ESTIMATING DYNAMIC ORIGIN-DESTINATION DEMANDS FROM LINK AND PROBE COUNTS

by

Bruce R. Hellinga

A thesis submitted to the Department of Civil Engineering in Conformity with the requirements for the degree of Doctor of Philosophy

Queen's University Kingston, Ontario, Canada August 1994

Copyright © Bruce R. Hellinga, 1994

Abstract

Many traffic engineering problems consist of evaluating a number of alternate schemes in terms of some measures of performance, such as total travel time, volume to capacity ratio, average speed, total fuel consumption, total emissions, etc. These evaluations generally require prior knowledge of the temporal trip making behaviour of drivers by origin and destination. This thesis presents the development, application, and evaluation of two models capable of inferring these temporal origin - destination (O-D) traffic demands on the basis of observed link traffic flows and assumptions regarding drivers' route choices. In particular, this thesis presents the development and evaluation of a Least Squared Error (LSE) and a Least Relative Error (LRE) model, each of which is capable of estimating either static demands, a time series of static demands, or dynamic demands. Furthermore, the potential of using probe data, from route guidance system (RGS) equipped vehicles, to enhance these estimated dynamic O-D demands is examined.

Both of the LSE and LRE models' mathematical formulations are presented. The LSE model formulation parallels that of a least squared regression as the error function is composed of the sum of the squared absolute difference between the observed and estimated link flows. In contrast, the LRE model is formulated on the basis that the link flow error, when measured relative to the observed flow, is to be minimized instead. Iterative solution algorithms, that are modifications of the Jacobi and Gauss-Seidel techniques, are proposed to solve each of the model formulations. It is shown, by way of the application of these iterative algorithms to several example networks, that the estimated O-D demands, which result from these iterative solution techniques, are consistent with the model formulations and with the analytical solutions. Furthermore, it is shown for several examples that, when multiple solutions exist which each exactly replicate the observed link flows, and no prior O-D demand information is specified, both the LSE and LRE models estimate demands that closely approximate the maximum likelihood solution.

The proposed iterative solution algorithms have been incorporated into a computer model called QUEENSOD. This model can be practically applied to real networks using current computer memory constraints. This thesis describes the application of the LSE and LRE models to a 35 km section of multilane urban freeway in Toronto, Canada, in which alternate parallel routes exist. Dynamic 15 minute O-D demands were estimated for the eastbound direction for the period from 5 am to 11 am. Despite FTMS detector data being available for only 45% of the network, a correlation coefficient of approximately 98% was obtained for both models. This value reflects the high linear correlation between estimated and observed link traffic flow data for this network. The statistical analysis of the expected quality of O-D demands, which are estimated solely on the basis of RGS probe vehicle data, indicated that even for levels of market penetration of 30 - 50%, the O-D estimates are unlikely to be of sufficient quality to be of practical benefit.

Acknowledgements

Many people have contributed to this work in direct and indirect ways. I am gratefully indebted to my colleagues for their helpful discussions and constructive criticisms; the support staff for cheerfully helping me complete cryptic forms and innumerable other details that threatened to impinge on my sanity; faculty members of the Department of Civil Engineering for their interest; Dr. Michel Van Aerde, my supervisor, for his advice, direction, optimism, and his infectious enthusiasm; my parents and brothers for caring; and above all, my wife, for her constant support, encouragement and patience.

Appreciation is extended to Phil Masters and David Tsui of the Freeway Traffic Management Systems Section of the Ontario Ministry of Transportation for supporting this research and for making the FTMS data available.

The following organizations are gratefully acknowledged for their contributions in funding the research described in this thesis: Natural Sciences and Engineering Research Council of Canada, Ontario Ministry of Colleges and Universities, Transportation Association of Canada, Canadian Transportation Research Forum, Ontario Ministry of Transportation of Ontario, and Queen's University.

to Lorraine

Table of Contents

xii

Abstract	ii
Acknowledgements	iii
List of Figures	
List of Tables	xviii
Nomenclature	XX
Glossary	xxiii
CHAPTER 1 INTRODUCTION	1
1.1 What are Origin-Destination Demands and Why are They Needed?	1
1.2 How can O-D Demands be Obtained?	4
1.3 What are Dynamic O-D Demands?	5
1.4 What is the Current Status of Synthetic O-D Demand Estimation?	6
1.5 What are some of the Theoretical and Practical Problems of O-D Estimation?	7
1.6 What is the Scope and Approach of this Thesis?	9
1.6.1 Thesis objectives.	9
1.6.2 Problems addressed by this thesis	9
1.6.3 Problems not addressed by this thesis	10
1.6.4 Similarities and differences to other approaches	10
1.6.5 Method of approach	11
1.6.6 Method of validation	11
1.6.7 Method of presentation	12
CHAPTER 2 ASSESSMENT OF THE CURRENT STATE-OF-THE-ART	14
2.1 Introduction	14
2.2 General Solution Approach Categories	15
2.2.1 O-D estimates from transportation planning models	15
2.2.2 O-D estimates from direct sampling	17
2.2.3 O-D estimates from link counts	18
2.3 Definition of Link-Count Based O-D Estimation Problem	19
2.4 Categorization of Link-Count Based Approaches	20
2.4.1 Heuristic vs. mathematical approaches	21
2.4.2 Intersection vs. network approaches	22
2.4.3 Pre-specified vs. variable routes	23

2.4.4 Static vs. dynamic approaches	24
2.5 Additional Issues	26
2.5.1 Under- and over- specification	26
2.5.2 Accuracy of estimated O-D demands	28
2.5.3 Use of <i>a priori</i> information	29
2.6 Summary	30
CHAPTER 3 ANALYSIS AND COMPARISON OF TWO MAXIMUM	
LIKELIHOOD APPROACHES	31
3.1 Introduction	31
3.1.1 Background	
3.1.2 Structure of chapter	32
3.2 Generalized Model	32
3.2.1 Introduction	32
3.2.2 Mathematical derivation	33
3.3 Vehicle Link-Count Based Model	38
3.3.1 Model formulation	38
3.3.2 Analytical solution for a simple network	40
3.3.3 Formal solution	43
3.4 Trip-Count Based Model	45
3.4.1 Model formulation	45
3.4.2 Formal solution	47
3.5 Examination of Model Assumptions	47
3.5.1 Stirling's approximation	48
3.5.2 Total number of trips	49
3.6 Comparison of Methods	50
3.6.1 Example networks	50
3.6.2 Effect of prior information	51
3.6.3 Effect of redundant link flows	53
3.7 Summary	54
CHAPTER 4 DEVELOPMENT OF A LEAST SQUARES O-D ESTIMATION	
MODEL	56
4.1 Introduction	56

	4.2 Development of the LSE Model	57
	4.2.1 Mathematical basis	57
	4.2.2 Example with link flow continuity	58
	4.2.3 Example with link flow discontinuity	59
	4.2.4 Example with multipath routes	60
	4.3 Incorporation of Link Flow Reliability	61
	4.3.1 Modified model formulation	61
	4.3.2 Example considering relative reliability of link flows	62
	4.3.3 Determining appropriate link flow reliability factors	65
	4.4 Characteristics of Formulation	69
	4.4.1 Negative demand estimates	69
	4.4.2 Suitability of error function	71
	4.4.3 Selecting a solution when multiple solutions exist	71
	4.5 Development of a Solution Algorithm	73
	4.5.1 Formal solution	73
	4.5.2 Limitations of the formal solution	74
	4.5.3 Iterative solution algorithm	75
	4.5.4 Algorithm stopping criteria	80
	4.5.5 Application of iterative algorithm to a simple network having multiple solutions	82
	4.5.6 Application of iterative algorithm to a second simple network	85
	4.5.7 Application of iterative algorithm to a network with non-uniform link reliability factors	86
	4.6 Incorporation of Seed O-D Demand Reliability	
	4.6.1 Introduction	
	4.6.2 Modification to the solution algorithm	89
	4.6.3 Example application considering the relative reliability of seed O-D demands	
	4.7 Impact of Redundant Link Flows	
	4.8 Summary	
C	HAPTER 5 DEVELOPMENT OF A LEAST RELATIVE ERROR O-D	
	ESTIMATION MODEL 5.1 Introduction	95
	2 1 HH COURTER (1881)	91

5.2 Development of the LRE Model	95
5.2.1 Mathematical basis	96
5.2.2 Example with link flow continuity	99
5.2.3 Example with link flow discontinuity	101
5.2.4 Example with multipath routes	104
5.3 Incorporation of Link Flow Reliability	105
5.4 Characteristics of Formulation	105
5.5 Development of a Solution Algorithm	106
5.5.1 Formal solution	106
5.5.2 Iterative solution algorithm	107
5.5.3 Algorithm stopping criteria	108
5.5.4 Application of iterative algorithm to a simple network having multiple solutions	111
5.5.5 Application of iterative algorithm to a simple network having a single solution with multiple paths	114
5.6 Incorporation of Seed O-D Demand Reliability	117
5.6.1 Modification to algorithm	117
5.6.2 Example application considering the relative reliability of seed O-D demands	118
5.7 Summary	124
CHAPTER 6 DYNAMIC EXTENSION OF STATIC LSE AND LRE MODELS	126
6.1 Introduction	126
6.2 Establishing that a Need for Dynamic Demands Exists	126
6.2.1 Spatial and temporal variations evident in field O-D patterns	127
6.2.2 Effects of violating static estimation assumptions	130
6.3 Extensions to LSE and LRE Static Models	132
6.3.1 Formulation of dynamic LSE and LRE models	133
6.3.2 Characteristics of dynamic estimation.	134
6.3.3 Generating link use probabilities from known route trees and link travel times	135
6.3.4 Identifying extraneous link flows	137
6.4 Application of Dynamic Models to a Small Hypothetical Network	
6.4.1 Network characteristics and scenario descriptions	139

6.4.2 Measures for comparison	140
6.4.3 Analysis of model results	142
6.5 Limitations of Dynamic Models	143
6.6 Application of Dynamic Models to a Larger Hypothetical Network	144
6.6.1 Network characteristics	145
6.6.2 Description of the INTEGRATION model	146
6.6.3 Scenarios being examined	146
6.6.4 Analysis of model results	147
6.7 Summary	152
CHAPTER 7 ESTIMATING O-D DEMANDS FOR A FREEWAY CORRIDOR IN TORONTO, CANADA	153
7.1 Introduction	153
7.1.1 Description of Highway 401	154
7.1.2 Description of the freeway traffic management system	155
7.2 Acquisition of Data Necessary for O-D Estimation	155
7.2.1 Network representation	155
7.2.2 Link flows	156
7.2.3 Route determination	157
7.2.4 Link travel times	158
7.3 Determination of Time Varying Demands	161
7.3.1 Introduction	161
7.3.2 Evaluation of estimated demands with respect to aggregated observed link flows	162
7.3.3 Evaluation of estimated demands with respect to individual observed link flows	164
7.3.4 Evaluation of estimated demands with respect to observed origin productions and destination attractions	
7.3.5 Further characteristics of the estimated O-D demands	172
7.4 Summary	174
CHAPTER 8 RGS VEHICLE PROBES AS ESTIMATORS OF DYNAMIC O-D DEMANDS	175
8.1 Introduction	175
8.1.1 Vehicle probe data	175

	8.1.2 Focus of this chapter	176
	8.1.3 Nomenclature	177
	8.2 Derivation of Estimate of Population O-D Demands	178
	8.2.1 Estimating the population O-D demand D_{ij} - Approach 1	178
	8.2.2 Estimating the population O-D demand D_{ij} - Approach 2	179
	8.2.3 Estimating the population probability P_{ij} and total departures D	179
	8.2.4 Estimating the exact level of market penetration from route flows	180
	8.2.5 Estimating the approximate level of market penetration from link flows	180
	8.2.6 Summary	181
	8.3 Derivation of Estimate of Reliability of O-D Demand Estimates	182
	8.3.1 Estimating the reliability of the O-D demand estimate	182
	8.3.2 Estimating the reliability of the population probability	183
	8.3.3 Estimating the reliability of the total number of network departures	183
	8.3.4 Summary	184
	8.4 Practical Considerations	184
	8.4.1 Population size	185
	8.4.2 Level of surveillance	185
	8.4.3 Impact on O-D reliability	186
	8.4.4 Example illustration	186
	8.5 Evaluation of Network Level of Surveillance	188
	8.6 Example Network Description and Results	
	8.6.1 Network structure	190
	8.6.2 Results for example network	192
	8.7 Summary	195
\mathbf{C}	HAPTER 9 CONCLUSIONS AND RECOMMENDATIONS	197
	9.1 Overview	197
	9.2 Conclusions	197
	9.2.1 Proposed models	198
	9.2.2 Application of proposed models	198
	9.2.3 Potential of probe data	199
	9.3 Limitations of this Research.	200

9.3.1 Limitations of the LSE and LRE models	200
9.3.2 Limitations of the statistical evaluation of probe data	200
9.4 Significance of the Thesis Contributions	201
9.4.1 Significance of the LSE and LRE models	201
9.4.2 Significance of the evaluation of vehicle probe data	201
9.5 Recommendations for Further Work	202
9.5.1 Further development of the LSE and LRE models	202
9.5.2 Further development of the use of probe data in estimating dynamic demands	202
REFERENCES	204

List of Figures

Figure	Title	Page
1-1	Simple network used to illustrate the impact of assuming link flows versus O-D demands as independent inputs to a typical traffic engineering evaluation	3
2-1	Use of the trip generation and trip distribution phases of the transportation planning process to estimate O-D demands	17
2-2	Use of partial license plate number sampling to estimate O-D demands	18
2-3	Use of observed link flows to estimate O-D demands	19
2-4	Estimating static turning movements for an intersection	23
2-5	Estimating time varying turning movements at intersections	26
2-6	Graphical representation of maximum possible relative O-D error	29
3-1	Illustration of a simple example having two possible outcomes	34
3-2	Comparison of the information content contained within a possible outcome under the assumption that outcomes are equally likely $(\ln(G))$ and that outcomes are weighted by their likelihood of occurring $(\ln(G'))$	35
3-3	Comparison of the information content contained within possible outcomes under the assumption that the two outcomes are equally likely and under the assumption that state <i>A</i> has a 75% likelihood of occurring	36
3-4	Simple example two-link network	40
3-5	Information objective function as a function of estimated demand when no prior information exists	42
3-6	Information objective function magnitude as a function of feasible <i>a priori</i> matrices and estimated demand	42
3-7	Information objective function as a function of estimated demand when prior information exists but this prior demand is not a feasible solution	43
3-8	Quality of two different versions of Stirling's approximation	48
3-9	Comparison of alternative assumptions about the total number of trips	50
3-10	Two simple networks used to illustrate several characteristics of the link based entropy and link-count based information model formulations	51
4-1	Example three-link network	59
4-2	Example four-link network exhibiting multiple paths between an origin and destination zone	61
4-3	Example line three-link network used to illustrate the impact of link flow reliability	63
4-4	Variation in link flow error as a function of link flow reliability	64

Figure	Title	Page
4-5	Effect of link flow reliability on optimal O-D demand	64
4-6	Illustration of the impact of variance on the validity of the mean	65
4-7	Variance of vehicle counts observed on a three lane section of Highway 401 Eastbound Collectors at Avenue Road	66
4-8	Coefficient of variation of vehicle counts observed on a three lane section of Highway 401 Eastbound Collectors at Avenue Road	67
4-9	Example three-link network used to illustrate a heuristic method of determining link flow reliability	68
4-10	Impact of truncation on solution optimality	70
4-11	Example two-link network	72
4-12	Illustration of the feasible O-D solution space for a two-link example network	72
4-13	Effect of the relaxation factor and the number of iterations on the squared link flow error (Seed matrix = $(5,5,5)$)	83
4-14	Effect of the relaxation factor and the number of iterations on a selected O-D demand - T_{AC} (Seed matrix = (5,5,5))	84
4-15	Demand estimates and associated squared link flow error as a function of the iteration for a situation in which the unconstrained optimum results in a negative O-D estimate	86
4-16	Demand estimates and associated squared link flow error as a function of the iteration for a situation in which non-uniform link flow reliability factors exist	87
4-17	Demand estimates and associated squared link flow error as a function of the link flow reliability	88
4-18	Example four-link network used to illustrate the impact of the relative reliability of seed O-D demands	90
4-19	Intermediate estimates of O-D demand T_{AD} for different levels of seed O-D demand reliability	92
4-20	Similarity of estimated demand (T_{AD}) to seed demand (t_{AD}) as a function of seed reliability factor (R_{AD})	92
4-21	Final estimates of O-D demands for different levels of seed O-D demand reliability (t_{AD})	93
5-1	Illustration of the functional form of the LRE model's error function	97
5-2	Example three-link network	100

Figure	Title	Page
5-3	Graphical illustration of the concavity of the relative link flow error function in the vicinity of the optimal solution	103
5-4	Example four-link network exhibiting multiple paths between an origin and destination zone	104
5-5	Comparison of actual average relative error and average relative error estimated using Equation [5-14] over a range of flow errors	110
5-6	Comparison of actual absolute error and absolute error estimated using Equation [5-16] over a range of flow errors	110
5-7	Example four-link network in which multiple feasible O-D solutions are possible	111
5-8	Iterative estimates of O-D demands for simple example network having all-or-nothing routes	113
5-9	Link flows resulting from estimates of O-D demands for simple example network having all-or-nothing routes	113
5-10	Rate of convergence of demands estimated by the LRE model for a simple network having a single solution and multiple paths	116
5-11	Relative link flow error resulting from the application of the LRE algorithm to a simple network having multiple path routes	116
5-12	Marginal improvement in relative link flow error (E_R) resulting from the application of the LRE algorithm to a simple network having multiple paths	117
5-13	Example four-link network used to examine the relative reliability of seed O-D demands	118
5-14	Similarity of estimated demand (T_{AD}) to seed demand (t_{AD}) as a function of seed reliability factor (R_{AD})	120
5-15	Intermediate estimates of O-D demands for Scenario 2	121
5-16	Intermediate estimates of O-D demands for Scenario 3a	121
5-17	Relative error (E_R) of intermediate O-D demand estimates for Scenario 3a	122
5-18	Intermediate estimates of O-D demand T_{AD} for different levels of seed O-D demand reliability	123
5-19	Final estimates of O-D demands for different levels of seed O-D demand reliability (t_{AD})	124
6-1	Typical observed aggregated 15 minute link volumes by time of day	128
6-2	Typical temporal and spatial variation in O-D demands as depicted by total observed flow from specific origin zones	129

Figure	Title	Page
6-3	Typical temporal and spatial variation in O-D demands as depicted by total observed flow to specific destination zones	129
6-4	Approximating temporal variations in link flows through the use of a time series of static O-D demands	130
6-5	Example linear five-link network used to illustrate the relative accuracy of a static demand, a time-series of static demands, and a dynamic demand	130
6-6	O-D estimation results as a function of the estimation process utilized	132
6-7	Discretized representation of demand propagation over time and space	136
6-8	Computation of observed link flows based on demand departure time and link travel times	136
6-9	Illustration of occurrence of flows that do not result from demands considered within analysis period	138
6-10	Illustration of a more complex situation for the determination of extraneous link flows	138
6-11	Illustration of the occurrence of partially extraneous flows	139
6-12	Example five-link network used to illustrate the performance characteristics of the dynamic LSE and LRE models	139
6-13	Illustration of the sensitivity of the normalized link flow error (\overline{E}^n) and the correlation coefficient to errors link flows	142
6-14	Illustration of possible discrepancies between true demand profile and observed capacity constrained flows	144
6-15	Process used to generate an appropriate time-series of observed link flows and routes using the INTEGRATION simulation model	144
6-16	Hypothetical network used to illustrate the dynamic estimation abilities of the LSE and LRE models	145
6-17	Observed link flows and flows estimated by the dynamic LRE model from the true dynamic demand (Scenario 1a)	150
6-18	Selected O-D demands estimated by the dynamic LRE model during each iteration (Scenario 2b)	150
6-19	Normalized link flow error resulting from the LRE model as a function of the iteration number	151
6-20	Normalized link flow error resulting from the LSE model as a function of the iteration number	151
7-1	Location of Highway 401 study site	154

Figure	Title	Page
7-2	Link and node representation of the Highway 401 eastbound network	156
7-3	Typical 20 second and aggregated 15 minute FTMS detector volume data reflecting conditions on a 3 lane section in the eastbound collector at	
	Avenue Road on May 1, 1991	157
7-4a	Normalized link error resulting from the LSE model	
7-4b	Normalized link error resulting from the LRE model	163
7-5a	Comparison of observed and estimated flows by time of day for link 8	165
7-5b	Comparison of observed and estimated flows by time of day for link 12	165
7-6a	Correlation between observed link flows and those estimated by the dynamic LSE model	166
7-6b	Correlation between observed link flows and those estimated by the dynamic LRE model	166
7-7a	RMS error between estimated and observed total origin productions presented as a percent of the average observed total	170
7-7b	RMS error between estimated and observed total destination attractions presented as a percent of the average observed total	170
7-8a	Correlation between observed total destination zone flows and those estimated by the LSE model	171
7-8b	Correlation between observed total destination zone flows and those estimated by the LRE model	171
7-9	Average trip length implied by the time varying demand estimated by the LSE and LRE models as a function of iteration number	173
7-10	Total number of trips contained within the time varying demand estimated by LSE and LRE models as a function of iteration number	173
8-1	Schematic representation of relationships between several key variables which were defined	178
8-2	Illustration of the coefficient of variation of estimated O-D demand as a function of level of market penetration assuming all links in network are under surveillance	187
8-3	Illustration of the coefficient of variation of estimated O-D demand as a function of total number of trips assuming all links in network are under surveillance	188
8-4	Configuration of example traffic network	191
8-5	Process used to produce realistic traffic conditions on example network	191

