
3) If the set of O-D pairs is much larger (relative to the set of link observations available), the quality of the estimated O-D tables degrades considerably because the total number of unknowns approaches or exceeds the information available in the observations.
4) The O-D estimates for the predominant vehicle class (in most cases, automobiles) are more reliable than the estimates for classes with fewer vehicles (e.g., trucks).
5) Without classified link counts, the estimation of truck O-D volumes is completely ineffective. Even if all sensors can differentiate between automobile and trucks (i.e. dual loop counters), the estimates of truck O-D volumes are nearly all incorrect. It is preferable to have 50% classification counts than to have 100% dual loop counts. Using dual loop detectors to substitute for classification counts in estimating O-D tables with more than one truck class is not likely to prove useful.
6) When only single loop detectors exist, even with 100% link coverage, none of the multiclass O-D entries (automobile or truck) is estimated usefully. This is important because single loop counters are currently the dominant source of traffic counts in most urban areas. Moreover, the link coverage in practice is much less than 100%,

and under these conditions, it is likely to be a futile exercise to attempt estimation of multiclass O-D tables.

7) Adding turning movement data to link counts is an effective way of increasing the quality of O-D estimates. When only total vehicle turning movements are available, the primary effect is on estimates of automobile (class 1) O-D volumes. If classification counts are available on links, addition of total turning movements at intersections improves the automobile estimation accuracy, although it provides little benefit for the truck classes.
8) When classification information is not available in the link count data, classified turning movements can improve estimation of truck O-D tables dramatically.
9) If vehicle classification counts are available from the link data, classified turning movements can be substituted by cheaper dual turning movements and have similar overall O-D estimation accuracy.
10) There is marked deterioration in the quality of the estimated O-D tables when the observed data contains errors of the magnitude that might normally be expected in practice. This is not unexpected, but this emphasizes two important aspects of the multiclass O-D estimation problem. First, it is important to reduce the errors in the observed data whenever possible. This may mean using better counting technology, or doing more filtering of the data before it is used. Second, in the presence of errors it is necessary to have far more observations to support O-D estimation than would be necessary in an error-free environment. This has implications for sensor location, which includes consideration of both how many sensors, and where they should be located.

It is important to verify these conclusions using additional test networks, but even on the basis of this one test network the experimental results point to important directions for further study. These directions, along with overall conclusions from the investigation described here, are discussed in Chapter 5.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

Since the mid-1970's, there has been interest in estimating origin-destination matrices from link count (or other) data. For the most part, these efforts have focused on creating O-D tables for passenger movements by automobile, using a single vehicle class. The interest in treating multiple vehicle classes separately is growing, especially in urban areas where truck flows are of increasing interest. It has been common in transportation planning studies to assume that truck flows are a constant (and generally small) fraction of automobile flows. However, this assumption ignores the substantial differences that may exist between the origins and destinations for trucks and those for automobiles.

Estimating O-D tables for trucks is of substantial interest to transportation planners because trucks impose different levels of pavement damage than cars, they have different emission characteristics, different accident patterns, and may be the subject of different types of flow controls. However, truck O-D estimation suffers from two fundamental challenges. First, it is inherently a multi-class problem because there are several different size classes of trucks, and this makes the estimation problem more difficult. Second, the available data on truck movements is very sparse, particularly data that separates different classes of trucks. Most of the data on which O-D estimation usually depends is link counts, and these data are normally collected by loop detectors buried in the roadway. However, single-loop detectors, which are ubiquitous in practice, cannot reliably distinguish a truck from a car – they simply can count the total number of vehicles passing over the detector during a specified time interval. Because the traffic stream in most locations is heavily dominated by automobiles, these total counts have been considered usable for single-class O-D estimation with the implicit assumption that trucks follow the same pattern of movements as cars. As urban areas become more concerned about truck movements, the long-used assumption that trucks are just an “add-on” to the automobile flow has become less acceptable, but the lack of truck movement data greatly hinders the process of estimating truck O-D tables separately from automobiles.

This thesis proposes a bilevel optimization model and corresponding solution method for static multi-class O-D estimation using various types of data. The model is an iterative combination of an optimization on the upper level and stochastic network loading on the lower level. The optimization model in the upper level is used to derive the O-D table entries that can minimize the sum of squared differences between observations from different data sources and the predictions of those values based on the network loading of estimated O-D tables.