Figure	Title	Page
8-6	Sample of typical probe information provided by simulation model	192
8-7	Random variation in the estimated level of market penetration (m) for a global level of market penetration (M) of 5% and 20%	193
8-8	Analytical estimates and actual simulation results for a selected O-D demand over time of day for an average level of market penetration of 20% (O-D pair 1-3)	194
8-9	Analytical estimates and simulation results for a selected O-D demand for various levels of market penetration (O-D pair 1-3 for time period from 7:15 - 7:30 AM)	194
8-10	Effect of level of market penetration on the accuracy of O-D estimation	196

List of Tables

Table	Title	Page
1-1	Characteristics of example network	3
1-2	Illustration of several theoretical difficulties that can be encountered when estimating O-D demands from link flows	8
3-1	Characteristics of example networks	51
3-2	Maximum likelihood O-D estimates for a two-link example network	53
3-3	Maximum likelihood O-D estimates for a four-link network with redundant link flows	54
4-1	Squared link flow error for all potential O-D solutions in the vicinity of the analytical solution (vph ²)	60
4-2	Example application of a heuristic approach to determining link flow reliability factors	69
4-3	Gauss-Seidel and Jacobi iterative method results from a simple network having multiple solutions that satisfy the system of linear equations	77
4-4	Iterative solution algorithm for estimating O-D demands that minimize the squared link flow errors	80
4-5	Iterative algorithm results for a simple example network	82
4-6	Scenario configurations and estimated demands used to illustrate the impact of seed O-D demand reliability factors	90
4-7	LSE model O-D estimates for a four-link network with redundant link flows	93
5-1a	Left hand side of non-linear constraints computed for several O-D demands (vph)	101
5-1b	Relative link flow error (E_R) for all potential O-D solutions in the vicinity of the analytical solution.	101
5-2a	Left hand side of non-linear constraints computed for several O-D demands for a network exhibiting link flow discontinuity (vph)	102
5-2b	Relative link flow error (E_R) for all potential O-D solutions in the vicinity of the analytical solution for a network exhibiting link flow discontinuity	102
5-3	Left hand side of non-linear constraints computed for several O-D demands for a network exhibiting multipath routes	105
5-4	LRE iterative solution algorithm	108
5-5	Iterative algorithm results of LRE model for a simple example network with multiple solutions and all-or-nothing routings	112

List of Tables (continued)

	3-0	which multiple paths exist	114
_	Table	Title	Page
	5-7	Scenario configurations and estimated demands used to illustrate the impact of seed O-D demand reliability factors	118
	6-1	Link counts observed on example linear five-link network	131
	6-2	Actual network O-D demands and observed link flows	140
	6-3	Description of scenarios evaluated to illustrate the dynamic estimation abilities of the LSE and the LRE models	140
	6-4	Results from the application of the dynamic LSE and LRE models to a simple five-link network	143
	6-5	Actual time varying demand used to determine observed flows	146
	6-6	Description of scenarios evaluated to illustrate the dynamic estimation abilities of LSE and LRE models	147
	6-7	LSE and LRE dynamic model results for a hypothetical integrated freeway/arterial network	149
	7-1	Number of links and nodes used to represent the eastbound direction of the Highway 401 network	156
	7-2	Truncation effect on recorded station occupancies	159
	7-3	Measured trip duration as a function of day, departure time, and route taken, for the eastbound direction of Highway 401	161
	7-4	Definition of zones and the links that constitute the zonal flow	169

Nomenclature

Term	Description
α	relaxation factor used in the LSE model's iterative solution algorithm
β	dimensionless parameter used in BPR travel time function
φ	dimensionless parameter used in BPR travel time function
$\Delta \overline{E}^{n}$	marginal reduction in average relative link flow error (LSE model)
$\Delta \overline{E}_{\scriptscriptstyle R}^{n}$	marginal reduction in average relative link flow error (LRE model)
λ_a	Lagrange multiplier for link a
а	link number
C	link capacity (vph)
c_{ij}	the number of probe departure calls received for origin i and destination j during the given time slice
C_{ij}	value of the constraint equation for origin i and destination j
d	time period when demand departs origin i
d	an estimate of D based, in part, on probe and standard loop detector data
D	the true total number of trip departures initiated in the entire network during the given time slice
d_{ij}	an estimate of the number of trip departures between origin i and destination j
D_{ij}	the true number of trip departures between origin i and destination j in the given time period
$D^{I}{}_{a}$	discontinuity of upstream node of link a
D^2_a	discontinuity of downstream node of link a
E	total squared link flow error (LSE model)
E_R	total relative link flow error (LRE model)
\overline{E}	average absolute link flow error estimated from E
Eabs	true average absolute link flow error
\overline{E}^{n}	average absolute link flow error estimated as a proportion of the observed flow
$\overline{E}^{\scriptscriptstyle rel}$	true average absolute link flow error computed as a proportion of the observed flow
$\overline{E}_{\scriptscriptstyle R}$	average absolute link flow error estimated from E_R
$\overline{E}_{\scriptscriptstyle R}^{\scriptscriptstyle n}$	average relative link flow error estimated as a proportion of the observed flow
G	number of equally probably outcomes associated with a solution
i	origin zone

Nomenclature (continued)

Term	Description
I	quantity of information
j	destination zone
K	constant used in information theory
m	an estimate of M based on a sample of the RGS equipped and non-equipped vehicles observed on selected links on the network
M	the population network wide level of market penetration defined as the proportion of all trips, departing at some time t , that are RGS-equipped
M_{ij}	the population level of market penetration for origin i and destination j
n_p	the total number of probe vehicles initiating trips in the network during the current period
0	time period when demand is observed arriving on link a
P	matrix of link use probabilities
$\mathbf{P}^{\mathbf{t}}$	transpose of the matrix P
P_a^k	proportion of the flow entering on $\log a$ that exits at $\log k$
p_{ij}	an estimate of P_{ij} made from some sample of the total population (the probes)
P_{ij}	the population probability of any given trip being between origin i and destination j
P^{ao}_{ijd}	proportion of demand departing origin i at time d en route to destination j , that will arrive on link a at time o
q_{ij}^{a}	prior probability that counts on link a are associated with origin i and destination j
r_{od}	sample correlation coefficient computed between estimated and true O-D demands
$r_{\mathcal{V}}$	sample correlation coefficient computed between estimated and observed flows
R_a	relative reliability of the flow observed on link a
t	total travel time (seconds)
T	column vector of unknown O-D traffic demands
t_0	free speed travel time (seconds)
t_{ij}	prior estimate of the traffic demand between origin i and destination j (vph)
T_{ij}	traffic demand departing from origin i destined for zone j (vph)
T_{ijd}	traffic demand departing from origin i at time d destined for zone j (vph)
T_{ij}^{l+I}	new estimate of the true traffic demand between origin i and destination j (vph)
V	link flow (vph)
V	column vector of observed link flows

Nomenclature (continued)

Term	Description			
V_a	flow estimated to traverse link a (vph)			
V_a^{i}	observed flow entering the intersection at $leg k$ (vph)			
V_{ao}	flow estimated to arrive on link a during time period o (vph)			
V_k^o	observed flow exiting the intersection at leg k (vph)			
V'_a	flow observed to traverse link a (vph)			
V'_{ao}	flow observed to arrive on link a during time period o (vph)			
W	measure of entropy			

Glossary

Term	Description				
AADT	average annual daily traffic				
ATMS	advanced traffic management systems				
BPR	Bureau of Public Roads				
CCTV	closed circuit television cameras				
COMPASS	FTMS operating on Highway 401 in Toronto, Canada				
continuity	Flow continuity at nodes is used in this thesis to imply that total flow into the node is equal to the total flow out of the node. The term consistency may be substituted for continuity.				
CONTRAM	route choice and traffic simulation model				
COV	coefficient of variation (standard deviation divided by the mean)				
FHWA	Federal Highway Administration				
FREQ	traffic simulation model based on shock wave analysis				
FTMS	freeway traffic management system				
INTEGRATION	microscopic integrated network traffic simulation model				
ITE	Institute of Transportation Engineers				
IVHS	intelligent vehicle highway systems				
LRE Model	least relative error O-D estimation model				
LSE Model	least squares error O-D estimation model				
MOP	measure of performance				
MTO	Ministry of Transportation of Ontario				
O-D	origin - destination				
RGS	route guidance system				
Seed	seed matrix is used interchangeably with prior matrix				
TMC	traffic management centre				
UTCS	urban traffic control systems				

CHAPTER 1

INTRODUCTION

Potential solutions to many traffic engineering problems are typically evaluated on the basis of aggregated link-level measures of performance such as average speed, average delay, and total fuel consumption. In many urban centres, link-level traffic data, such as traffic volume, lane occupancy, and traffic speed, are readily available from existing traffic surveillance systems. Unfortunately, since link-level data are often impacted by traffic control systems, such as ramp meters and traffic signals, the evaluation of many potential traffic engineering solutions requires that trip-level data, such as trip origin, trip destination, and trip departure time, be known. As these trip-level data are often independent, in the short-term, of traffic control systems, they can represent the independent input into the traffic engineering evaluation process. Unfortunately, trip-level data are typically not readily available and are usually very costly to obtain through direct measures. This thesis examines the problem of estimating generally unknown trip-level data (origin - destination traffic demands) from readily available link-level data.

This chapter introduces the problem of estimating static and dynamic origin - destination traffic demands and briefly discusses approaches to solving this problem. First, four questions are posed and answered; "What are O-D demands?", "How can they be obtained?", "What are dynamic O-D demands?", and "What is the current status of O-D demand estimation?". Subsequently, several theoretical and practical difficulties are identified. Finally, the scope and approach of this thesis are defined and presented.

1.1 What are Origin-Destination Demands and Why are They Needed?

Many traffic engineering problems consist of evaluating a number of alternative schemes in terms of some measures of performance, such as total travel time, volume to capacity ratio, average speed, total fuel consumption, total emissions, etc. Typically, when making these evaluations, it is assumed that traffic flows on links are known and that the magnitudes of

these traffic flows are not affected by the performance of the scheme being evaluated. For example, the analysis of isolated traffic signals assumes that the traffic flows on each approach link are known *a priori*. It is further assumed that these approach flows remain unaffected by the performance of the traffic signal – regardless of how well or how poorly the traffic signal operates. This assumption is not limited to the analysis of isolated intersections. The TRANSYT model (Robertson, 1969; FHWA, 1984), which is capable of optimizing coordinated signal networks, also assumes that the magnitude of all approach flows are known and remain unchanged by the performance of the signals.

For these evaluation approaches, link flows are considered to be independent input data. However, traffic networks are dynamic environments in which multiple feedback loops exist. Individual drivers may alter their routing patterns in response to the traffic conditions that they experience, anticipate, or are informed of. Thus, the magnitude of the link traffic flows are actually outcomes of a process that may be dependent on the network traffic conditions that are experienced. The obvious question then, is "What are the independent inputs?".

In the context of traffic routing analyses, which are conducted for the purpose of evaluating operational control strategies, it can often be assumed that trip origin and trip destination locations are fixed and are therefore the independent inputs. Consider an automobile trip made for the purpose of commuting to work. The origin of the trip is the driver's residence, while the destination is the driver's place of employment. In the short term, both of these locations can often be assumed to be fixed and thus they are not affected by the existing traffic conditions.

One may now consider all of the drivers within an urban area, each departing from their own residence (origin) at some distinct time, *en route* to their own place of employment (destination). This is the most disaggregate level at which the system can be represented. By nature, origin-destination (O-D) demands are discrete rather than continuous quantities as vehicles are discrete entities, for which fractional values have no physical meaning. Of course, even as discrete quantities, it is not often practical, for even small networks, to analyze the trip making problem at this micro level. Rather, it is usually more convenient to aggregate both time and space in some manner. First, the network is divided into several geographic zones. Any trip beginning from a location within the boundaries of that zone, is considered to originate from that zone. Similarly, a trip's destination zone is that zone whose boundaries encompass the specific location of the trip's destination.

Trip departure time is usually also aggregated into a number of periods of finite duration. All trips that have departed after the start of a period and prior to the end of that same period, are often aggregated as if they had departed at a uniform rate throughout that time period.

The aggregated trip making behaviour, of all of the individual drivers, is referred to as the origin-destination demand for a particular network within a particular time frame. For each origin-destination pair, there will exist some number of vehicles that will, on average, make a trip from the origin zone to the destination zone. Usually this demand is represented as the number of vehicles that wish to make a trip per unit of time (i.e. vehicles per hour).

It is useful to illustrate the potential impact of considering link flows as the independent inputs versus considering O-D demands as the independent inputs. Consider the simple network illustrated in Figure 1-1, having two origins, one destination, and six links. A bridge, having a fixed capacity of 4000 vph (vehicles per hour) exists on link 5. The existing O-D demands and resulting link flows are provided in Table 1-1. It is desired that the impact on

the system of increasing the bridge capacity to 6000 vph be estimated. It is assumed that link travel time increases linearly with flow, so that at capacity, travel times are twice as long as the free flow travel time. It is further assumed that all drivers choose their routes so as to minimize their own travel time.

If it is assumed that link flows are independent inputs, then link flows are fixed and the only impact of increasing the capacity of the bridge is a decrease in travel time across the bridge as the volume to capacity ratio is reduced from 1.0 to 0.66. If, however, link flows are assumed to be variable and O-D demands are considered to be the independent inputs, then in response to the increased capacity on the bridge, drivers could alter their routes to minimize their own travel times and the flows on the links could change. As indicated, the magnitude of the expected impact on system delay, of increasing the capacity of the bridge, can be significantly different depending on whether link flows or O-D demands are assumed as the independent input.

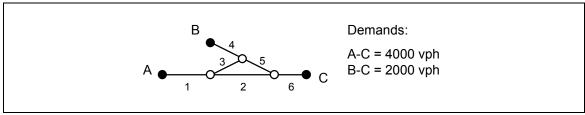


Figure 1-1: Simple network used to illustrate the impact of assuming link flows versus O-D demands as independent inputs to a typical traffic engineering evaluation

Table 1-1. Characteristics of example network							
Link	Capacity	Free Flow	Existing	Flow Independent		O-D Independent	
	(vph)	Time	Flows	Link Flows	Link Travel	Link Flows	Link Travel
		(min.)	(vph)		Time		Time
1	4000	10	4000	4000	20.0	4000	20.0
2	4000	20	2500	2500	32.5	2182	30.9
3	4000	10	1500	1500	13.8	1818	14.5
4	4000	10	2000	2000	15.0	2000	15.0
5	4000	10	3500	3500	18.8	3818	16.4
6	6000	10	6000	6000	20.0	6000	20.0
System Delay				6625 hours		6440 hours	_

Table 1-1: Characteristics of example network

The assumption that trip origins and destinations are fixed is not strictly true in the context of a regional planning analysis, in which long term interactions between land use development, employment opportunities, residential locations, and transportation infrastructure, are all inter-related. For example, the creation of lower cost suburban housing permits people to change the location of their residence. Alternatively, the supply of new transportation infrastructure may stimulate the development of new businesses, which in turn provide new trip destinations at these new employment opportunities. However, these changes generally occur slowly, such that O-D demands can be considered to remain constant within the much shorter time span that is considered in the evaluation of most traffic control strategies.

For O-D demands to be considered the independent inputs, it must also be assumed that alternative control strategies do not significantly alter the demands. Depending on the control strategy being investigated, O-D demands may change in several ways. For example, the

introduction of flexible work hours will affect trip departure time. Similarly, any alterations to the utility, or relative attractiveness, of one or more of the available transportation modes may affect trip mode choice between the use of private automobiles versus public transit. Finally, the relocation of large employment centers, such as government offices, will affect the location of some work trip destinations. Any of these changes may also cause some trip makers to decide not to make the trip, or cause additional trips to be made. All of these scenarios reflect significant changes to land use or transportation infrastructure, and as such are usually not analyzed using traffic engineering techniques. These situations are therefore not examined in this thesis.

1.2 How can O-D Demands be Obtained?

O-D demands can be observed directly only through some form of survey, in which the actual origin and destination of a vehicle or trip maker are obtained. Typically, one of three methods is used to carry out an O-D survey. Within the first method, a subset of households are identified in such a way as to capture a representative sample of the population. Each household is then telephoned and the respondent is asked to indicate the origin and destination of their typical home-to-work and other types of trips. A second approach is to stop some portion of *en route* vehicles for a roadside survey, in which the driver of each vehicle is asked to indicate the origin and destination of her/his current trip. A third approach relies on recording vehicle license plate numbers at several strategic sites within a study area. The license numbers recorded at one site are subsequently matched against observations at other sites to determine the origin and destination zone of each vehicle.

Each of these approaches provides a direct observation of the O-D demands for a sample of the population. However, these approaches are also expensive, and are susceptible to various sampling errors as well as systematic errors. Furthermore, it is rarely possible to extract dynamic demands that are applicable for short periods, of say 5 to 15 minutes, from direct survey data.

In the absence of direct observation, O-D demands are sometimes also estimated based on land use data. Typically, within the traditional four stage transportation planning process (trip generation, mode split, trip distribution, and assignment), O-D demands are estimated based on various zonal land use data, such as population, employment opportunities, and measures of affluence. Zonal statistics are obtained from land use plans and census data, and trip productions and attractions are obtained through household surveys. Relationships between the number of trips produced or attracted are then typically determined through the calibration of multiple linear regression models. These total zonal attractions and productions are subsequently distributed into individual O-D demands, typically through the use of a gravity type of model.

Conversely, O-D demands may be estimated through the use of a direct demand planning model, in which all four elements of the transportation process are integrated into a single model. Unfortunately, both approaches rely on aggregate zonal characteristics to determine O-D demands. This level of analysis may be well suited to aggregate planning analyses, but it is not appropriate for the estimation of O-D demands for traffic control strategy evaluations. The relationships between socio-economic factors and trip generation rates are insufficiently strong to permit the estimation of the time of trip departure – data that are necessary for the estimation of dynamic demands. Furthermore, the level of detail considered in most traffic

analyses, precludes the formation of large zones, for which trip generation rates could be adequately determined based only on the aggregate zonal characteristics.

A third approach attempts to infer the unknown O-D demand from observed link traffic flow data. This method circumvents the difficulty of adequately estimating zonal trip productions and attractions from aggregate zonal characteristics, as the traffic flows can be observed directly on the links within the network using vehicle loop detectors. Typically, the objective of the O-D estimation technique is to estimate an O-D demand that replicates the observed flows as closely as possible. Within these techniques, there is no need to calibrate relationships linking trip making behaviour to socio-economic factors, but unfortunately, there are several other difficulties that must be overcome, which will be discussed in detail later.

Recently, demonstration projects have commenced using systems that attempt to improve the efficiency of traffic networks by transmitting near real-time traffic data (i.e. link travel times) to suitably equipped vehicles. Computers on-board these vehicles can then utilize these data to plan a route towards the driver's stated destination. In addition to receiving information, such equipped vehicles also transmit data, including the vehicle's origin and destination, back to the traffic control centre. This two-way communication provides an opportunity to directly collect, in real-time, time-varying demands for a sample of the total vehicle population. The statistical reliability of such data as estimators of population characteristics, is susceptible to some of the same sampling errors as are other forms of direct survey, however, probe based O-D estimates may be obtained economically if a two-way communication link exists, and these estimates may be dynamic.

These last two approaches are expected to hold the most promise for the reliable estimation of dynamic O-D demands that are suitable for traffic analyses. As such, this thesis focuses on examining the current state-of-the-art of these approaches, identifying remaining shortcomings, and developing and evaluating two estimation methods that fall within these approach categories.

1.3 What are Dynamic O-D Demands?

Earlier sections of this thesis have determined what O-D demands are, and why they are needed. In addition, four potential methods of obtaining O-D demands have been identified. At this point it is necessary to more clearly differentiate between what are considered to be static and dynamic O-D demands.

Dynamic demands are considered to differ from static demands in that static demands assume that every trip is, or can be approximated to be, completed within a single analysis time slice such that any temporal and spatial interactions between consecutive time slices can be ignored. In order to avoid violating this assumption, static demands are often developed to be applicable for a relatively long period, say from 30 minutes up to several hours. Within this time period, vehicles are further assumed to depart their origins at a constant rate.

However, as will be illustrated later in this thesis, field data show that O-D demands can vary significantly with time of day. Static demands, with application periods of as much as several hours, are therefore often not adequate to reflect these temporal variations. It is possible, however, to develop a time series of static demands, which reflect to some extent the temporal variations observed in the actual field demands. In creating this time series of

demands, as the duration over which each individual static demand is active becomes shorter, the assumption, that all trips can be completed within the departure time slice, becomes increasingly more violated. Dynamic O-D demand estimates, which do not require trips to be completed within their departure time slices, may have departure time slices of any duration without violating these assumptions.

The static approach of inferring O-D demands from link flows is easily adapted to the estimation of dynamic O-D demands. Link flows are usually obtained through the use of induction loop detectors and are typically available on a 20 second to one minute basis from urban traffic control systems (UTCS) and freeway traffic management systems (FTMS). Vehicle probe data also provide an opportunity to obtain efficiently, in near real-time, O-D data for a sample of the vehicle population. These O-D estimation approaches, therefore, lend themselves rather well to the task of estimating dynamic O-D demands.

1.4 What is the Current Status of Synthetic O-D Demand Estimation?

Over the last four decades, much work has been carried out to develop systematic approaches for reliably estimating O-D demands. Initially, these efforts were limited to developing methods appropriate for longer term transportation planning purposes as the highway network went through a rapid expansion phase. More recently, due in large part to the need to manage existing automobile transportation infrastructure more efficiently, there has been greater emphasis on estimating O-D demands that may be used for more short term traffic analyses.

In discussing the current status of O-D estimation approaches, it is appropriate to introduce a system for classifying existing and newly proposed approaches. This system is intended only to provide a structure within which an assessment of different techniques can be made, and is not intended to be a definitive categorization of all proposed methods. In fact, some approaches can arguably be assigned to more than one category. However, the chosen classification permits several important fundamental distinctions between different approaches to be made more easily and concisely. In addition to the four classifications that are defined, several specific problems, that must be addressed in estimating O-D demands, are identified and discussed

The first distinction that can be made is between heuristic and mathematical approaches. Several heuristic approaches have been proposed, however, since they often lack a sound mathematical foundation, less confidence can be placed in their estimation abilities. In general the preferred approach has a mathematical basis that has more clearly defined theoretical properties and may also be practically applied.

A second distinction is made based on the scope of the approach's applicability. One can generalize this distinction to classify approaches as being either applicable to turning movements at isolated intersections only, or to more general networks. As the estimation of intersection turning movements is simpler than the estimation of O-D demands for general networks, much of the earlier research reported on in the literature has examined only intersections, but is not necessarily valid for the purposes of this thesis.

The third distinction, that is made between different techniques, is based on the approaches' assumption regarding the types of routes that drivers utilize. Routes must usually be known in order to convert estimates of O-D demands into estimates of link flows. The vast majority of approaches therefore assume that routes are known *a priori* and do not vary with levels of O-D demand. A few methods have explicitly recognized that the routes that drivers utilize are dependent on the as yet unknown O-D demands. These approaches must attempt to estimate simultaneously the O-D demands and the routes. Several such mathematical models have been formulated, however, adequate solution algorithms for these models have not yet been fully developed.

The last distinction, that is made between the capabilities of different approaches, is to distinguish if the techniques estimate static and/or dynamic O-D demands. It should be noted that the problem of estimating dynamic demands is considerably more demanding than that of estimating static demands. In addition, even though static estimation is considerably less challenging, no single static estimation method has been shown to be both theoretically and practically superior to all others. Perhaps the greatest impact on the current state of static estimation has been made by Van Zuylen and Willumsen (1980; Van Zuylen, 1981; Willumsen, 1992), with their development of compatible yet contrasting trip-count based entropy maximizing and link-count based information minimizing models.

In general, much less has been accomplished in developing dynamic O-D demand estimation methods. Significant contributions have been made by Cremer and Keller (1987; 1984; 1981; Cremer 1991) and others (Kessaci, *et al.*, 1989) to estimating dynamic intersection turning movements, but to-date it has either not been possible to extend the formulation of these methods of application to general networks, or their performance upon application to general networks has proved unsatisfactory.

1.5 What are some of the Theoretical and Practical Problems of O-D Estimation?

Lastly, several theoretical and practical problems, that are encountered when estimating network demands, are used to provide another means of comparing different approach strategies. These difficulties, illustrated in Table 1-2, include the theoretical possibility that multiple O-D demands may exist that exactly replicate the observed link flows. In practice, link flow continuity at nodes rarely exists within a set of observed data, with the result that often no O-D demand exists that exactly replicates the observed flows. However, multiple solutions, that have equal levels of link flow error associated with them, may still exist. It is usually not known *a priori* into which of the four categories a particular estimation problem can be classified.

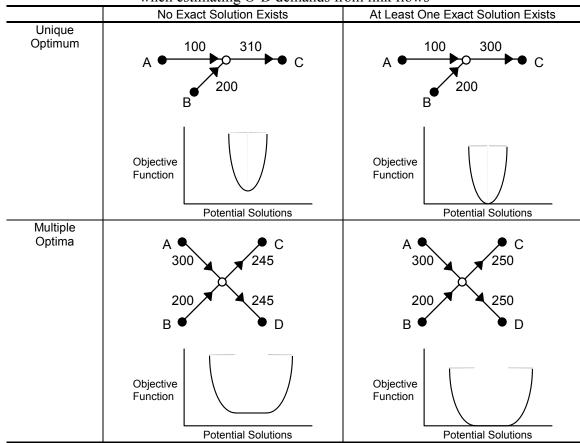


Table 1-2: Illustration of several theoretical difficulties that can be encountered when estimating O-D demands from link flows

Further problems arise due to the fact that the routes, that are utilized by the O-D demands, are usually not known. Furthermore, even if the routes are known, the proportion of the demand that utilized each available route is often still unknown. Lastly, but possibly of greatest significance, usually no comparisons can be made between the estimated and the true demands, as these true demands are rarely known. It is possible to carry out controlled experiments in which the true O-D demands are known, however, it is usually not possible to transfer the conclusions that are made under these conditions to the general case, as the performance of the estimation approach is usually a function of the network's characteristics.

1.6 What is the Scope and Approach of this Thesis?

1.6.1 Thesis objectives

This thesis has the following eight specific objectives:

- 1. To develop a theoretically sound static and dynamic network O-D estimation model formulation that is capable of being applied to real world traffic networks.
- 2. To incorporate explicitly into the model formulation an ability to reflect the relative reliability of the different sources of observed link traffic flows.

- 3. To add to this model a feature that explicitly permits a consideration of the relative reliability of the prior O-D information that is specified for each individual O-D cell.
- 4. To develop a model solution procedure that is responsive to both theoretical and field requirements. For example, the model must be able to estimate the most likely O-D when either multiple solutions exist that exactly replicate the observed link flows, or when no feasible solution exists that replicates the observed flows. The model must accomplish this without any pre-processing of the observed link flow data to ensure flow continuity, or to remove redundancies.
- 5. To illustrate the performance characteristics of the model for a number of simple hypothetical example networks for which true demands are known and analytical solutions can be applied.
- 6. To explore the viability of applying the model to a 35 kilometre section of a multilane urban freeway corridor in Toronto, Canada, using aggregated 15 minute FTMS link flow data from 5 AM to 11 AM. This system is one of the most heavily traveled freeway corridors in North America, and provides a challenging field test for the O-D estimation model.
- 7. To examine statistically the expected reliability of O-D demand estimates made solely on the basis of probe vehicle data.
- 8. To develop expressions that can be used to quantify the reliability of O-D demands that have been made solely on the basis of probe vehicle data.

1.6.2 Problems addressed by this thesis

It is prudent to state explicitly the problems, associated with estimating O-D demands, that are addressed within this thesis.

- 1. This thesis specifically examines the problem of estimating static and dynamic network O-D demands based on observed link flow data.
- 2. The theoretical problem of choosing one specific solution when multiple solutions may exist is also addressed. The problem of estimating an O-D demand when the presence of flow discontinuity precludes the existence of an O-D demand, that can exactly replicate the observed link flows, is also addressed. Since no direct measure can be estimated of how well the estimated demand replicates the unknown true demand, several surrogate measures of performance are developed.
- 3. The problem of incorporating the relative reliability of link flows and prior information within the mathematical formulation is addressed.
- 4. The problem of developing a solution algorithm that can be implemented using available technology for application to real world networks is addressed.
- 5. This thesis addresses the problem of quantifying the reliability of O-D estimates made solely on the basis of probe vehicle data.

1.6.3 Problems not addressed by this thesis

In contrast to the previous section, it is useful to state explicitly which aspects of the O-D demand estimation problem are not addressed by this thesis.

- 1. This thesis does not address the problem of determining vehicle routes and route weights. It is assumed throughout this thesis that adequate vehicle routes and route weights will be exogeneously defined.
- 2. This thesis does not examine the relative benefits or dis-benefits of using RGS equipped O-D estimates as prior information to a synthetic O-D estimation model.

1.6.4 Similarities and differences to other approaches

This thesis develops two O-D estimation models. The first, the least squared error (LSE) model adopts, as its objective function, the sum of the squared link flow differences. The LSE model objective function is similar to a least squares regression, but the iterative LSE solution algorithm constrains the O-D estimates to non-negative values. The LSE model provides an estimate that is closer to the true constrained global optimum than does truncation of the unconstrained estimate. The LSE model is also able to estimate an O-D demand when the problem is under-specified and multiple solutions exist. The formal solution to constrained or unconstrained regression is not valid under these conditions. Furthermore, it is shown for several examples that, when multiple solutions exist, and no prior information is available, the LSE model provides an O-D estimate that approximates the most likely solution.

The second proposed model, the least relative error model (LRE), is based on the assumption that the error function should consider the magnitude of the link flow error relative to the observed flow. Again, it is shown for several examples that, when at least one feasible solution exists, the LRE model results in O-D estimates that are similar to the information minimization model proposed by Van Zuylen and Willumsen (1980). When no feasible solution exists that exactly replicates the observed flows, the information minimization model cannot be applied.

The estimation of O-D demands solely on the basis of probe vehicle data has few similarities to other existing work, except that it may be considered to be somewhat similar to that of estimating O-D demands from survey methods. However, several characteristics unique to vehicle probe data decrease this level of similarity. First, probe data reliably provide the origin and destination for each probe vehicle, while license plate surveys often rely on only a portion of the license plate identifier, introducing the possibility that the recorded vehicle identifier is no longer unique. Additionally, license plate and road side surveys cannot ensure that the same vehicle does not pass through several recording stations within the same trip. The primary difficulty that is faced when making probe based O-D estimates is determining the proportion of the total population that is being sampled.

1.6.5 Method of approach

Two similar, but different, network O-D estimation models are formulated. Initially, the models are formulated to estimate static demands only, but after ensuring that these models are well behaved and exhibit desirable characteristics for the static problem, their formulations are extended to the dynamic case. The first model that is proposed assumes the sum of the squared deviations between the observed and estimated link flows as the error function that is to be minimized, while the model's algorithm ensures that O-D estimates are non-negative. This model is referred to as the Least Squared Error (LSE) model.

The second proposed model is based on the assumption that the most appropriate error function is a relative measure of the link flow error, where any errors are expressed as a proportion of the observed flow. This model is referred to as the Least Relative Error (LRE) model.

1.6.6 Method of validation

The proposed LSE and LRE models are validated in several ways, as described below.

- First, the models' mathematical formulations are completely described and any assumptions are stated and justified. Limitations of both the formulations and the solution algorithms are noted and their impact on the estimated O-D demand is explored.
- 2. The model formulations are solved analytically for several small hypothetical networks, to demonstrate the performance and validity of the models.
- 3. The iterative solution algorithms developed to solve efficiently the formulations are presented and are used to estimate O-D demands for the same hypothetical networks that are used to demonstrate the model formulations. It is shown that the solution algorithm results are consistent with the model formulations.
- 4. The model estimates are compared, for two separate networks, with estimates from the information minimizing and entropy maximizing models proposed by Willumsen and Van Zuylen.
- 5. Both models are applied to dynamic O-D estimation problems on hypothetical networks for which an earlier O-D loading is performed, such that the true O-D demand is known
- 6. The LSE and LRE models are used to estimate 15 minute dynamic O-D demands based on FTMS data from 5 AM to 11 AM, for a 35 kilometre section of a multi-lane urban freeway network in Toronto, Canada. Since the true demands are not known for this real network, only surrogate measures of performance are examined.
- 7. In each validation case, several measures of performance are examined, including the average root-mean-squared (RMS) deviation between the estimated and true O-D demand (when the true demand is known), the RMS deviation between estimated and observed link flows, the correlation coefficient based on observed and estimated flows, and the marginal reduction in link flow error associated with each iteration of the solution algorithm.
- 8. Finally, the expressions for the reliability of RGS probe based O-D estimates, derived from statistical sampling theory, are compared with simulation model results. These comparisons serve two purposes. First, the similarity between the analytical estimates and the simulation results provide confidence in the appropriateness of the approximations made in deriving the analytical expressions. Second, the comparisons provide a level of confidence in using the simulation model to evaluate networks that are too complex to easily permit the application of the analytical expressions.

1.6.7 Method of presentation

Chapter 2 provides a detailed review and assessment of the current state of O-D estimation techniques. This chapter briefly examines each of the four general O-D estimation approaches that have been proposed in the literature. It is demonstrated that the use of link flows is the most promising of these approaches. Further examination of the existing literature is limited to those techniques that rely on the use of observed link flows to estimate the demands. As there exists a large body of relevant literature, the presentation of this assessment is organized in accordance with the O-D estimation approach categories that were defined in Section 1.4.

In Chapter 2, it is found that though no single estimation technique has been shown to be superior to all others, the maximum likelihood models that have been proposed by Van Zuylen and Willumsen, seem to serve as a benchmark against which many other models are typically compared. Chapter 3, then, examines in more detail these two models, in order that the theoretical basis, assumptions, and limitations can be fully understood. The performance characteristics of these two models are examined for several example networks.

Chapter 4 provides the formulation and performance characteristics of the LSE model, which is the first of the two O-D estimation models proposed in this thesis. The formulation of the LSE model is illustrated for several network traffic conditions, including situations in which flow continuity, flow discontinuity, and multi-path route issues exist. The incorporation of link flow reliability into the model formulation is described and illustrated, and the issue of determining appropriate link flow reliability factors is examined. Potential factors which are explored to incorporate the latter, include the inverse of the flow variance, the inverse of the coefficient of variation (COV), and a heuristically determined factor based on the magnitude of flow discontinuity. An iterative solution approach, that is used to solve the LSE model formulation, is developed on the basis of the Gauss-Seidel and Jacobi iterative solution techniques. Conclusions regarding the LSE model's theoretical appropriateness and its performance characteristics are made.

Chapter 5 presents the formulation of the second proposed model, the LRE model. The presentation of this model parallels the presentation of the LSE model in Chapter 4. Example networks which were used to illustrate the performance characteristics of the LSE model in Chapter 4 are also used here to illustrate the performance of the LRE model. Conclusions are made regarding the appropriateness of the LRE model.

Chapter 6 provides the dynamic extension to both the LSE and LRE static model formulations. The need for dynamic demands, as opposed to static or a time-series of static demands, is demonstrated using field data. The computational implications of extending the static formulations to permit dynamic estimation are examined. A method for determining link use probabilities from known routes, route weights, and link travel times, is also presented.

Chapter 7 describes the application of the LSE and LRE dynamic models to a 35 kilometre section of a multi-lane freeway network in Toronto, Canada. This application relies on the use of FTMS link flow and link travel time data. The true demands, routes and route weights are not known. Dynamic 15 minute demands are estimated for the 6 hour period from 5 AM to 11 AM.

Chapter 8 presents a statistical analysis of the reliability of O-D demands estimated solely on the basis of RGS probe data. This statistical analysis is compared with a simulation approach for a hypothetical network.

Conclusions and recommendations are made in Chapter 9.

CHAPTER 2

Assessment of the current state-of-the-art

2.1 Introduction

The problem of obtaining accurate and reliable O-D demands has challenged traffic engineers and planners for several decades. During this time, considerable research effort has been directed at devising methods for solving this problem by means of various direct survey and indirect analytical methods. More recently, due to the need to have adequate knowledge of dynamic O-D demands for the operation and evaluation of FTMS, UTCS, and RGS, the desired degree of aggregation and resolution of the estimated O-D demands has also changed. It is therefore instructive to review the range of relevant approaches that has been developed to-date, and then to utilize the gap between the current state-of-the-art and current requirements to indicate the present needs for further research.

This chapter is organized as follows. Section 2.2 identifies and briefly describes the four basic categories of approaches that can be used to estimate O-D demands. As this thesis addresses the issue of estimating O-D demands from link traffic flows, Section 2.3 describes in greater detail that specific fundamental problem and solution approach. Section 2.4 provides a comprehensive examination of previous research with respect to this link flow based approach. The existing body of literature is categorized by approach methodology, permitting efficient dissemination. It also provides an organized structure that facilitates the identification and evaluation of the various link based O-D estimation approaches that have been previously proposed. Section 2.5 provides a summary of the state-of-the-art and as such provides a logical reference point from which the current thesis contributions can be examined.

2.2 General Solution Approach Categories

During the course of the past four decades, many different approaches have been utilized in attempting to estimate O-D demands. Part of the reason that such a variety of approaches

exists, is the changing requirements that have emerged. These changes result from a shift in emphasis from requiring strictly static O-D demand forecasts for regional planning purposes, to requiring fully dynamic O-D demands for the operation and evaluation of RGS and other traffic management systems.

The majority of the existing O-D solution approaches can be classified as belonging to at least one of the following three categories;

- 1. Transportation planning models
- 2. Direct sampling of individual trip makers
- 3. Indirect estimation from link flows

In some instances the allocation of a particular solution approach to one of the above three categories is somewhat arbitrary as the final approach may contain elements of more than one category. However, the identification of the above three categories still serves as a convenient overall structure from which to begin the discussion.

The following three subsections provide a brief description of each of the above three categories and outline the relevance of each to the problem of estimating O-D demands for the operation and evaluation of RGS and other traffic management systems.

2.2.1 O-D estimates from transportation planning models

Historically, the focus of urban transport systems planning was primarily on estimating expected travel demands for the purpose of designing future transportation infrastructure systems. From this need, the typical four stage travel demand forecasting process, consisting of trip generation, mode split, trip distribution, and route assignment, evolved.

Two of the stages of this process, trip generation and distribution, combine to estimate expected O-D demands, and as such, are of relevance to this thesis (Figure 2-1). The following is a brief description of the trip generation and distribution processes along with a critique of the applicability of this approach to estimating O-D demands for operation and evaluation of traffic management systems. Many quality reference sources (including Hutchinson, 1974 and Kanafani, 1983) are available that describe trip generation and distribution in much more detail than will be presented here.

The process of trip generation involves the estimation of zonal trip productions and attractions. Typically, this is accomplished by utilizing a calibrated regression model that forecasts the number of zonal trip productions and trip attractions as some function of a number of independent zonal land use descriptors.

In order to estimate the zonal productions and/or attractions, the coefficients of the trip generation model need to be determined. This calibration requires a knowledge of the relevant zonal characteristics as well as the actual trip productions and attractions. Using these known base year conditions, the model parameters are typically estimated using regression. Having calibrated the demand equation for the base year conditions and if estimates of future zonal characteristics are available, the equation can then be used to estimate future demands. The result of this demand model is an estimate of the total future productions and attractions for each zone in the network. In order to determine the interaction between zones, a distribution of these total productions and attractions is necessary.

In order to resolve the underspecifications that exist, the relationship that $T_{ij} \propto 1/d_{ij}^n$ is introduced, where T_{ij} is the demand between origin i and destination j, and d^n is some measure of the separation distance between i and j. The similarity to Newton's gravitational model has resulted in the common reference to the above model as the gravity model. The separation distance is referred to as the deterrence or impedance function and is often measured by travel time.

The gravity model is a subset of the more general synthetic O-D problem if the row and column totals are considered to be observed link flows on the zone connectors only. Wilson (1970) has shown how the gravity model can be derived based on the entropy-maximization principle of information theory. This formulation provides the most probable O-D trip matrix given the zonal trip productions and attractions.

The transportation planning approach to estimating O-D demands exhibits at least three characteristics that make it undesirable for the estimation of O-D demands for the operation and evaluation of traffic management systems. First, the calibration and subsequent use of the model require that trip attractions, productions, zonal characteristics, and trip length frequency be known for some base year conditions. These data and regression models reflect aggregate zonal socio-economic characteristics that are costly to obtain and cannot reflect the temporal level of detail of trip making necessary for traffic management use.

Second, the traditional transport planning approach is generally only capable of capturing the aggregate changes in trip making behaviors that are reflected by changes in aggregate zonal characteristics. It is not possible to capture the daily or hourly variations that are a part of the dynamic nature of O-D traffic demands.

Third, the use of a gravity model to resolve the problem of underspecification can be limiting as it assumes that only traffic flows on zone connector links are available, and that trip making behaviour is adequately described by the gravity model. This structure may suitably describe trip making behaviour at a regional level, but is not likely to be adequate for describing trip making behaviour within urban centres.

Direct demand models differ from the four quasi-independent stages of urban transport systems process in that they incorporate O-D estimation, model split, and assignment all within the same model. The potential advantage of direct demand models is that their unified structure provides for a more sound conceptual basis. However, it would appear that these models have little more of practical value to offer for estimating O-D demands for the operation and evaluation of traffic management systems than the four stage transportation planning models already examined.

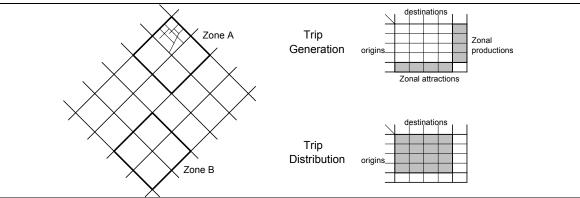


Figure 2-1: Use of the trip generation and trip distribution phases of the transportation planning process to estimate O-D demands

2.2.2 O-D estimates from direct sampling

One method of obtaining an O-D for a region is to directly measure the O-D for some random sample of the population. Then, using statistical sampling theory, conclusions can be drawn regarding the O-D of the entire population.

As illustrated in Figure 2-2, O-D demands are typically observed, by way of a license plate survey, for a sample of the population. Observers are located at a number of *zones* in the network and license plate numbers are recorded for a portion of the total traffic passing the observation station. After the observation period is complete, license plate numbers observed at different zones are used to match vehicles' origins and destinations.

In practice, due to observation time constraints, usually only a portion of the license plate number is able to be recorded. In this situation, not only must standard sampling error be considered, but as the recorded portion of the license number may no longer be a unique vehicle identifier, errors can arise due to incorrect vehicle identification.

It has been reported in the literature (Hauer, 1978) that the presence of spurious matches may in practice lead to significant bias. As a result, a number of statistically based approaches have been proposed to minimize this bias. One of the most recent contributions proposes that travel time be used to identify physically infeasible matches (Watling, *et al.*, 1992). Watling (1990) also provides a review of earlier methods.

Despite the use of such methods to minimize the impact of spurious matches, direct sampling by way of license plate surveys is costly, time consuming, and usually only provides an aggregate O-D, reflecting average O-D demands over a commuting peak period. Even though license plate surveys can often provide better estimates of O-D demands than the transportation planning process, they are still usually inadequate for developing O-D demands suitable for the operation and evaluation of traffic management systems.

Though direct sampling, by way of license plate registration matching, is not an adequate approach for estimating dynamic O-D demands, the use of statistical sampling methods will be utilized later in this thesis to provide estimates of the O-D departure rates, and the reliability of these rates, based on information obtained from RGS-equipped vehicles.