The upper-level optimization uses the limited memory Broyden–Fletcher–Goldfarb–Shanno method with bounded constraints (LBFGSB) implemented in java. A probit model is assumed in the lower-level stochastic user equilibrium (SUE) problem and a computational approach proposed by Maher and Hughes (1997) is used to construct the SUE solutions.

Extensive computational experiments have been performed on a small test network with 24 nodes and 76 links. These tests verify the effectiveness of the problem formulation and solution algorithm, and also offer important insights into the multiclass O-D estimation process with different types of data available. The tests have evaluated data from different types of link sensors (including single and dual loop counters, and methods of collecting link classification data). The potential addition of turning movement data has also been evaluated. The conclusions from these experiments are listed at the end of Chapter 4 and will not be repeated here. However, these conclusions indicate important areas for future development and practical implementation of multiclass O-D estimation.

Multiclass O-D estimation is computationally feasible and an effective method for doing it has been developed in this research. However, the quality of the resulting O-D tables is dependent on having good data that separates the vehicle classes. There is no “computational trick” that will substitute for such data. The practical implication of this finding is that traffic management and transportation planning agencies that desire to better estimate truck O-D movements (in addition to automobile O-D tables) need to prepare to collect much better vehicle classification data than is currently collected in most urban areas.

As a consequence, there is significant value to development of innovative and easily deployable sensors to collect vehicle classification data. These sensors may go beyond the traditional focus on collecting link counts, and the O-D estimation method developed here is capable of accepting a wide variety of types of data.

Inclusion of turning movement data can improve the O-D estimates significantly, and such data can be incorporated easily into the estimation framework developed here. Furthermore, classification data for turning movements at intersections and classification data in link counts are somewhat substitutable. It is not necessary to have classification counts everywhere, as long as there are enough of them somewhere. In current practice, turning movements are usually collected only occasionally at a given intersection, as part of a signal timing study, for example. The findings here indicate that more regular collection of turning movements, particularly if they contain vehicle classification information, is likely to be important for better estimation of multiclass O-D movements.

Because collecting vehicle classification data is considerably more expensive than collecting just total vehicle counts, and classification data need not be collected everywhere, there is an important direction for further research to address the sensor location issue. That is, given some set of sensor technologies (with costs, capabilities and error characteristics), how many of what types of sensors should be used, and where should they be deployed to allow most effective estimation of multiclass O-D movements? This is a central question to be addressed in the continuation of the current research reported here.

The presence of reporting errors of the magnitude currently experienced in practice for count data causes severe deterioration in the quality of the estimated multiclass O-D tables. This implies that there is significant value in finding ways to reduce the error magnitudes. This may be accomplished by technological improvements in the sensors themselves. It may also be possible to reduce the effective errors by comparing data from sensors that are in close proximity to one another, or finding other ways to filter the data. The model structure developed here contains weighting factors that allow greater weight to be placed on data from more accurate sensors. In the experiments reported here, that weighting capability has not been used, but the framework exists for doing so.

There is also interest in incorporating data into the multiclass O-D estimation process that may not come directly from traffic sensors at all. A variety of data are collected for purposes other than traffic management and control, and by agencies or organizations that have very different objectives. For example, cell phone carriers know how individual phones have moved from one cell to another. This is potentially useful information for identifying partial paths for vehicles. In many cases, even more detailed information is available from GPS traces on individual cell phones or vehicles, but use of this data may create more serious privacy concerns.

Another type of data potentially useful, especially for truck movements, is economic activity data in various zones. An example of using this type of information to estimate total truck origins and destinations was discussed briefly in Chapter 3. The tests conducted in this thesis have not used that type of data, but it represents another example of non-traffic data that may be useful.

Finally, the effort represented in this thesis is directed at the static O-D estimation problem, where average trip rates over a period like a peak hour or multi-hour period are estimated. An obvious extension of this work is to the case where dynamic estimates are made, representing trips being originated or terminated over much shorter periods (e.g., a few minutes). The dynamic capability would create an opportunity to implement short-term control actions on truck movements, for example. However, the data and computational requirements to support a dynamic O-D estimation process would be considerable.

The research in this thesis contributes an important capability for multiclass O-D estimation. This is likely to be of particular interest in urban areas where there is a desire to better understand truck movements so that more effective policies on emissions, pavement maintenance and energy consumption can be designed. This work has also opened up a set of promising directions for further research and development to be pursued in future efforts.

REFERENCES

Akcelik, R., 1991. Travel time functions for transport planning purposes: Davidson's function, its time-dependent form and an alternative travel time function. *Australian Road Research*, 21 (3), 49-59.