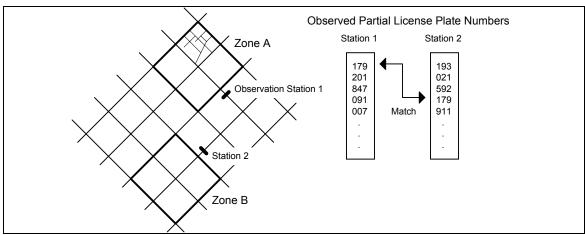


Figure 2-2: Use of partial license plate number sampling to estimate O-D demands

2.2.3 O-D estimates from link counts

The last category of approach for estimating O-D demands, and the one of most interest to this thesis, is the estimation of O-D demands based on observed link traffic counts. This approach, illustrated in Figure 2-3, has a number of advantages over the previous three approach categories.

- 1. With the increasingly popular use of FTMS and UTCS, most urban centers have electronic surveillance on at least some portion of the road network. This electronic surveillance, typically in the form of induction loop detectors, is able to provide continuous centralized link traffic counts in near real-time. Being able to utilize these freely available data would be advantageous from a data collection viewpoint alone.
- 2. Link traffic flows are the manifestation of O-D demands interacting within the network. Thus, link counts provide a direct reflection of the unknown O-D demands. Furthermore, these counts reflect the dynamic spatial and temporal behavior of these demands.
- 3. Link counts provide link specific constraints to the O-D estimation problem, as opposed to aggregate zonal constraints. The degree of aggregation is determined by the level of resolution of the traffic network representation.

It is primarily for these three reasons that estimating O-D demands from link flows has attracted so much attention in the past 20 years.

Unfortunately, despite the aforementioned advantages, the best manner in which the information, provided by link counts, can be used to estimate O-D demands still remains to be proven. In order to present the different philosophies evident in the existing literature, and to provide some evaluation of the state-of-the-art, it is useful to first examine in some detail the basic problem of estimating O-D demands from link counts. This examination will clearly reveal several difficulties that must be addressed when estimating O-D demands from link counts. These difficulties are identified and the existing literature is assessed according to the philosophical approach adopted to deal with these difficulties.

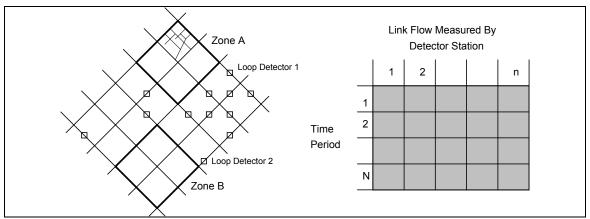


Figure 2-3: Use of observed link flows to estimate O-D demands

2.3 Definition of Link-Count Based O-D Estimation Problem

Before discussing the problem of synthetically estimating O-D demands, it is instructive to investigate first the inverse or complementary relationship between O-D demands and link traffic flows. The reason for this is that many synthetic O-D estimation models involve a loading model in which a known O-D matrix is applied to a network in order to estimate the associated link flows. The correlation between such estimated and observed flows is then utilized to refine the initial O-D matrix.

Traditionally, in the transportation planning context, a static O-D matrix is assumed to be a known input to the traffic assignment process. In order to determine the resulting link flows, the O-D demand is assigned to the network. This requires the simultaneous determination of the routes that traffic is likely to utilize and the link flows and travel times that occur on these routes, as these latter link times will impact the above route choices. Various techniques have been used to assign the traffic demand, but regardless of the method, the result can be expressed as a three-dimensional matrix describing the proportion of demand departing origin i, en route to destination j, using directed arc (link) a. For time varying O-D demands, this link use proportion matrix includes two more indices; namely the time at which demand departs the origin (a), and the time at which the demand is observed on link a (a).

The flow estimated to arrive on a link at any time is then given by Equation [2-1].

$$V_{ao} = \sum_{d} \sum_{ii} \left(T_{ijd} \cdot P_{ijd}^{ao} \right) \tag{2-1}$$

where: i = origin zone number

j = destination zone number

a = link number

d = time when demand departs origin i

o = time when demand is observed arriving on link a

 V_{ao} = flow observed arriving on link a at time o

 P_{iid}^{ao} = proportion of demand departing origin *i* at time *d* en route to

destination j, that will arrive on link a at time o

 T_{iid} = demand departing from origin *i* at time *d* destined for *j*

Equation [2-1] consists of three quantities; link flows, O-D demands, and link use proportions. In the O-D estimation process, link flows are directly observable, and as such can be considered to be known, though this knowledge is usually not without error. The O-D demands are unknown, and are the objectives of the estimation process. The link use proportions, which themselves may be a function of the unknown demands, are also unknown.

Typically, it is either assumed that the network is of such a configuration that the link use proportions are obvious (i.e. a static scenario on a linear network), or it is assumed that the link use proportions are constant, and that they can be determined exogeneously to the O-D estimation problem. In this way, Equation [2-1] contains only one unknown quantity, namely the O-D demand.

For the estimation of dynamic O-D demands, $(a \cdot o)$ equations and $(i \cdot j \cdot d)$ unknown demands exist. In theory, when node and path flow continuity exist, the number of unknowns is generally greater than the number of independent equations. The problem is then underspecified and multiple solutions often exist that each exactly replicate the observed link flows.

In practice, due to measurement error and the temporally or spatially varying nature of traffic flow, node and path flow continuity rarely exist. Thus, few, if any, of the $(a \cdot o)$ equations are redundant, and there are often more equations than unknowns. The problem is now overspecified, and there is, in general, no solution that exactly replicates the observed link flows.

The approach taken to solving the O-D estimation problem depends a great deal on the assumptions made concerning the character and availability of the link use proportions, the existence of node and path flow continuity, the static or dynamic nature of demands, the presence and quality of additional information, and the scope of the network (intersection, linear corridor, or general network). In the next section, the existing literature will be examined to determine how other researchers have approached the problem. Particular attention will be paid to understanding the impacts that their assumptions have on the ultimate performance of their proposed method.

2.4 Categorization of Link-Count Based Approaches

This section provides the results of an examination of approaches to O-D estimation from link flows as reported in the literature. This section is not intended to be a comprehensive summary of all previous work, as many such summaries are already available (Cascetta and Nguyen, 1988; Willumsen, 1992; McNeil and Hendrickson, 1985b). Instead, the intent here

is to categorize the rather large body of previous work in order to identify the major philosophical approaches that are prevalent. Having identified these different approaches, comments are then made concerning their relative strengths and weaknesses. This structure permits the reader to assimilate the current state-of-the-art, and provides an appropriate context in which the contributions of this thesis can be presented.

As noted in Chapter 1, four classifying distinctions are made between existing link flow based O-D estimation approaches. These four distinctions are:

- 1. Heuristic vs. mathematical approaches
- 2. Intersection vs. network O-D demands
- 3. Pre-specified vs. variable routes
- 4. Static vs. dynamic O-D demands

The following subsections discuss representative proposed models for each of the classifications. Within each classification, the state-of-the-art is assessed and comments are made.

2.4.1 Heuristic vs. mathematical approaches

In contrast to the typical approach, in which a rigorous mathematical formulation is first developed, a number of authors have proposed heuristic approaches to the estimation of O-D demands. Hauer *et al.* (1981) proposed that intersection turning movements proportions, could be considered a function of intersection type, where five different intersection types were identified. On the basis of data for 145 intersections, average turning movement proportions were then computed. These proportions could then be used to estimate turning movements at any other intersection.

Other heuristic approaches that have been proposed include those that are only appropriate for estimating turning movements at intersections (Mountain and Westwell, 1983; Mekky, 1979), and those that are appropriate for estimating static demands in more general networks. Yagar (1988) proposed a method that utilizes roadside survey data to obtain static O-D information for a sample of the population. The responses from each survey station can then be scaled up to reflect the total link flow observed on that link. Any O-D pairs that are not represented within the sample, and any link flows that are underestimated by the estimated O-D demand, are assumed to be insensitive to the impact of control strategies and as such are represented as deterministic pre-load flows.

Barbour *et al.* (1991) proposed a method that is based on the incremental determination of static demands on the basis of the flows on the shortest paths between each O-D pair. This method does not require prior knowledge of either routes or O-D demands, however, the shortest path algorithm requires that delay-flow relationships be known for each link in the network.

The most significant criticism of heuristic approaches, is that they lack sufficient mathematical basis to instill confidence in the properties of the estimated solution. Furthermore, few of these heuristic approaches have been shown, by way of application, to be reliable methods for the practical estimation of dynamic O-D demands in a general network.

2.4.2 Intersection vs. network approaches

The problem of estimating O-D demands for a single intersection simplifies to that of estimating the proportion of the approach flow that exits the intersection by each possible exit. These proportions are generally referred to as the turning proportions. The constrained problem is simpler to solve than the more general network O-D estimation problem, particularly for the static case. To illustrate this, consider Figure 2-4, which depicts a typical four leg intersection. It is desired to determine the proportion of the flow arriving at approach 1, that exits the intersection at each of the other three legs. This situation can be mathematically stated as in Equation [2-2].

$$V_k^o = \sum_a \left(P_a^k \cdot V_a^i \right), \, a \neq k \tag{2-2}$$

where: $V_k^o = \text{observed flow exiting at leg } k$ $V_a^i = \text{observed flow entering at leg } a$ $P_a^k = \text{proportion of flow entering on leg } a \text{ that exits at leg } k$

One notable difference between Equation [2-2] and Equation [2-1] is the absence of any O-D demand term within Equation [2-2]. The only unknown quantities in Equation [2-2] are the turning proportions. If flow continuity exists, then at least one exact solution also exists. Furthermore, as there is only a single equation per approach leg, there are no redundant equations. For this four leg intersection, there are three unknown turning movement proportions per approach leg, or a total then of 12. Equation [2-2] provides four constraint equations. An additional four constraints are provided by the fact that the sum of the turning proportions for each entrance leg a, across each exit leg k, must be equal to one. Thus, there exists 12 unknowns, but only 8 independent constraint equations. This system is underspecified, and will usually have multiple solutions that satisfy the constraint equations.

This problem is simpler to solve than the more general problem of estimating O-D demands for a network, as the problem of obtaining link use probabilities is avoided. The routes are also known (i.e. each existing intersection leg, except the one by which the demand enters the intersection, is a potential route), and it is the route weights (turning movements) that are being solved for. The actual O-D demands are then determined by applying these turning movements to the observed entrance flows.

Due to the simpler problem that isolated intersections pose, a significant amount of the initial O-D estimation research focused on the estimation of static intersection turning movements (Jeffreys and Norman, 1977; Van Zuylen, 1979; Norman et al., 1979; Bell, 1984b, and Martin *et al.*, 1992)

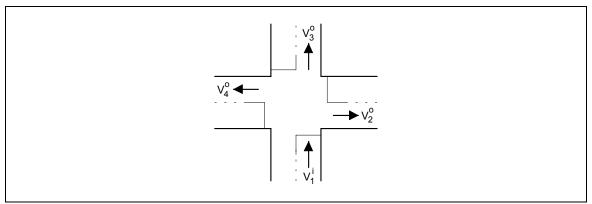


Figure 2-4: Estimating static turning movements for an intersection

2.4.3 Pre-specified vs. variable routes

Most existing solution approaches assume that the proportion of drivers between any origin i and destination j using a particular route r is known and is independent of the amount of traffic flow on route r. This assumption permits the link use proportions to be specified independently from the O-D demands. Having assumed that routes are independent of flow, most authors further assume that these unknown link use proportions have been exogeneously defined. The problem then becomes one of estimating the unknown O-D demands on the basis of the known link use proportions and the observed link traffic flows.

In reality, routes are typically not independent of the prevailing demand. Therefore, as demand increases, congestion and delay on the selected route generally increase until some point at which drivers will switch to another route having a lower travel cost. When O-D demands are known, equilibrium and stochastic equilibrium assignment methods can be used to develop routes that are consistent with these relationships.

If routes are not assumed to be fixed, link use proportions can no longer be determined independent from the O-D demands. Due to this additional complexity, most O-D estimation techniques assume fixed routes. Notable exceptions include a mathematical model proposed by Fisk (1988 and 1989) that is based on game theory (Fisk, 1984) and combines user optimum assignment with maximum entropy. This approach results in a bi-level mathematical program for which an efficient solution algorithm does not yet exist.

Nguyen (1977 and 1984) also formulated an O-D estimation model based on the assumption that routes were a function of the unknown demands. Unfortunately, the model did not provide a unique solution. This method was subsequently extended (Turnquist *et al.*, 1979; LeBlanc *et al.*, 1982; Sheffi *et al.*, 1987), however, the extended method requires accurate knowledge of the link travel times and of the relationship between link volume and link travel time.

All current user and system equilibrium assignment formulations require that the relationships of link travel time and link flow be known *a priori* and be a function. Most researchers have assumed a BPR (Bureau of Public Roads, 1965) form of the travel time – flow relationship (Equation [2-3]). However, for congested traffic networks, this relationship

often does not accurately reflect the travel times experienced by drivers. This error is particularly evident when considering the problem of estimating dynamic O-D demands.

$$t = t_0 \left(1 + \varphi \left(\frac{V}{C} \right)^{\beta} \right) \tag{2-3}$$

where: t = total travel time (seconds)

 t_0 = free speed travel time (seconds)

 φ = dimensionless parameter

 β = dimensionless parameter

V = link flow (vph)

C = link capacity (vph)

A more recent approach (Yang *et al.*, 1992) integrates generalized least squares with equilibrium assignment. Again, this formulation leads to a bi-level mathematical program. Recognizing that the bi-level programming problem cannot be efficiently solved, the authors propose a heuristic iterative solution similar to that originally proposed by Hall *et al.* (1980). However, there is no guarantee that this iterative heuristic provides a solution that satisfies the constraints of the formulation.

It appears that, though a number of mathematical models, which consider routes to be a function of demand, have been formulated, these models are difficult to solve for practical sized networks. Approximate solution approaches can be used for practical sized networks, however, these solution algorithms no longer guarantee that the constraints in the formulation are met.

2.4.4 Static vs. dynamic approaches

The research reported in the literature, that has been discussed in this thesis thus far, has been confined to methods for estimating static O-D demands. For operation and evaluation of traffic management systems, it is beneficial to have knowledge of time varying demands.

Given the existing diversity of static demand estimation approaches and the lack of a clearly superior method, it is not surprising that there has been much less work conducted investigating the more complex task of estimating time varying demands. A comprehensive summary of existing literature is provided by Cremer (1992).

Much of the work, that has been conducted to-date on estimating time varying demands, has focused on the simpler task of estimating time varying turning movements at intersections (Figure 2-5). This work has included using least squares (Cremer *et al.*, 1987), constrained optimization, recursive parameter estimation (Cremer, 1981 and 1984), and Kalman filtering (Kessaci *et al.*, 1989). Keller and Ploss (1987) extended the concept of utilizing the coefficient of correlation by creating a methodology that combined static and dynamic information. Results were generally favorable for intersections in which there existed a high degree of correlation between entrance and exit flows. However, as expected, when this degree of correlation decreases, accuracy of these methods rapidly deteriorates. Furthermore, some methods experience problems with computational inefficiencies, or lack the guarantee that formulated constraints are satisfied by the solution.

Procedures for estimating time varying demands for linear systems have also been proposed (Chang *et al.*, 1992). This method is based on using the extended Kalman filter, however, the solution algorithm does not ensure that formulated constraints are met.

Janson and Southworth (1992) have utilized a maximum entropy formulation to determine the distribution of departure times for a known static matrix. Bell (1991b) has proposed a method of estimating time varying demands in the presence of platoon dispersion. This method utilizes constrained recursive least squares but the solution algorithm does not ensure that estimates satisfy formulated constraints.

One of the only practical dynamic estimation methods proposed to-date that is applicable to networks is that of Willumsen (1984). The entropy maximization formulation requires that a feasible solution exists and that the link use probabilities be exogeneously defined. However, due to the additional time considerations, the link use probability matrix becomes five dimensional. For even moderate sized networks, computer memory requirements become excessive. To avoid this limitation, Willumsen devises a heuristic solution approach that utilizes the CONTRAM route choice and simulation model to provide the link use proportions.

Recently, Cascetta *et al.* (1993) has proposed a generalized least squares dynamic O-D estimator. This model was tested on a linear section of freeway in Italy, having 19 zones, 19 nodes, and 54 links. The actual demand was known, so comparisons between the estimated demand and the true demand was possible. Observed flows were, however, generated from the known demands, rather than observed on the links, eliminating the possibility of inconsistencies in the observed flows. Furthermore, as the network was linear, the routes were known exactly. The authors report that "results were generally satisfactory", however, the known true demands are not provided, so it is not possible to replicate their results, nor compare the performance of other methodologies against their results.

This lack of quantitative details is common in the research reported on in the literature. Many approaches are formulated, but not illustrated, even for a very simple network. Furthermore, for those approaches that are tested on a network, seldom is sufficient information given to permit other researchers to replicate the reported results, or to apply alternative methodologies to the same network in order to provide direct comparison of performance. It is largely for this reason that few quantitative comparisons of models' performances appear in this literature review. It is also likely for this reason that the objective comparison of the performance of more of the various proposed models has not been previously undertaken.

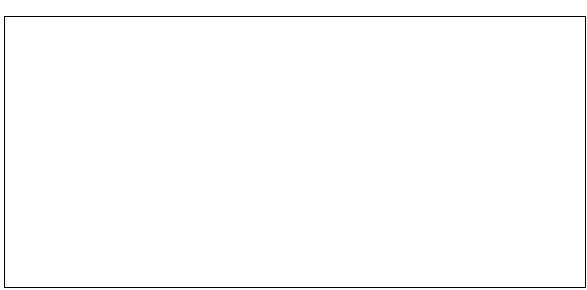


Figure 2-5: Estimating time varying turning movements at intersections (Source: Cremer, 1992)

2.5 Additional Issues

In addition to the four classifications discussed in Section 2.4, three additional issues, that arise in the estimation of O-D demands, need be examined. The extent to which existing models can approach each of these three issues is examined next.

2.5.1 Under- and over- specification

In theory, the problem of estimating O-D demands from link traffic counts is often underspecified, resulting in multiple solutions that satisfy the observed link flow constraints. Under these conditions, the problem of estimating O-D demands consists of two elements; defining a set of feasible solutions, and selecting a single solution from the multiple solutions that exist.

a. Multiple solutions exist that exactly replicate the observed link flows

One common method relies on the introduction of a convex likelihood function which is then minimized. This approach has been adopted by a large number of authors but was most prominently introduced by Van Zuylen and Willumsen (1980). Specifically, Van Zuylen and Willumsen proposed two different, but related, O-D estimation formulations. Each formulation was based on maximizing some measure of the likelihood of an O-D demand matrix. One of the formulations was derived under the assumption that the unit of measure was a trip, while the other assumed a link-count as the unit of measure. Since these two models have made a significant impact on the current state-of-the-art of O-D estimation, and because these two methods can be used as a starting point for the development of the two models proposed in this thesis, the models developed by Van Zuylen and Willumsen will be examined more closely in Chapter 3.

The definition of a likelihood function relies on a mathematical statement of what outcomes are most probable. However, reflecting actual driver behaviour trends within these statements is often difficult. As the use of additional high quality information usually improves the O-D estimate, researchers have incorporated knowledge of these patterns into the estimation process. Robillard (1975) and others (Högberg, 1979; and Irving *et al.*, 1986) have imposed a structure to the estimated O-D by introducing a distribution model. Still others have utilized additional information in order to select an appropriate matrix, such as a trip length distribution (Chan, 1986), license plate survey data (Geva, 1983), cordon counts (Landau, 1982), or a prior matrix (Van Zuylen, 1981). Usually, the use of additional information has improved the accuracy of the estimated demand. Unfortunately, the introduction of additional information requires that, for these methods to be of practical use, these data be readily available. Differences between the solutions generated by different likelihood functions will be examined in Chapter 3 of this thesis.

b. No solution exists that exactly replicates the observed link flows

In practice, due to the temporal nature of traffic flow and measurement error, observed link flows are rarely consistent. Thus, there usually is no feasible O-D solution that can exactly replicate the observed link flows. Methods that require at least one feasible solution either assume consistent flows are available, or preprocess the observed link flows to create consistent flows for which multiple solutions may exist (Van Zuylen and Branston, 1982). Others, recognizing that often no one exact solution exists, have applied constrained and unconstrained generalized least squares regression (Bell, 1991a; Cascetta, 1984; Hendrickson *et al.*, 1984a and 1984b), unconstrained optimization (Mekky, 1979), and log-linear models (Bell, 1984a).

Unfortunately, some of these approaches cannot guarantee that the estimated O-D demands will be non-negative, as the physical interpretation of the decision variables requires. Others ensure that the solution is positive by truncation and projection (Hendrickson and McNeil, 1984a and 1984b), however, the impact on the objective function value of such intervention is not known. Though a feasible solution is achieved, it can no longer be stated with certainty that the solution is optimal, nor is it known how close the proposed solution is to the optimal solution. Furthermore, some of the proposed methods are, by their authors' own admission, inappropriate for practical use, as the solution algorithm requires the manipulation of very large matrices (Bell, 1991a).

Still other authors have explicitly treated link counts as random variables, usually assuming that they have a Poisson distribution in order to take advantage of the characteristic that the mean and variance are equal (Mauer, 1983; Davis and Nihan, 1991). Jörnsten and Wallace (1993) propose a stochastic programming approach that explicitly recognizes that the inconsistencies within the observed link flow data result from some unknown underlying distribution of traffic flows. Unfortunately, when examining a time series of traffic flows, the flow variance within each time period is likely to be a function of the level of demand, and the traffic conditions on the network. One clearly expects to observe greater variance in flow at high levels of flow, than at low levels of flow. However, under congested conditions, observed variations in flow may be due solely to the movement of shockwaves, and not reflect in any way, the magnitude, or reliability, of the true demands. An alternative approach for dealing with link flow inconsistencies will be presented in Section 4.3.3.

Despite the significant effort and the variety of approaches proposed, no one method has been shown to be consistently significantly superior. However, a review of the literature indicates that the use of maximum likelihood within model formulations is rather prevalent. As such, it would seem prudent to examine likelihood more closely as it pertains to O-D estimation. The maximum entropy and minimum information likelihood formulations, which were introduced by Van Zuylen and Willumsen, are therefore examined in more detail in Chapter 3.

2.5.2 Accuracy of estimated O-D demands

One difficulty that plagues all O-D estimation techniques is that of quantifying the accuracy of the estimated O-D demand. This difficulty arises because the true O-D is, by definition, usually unknown.

Empirical studies have been carried out in which the estimated accuracies of various O-D estimation techniques are compared using either synthetic O-D data or O-D data gathered by means of a survey (Han, 1983). The reported accuracies vary significantly, depending on the network size, traffic conditions, and information made available to the estimation technique. Unfortunately, the results of these empirical studies cannot be generalized to be applicable to other networks, or even to different traffic conditions on the same networks.

Bell (1983a) has investigated the use of the variance in observed link counts in order to determine confidence limits about the estimated static O-D. However, the method assumes that the probability that a trip is between a specific O-D pair, given that a trip occurs, is known. It is further assumed that link use probabilities are known exactly so that any ensuing link flow error can be solely attributed to inaccuracies in the estimated demand. Neither of these assumptions are likely to be met in practice. A second method proposed by Bell (1983b) approximates the variance of static O-D estimates on the basis of the mean and variance of link flows and by assuming that demands are log-normally distributed. It is not clear that either of these two approaches can be generalized to the dynamic problem.

More recently, Bell (1985) has proposed that confidence intervals about the estimated O-D can be computed assuming that demand follows a log-normal distribution and that the variance and covariance of the prior matrix and link flows are known. Again, in practice, these data are not likely to be known. In addition, concern has been recently raised as to the appropriate manner in which variance - covariance matrices should be computed and how they should be interpreted (Jörnsten and Wallace, 1994). In light of the temporal and spatial variations in traffic flows, it is not clear what the variance - covariance matrices ultimately mean.

Another approach estimates a quantity termed the *maximum possible relative error* (Yang *et al.*, 1991). This quantity is a measure of the maximum distance from the chosen most likely solution to some other solution within the feasible region (Figure 2-6). This approach assumes that link flows and route choice proportions are known exactly such that any error is due only to choosing the incorrect solution from those within the feasible region.

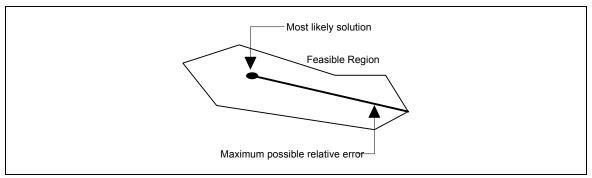


Figure 2-6: Graphical representation of maximum possible relative O-D error

Despite these contributions, there remains no means of determining, in practice, the accuracy of an estimated O-D demand. This lack of ability to quantify the accuracy of the estimated O-D is only important if knowing an accurate O-D is important. Stokes and Morris (1986) examined the effects of O-D accuracy on several network measures of performance (MOEs). The FREQ traffic simulation model was used to simulate a freeway section for which morning peak 15 minute O-D demands were first obtained through a survey and then through synthetic estimation. The study concluded that total travel time, total travel distance, and average network speed, were all insensitive to the distribution of the O-D demands. However, in this study, only network aggregate measures of effectiveness were compared. Also, since the total ramp flows were the same for both the observed demands and the synthetic demands, and since in a linear system there are no re-routing opportunities, the only impact that the distribution of demands has on the network is where congestion occurs. The effects of demand distribution on network aggregate measures of effectiveness will then be minimal. Thus, the results from Stokes et al., should not be generalized to be true for all networks. Clearly, for non-linear networks, drivers' re-routing opportunities depend upon their desired destinations.

It would appear then, that further research is required to examine in more detail the impacts of inaccurate O-D demands on the operation and evaluation of traffic management systems. Such an analysis is not performed in this thesis.

2.5.3 Use of *a priori* information

Many proposed approaches require the use of *a priori* information in the form of a seed matrix in order to solve the multiple solution problem. It has been clearly demonstrated that for these solution methods, the structure of the seed matrix has a significant impact on the nature and accuracy of the estimated matrix (Lam and Lo, 1991). Unfortunately, few guidelines exist to help practitioners select an appropriate seed matrix.

Bell (1984b) has proposed a method for obtaining a seed matrix and quantifying its variance, but this method is only applicable for turning movements at intersections. Spiess (1987) has proposed a method that treats the elements of the seed matrix as observations from independent Poisson distributions with unknown means. Assuming that link flows are consistent and known exactly, Spiess presents a method that utilizes the maximum likelihood

concept to determine the unknown means of the cells within the seed matrix. It can then be statistically determined if the estimated matrix satisfies the observed link flows.

Van Zuylen and Willumsen (1980) have proposed two maximum likelihood models that both rely on the use of a seed matrix. The formulations use the seed matrix as an initial target, but, as noted by other researches (Bell, 1984a; Jörnsten, 1994; Maher, 1983), maximal belief is placed in the observed flows and minimal belief is placed in the seed matrix.

It would appear from the literature, that provision for the use of a seed matrix is desirable, as it permits the updating of an out-dated matrix, or the incorporation of information from a partial matrix. However, there is a need to permit the specification of the reliability of each cell within the seed matrix when compared with other cells and when compared with the observed link flows, and to explicitly accommodate these relative levels of reliability within the O-D estimation process.

2.6 Summary

The previous examination of the existing body of literature has shown that many different approaches have been proposed for the estimation of static O-D demands. From this range of proposed methods it can also be concluded that no one single method is clearly superior to all others. It would appear, however, that despite some undesirable characteristics, the use of a maximum likelihood formulation is not only common to many approaches, but it is also preferred by many.

Significant concerns have been raised over the lack of control which practitioners have over specifying the impact of the accuracy of seed matrices, particularly in view of how significantly the seed matrix affects the final estimated O-D demand.

Little work has also been conducted investigating the impacts of utilizing inaccurate O-D demands for network control and/or evaluation. Some reported research indicates that MOEs are insensitive to changes in O-D in a linear corridor, however, these results cannot likely be generalized to a complete two dimensional network.

It is clear, that any successful O-D estimation model must have a sound mathematical basis, such that the objective of the model, and the method used to achieve this objective, can be clearly and concisely stated. Furthermore, the assumptions made within the mathematical formulation must be reasonable with respect to known field conditions. There is little value in formulating a model on the basis of an assumption, that must be subsequently violated in order to apply the model in practice.

Many of the proposed static and dynamic O-D estimation methodologies presented in the literature are mathematically formulated but are often only solved using algorithms that provide approximate solutions. The magnitude of the error that is introduced by these approximations is seldom known, estimated or even considered. In some cases, it is not clear if the benefits being sought by these complex mathematical formulations may not be negated by the approximations made in the solution algorithms.

CHAPTER 3

ANALYSIS AND COMPARISON OF TWO MAXIMUM LIKELIHOOD APPROACHES

3.1 Introduction

The literature review in Chapter 2 has shown that numerous O-D estimation approaches have been proposed. Despite the number of proposed methods, no single method has proved to be conclusively or consistently superior to all others. However, an element common to many of the proposed estimation methods, is the use of a maximum likelihood objective function in order to select a preferred solution when many solutions are feasible. Two significant maximum likelihood models, one which is trip based and the other which is vehicle-count based, will be examined in this chapter.

3.1.1 Background

Two of the best known O-D demand estimation models are the maximum likelihood models originally developed by Van Zuylen and Willumsen (1980). Their derivation is examined in detail here for several reasons.

First, these two models are considered by many to be the standard against which most other approaches are compared. In order to understand the important characteristics of these two models, the various explicit and implicit assumptions that are made during the models' derivations must first be identified and understood.

Second, the two synthetic O-D estimation methods proposed in Chapters 4 and 5 will be compared with these two approaches at both a theoretical and performance level. To carry out a theoretical comparison, the theory on which these two maximum likelihood methods are based must therefore also be fully understood.

Third, the literature describing the models proposed by Van Zuylen and Willumsen provide only a brief description of the derivation of the models. While some mathematicians might find this level of detail adequate, it is likely that most engineers would need to invest a significant amount of time and effort before the practical impacts of the theoretical bases, assumptions, and approximations could be fully appreciated.

Therefore, the ultimate merit in carrying out a complete derivation is twofold. A complete derivation will show conclusively the mathematical consistency (or inconsistency) of the formulated model, and second, all assumptions and approximations within the derivation can be more clearly identified and subsequently challenged.

3.1.2 Structure of chapter

Two separate, but similar models are described in this chapter. The first model, derived using information theory, is based on the premise that vehicle counts on links are the fundamental unit of observation. The second model, derived from entropy considerations, is based on the premise that trip counts are the fundamental unit of observation. That one model is developed based on information theory, while the other is based on entropy theory, is not critical, since both theories are intimately related. The prime differences in the models arise from the differences in the observation unit adopted — a trip in the entropy model and a vehicle count on a link in the information model, and differences in their solutions arise when networks exhibit biases that affect one measurement unit differently from the other.

A general form of the information minimization model is developed first in Section 3.2, and then adapted in Section 3.3 to the specific task of estimating O-D demands from vehicle counts on links. The complete mathematical derivation and formulation of the objective function and constraints is then provided in Section 3.3.1. The characteristics of the formulation are also illustrated through the use of a simple example network in Section 3.3.2. A similar development of the entropy maximization trip-count based model is subsequently presented in Section 3.4. Again, the mathematical formulation is provided, and illustrated using a simple example network.

A comparison is made between the entropy maximization and information minimization approaches. Theoretical differences are discussed as well as the practical implications of these differences. These implications are illustrated for a variety of scenarios using a simple example network. Finally, conclusions are made regarding the theoretical and practical strengths and weaknesses of the models, as well as the suitability of the two approaches for estimating O-D demands.

3.2 Generalized Model

3.2.1 Introduction

Approximately four decades ago, a new theory of information, having much in common with statistical thermodynamics, was developed. Both information theory and thermodynamics

have been applied primarily to the fields of chemistry and physics, however, these concepts have also been found to be useful for the estimation of O-D demands in a traffic network context.

Without presenting a lengthy discussion on the theoretical and philosophical aspects of information and entropy, it is sufficient to state that information is equal to negative entropy (Brillouin, 1964). Of course, this definition is only useful if entropy is already defined. Entropy, in the context of thermodynamics, is a measure of the quality of energy in a system. Entropy increases as the quality of energy decreases. For the purposes of this research, it is more appropriate to view entropy as a measure of the likelihood of the system's state. This definition is consistent with the formal thermodynamic definition, as poor quality energy sources (i.e. heat) are more likely than high quality sources (i.e. electrical energy).

Starting from this conceptual framework, one can then proceed to develop a generalized mathematical formulation based on information theory. In the next section this general formulation is adopted for the express purpose of estimating O-D demands.

3.2.2 **Mathematical derivation**

Brillouin (1962) defines the quantity of information that is contained within a sequence of entity states as being

$$I = K \ln G \tag{3-1}$$

where: I = the quantity of information

K = a constant

G = the number of equally probable outcomes

In information theory, the constant K, is used to transform the dimensionless information quantity into a quantity having some units. However, as for optimization purposes, the objective function which contains the expression of information is maximized, constants within the objective function do not impact the optimum. Therefore, it is sufficient to define information as

$$I = \ln G \tag{3-2}$$

In order to quantify the information content in a problem, it is necessary to derive an expression for G, the number of equally possible outcomes.

Brillouin (1962) indicates that an expression for G can be derived on the basis of the number of ways in which each possible state outcome can occur. To illustrate this approach, consider the simple example illustrated in Figure 3-1 in which only two states exist; represented by A and B. Consider a total of N observations of which n_A are state A and n_B are state B. Then, as expressed in Equation [3-3a], the number of ways that these n_A state A observations can occur is just the number of combinations of N items taken n_A at a time.

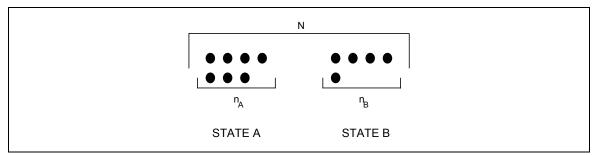


Figure 3-1: Illustration of a simple example having two possible outcomes

Equation [3-3a] is only valid if each outcome is equally likely. To relax this constraint, it is desirable to explicitly incorporate the probability of each outcome into this definition of G. Using the current example, let q_A and q_B be the probability that state A and state B respectively occur. Then, the probability of observing exactly one state A outcome is, by definition, q_A . Assuming that events are independent, the probability of observing two state A outcomes is the product of their respective probabilities $(q_A \cdot q_A)$. It follows then, that the probability of observing n_A state A outcomes must be equal to $q_A^{n_A}$.

The expression for G in Equation [3-3a] becomes the number of ways in which the outcome could occur, weighted by the probability of each outcome occurring (Equation [3-3b]), which can be algebraically simplified to produce Equation [3-3c].

$$G = C_{n_A}^N = \frac{N!}{(N - n_A)! n_A!} = \frac{N!}{n_B! n_A!}$$
 [3-3a]

$$G' = \frac{N!(q_A^{n_A})}{n_A!(N - n_A)!} \times \frac{(N - n_A)!(q_B^{n_B})}{n_B!(N - n_A - n_B)!}$$
 [3-3b]

$$G' = \frac{N!q_A^{n_A}q_B^{n_B}}{n_A!n_B!}$$
 [3-3c]

where: G = number of equally probable outcomes

G' = number of outcomes weighted by the probability of occurrence

N = total number of observations

 n_A = number of state A outcomes

 n_B = number of state B outcomes

 q_A = probability that state A occurs

 q_B = probability that state B occurs

It is of interest to examine Equations [3-3a] and [3-3c] by way of a simple example in which there are two outcomes and 40 observations. Figure 3-2 illustrates the values of the natural logarithm of G and G' for all possible outcomes when the probability of each outcome occurring is equal $(q_A = q_B = 0.5)$. The two equations provide different estimates of the information contained in each outcome, however, the maximum of each expression

occurs for the same outcome, namely $n_A = n_B = 20$. This result is consistent with intuition, as the most likely split between two equally likely states is 50% of each state.

Having shown mathematically, and illustrated by way of example, the relationship between Equations [3-3a] and [3-3c], it is useful to illustrate the effect of outcomes that are not equally likely. Consider an example in which there are two outcomes and 40 observations. In the first instance, both outcomes are considered to be equally likely ($q_A = q_B = 0.5$), while in the second instance, state A has a 75% likelihood of occurring ($q_A = 0.75$; $q_B = 0.25$). For each scenario, Figure 3-3 illustrates the measure of information contained in each possible outcome, computed using Equation [3-3c]. When the outcomes are equally likely, the information content is maximized when the number of state A and B outcomes are equal. When state A has a 75% likelihood of occurring, the information content is maximized when 30 state A and 10 state B outcomes are observed. The 30 state A outcomes represent 75% of the total 40 observations, while the 10 state B outcomes represent 25% of the 40 observations. Thus, it has been shown, by way of a simple example, that Equation [3-3c] is appropriate for situations in which outcomes are equally probable and when outcomes are not equally probable.

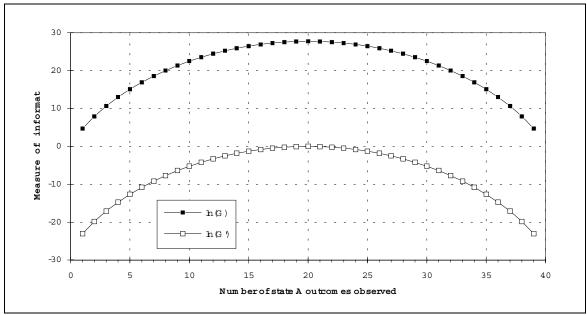


Figure 3-2: Comparison of the information content contained within a possible outcome under the assumption that outcomes are equally likely $(\ln(G))$ and that outcomes are weighted by their likelihood of occurring $(\ln(G'))$

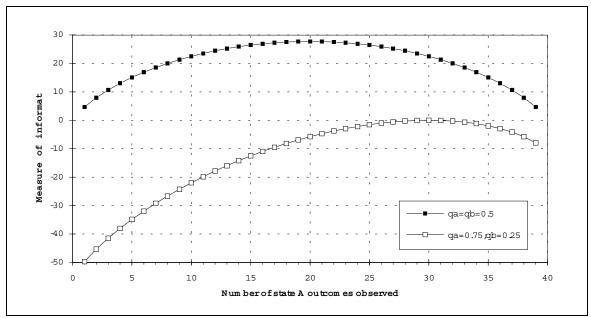


Figure 3-3: Comparison of the information content contained within possible outcomes under the assumption that the two outcomes are equally likely and under the assumption that state *A* has a 75% likelihood of occurring

Equation [3-3c] has been developed under the assumption that only two states exist. This expression can be generalized, as indicated in Equation [3-4a], to describe situations in which Z possible states exist. For each term in Equation [3-4a], the factorial $(N-n_i)!$ in the numerator, can be divided by the same factorial in the denominator of the previous term, to form Equation [3-4b]. The remaining terms can then be simplified to form Equation [3-4c].

$$G' = \frac{N!q_1^{n_1}}{n_1!(N-n_1)!} \times \frac{(N-n_1)!q_2^{n_2}}{n_2!(N-n_1-n_2)!} \cdots \times \frac{\left(N-\sum_{i=1}^{Z-1}n_i\right)q_Z^{n_Z}}{n_Z\left(N-\sum_{i=1}^{Z}n_i\right)}$$
[3-4a]

$$G' = N! \frac{q_1^{n_1}}{n_1!} \times \frac{q_2^{n_2}}{n_2!} \times \frac{q_3^{n_3}}{n_3!} \cdots \times \frac{q_Z^{n_Z}}{n_Z!}$$
 [3-4b]

$$G' = N! \pi \left(\frac{q_k^{n_k}}{n_k!} \right)$$
 [3-4c]

where: G' = number of outcomes weighted by the probability of occurrence

N = total number of observations

 n_1 = number of state 1 outcomes

 n_2 = number of state 2 outcomes

 n_Z = number of state Z outcomes

 q_I = probability that state I occurs

 q_2 = probability that state 2 occurs

 q_Z = probability that state Z occurs

Substituting Equation [3-4c] into Equation [3-2] results in a measure of information as expressed by Equation [3-5a], which can be expanded to produce Equation [3-5b].

$$I = \ln \left[N \frac{\pi \left(q_k^{n_k} \right)}{\pi \left(n_k! \right)} \right]$$
 [3-5a]

$$I = \ln(N!) + \sum_{k} (n_k \ln(q_k)) - \sum_{k} (\ln(n_k!))$$
 [3-5b]

where: N = number of observations

k = a particular state

 n_k = number of times state k outcomes are observed

Since factorials are difficult to compute, even for relatively small numbers, Stirling's formula (Wilson and Kirkby, 1975), expressed in Equation [3-6], may be used as an approximation. Substituting Equation [3-6] into Equation [3-5b] results in Equation [3-7].

$$\ln(X!) \approx X \ln(X) - X \tag{3-6}$$

$$I \approx N \ln N - N + \sum_{k} (n_k \ln q_k) - \sum_{k} (n_k \ln n_k - n_k)$$
 [3-7a]

$$I \approx N \ln N + \sum_{k} (n_k \ln q_k) - \sum_{k} (n_k \ln n_k) + \sum_{k} n_k - N$$
 [3-7b]

where: N = number of observations

k = a particular state

 n_k = number of times state k outcomes are observed

The sum across all states k, of the number of observed outcomes of each state (n_k) , must be equal to the total number of N observations (Equation [3-8]). Substituting Equation [3-8] into Equation [3-7b] provides Equation [3-9], which, after algebraically simplifying, provides Equation [3-10], the generalized form of the measure of information contained in a set of N observations.

In Equation [3-10], it is convenient to have the term n_k in the numerator within the natural logarithm rather than in the denominator. This inversion is possible as ln(x) = -ln(1/x), leading to Equation [3-11], the final generalized expression for information contained within N counts.

$$\sum_{k} n_k = N \tag{3-8}$$

$$I \approx \sum_{k} (n_k) \ln N + \sum_{k} (n_k \ln q_k) - \sum_{k} (n_k \ln n_k)$$
 [3-9]

$$I \approx \sum_{k} \left(n_k \ln \left(\frac{q_k N}{n_k} \right) \right)$$
 [3-10]

$$I \approx -\sum_{k} \left(n_k \ln \left(\frac{n_k}{q_k N} \right) \right)$$
 [3-11]

where: N = number of observations

k = a particular state

 n_k = number of times state k outcomes are observed

3.3 Vehicle Link-Count Based Model

3.3.1 **Model formulation**

The general expression of information, provided in Equation [3-11], must be further specialized for the task of estimating O-D demands. Following the vehicle link-count based derivation presented by Van Zuylen and Willumsen (1980), consider the state k to be a particular origin - destination pair that a vehicle is associated with. Then, the number of vehicles observed on link a, that are traveling between origin i and destination j, can be expressed as the total demand between i and j multiplied by the probability that demand between i and j uses link a (Equation [3-12]).

$$n_{ij}^a = T_{ij} P_{ij}^a \tag{3-12}$$

where: n_{ii}^{a} = number of vehicles counted (outcomes) on link a that are associated with origin i and destination j (state k) (vph)

 T_{ij} = true demand between origin i and destination j (vph) P_{ij}^{a} = probability that demand between i and j will use link a

Without any information to the contrary, one can assume that a vehicle observed on a link is as likely to be traveling between one O-D pair as any other. However, if information regarding the current O-D demand patterns is available, then it is possible to compute this probability directly. Following Van Zuylen and Willumsen (1980), it is assumed that the probability of observing a vehicle, traveling between zones i and j, on link a, can be estimated to be equal to the proportion of the prior link flow associated with a specific O-D.

$$q_{ij}^{a} = \frac{t_{ij} P_{ij}^{a}}{\sum_{ij} \left(t_{ij} P_{ij}^{a} \right)}$$
 [3-13]

 t_{ij} = prior estimate of demand between origin i and destination j (vph) $P_{ij}{}^a$ = probability that demand between i and j will use link a $q_{ij}{}^a$ = prior probability that counts on link a are associated with origin i

If, the denominator of Equation [3-13], is represented by S_a , as defined by Equation [3-14], then, the information contained in V'_a vehicle counts, observed on a single link a, can be determined by substituting Equations [3-12] and [3-13] into Equation [3-11] to produce Equation [3-15]

$$S_a = \sum_{ij} t_{ij} \ P_{ij}^a$$
 [3-14]

$$I_a \approx -\sum_{ij} \left(T_{ij} P_{ij}^a \ln \left(\frac{S_a T_{ij}}{V_a' t_{ij}} \right) \right)$$
 [3-15]

 V'_a = total number of observed vehicle counts on link a (vph) t_{ij} = prior estimate of demand between origin i and destination j (vph) T_{ij} = unknown current demand between origin i and destination j (vph) P_{ij}^a = probability that demand between i and j will use link a

The information contained in the link traffic counts on a number of links, which is expressed by Equation [3-16], is simply the sum of the information contained within the individual link counts.