Al-Battaineh, O., Kaysi, I.A., 2007. Genetically-optimized origin-destination estimation (GOODE) model: application to regional commodity movements in Ontario. *Canadian Journal of Civil Engineering*, 34, 228-238.

Avriel, M., 1976. *Nonlinear Programming: Analysis and Methods*, Prentice-Hall, Englewood Cliffs, NJ.

Brenninger-Göthe, M., Jörnsten, K.O., Lundgren, J.T., 1989. Estimation of origin-destination matrices from traffic counts using multiobjective programming formulations. *Transportation Research, Part B*, 23, 257-269.

Byrd, R.H., Lu, P., Nocedal, J., Zhu, C. 1995. A limited memory algorithm for bound constrained optimization. *SIAM J. SCI. COMPUT.*, 16(5), 1190-1208.

Byrd, R.H., Lu, P., Nocedal, J., Zhu, C. 1997. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software*, 23(4), 550-560.

Carey, M., Hendrickson, C., Siddharthan, K., 1981. A method for direct estimation of origin/destination trip matrices. *Transportation Science*, 15, 32-49.

Cascetta, E., 1984. Estimation of Trip Matrices from Traffic Counts and Survey Data: A Generalized Least Squares Estimator, *Transportation Research, Part B*, 18, 289-299.

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

Cascetta, E., Nuzzolo, A., Russo, F., Vitetta, A. 1996. A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. In: Lesort, J.B. (Ed.), *Transportation and Traffic Theory*, Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Lyon, France, 697-711.

Clark, C.E., 1961. The greatest of a finite set of random variables. *Operations Research*, 9, 145-162.

Connors, R.D., Sumalee, A., Watling, D.P., 2007. Sensitivity Analysis of the Variable Demand Probit Stochastic User Equilibrium with Multiple User-Classes, *Transportation Research, Part B*, 41, 593-615.

Daganzo, C.F., 1982. Unconstrained extremal formulation of some transportation equilibrium problems. *Transportation Science*, 16, 332-360.

Dennis, J.E., Schnabel, R.S., 1996. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Society for Industrial and Applied Mathematics, Philadelphia.

Dial, R.B., 1971. A probabilistic multipath traffic assignment algorithm which obviates path enumeration. *Transportation Research*, 5, 83-111.

Federal Highway Administration, (http://tmip.fhwa.dot.gov/resources/clearinghouse/docs/mvrcm/ch7.htm), accessed 8/1/2011.

Garcia, C.B., Zangwill, W.I., 1981. *Pathways to Solutions, Fixed Points and Equilibria*. Prentice-Hall, Englewood Cliffs, NJ.

Ha, T.-J., Lee, S., Kim, J., Lee, C., 2007. Gradient Method for the Estimation of Travel Demand Using Traffic Counts on a Large Scale Network, in T.-J. Cham, *et al*. (eds), *MMM 2007, Lecture Notes in Computer Science 4352*, 599-605.

Jeng, S.-T., Ritchie, S.G., 2008. Real-Time Vehicle Classification Using Inductive Loop Signature Data, *Transportation Research Record 2086*, 8-22.

Kobos, M. (2012, August 3). Java wrapper for the Fortran L-BFGS-B algorithm, from GitHub, (https://github.com/mkobos/lbfgsb_wrapper), accessed 10/3/2012.

Lam, W.H.K., Lo, H.P., 1990. Accuracy of O-D Estimates from Traffic Counting Stations, *Traffic Engineering and Control*, 31, 358-367.

LeBlanc, L.J., Morlok, E.K., Pierskalla, W., 1975. An efficient approach to solving the road network equilibrium traffic assignment problem, *Transportation Research*, 9:5, 309-318.

List, G.F., Turnquist, M.A., 1994. Estimating truck travel patterns in urban areas, *Transportation Research Record 1430*, 1-9.

List, G.F., Turnquist, M.A., 1995. A GIS-based approach for estimating truck flow patterns in urban settings. *Journal of Advanced Transportation*, 29, 281-298.

List, G.F., Konieczny, L.A., Durnford, C.L., Papayanoulis, V., 2002. Best-practice truck-flow estimation model for the New York City region. *Transportation Research Record 1790*, 97-103.

Luenberger, D.G. and Y. Ye. 2008. *Linear and Nonlinear Programming, 3rd Edition*. Springer, New York.

Maher, M.J., 1983. Inferences on Trip Matrices from Observations on Link Volumes: A Bayesian Statistical Approach, *Transportation Research, Part B*, 17, 435-447.