The objective is to determine values for T_{ij} that will maximize the information content of the estimated demand. Formally, this is accomplished by maximizing the total information subject to the constraint that the estimated O-D demands satisfy the observed link flow constraints. Since the expression for information is a negative quantity, it is convenient to multiply the expression by -1 and then to minimize I, as shown in Equations [3-17] and [3-18]. Equations [3-17] and [3-18] represent the final form of the model formulation originally presented by Van Zuylen and Willumsen (1980).

$$I \approx -\sum_{a} \sum_{ij} \left[T_{ij} P_{ij}^{a} \ln \left(\frac{T_{ij} S_{a}}{t_{ij} V_{a}'} \right) \right]$$
 [3-16]

minimize
$$I \approx \sum_{a} \sum_{ij} \left[T_{ij} P_{ij}^{a} \ln \left(\frac{S_{a} T_{ij}}{V_{a}' t_{ij}} \right) \right]$$
 [3-17]

subject to
$$V_a' = \sum_{ij} T_{ij} P_{ij}^a$$
 [3-18]

 V'_a = total number of observed vehicle counts on link a (vph) t_{ij} = prior estimate of demand between origin i and destination j (vph) T_{ij} = unknown current demand between origin i and destination j (vph) P_{ij}^a = probability that demand between i and j will use link a

3.3.2 Analytical solution for a simple network

A number of approaches may be used to solve the mathematical formulation expressed by Equations [3-17] and [3-18]. First, the constraints expressed in Equation [3-18] may be algebraically incorporated into the objective function, such that the objective function is expressed as a function of a single unknown variable. Then, since the objective function is convex, and the constraints are linear, the first order derivative may be used to find the global minimum or maximum.

A second approach is to form the Lagrangian in order to combine the constraints and the objective function into a single expression. This method avoids having to explicitly express the constraints in terms of a single unknown, and so can be used even when the objective function and constraints are complex expressions. The partial derivatives of the Lagrangian with respect to the unknown quantities and the Lagrangian multipliers provide the global minimum and maximum.

Both of the above two approaches are of limited practical use when dealing with even medium sized traffic networks, as they require a large number of equations to be algebraically solved in order to determine the optimal solution. A third approach that attempts to avoid these problems is the utilization of an iterative algorithm that solves the mathematical model formulation numerically. Ideally, this algorithm will exactly meet all of the mathematically formulated constraints and is guaranteed to converge to a unique solution. However, in practice, the algorithms that are available often do not necessarily meet these idealized goals.

An algebraic solution can be examined in terms of the simple two-link network presented in Figure 3-4. A general formal solution, obtained by forming the Lagrangian, is developed in Section 3.3.3.

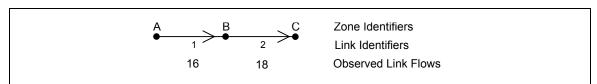


Figure 3-4: Simple example two-link network

For this network, only three O-D demands $(T_{AB}, T_{AC}, \text{ and } T_{BC})$ are feasible. If the unknown demand from zone A to zone C is represented by x, then for this network, the remaining two demands that satisfy the link flow constraints can be explicitly defined in terms of x ($T_{AB} = 16$ -x, $T_{AC} = x$, and $T_{BC} = 18$ -x; $T = \sum T_{ij} = 34$ -x).

To determine the most likely demands, we find those demands that minimize Equation [3-17]. First, we define S_a for both of the links in the network using Equation [3-14].

$$S_1 = t_{AB} + t_{AC}$$
 [3-19a]

$$S_2 = t_{AC} + t_{BC}$$
 [3-19b]

where: t_{ij} = prior demand between origin i and destination j (vph)

Substituting the expressions for t_{ij} and S_a into Equation [3-17] for each link separately results in Equations [3-20a] and [3-20b].

$$I_1 \approx x \ln \left(\frac{x(t_{AC} + t_{AB})}{16t_{AC}} \right) + (16 - x) \ln \left(\frac{(16 - x)(t_{AC} + t_{AB})}{16t_{AB}} \right)$$
 [3-20a]

$$I_2 \approx x \ln \left(\frac{x(t_{AC} + t_{BC})}{18t_{AC}} \right) + (18 - x) \ln \left(\frac{(18 - x)(t_{AC} + t_{BC})}{18t_{BC}} \right)$$
 [3-20b]

To determine the demand that results in the minimum information, it is necessary to differentiate I with respect to x, the only remaining unknown, as t_{ij} , the a priori information, must already be known.

$$\frac{\partial}{\partial x} = 0 = \frac{\partial}{\partial x} = \frac{I_1}{x} + \frac{\partial}{\partial x} = \frac{I_2}{x}$$
 [3-21a]

$$\frac{\partial}{\partial} \frac{I_1}{x} \approx \ln\left(\frac{x(t_{AC} + t_{AB})}{16t_{AC}}\right) - \ln\left(\frac{(16 - x)(t_{AC} + t_{AB})}{16t_{AB}}\right)$$
[3-21b]

$$\frac{\partial}{\partial} \frac{I_2}{x} \approx \ln \left(\frac{x(t_{AC} + t_{BC})}{18t_{AC}} \right) - \ln \left(\frac{(18 - x)(t_{AC} + t_{BC})}{18t_{BC}} \right)$$
 [3-21c]

Substituting Equations [3-21b] and [3-21c] into Equation [3-21a] and simplifying, results in the following quadratic equation.

$$(t_{AC}^2 - t_{AB}t_{BC})x^2 - (34t_{AC}^2)x + 288t_{AC}^2 = 0$$
 [3-22]

Examination of Equation [3-22] reveals that, when no *a priori* information is available (i.e. when $t_{AB} = t_{AC} = t_{BC}$), then the equation simplifies to

$$x \approx 288/34 = 8.47$$
 [3-23]

For this example network, Equation [3-23] holds true whenever the prior matrix is uniform, regardless of the magnitude of the matrix. However, when prior information exists, the prior matrix is not uniform, and the final solution is dependent on the *a priori* information.

For this example then, the demand matrix which minimizes the objective function, as defined in Equation [3-17], is $T_{AB} = 7.53$, $T_{AC} = 8.47$, $T_{BC} = 9.53$, and $T = \sum T_{ij} = 25.53$.

Figure 3-5 illustrates the objective function (Equation [3-17]) for various estimated O-D demands when no prior information is available. As expected, the objective function is convex and its minimum is located at $T_{AC} = 8.47$.

Consider another case in which there is prior information, but this information is in the form of a matrix which is already a feasible solution. It is desirable to determine what the optimal solution is when the prior matrix is already a feasible solution. Figure 3-6 provides the surface reflecting the behaviour of the objective function over the feasible range of the prior matrix and the estimated demand matrix. It can be seen that for any feasible prior

matrix, the minimum objective function value occurs for a matrix that is exactly equal to the prior matrix. This means that any prior matrix that is feasible, is also the most likely. It should be noted that for both the prior matrix and the estimated matrix, the feasible solution space is defined as the set of O-D demands that exactly replicate the observed link flows.

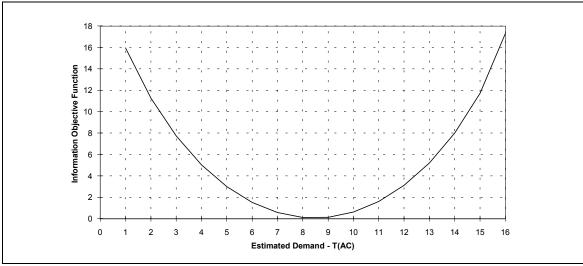


Figure 3-5: Information objective function as a function of estimated demand when no prior information exists

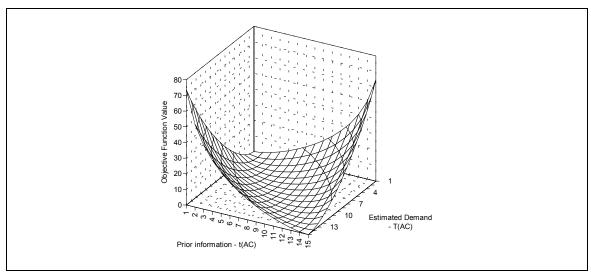
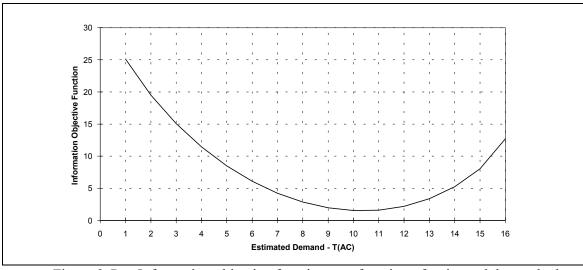


Figure 3-6: Information objective function magnitude as a function of feasible *a priori* matrices and estimated demand



Information objective function as a function of estimated demand when Figure 3-7: prior information exists but this prior demand is not a feasible solution.

When a priori information exists but is not in the form of a feasible solution, then the solution which minimizes the objective function is different from the a priori information and different from the solution that would be optimal if no a priori information were available. Consider again the previous example network. Assume that the prior information indicated that $t_{AC}=10$, $t_{AB}=10$, and $t_{BC}=4$. This demand matrix is not a feasible solution in this case, so it also cannot be the demand which minimizes the objective function. Figure 3-7 illustrates the objective function curve for the range of feasible estimated demands. Analytically, this situation can also be solved using Equation [3-22], resulting in T_{AC} =10.37, T_{AB} =5.63, T_{BC} =7.63 and T= ΣT_{ii} = 23.63.

3.3.3 Formal solution

The solution to the formulated model can be obtained by differentiation of the Lagrangian. The Lagrangian permits the incorporation of the constraints and objective function into a single expression, for which the minimum can be found.

$$L(T_{ij}) = \sum_{a} \sum_{ij} \left[T_{ij} P_{ij}^{a} \ln \left(\frac{S_{a} T_{ij}}{V_{a}' t_{ij}} \right) \right] + \sum_{a} \left[\lambda_{a} \left(\sum_{ij} \left(T_{ij} P_{ij}^{a} \right) - V_{a}' \right) \right]$$
 [3-24]

 V'_a = total number of observed vehicle counts on link a (vph)

 t_{ij}^{a} = prior estimate of demand between origin i and destination j (vph) T_{ij} = unknown current demand between origin i and destination j (vph) P_{ij}^{a} = probability that demand between i and j will use link a

 λ_a = Lagrange multiplier for link a

We can determine the global minimum by taking the partial derivatives of $L(T_{ii})$ with respect to each of the unknown O-D demands (T_{ii}) .

$$\frac{\partial L}{\partial T_{ij}} = \sum_{a} \left[P_{ij}^{a} \ln \left(\frac{S_a T_{ij}}{V_a' t_{ij}} \right) \right] + \sum_{a} \left[T_{ij} P_{ij}^{a} \left(\frac{1}{T_{ij}} \right) \right] + \sum_{a} \left[\lambda_a P_{ij}^{a} \right] = 0$$
 [3-25a]

$$\sum_{a} \left[\ln \left(\frac{S_a T_{ij}}{V_a' t_{ij}} \right)^{P_{ij}^a} \right] = -\sum_{a} \left[P_{ij}^a \left(1 + \lambda_a \right) \right]$$
 [3-25b]

$$\prod_{a} \left(\frac{S_a T_{ij}}{V_a' t_{ij}} \right)^{P_{ij}^a} = e^{-\sum_{a} \left[P_{ij}^a (1 + \lambda_a) \right]}$$
 [3-25c]

$$\prod_{a} \left(\frac{T_{ij}}{t_{ij}} \right)^{P_{ij}^{a}} = \prod_{a} \left(\frac{V_{a}'}{S_{a}} \right)^{P_{ij}^{a}} e^{-\sum_{a} \left[P_{ij}^{a} (1 + \lambda_{a}) \right]}$$
 [3-25d]

$$\left(\frac{T_{ij}}{t_{ij}}\right)_{a}^{\sum P_{ij}^{a}} = \prod_{a} \left(\frac{V_{a}'}{S_{a}}\right)_{ij}^{P_{ij}^{a}} e^{-\sum_{a} \left[P_{ij}^{a}(1+\lambda_{a})\right]}$$
[3-25e]

Setting $g_{ij} = \sum_{a} P_{ij}^{a}$ results in,

$$T_{ij}^{g_{ij}} = t_{ij}^{g_{ij}} \prod_{a} \left(\frac{V_a'}{S_a}\right)^{P_{ij}^a} e^{-\sum_{a} \left[P_{ij}^a (1+\lambda_a)\right]}$$
[3-25f]

but,

$$e^{-\sum_{a} \left[P_{ij}^{a}(1+\lambda_{a})\right]} = \left[e^{-(1+\lambda_{a})}\right]^{-\sum_{a} P_{ij}^{a}}$$
 [3-26]

and, since

$$x^{\sum_{i} y_i} = \prod_{i} x^{y_i}$$
 [3-27]

then,

$$\left[e^{-(1+\lambda_a)}\right]^{\sum P_{ij}^a} = \prod_a \left(e^{-(1+\lambda_a)}\right)^{p_{ij}^a}$$
 [3-28]

Substituting Equation [3-28] into Equation [3-25f] produces,

$$T_{ij} = t_{ij} \prod_{a} \left[\left(\frac{V_a'}{S_a} \right) e^{-(1+\lambda_a)} \right]^{P_{ij}^a/g_{ij}}$$
 [3-29]

If,

$$X_a = \left(\frac{V_a'}{S_a}\right) e^{-(1+\lambda_a)}$$
 [3-30a]

then,

$$T_{ij} = t_{ij} \prod_{a} (X_a)^{\frac{p_a^a}{p_{ij}^a}}$$
 [3-30b]

where: V'_a = total number of observed vehicle counts on link a (vph) t_{ij} = prior estimate of demand between origin i and destination j (vph) T_{ij} = unknown current demand between origin i and destination j (vph) P_{ij}^a = probability that demand between i and j will use link a

 $g_{ij} = \sum_{a} P_{ij}^{a}$

 X_a = defined in Equation [3-30a]

Equation [3-30b] is the same as Equation [21] provided by Van Zuylen and Willumsen (1980).

3.4 **Trip-Count Based Model**

As indicated in Section 3.2, entropy is equivalent to negative information. As such, the mathematical formulation for the estimation of O-D demands based on entropy is very similar to that based on information theory. However, in this section we examine, and present, the development of a trip-count based model originally developed by Van Zuylen and Willumsen (1980) on the basis of entropy considerations.

3.4.1 Model formulation

Since entropy is a measure of negative information, we may begin the development of the trip based model from Equation [3-11] which is reproduced below.

$$W = -I \approx \sum_{k} \left(n_k \ln \left(\frac{n_k}{q_k N} \right) \right)$$
 [3-31]

where: W = measure of entropy

I = measure of information

N = number of observations

k = a particular state

 n_k = number of times state k outcomes are observed

 q_k = the probability that state k occurs

For this model, we consider state k to be the origin - destination pair that a trip (not a vehicle count) is associated with. It follows then, that the number of times that state k outcomes are observed, is equal to the demand between i and j (Equation [3-32]). As indicated in Equation [3-33], the total number of observations is equal to the total number of trips.

$$n_k = T_{ij} ag{3-32}$$

$$T = \sum_{ij} T_{ij} = N = \sum_{k} n_k$$
 [3-33]

where: T = total number of trips

 T_{ij} = Demand between origin *i* and destination *j*

N =number of observations

k = a particular state

 n_k = number of times state k outcomes are observed

The prior probability that a trip is associated with a particular O-D pair can be determined from a prior demand matrix. The probability that a trip is associated with origin i and destination j is equal to the proportion of the total prior demand that was associated with i and j (Equation [3-34]).

$$q_{ij} = \frac{t_{ij}}{\sum_{ij} t_{ij}}$$
 [3-34]

where: q_{ij} = the probability that a trip is between origin i and destination j t_{ij} = prior demand between origin i and destination j

The trip based model is determined by substituting Equations [3-34], [3-33] and [3-32] into Equation [3-31] to produce Equation [3-35].

$$W \approx \sum_{ij} \left(T_{ij} \ln \left(\frac{T_{ij}}{T} \frac{\sum_{ij} t_{ij}}{t_{ij}} \right) \right)$$
 [3-35a]

$$W \approx \sum_{ij} \left(T_{ij} \ln \left(\frac{T_{ij}}{q_{ij}} \right) - T_{ij} \ln T \right)$$
 [3-35b]

$$W \approx \sum_{ij} \left(T_{ij} \ln \left(\frac{T_{ij}}{q_{ij}} \right) \right) - T \ln T$$
 [3-35c]

The objective is to maximize Equation [3-36] subject to the constraint that the estimated O-D demand satisfies the observed link flow constraints.

If the total number of trips in the network (T) is known, then the terms involving T are constant and do not affect the location of the maximum. Van Zuylen and Willumsen (1980) make this assumption with the result that they minimize Equation [3-38].

maximize
$$W \approx T \ln(T) - \sum_{ij} \left(T_{ij} \ln \left(\frac{T_{ij}}{q_{ij}} \right) \right)$$
 [3-36]

subject to
$$V_a = \sum_{ij} T_{ij} P_{ij}^a$$
 [3-37]

$$W \approx \sum_{ij} \left(T_{ij} \ln \left(\frac{T_{ij}}{q_{ij}} \right) \right)$$
 [3-38]

where: W = measure of entropy

T = total number of trips

 T_{ii} = demand between origin i and destination j

 q_{ij} = the probability that a trip is between origin i and destination j P_{ij}^{a} = the probability that demand between i and j will use link a V_a = the observed flow on link a

Van Zuylen and Willumsen do not, however, examine the validity of this assumption, or the potential impacts that it may have on the selected optimum. Their lack of investigation implies that the objective function expressed in Equation [3-38] produces the same solution as Equation [3-36]. This is indeed not true, as will be demonstrated in Section 3.5.

3.4.2 Formal solution

The formal solution to the trip based model is similar to that of the link-count based model presented in Section 3.3.3. The complete derivation of the formal solution is not shown here, but the resulting solution is given by Van Zuylen and Willumsen (1980) as Equations [3-39a] and [3-39b].

$$X_a = \left(\sum_{ij}^{L} t_{ij}\right)^{1/L} e^{-\lambda_a}$$
 [3-39a]

$$T_{ij} = t_{ij} \prod_{a} (X_a)^{P_{ij}^a}$$
 [3-39b]

where: $T_{ij} = \text{demand between origin } i \text{ and destination } j$ $t_{ij} = \text{prior demand between origin } i \text{ and destination } j$ $P_{ij}^{\ a} = \text{the probability that demand between } i \text{ and } j \text{ will use link } a$

 λ_a = Legrange multiplier for link a

A comparison between the formal solution of the trip based entropy model (Equations [3-39a] and [3-39b])and the link-count based information model (Equations [3-30a] and [3-30b]), indicates that the two formal solutions are very similar.

3.5 **Examination of Model Assumptions**

In this section, two of the assumptions made in the development of the formulation of the likelihood models discussed in this chapter, will be examined. The first assumption, that Stirling's approximation is sufficiently accurate, is made in the development of both the vehicle-count and the trip-count based models. The second assumption, that the total number of trips is constant, is only made for the trip based model. Each of these assumptions is discussed in detail below.

3.5.1 Stirling's approximation

Van Zuylen and Willumsen (1980) used the following form of Stirling's approximation in the derivation of their link-count based information minimization model and their trip-count based entropy maximization model.

$$\ln(X!) \approx X \ln(X) - X \tag{3-40}$$

However, a more accurate representation of Stirling's approximation is provided in Equation [3-41] (Feller, 1966).

$$\ln(X!) \approx 0.5 \ln(2\pi) + (X+0.5) \ln(X) - X$$
 [3-41]

A comparison of the accuracy of each of these approximations is provided in Figure 3-8. It is clear from this figure that Equation [3-41] provides a better estimate of ln X! than does Equation [3-40]. However, as the ultimate goal is to find the O-D demand that maximizes the objective function, it is the derivative, not the absolute magnitude of the objective function that is of interest.

The derivative of Equation [3-40] with respect to X is simply $\ln(X)$. The derivative of Equation [3-41] is $\ln(X) + 1/(2X)$. The difference between these two derivatives is the term 1/(2X) which becomes increasingly smaller as the magnitude of X increases. As illustrated by Figure 3-8, when X is larger than 60, the additional approximation error incurred by using Equation [3-40], instead of the more accurate Equation [3-41], is approximately 2%. If the estimated demands between each origin and destination are sufficiently large (i.e., greater than 60 vehicles), then it can be reasonably assumed that the error incurred by using the less accurate representation of Stirling's approximation will not have a significant impact on the resulting solution.

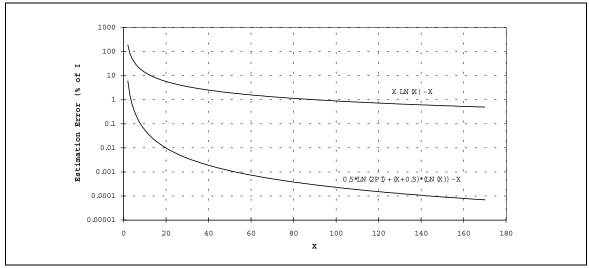


Figure 3-8: Quality of two different versions of Stirling's approximation

3.5.2 **Total number of trips**

The second area of concern is the assumption that the total number of trips is constant, and therefore can be eliminated from the objective function. Consider once again the two-link example network illustrated earlier in Figure 3-4. For this network, the entropy can be computed for each feasible O-D demand. We simplify the illustration by assuming that prior information is not considered so that the entropy can be defined by Equation [3-42a]. To facilitate computation, we take the natural logarithm of W and make use of Stirling's approximation (Equation [3-40]) to obtain Equation [3-42b].

If it is assumed that the total number of trips T, is constant, then T does not impact the determination of the O-D demand that maximizes the objective function, and as such can be omitted from the objective function (Equation [3-42c]).

$$W = \frac{T!}{\pi T_{ij}!}$$
 [3-42a]

$$S = \ln W = T \ln T - T - \sum_{ij} \left(T_{ij} \ln T_{ij} - T_{ij} \right)$$
 [3-42b]

$$S' = -\sum_{ij} (T_{ij} \ln T_{ij} - T_{ij})$$
 [3-42c]

where: W = measure of entropy

T = total number of trips (vph)

 T_{ij} = demand between origin i and destination j (vph) S = log transformation of W S' = log transform of W assuming T is constant and can be omitted

Figure 3-9 illustrates the values of the objective function when the total number of trips is considered variable (Equation [3-42b]) and when the total number of trips is considered to be constant (Equation [3-42c]). It is clear from Figure 3-9 that the two objective functions provide radically different solutions for this example. The feasible solution that maximizes the entropy when T is considered constant is $T_{AB} = 3$, $T_{AC} = 13$, and $T_{BC} = 5$. When T is not held constant a different O-D solution ($T_{AB} = 11$, $T_{AC} = 5$, and $T_{BC} = 13$) is found to maximize the entropy. The occurrence of two different O-D solutions, that are both considered to maximize the entropy, indicates that maximizing Equation [3-42b] is not equivalent to minimizing Equation [3-42c]. The occurrence of this discrepancy is not limited to this specific example. It can be shown that two different entropy maximizing O-D estimates exist for other networks as well, leading to the conclusion that this discrepancy may exist in any network.

This leads to somewhat of a quandary, as the model formulation requires that the total number of trips be known, yet in practice, this quantity is rarely known. It is suspected that many practitioners have utilized this trip based method, without knowledge of the total number of trips, and have accepted the demand estimates as being the *most likely*, when in fact, they were not.

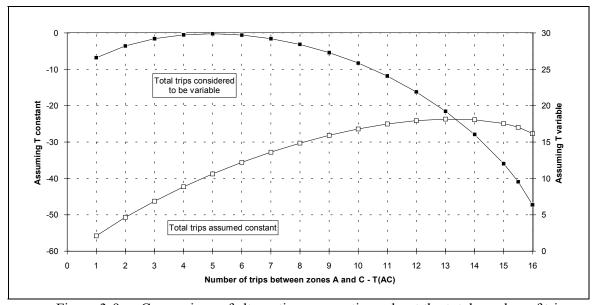


Figure 3-9: Comparison of alternative assumptions about the total number of trips

3.6 Comparison of Methods

The previous sections in this chapter have presented the derivation of a trip based and a link-count based maximum likelihood model. It has been shown that the theoretical bases of these two approaches are very similar, however, some differences do exist between them. These differences will be examined by comparing O-D estimates from the two approaches, for a variety of scenarios. Special attention will be given to assessing whether the solution algorithm is consistent with the formal solution. These comparisons are based on two simple networks.

3.6.1 Example networks

To facilitate comparisons between the two likelihood estimation methods, it is useful to define two simple networks, which may be used to illustrate the characteristics of the models. Figure 3-10 depicts two such simple networks; a three zone, two-link network, and a four zone, four-link network. These two networks were chosen, as they have also been used by Van Zuylen (1981), to illustrate several peculiar characteristics of the trip based and the link-count based models. Furthermore, these networks will be used in Chapters 4 and 5 to

illustrate the characteristics of the models proposed in this thesis and to compare their performance with that of the trip based maximum entropy model and the link-count based minimum information model.

It should be noted that, in general, multiple O-D solutions will exist for both of these networks, as no O-D demand has sole use of any single link. For the comparisons carried out in this chapter, it is assumed that link flows are known exactly. The network structure also precludes the analysis of the impacts of multiple routes, however, these effects will be examined later in this thesis.

The assumed prior matrix (seed matrix) and actual O-D demands and the observed flows for each network are provided in Table 3-1. These values are the same as those presented by Van Zuylen (1981).

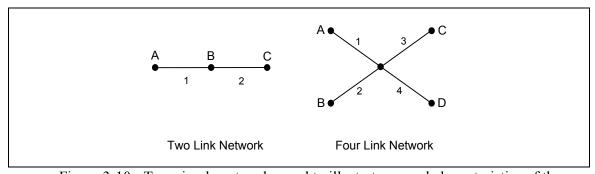


Figure 3-10: Two simple networks used to illustrate several characteristics of the link based entropy and link-count based information model formulations

	Two-Link Network			Four-Link Network			
O-D	A-B	A-C	B-C	A-C	A-D	B-C	B-D
Seed	3	5	4	10	15	8	20
True	6	10	8	15	10	8	20
Link	1	2		1	2	3	4
Obs. Flow	16	18		25	28	23	30

Table 3-1: Characteristics of example networks

3.6.2 Effect of prior information

The solution algorithms for the trip based and link-count based maximum likelihood models proposed by Van Zuylen and Willumsen (1980) require the presence of a prior matrix. This matrix need not contain any information regarding the true O-D demand patterns (i.e. all cells may have equal magnitude), but it must exist. The question of interest is whether the magnitude of this prior matrix has any impact on the estimated O-D solution. Van Zuylen (1981) has shown, using the two-link network illustrated in Figure 3-10, that the trip based model is sensitive to the magnitude of the prior matrix. Table 3-2 provides the demands resulting from both the trip based and link-count based maximum likelihood models for five different scenarios. Each scenario assumes a different seed matrix.

In Scenario 1, the seed matrix is 50% of the true demand. The link-count based model correctly estimates the true demand, however, the trip based model estimate is not equal to the true demand. This same scenario is examined by Van Zuylen (1981), who concluded that trips that are only counted once (i.e. demands A-B and B-C) are underestimated by the trip based model, while trips that are counted twice are overestimated. In fact, according to Van Zuylen, other researchers have also determined that the trip based maximum likelihood model described by Equation [3-39] tends to overestimate trips that are counted more frequently than the average (Smit, 1980; Willis and Chan, 1980).

Scenario 2 examines the impact of the magnitude of a uniform matrix on the most likely O-D matrix estimated by both of the maximum likelihood models. Scenario 2a has a uniform matrix with a magnitude of 10 in each cell. The total number of trips represented by this seed matrix is 30, which is greater than the 24 trips represented by the true demand matrix. The trip based and link-count based models provide different estimates of the *most likely* demand. Though the two solutions are different, it is not clear which of the two formulations actually provides the *most likely* demand from a practical point of view. Some insight is provided by the results from Scenarios 2b and 2c. The demand that is proposed by the link-count based model as the most likely demand does not vary with changes in the total number of trips represented by the uniform seed matrix. However, the results from the trip based model indicate that it is sensitive to the magnitude of the seed matrix. It is interesting to note, that for Scenario 2b, in which the seed matrix contains the same number of trips as the true matrix (24), the trip based model produces results that closely replicate those provided by the link-count based model. For Scenario 2c, the total number of trips in the seed is less than the number of trips in the true demand matrix. The trip based model demand estimate is again different from the link-count based model estimate.

The results from Scenarios 2a, 2b, and 2c indicate, that when the number of trips within the seed matrix is less than the number of trips within the true matrix, the trip based model underestimated short trips (O-D demand A-B and B-C) and overestimated long trips (A-C). When the number of trips in the seed matrix exceed the actual number of trips, the trip based model overestimated short trips and underestimated long trips.

The results from Scenario 2c may also be examined with respect to the trip based model formulation expressed earlier in Equation [3-38], which indicates that the entropy measure, W, is a function of only T_{ij} and q_{ij} . Prior information is incorporated into the term q_{ij} , which is the ratio of the prior demand between i and j to the total number of trips represented by the entire prior matrix. Since q_{ij} is a ratio, only the magnitude of t_{ij} relative to Σt_{ij} has an impact on q_{ij} , implying that Equation [3-38] is insensitive to the total number of trips represented by the seed matrix. This insensitivity of Equation [3-38] to the number of trips in the seed matrix is inconsistent with the properties of the algorithm used to solve Equation [3-38], as evident by the results in Table 3-2.

Van Zuylen (1981) reports that though Murchland (1977) has shown that the multiproportional problem has a unique solution, no mathematical proof is available that the solution algorithm used by Van Zuylen to solve Equation [3-38] ultimately converges.

Scenario 3 indicates that both models provide similar results for a non-uniform seed, which is not a multiple of the true demand, but has the same total number of trips as the true

demand. Since the seed matrix in Scenario 3 does not satisfy the link flow constraints, and since it is also not a multiple of the true demand, the solutions provided by both models are not equal to the true demand. This is to be expected, as the seed matrix provides an initial starting point for the search for a feasible solution.

Van Zuylen (1981), having recognized the inappropriate sensitivity of the trip based model to the number of trips in the seed matrix, proposed a modification to the original solution algorithm. This modification consists of incorporating a scaling factor based on the ratio of the total number of trips in the true matrix to the total in the seed matrix. Of course, in practice, the true matrix is not known. As a surrogate measure, Van Zuylen relies on the ratio of all of the observed link flows, to the total link flows resulting from the seed matrix. This modified trip based entropy model is stated to be insensitive to the multiplication of the seed matrix by a constant factor α . Van Zuylen states that the convergence of the modified solution algorithm is "reasonably good", however, no quantitative measures of performance are provided.

Table 3-2: Maximum likelihood O-D estimates for a two-link examp	ple network
--	-------------

Scenario	S	eed D	emar	nd	Estimated Demand							
				Total		Trip Based			Link-count based			Total
	A-B	A-C	B-C	Trips	A-B	A-C	B-C	Trips	A-B	A-C	B-C	Trips
1	3	5	4	12	4.38	11.62	6.38	22.38	6.00	10.0	8.00	24.00
2a	10	10	10	30	8.00	8.00	10.0	26.00	7.53	8.47	9.53	25.53
2b	8	8	8	24	7.37	8.63	9.37	25.37	7.53	8.47	9.53	25.53
2c	1	1	1	3	2.77	13.23	4.77	20.84	7.53	8.47	9.53	25.53
3	10	10	4	24	5.54	10.46	7.54	23.54	5.63	10.37	7.63	23.63

3.6.3 Effect of redundant link flows

A second concern, raised by Van Zuylen, is the effect of redundant link flows on the estimated O-D demand. Table 3-3 presents O-D demands estimated by the trip based and link-count based models, for a number of different scenarios. The results indicate that the link-count model is sensitive to redundant link flow information. For the four-link network, one of the four observed link flows is redundant, as it does not provide any additional information. Since flow continuity exists at the node, flow into the node must equal flow out of the node. Thus, if flows on three of the four links are known, the remaining flow can be computed. The trip based model is not affected by the presence of redundant information.

Van Zuylen indicates that this characteristic feature of the link-count based model is illogical, and undesirable. The link-count based model formulation assumes that all observed flows provide additional information. Van Zuylen (1981) presents a modified form of this formulation that he claims is insensitive to the introduction of redundant information. However, this characteristic is not illustrated by way of an example. It appears that this insensitivity is only achieved when the redundant flows are explicitly identified, and the quantity X_{ab} from Equation [3-42], is set to unity for all redundant links.

Aside from the uncertainty about the validity of both the manner in which this insensitivity is achieved, and the claim itself, a more fundamental question exists. In the absence of link flow continuity, are any link flows redundant? If link flow continuity does

not exist, then it is not possible to satisfy the continuity equation at each node, and it is also not possible to determine the flow on the n^{th} link, if the flows on links 1 through (n-1) are known. In this sense, there cannot be redundant flows. Furthermore, flow discontinuity generally results from the presence of natural temporal and spatial variations in link flows. One link flow observation represents a sample from some unknown distribution of link flows. It would be expected that as the number of observations increased, the accuracy in the estimate of the true mean link flow would also increase.

It would appear, that for the estimation of O-D demands based on field data, the problem of link flow redundancy is a non-issue. In practice, each observed link flow can be considered to provide additional information, even though some links may only be providing additional samples of a mean link flow.

In Chapters 4 and 5 the performance of the proposed LSE and LRE models will be examined for scenarios in which redundant link flow information exists, and will be compared with that of the trip based and link-count based maximum likelihood models presented in this chapter.

Table 3-3: Maximum likelihood O-D estimates for a four-link network with redundant link flows (Source: Van Zuylen, 1981)

O-D	Trip	Link-count based							
	Based	Links having observed flows							
		all	1,2,3	1,2,4	1,3,4	2,3,4			
A-C	12.5	12.5	12.60	12.39	11.64	13.36			
A-D	12.5	12.5	12.40	12.61	13.36	11.64			
B-C	10.5	10.5	10.40	10.61	11.36	9.64			
B-D	17.5	17.5	17.60	17.39	16.64	18.36			

3.7 Summary

In Chapter 2, a comprehensive review of the literature indicated that many approaches to the estimation of network O-D demands, have been proposed. No one method has been shown to be clearly superior to all others, however, the trip based and link-count based maximum likelihood methods proposed by Van Zuylen and Willumsen, have made a significant impact on the development of subsequent models. Furthermore, these models have been generally seen as the benchmark against which static estimation methods are compared.

In this chapter, the formulation of both the vehicle-count based and trip-count based models have been examined in detail. It has been shown that the link-count based model has several desirable characteristics. First, the model is insensitive to the number of trips represented by the seed demand. Second, when a feasible prior demand matrix exists, the objective function is minimized when the estimated demand is equal to the prior matrix, implying that any prior matrix, that satisfies the link flow constraints, is considered to be the most likely demand matrix. Third, when the prior matrix is not a feasible solution, the solution algorithm iteratively modifies the prior matrix, such that a feasible solution is found. Lastly, the model is based on a sound formulation which can be described mathematically.

The link-count based model is not, however, without weaknesses. Firstly, the model formulation requires that node and path flow continuity exist in the observed link flows. In practice this situation is rarely encountered. The implications, for the estimated solution, of pre-processing link flow data to ensure continuity are not well defined or understood. Second, the impact of the seed matrix is not yet well understood. The model formulation places minimal belief in the prior matrix, and maximum belief in the observed link flows. There is no opportunity to explicitly place a measure of reliability on the prior matrix, or on some portion of it. Given the current testing and probable future implementation of route guidance systems (RGS), direct sampling of O-D demands, via communication with RGSequipped vehicles, will be possible. These demands will represent but a sample of the population demand patterns. The use of statistical procedures, as developed in Chapter 8 of this thesis, permit the estimation of the reliability of these O-D demands. It is expected that these demands would serve as prior knowledge to the O-D estimation procedure, which must therefore, have the ability to consider explicitly the reliability of each prior O-D estimate. Third, there is no opportunity to incorporate explicitly the reliability of observed link flows. This is particularly desirable when link flow data are obtained from sources having different levels of reliability associated with them.

The trip based model, being very similar to the link-count based model, shares many of the same characteristics of the link-count based model, however, the trip based model solution algorithm provides estimates that are not consistent with the model's mathematical formulation. The solution algorithm is sensitive to the number of trips contained within the prior matrix, but the mathematical formulation is not. A modified form of the trip based model is claimed to be insensitive to the magnitude of the prior matrix, however, there is an associated penalty in the rate at which the iterative solution algorithm converges.

Several significant strengths and weaknesses of the trip based and link-count based maximum likelihood models have been identified. Based on these characteristics, two new O-D estimation models are proposed. Chapter 4 presents the development of the mathematical formulation and subsequent iterative solution algorithm for a link-count based least squared error (LSE) model. Chapter 5 presents a similar development of a link-count based model that minimizes the relative error (LRE). The performance of both of these models is compared with the performance of the maximum likelihood models presented in this chapter. In Chapter 6, the LRE and LSE models are extended to facilitate the estimation of time varying demands.

CHAPTER 4

DEVELOPMENT OF A LEAST SQUARES O-D ESTIMATION MODEL

4.1 Introduction

The least squared error (LSE) model described in this chapter is based on the premise that the estimated O-D demand should minimize the squared link flow differences. This formulation does not require observed link flows to exhibit either node or path flow continuity. An iterative solution algorithm, which is based on the Gauss-Seidel and Jacobi solution techniques, is proposed to solve the model formulation.

The mathematical formulation can be derived from the basic assumption that the estimated O-D demand matrix must minimize the squared link flow error. The algorithm used to solve the mathematical formulation ensures that, when multiple solutions exist, and no prior information is available, the solution chosen approximates the solution that is most likely, according to the definition given in Chapter 3.

In this chapter, the LSE objective function is derived. Since the least-squared objective function is convex and differentiable, it is possible to form the normal equations by computing the partial derivatives of the objective function with respect to the unknown demands. The O-D solution, or solutions, that satisfy this system of linear equations, then also minimize the objective function. Subsequently, in Section 4.5, an iterative solution technique is developed for solving the system of normal equations. This technique deals with the issue of singularity, as well as non-negativity of the estimated demands.

4.2 **Development of the LSE Model**

4.2.1 Mathematical basis

In many instances, a least biased estimator is considered to be the one that minimizes the squared errors. This is the basis of least squares regression and is also chosen as the basis for the objective function of the O-D demand estimation model described in this chapter.

The mathematical derivation begins with the definition of the link flow error function as the sum of the squared difference between estimated and observed link flows (Equation [4-1]).

The objective is to identify the set of solutions that minimize E. Since E is a convex function, it is possible to create a system of normal equations by computing the partial derivative of E with respect to each unknown O-D demand (T_{ij}) in turn, and setting each resulting equation equal to zero (Equation [4-2]).

Equation [4-3] expresses the estimated link flow V_a , in terms of the as yet unknown demands. Since each O-D can appear in Equation [4-3] no more than once, the partial derivative of V_a with respect to some unknown demand T_{ij} , is simply the proportion of that demand contributing to the flow on link a (Equation [4-4]).

$$E = \sum_{a} (V_a - V_a')^2$$
 [4-1]

$$\frac{\partial}{\partial} \frac{E}{T_{ii}} = \sum_{a} \left[2(V_a - V_a') \frac{\partial V_a}{\partial T_{ij}} \right] = 0, \forall_{ij}$$
 [4-2]

$$V_a = \sum_{ij} \left(T_{ij} P_{ij}^a \right) \tag{4-3}$$

$$\frac{\partial V_a}{\partial T_{ii}} = P_{ij}^a \tag{4-4}$$

where: a = unique link identifier

 V_a = estimated flow on link a (vph)

 V'_a = observed flow on link a (vph) P_{ij}^a = probability that demand between i and j will use link a

 T_{ii} = demand between origin i and destination j (vph)

Substituting Equations [4-4] and [4-3] into Equation [4-2] results in Equation [4-5a] which represents a system of N linear equations with N unknown O-D traffic demands. Note, however, that not all of these N equations are necessarily independent. Thus, the existence of a single solution that satisfies this system of equations is not guaranteed.

A closer inspection of Equation [4-5a] indicates that, when considering some link a which is not part of any route between origin i and destination j ($P_{ii}^a = 0$), there is no contribution to the summation. Therefore, Equation [4-5a] can be simplified to form Equation [4-5b] in which the summation is only performed across those links which are on a path between origin i and destination j.

The least squares model can be formulated as the set of unknown O-D demands that satisfy the system of linear equations expressed in Equation [4-6]. This system of linear constraints does not require that the estimated link flows exactly match those observed, implying that neither node flow nor path flow continuity need exist. This permits the above model to be applied directly to actual networks using raw field data, something which was not possible with the formulations of Van Zuylen and Willumsen (1980) examined earlier without some form of data pre-processing.

$$\frac{\partial}{\partial} \frac{E}{T_{ij}} = \sum_{a} \left[2 \left(\sum_{ij} \left(T_{ij} P_{ij}^{a} \right) - V_{a}' \right) P_{ij}^{a} \right] = 0 \forall_{ij}$$
 [4-5a]

$$\frac{\partial}{\partial} \frac{E}{T_{ij}} = \sum_{r} \left[\left(\sum_{ij} \left(T_{ij} P_{ij}^{r} \right) - V_{r}^{r} \right) P_{ij}^{r} \right] = 0 \forall_{ij}$$
 [4-5b]

$$\sum_{r} \left[\left(\sum_{ij} \left(T_{ij} P_{ij}^{r} \right) - V_{r}' \right) P_{ij}^{r} \right] = 0 \forall_{ij}$$

$$[4-6]$$

where: a = unique link identifier

r = any link on a route between origin i and destination j

 V_a = estimated flow on link a (vph) V'_a = observed flow on link a (vph) P_{ij}^a = probability that demand between i and j will use link a

 T_{ii} = demand between origin *i* and destination *j* (vph)

4.2.2 **Example with link flow continuity**

It is useful to illustrate the effect of the system of normal equations expressed in Equation [4-6] using the simple three-link example network presented in Figure 4-1. The application of Equation [4-6] to this example provides two linear constraints as expressed in Equations [4-7a] and [4-7b].

$$(T_{AB} + T_{AC} - V'_{I}) + (T_{AB} - V'_{2}) = 0$$
 [4-7a]

$$(T_{AB} + T_{AC} - V'_{I}) + (T_{AC} - V'_{3}) = 0$$
 [4-7b]

where: $T_{ij} = \text{demand between origin } i \text{ and destination } j \text{ (vph)}$ $V'_a = \text{observed flow on link } a \text{ (vph)}$

Algebraically solving these two constraints results in the following expressions for T_{AB} and T_{AC} :

$$T_{AB} = T_{AC} - V'_3 + V'_2$$
, and
 $T_{AC} = (2 \cdot V'_3 + V'_1 - V'_2) / 3$

Since the above network provides a uniquely utilized link for each of the two unknown O-D demands, it is a trivial task to determine the correct O-D matrix when node and path flow continuity exist. For example, if the observed link flows are V'_I =50 vph, V'_2 =20 vph, and V'_3 =30 vph, then intuitively it can be seen that the unknown demands must be T_{AB} =20 vph and T_{AC} =30 vph. Algebraically solving for T_{AB} and T_{AC} results in

$$T_{AC} = (2 \cdot 30 + 50 - 20) / 3 = 30$$
 vph, and $T_{AB} = 30 - 30 + 20 = 20$ vph

indicating that for this simple example, the LSE model formulation correctly estimates the unknown demand.