Maher, M.J., 1992. SAM - A stochastic assignment model. *Mathematics in Transport Planning and Control*, Oxford University Press, 121-133.

Maher, M.J., 1998. Algorithms for logit-based stochastic user equilibrium assignment, *Transportation Research*, 32B, 539-549.

Maher, M.J., Hughes, P.C., 1997. A probit-based stochastic user equilibrium assignment model. *Transportation Research*, 31B, 341-355.

Maher, M.J., Zhang, X., Van Vliet, D., 2001. A bi-level programming approach for trip matrix estimation and traffic control problems with stochastic user equilibrium link flows," *Transportation Research*, 35B, 23-40.

Morales, J. L., Nocedal, J. 2011. Remark on "algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization. *ACM Transactions on Mathematical Software*, 38(1), Article 7.

Nguyen, S., 1984. Estimating Origin-Destination Matrices from Observed Flows, in M. Florian (ed.), *Transportation Planning Models*, Elsevier, Amsterdam, 363-380.

Nihan, N.L., Zhang, X., Wang, Y., 2002. *Evaluation of Dual-Loop Data Accuracy Using Video Ground Truth Data*, Research Report TNW2002, University of Washington, Seattle.

Ping Yi, Chun Shao, Jialei Mao. 2010. Development and preliminary testing of an automatic turning movements identification system. *Ohio Transportation Consortium*.

Raothanachonkun, P., Sano, K. and Matsumoto, S., 2005. Estimation of Multiclass Time-Varying Origin-Destination Matrices from Traffic Counts,

Raothanachonkun, P., Sano, K. and Matsumoto, S., 2006. Estimation of Multiclass Origin-Destination Matrices Using Genetic Algorithm,

Rios, A., Nozick, L.K., Turnquist, M.A., 2002. Value of different categories of information in estimating freight O-D tables. *Transportation Research Record*, 1783, 42.48.

Robillard, P., 1975. Estimating the O-D matrix from observed link volumes. *Transportation Research*, 9, 123-128.

Schneider, R., 2011. *Comparison of Turning Movement Count Data Collection Methods for a Signal Optimization Study*, URS Corporation, Grand Rapids, Michigan.

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*, 28B, 213-233.

Spiess, H., 1987. A maximum likelihood model for estimating origin-destination matrices. *Transportation Research*, 21B, 395-412.

Sun, C., Ritchie,S.G., 1999. Individual vehicle speed estimation using single loop inductive waveforms, *J TRANSP E*, 125(6), 531-538

Swann, S., 2010. *Albury City Comparison of Traffic Data Collection Methods*, Albury City, New South Wales, Australia.

Turnquist, M.A., Gur, Y., 1979. Estimation of Trip Tables from Observed Link Volumes, *Transportation Research Record*, 730, 1-6.

Van Zuylen, H.J., Willumsen, L.G., 1980. The most likely trip matrix estimated from traffic counts. *Transportation Research*, 14B, 281-293.

Wang, Y., Nihan, N.L., 2003. Can Single-Loop Detectors Do the Work of Dual-Loop Detectors?, *ASCE Journal of Transportation Engineering*, 129:2, 169-176.

Wei, H., 2011. *Optimal Loop Placement and Models for Length-based Vehicle Classification and Stop-and-Go Traffic*, Ohio Transportation Consortium Research Report, University of Cincinnati.

Wong, S.C., Tong, C.O., Wong, K.I., Lam, W.H.K., Lo, H.K., Yang, H., Lo, H.P., 2005. Estimation of Multiclass Origin-Destination Matrices from Traffic Counts, *Journal of Urban Planning and Development*, ASCE, 131, 19-29.

Yang, H., 1995. Heuristic algorithms for the bilevel origin-destination matrix estimation problem. *Transportation Research*, 29B, 231-242.

Yang, H., Sasaki, T., Iida, Y., 1992. Estimation of origin-destination matrices from link traffic counts on congested networks. *Transportation Research*, 26B, 417-434.

Yu, X., Prevedouros, P.D., Sulijoadikusumo, G. 2010. Evaluation of Autoscope, SmartSensor HD, and Infra-Red Traffic Logger for Vehicle Classification. *Transportation Research Record*(2160), 77-86.

Zhang, G., Wang, Y., Wei, H., 2006. Artificial Neural Network Method for Length-Based Vehicle Classification Using Single-Loop Outputs, *Transportation Research Record 1945*, 100-108.