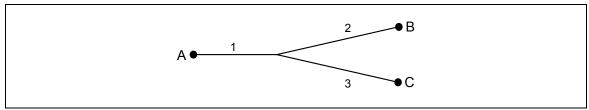


Figure 4-1: Example three-link network

4.2.3 Example with link flow discontinuity

A less trivial task is the determination of the most appropriate O-D matrix when path and node flow continuity do not exist. Once again using the three-link network illustrated in Figure 4-1, consider the observed link flows to be V'_I =50 vph, V'_2 =35 vph, and V'_3 =30 vph. Due to the lack of flow continuity, it is not immediately apparent what O-D matrix should be chosen. Algebraically solving the linear constraints of Equations [4-7a] and [4-7b] results in the following expressions for T_{AB} and T_{AC} :

$$T_{AC} = (2 \cdot 30 + 50 - 35) / 3 = 25 \text{ vph, and}$$

 $T_{AB} = 25 - 30 + 35 = 30 \text{ vph}$

The squared link flow error associated with the suggested solution is computed to be 75 vph². To illustrate that the above suggested solution does indeed minimize the squared link flow error, an enumeration of the link flow error of all potential O-D solutions located near the suggested optimum can be carried out. The results of this enumeration, presented in Table 4-1, indicate that the O-D demand proposed by the LSE formulation does minimize the squared link flow error.

the analytical solution (vpii)											
Demand	Demand between zones A and B										
between		(vph)									
A and C (vph)	25	26	27	28	29	30	31	32	33	34	35
20	225	197	173	153	137	125	117	113	113	117	125
21	197	171	149	131	117	107	101	99	101	107	117
22	173	149	129	113	101	93	89	89	93	101	113
23	153	131	113	99	89	83	81	83	89	99	113
24	137	117	101	89	81	77	77	81	89	101	117
25	125	107	93	83	77	75	77	83	93	107	125
26	117	101	89	81	77	77	81	89	101	117	137
27	113	99	89	83	81	83	89	99	113	131	153
28	113	101	93	89	89	93	101	113	129	149	173
29	117	107	101	99	101	107	117	131	149	171	197
30	125	117	113	113	117	125	137	153	173	197	225
31	137	131	129	131	137	147	161	179	201	227	257

Squared link flow error for all potential O-D solutions in the vicinity of Table 4-1: the analytical solution (vph²)

4.2.4 **Example with multipath routes**

The illustration of the appropriateness of the proposed LSE model formulation has been limited thus far to an example in which all-or-nothing routes have been assumed. It is instructive to demonstrate the appropriateness of the constraints for a simple example in which multipath routes exist. Consider the network illustrated in Figure 4-2 in which there exists two routes between origin zone A and destination zone C.

Application of Equation [4-6] to this example network provides the two linear constraints expressed in Equations [4-8a] and [4-8b].

$$(T_{AB} + T_{AC}P_{AC}^{1} - V_{1}') + (T_{AB} - V_{2}') = 0$$
 [4-8a]

$$(T_{AB} + T_{AC}P_{AC}^{1} - V_{1}') + (T_{AC}P_{AC}^{3} - V_{3}') + (T_{AC}P_{AC}^{4} - V_{4}') = 0$$
 [4-8b]

where: V'_{a} = observed flow on link a (vph) P_{ij}^{a} = probability that demand between i and j will use link a T_{ij} = demand between origin i and destination j (vph)

Algebraically solving Equations [4-8a] and [4-8b] results in the following expressions for T_{AB} and T_{AC} :

$$T_{AB} = \frac{(V_1' + V_2')(P_{AC}^1 + P_{AC}^3 + P_{AC}^4) - P_{AC}^1(V_1' + V_3' + V_4')}{2(P_{AC}^1 + P_{AC}^3 + P_{AC}^4) - P_{AB}^1P_{AC}^1}$$

$$T_{AC} = \frac{V_1' + V_3' + V_4' - T_{AB}}{P_{AC}^1 + P_{AC}^3 + P_{AC}^4}$$

If the observed link flows are V'_1 =30 vph, V'_2 =20 vph, V'_3 =10 vph, and V'_4 =20 vph, and it is assumed that 2/3 of the demand between zones A and C uses link 4, then T_{AB} and T_{AC} can be estimated to be:

$$T_{AB} = \frac{\left(30 + 20\right)\left(\frac{1}{3} + \frac{1}{3} + \frac{2}{3}\right) - \frac{1}{3}\left(30 + 10 + 20\right)}{2\left(\frac{1}{3} + \frac{1}{3} + \frac{2}{3}\right) - \left(1 \cdot \frac{1}{3}\right)} = 20 \text{ vph}$$

$$T_{AC} = (30+10+20-20)/(\frac{1}{3}+\frac{1}{3}+\frac{2}{3})=30$$
 vph

Since the proposed system of constraints can be solved analytically, the solution space contains only a single solution. As the above network provides a uniquely utilized link for each of the two unknown O-D demands, and node and path flow continuity exist in the observed link flow data, the single demand solution leading to the observed flows is clearly T_{AB} =20 vph and T_{AC} =30 vph. Once again, it is therefore shown that the proposed system of constraints leads to the selection of the correct O-D demand.

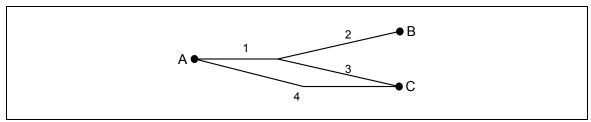


Figure 4-2: Example four-link network exhibiting multiple paths between an origin and destination zone

4.3 Incorporation of Link Flow Reliability

In practice, link flow data often are obtained from a variety of sources including induction loop detectors and video detectors operating as part of an FTMS or UTCS, manual traffic counts, or historical data. Rarely do these different sources have identical levels of reliability associated with them. Furthermore, varying levels of reliability often exist within the same data source. For example, at any given time, a number of the detector stations within an FTMS may be less accurate then the rest of the system due to a need for re-calibration, or due to malfunctions. It is desirable to reflect explicitly these different levels of reliability within the O-D estimation model.

4.3.1 Modified model formulation

From an engineering perspective, link flow errors associated with links for which observed data have a low level of reliability, should contribute less to the overall link flow error function, than errors associated with link data of a high degree of reliability. This goal is accomplished in the following manner:

First, define R_a as the relative reliability associated with the observed link flow data for link a. The relative reliability of any link flow may take on any positive value, however, we find it convenient to generally restrict the range from zero to one $(0 \le R_a \le 1.0)$. When R_a =0, then the observed link flow is completely unreliable, and any flow error resulting for this link should not be considered in the overall error function. Conversely, if R_a =1.0, then the observed link flow data are no less reliable than any other data, and any discrepancies between estimated and observed flows for this link should be considered in the error function

at full value. The relative reliability of observed link flows is incorporated into Equation [4-1] to create Equation [4-9]

$$E = \sum_{a} \left[(V_a - V_a') R_a \right]^2$$
 [4-9]

where: a = unique link identifier

 V_a = estimated flow on link a (vph) V'_a = observed flow on link a (vph)

 R_a = relative reliability of flow observed on link a

Following the same derivation as shown previously, the O-D demand, that minimizes E, can be found by computing the first partial derivative of E with respect to T_{ij} and setting it equal to zero. Finding this derivative for each unknown demand results in a set of linear equations defining the feasible region of solutions that minimizes the squared link flow error. The resulting set of linear equations is similar to those presented in Equation [4-5], except that the link flow reliability factor is now incorporated into Equation [4-10].

$$\frac{\partial}{\partial} \frac{E}{T_{ij}} = \sum_{r} \left[\left(\sum_{ij} \left(T_{ij} P_{ij}^{r} \right) - V_{r}' \right) R_{r}^{2} P_{ij}^{r} \right] = 0 \forall_{ij}$$
[4-10]

where: r = any link on a route between origin i and destination j

 V_a = estimated flow on link a (vph) V'_a = observed flow on link a (vph) P_{ij}^a = probability that demand between i and j will use link a

 T_{ii} = demand between origin *i* and destination *j* (vph)

 R_a = relative reliability of flow observed on link a

4.3.2 **Example considering relative reliability of link flows**

It is useful to examine, by way of example, how the relative reliability of link flows affects the feasible region defined by the system of normal equations. Consider the simple three-link linear network illustrated in Figure 4-3. Initially assume that all link flows are known with equal certainty $(R_1 = R_2 = R_3 = 1.0)$. If the flows observed are $V'_1 = 50$ vph, $V'_2 = 100$ vph, and $V'_3 = 50$ vph, then the link flow error associated with a demand estimate of $T_{AB} = 50$ vph, is computed as

$$E = [(50-50)\cdot 1.0]^2 + [(100-50)\cdot 1.0]^2 + [(50-50)\cdot 1.0]^2 = 2500 \text{ vph}^2$$

If, however, there is reason to believe that the flow obtained for link 2 is erroneous to such a degree that it is totally unreliable, then the relative reliability of link 2 is zero $(R_2=0.0)$. Now, for the same demand estimate $(T_{AB}=50 \text{ vph})$, the associated link flow error is computed as

$$E = [(50-50)\cdot 1.0]^2 + [(100-50)\cdot 0.0]^2 + [(50-50)\cdot 1.0]^2 = 0 \text{ vph}^2$$

Clearly, for the extreme cases in which either all information is equally reliable, or some information is totally unreliable and thus not considered, the use of relative link flow reliability factors is consistent with engineering intuition. A scenario more likely to be encountered in practice is one in which all data are not equally reliable, but where no data are totally unreliable. Consider, once again, the simple linear network example. Assume that the flows obtained for link 2 are half as reliable as the other flow data (R_2 =0.5). Then, the link flow error associated with T_{4R} = 50 vph is

$$E = [(50-50)\cdot 1.0]^2 + [(100-50)\cdot 0.5]^2 + [(50-50)\cdot 1.0]^2 = 625 \text{ vph}^2$$

Consider, Figure 4-4 which illustrates the variation of E in response to changes in the relative reliability of flow data for link 2 only. It is assumed that the demand remains at 50 vph and that the relative reliability of links 1 and 3 remain constant at 1.0. It is evident from Figure 4-4 that the reliability of observed flows can have a significant impact on the total link flow error. It is also evident that the convexity and functionality of E are maintained despite the introduction of the link flow reliability factors.

Since the objective is to estimate the demand that minimizes link flow error, it is of interest to examine the impact that changes in the relative reliability of link flow data have on the perceived optimum demand. Figure 4-5 illustrates the demand that minimizes link flow error for the complete range of relative reliability of flow for link 2. As determined earlier, when the observed flow for link 2 is regarded as being totally unreliable, the optimal demand is 50 vph. As more and more confidence is placed in the flow observed on link 2, the demand that minimizes the total link flow error also increases. This behaviour is consistent with engineering intuition. Flow continuity exists only when the flow on link 2 is not considered. Only under these conditions does an O-D demand exist that exactly replicates the observed link flows (T_{AB} =50 vph). As more confidence is placed in the flow of 100 vph observed on link 2, the error function is minimized by increasing the estimated demand.

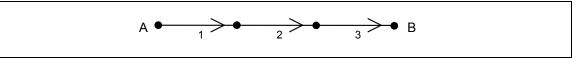


Figure 4-3: Example line three-link network used to illustrate the impact of link flow reliability

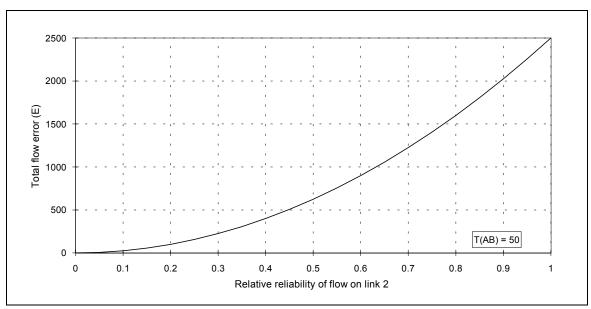


Figure 4-4: Variation in link flow error as a function of link flow reliability

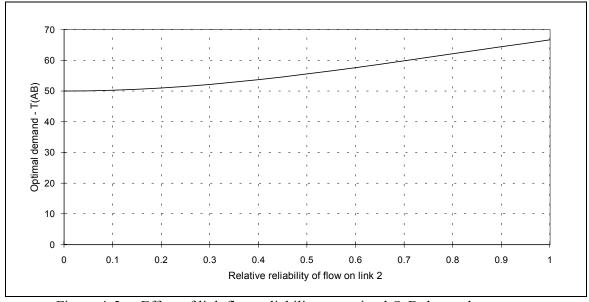


Figure 4-5: Effect of link flow reliability on optimal O-D demand

From the previous examination of a simple linear network example, it is evident that the relative reliability of link flow data can have a significant impact on the overall link flow error function, and consequently, on the perceived optimum O-D demand estimate. Since, in reality, most network demand estimation situations rely on data obtained from many sources having different levels of reliability, the incorporation of link flow reliability is a useful characteristic of a synthetic O-D estimation model.

It is also clear that since the degree of reliability of link flows influences the estimated demand, appropriate link flow reliability factors must be selected. Methods for determining appropriate reliability factors are examined in the next section.

4.3.3 Determining appropriate link flow reliability factors

Bell (1984) has suggested that observed link flows can be considered as random variables. Then, from a time series of link counts, the mean and variance of each link's flow can be computed. Under steady-state conditions, variances in observed flow result from randomness, and represent how much confidence can be placed in the validity of the mean flow value. As illustrated in Figure 4-6, when the variance is large, less confidence can be placed in the computed mean, as many of the individual observed flows are quite different from the mean value.

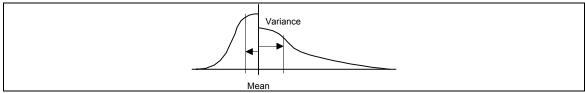


Figure 4-6: Illustration of the impact of variance on the validity of the mean

Using this argument, Bell proposes that weighted least squares formulations use the inverse of the variance as the link flow reliability factor. Thus, if the variance is small, the error associated with that link will be weighted more heavily, and consequently, the algorithm will attempt to replicate the observed flow more closely, implying that the observed flow is quite reliable. Conversely, if the variance is large, then the weighting will be small, and the algorithm will be less sensitive to the associated link flow error.

We illustrate this approach using data obtained from loop detectors located on Highway 401 in Toronto, Canada. The detectors report the number of vehicles detected crossing the detection zone in any of the three lanes, during the previous 20 second period. These data reflect freeway conditions from 4 AM to 9 AM on May 1, 1991. Figure 4-7 illustrates the computed variance of observed vehicle counts by time of day. Each data point represents the variance within the previous five minutes of 20 second observations. Figure 4-7 indicates that the variance remains quite small until approximately 5:45 AM, at which time, the variance begins to increase. This increase is not constant, but continues until approximately 7:15 AM, at which time the variance is almost 70 (veh/20 seconds)². After 7:15 AM the magnitude of the variance varies significantly. If we assume that the inverse of variance provides a measure of the reliability of the link counts, then we must conclude that the link counts observed prior to 6 AM are rather reliable, and most of those after 7 AM are very unreliable.

The variance must, however, be considered in relation to the magnitude of the mean. If the variance is large, but the mean is also large, then the relative accuracy of the mean may still be quite high. A convenient measure of the variance of data with respect to the magnitude of the mean is the coefficient of variation (COV). The COV is computed as the

standard deviation divided by the mean. Figure 4-8, illustrating the COV of vehicle counts computed from the same data as illustrated in Figure 4-7, indicates that the conclusions made based on Figure 4-7 are incorrect. The most reliable link counts are those observed after 7 AM, when mean link flows are quite large in comparison to the variance.

However, the use of the inverse of the variance as a weighting factor does not directly reflect the relative reliability of link flow. Rather, the weighting factor attempts to normalize the contribution made by each link to the error function. As it is usually expected that link flows with very large variances, also have very large link flow errors associated with them, their contribution to the error function may be so large as to overwhelm the contributions made by links having a smaller variance (and thus usually a smaller link flow error). Utilizing the inverse of the variance as a weighting factor would tend to reduce the contribution of links having a large variance.

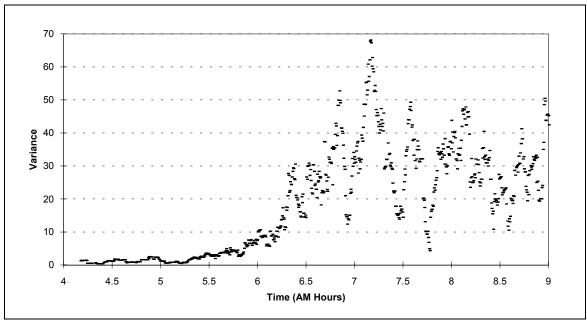


Figure 4-7: Variance of vehicle counts observed on a three lane section of Highway 401 Eastbound Collectors at Avenue Road. Each point represents the variance over the previous five minutes of 20 second link counts

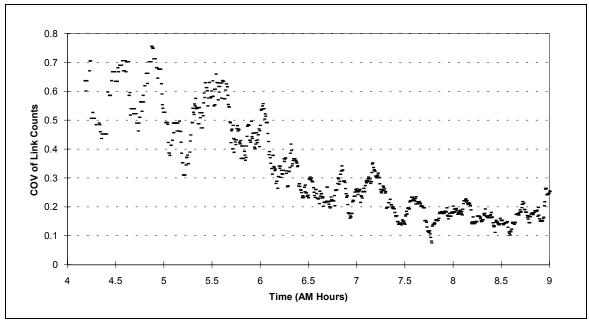


Figure 4-8: Coefficient of variation of vehicle counts observed on a three lane section of Highway 401 Eastbound Collectors at Avenue Road. Each point represents the COV computed over the previous five minutes of 20 second link counts

An alternative method of determining link flow reliability factors is based on the conservation of flow. This method is particularly useful when estimating static demands from a single observed flow for each link. Clearly, when only a single aggregate link flow is available, the variance cannot be computed, and the use of the inverse of the variance as a measure of reliability is not applicable. The method relies on the recognition that since flow cannot be added (except at origin zones), and cannot be removed (except at destination zones) from the network, flow discontinuity at a node must be due to measurement error, or temporal fluctuations in flow. If detector stations are relatively close together (i.e. 1 km) and the observation period is suitably long (i.e. 10 minutes) then in the absence of severe congestion, the effects of temporal perturbations will generally not be large. The existence of a large discontinuity in flow implies an erroneous observed link flow, and consequently, the reliability of such a flow is rather low.

Two difficulties exist with the use of this method of determining link flow reliability factors. First, the relationship between the magnitude of the flow discontinuity and the link flow reliability is not known quantitatively. For example, if two connected links experience flow of 100 vph and 200 vph respectively, then a flow discontinuity of 100 vph exists at the intermediate node. Though it is evident that at least one of the observed link flows is in error, it is not clear what the appropriate link flow reliability factor should be.

The second difficulty lies with the identification of suspect observed link flows. It is possible to identify nodes having flow discontinuity, but it is often not possible to determine which link, or links, are actually responsible for this discontinuity.

Despite these difficulties, a heuristic approach for determining suspect link flows can be defined. The magnitude of the flow discontinuity at each node (not zone) in the network is determined in turn using Equation [4-11]. If the magnitude of the discontinuity exceeds a user specified relative level, then each link connected to this node is tagged. Any link that is tagged twice must be connected to two nodes exhibiting discontinuity in excess of the threshold values and is considered to be suspect. Any link that is tagged only once, but is connected at one end to either an origin or a destination zone, is also considered suspect. The user may control whether suspect link flows are considered to be totally unreliable, or assigned a link flow reliability factor based on the average relative flow discontinuity (Equation [4-12]).

$$D_{node} = \left| \sum_{a} V_a' \right| / \sum_{a} |V_a'| \tag{4-11}$$

where: V'_a = observed flow on link a. Note that flow into the node is considered positive, while flow out of the node is negative

$$R_a = 1 - \frac{1}{n} \left(D_a^1 + D_a^2 \right) \tag{4-12}$$

where: D_a^l discontinuity of upstream node of link a

 D_a^2 discontinuity of downstream node of link a

n = number of nodes that link a is connected to (n=1 or n=2)

 R_a = relative reliability of flow on link a

To illustrate this heuristic approach, consider the three-link network illustrated in Figure 4-9. As indicated, the network consists of two intermediate nodes, an origin zone, and a destination zone. Table 4-2 provides the associated calculations based on Equations [4-11] and [4-12]. Table 4-2 indicates that the flow on link 1 is the least reliable, while the flow on link 3 is the most reliable. These results are consistent with engineering intuition, indicating that the heuristic approach provides a practical alternative to the use of the inverse of COV.

It must be noted, however, that when flows are not known for all links connected to a node, it is no longer possible to calculate the flow discontinuity at the node using Equation [4-11].

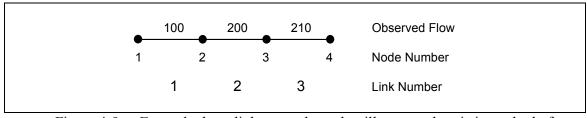


Figure 4-9: Example three-link network used to illustrate a heuristic method of determining link flow reliability

1	chability factors								
Discontinuity at Nodes (D_{node})									
Node	Flow In	Flow Out	D _{node}						
1	N.A.	100	N.A.*						
2	100	200	0.3333						
3	200	210	0.0244						
4	210	N.A.	N.A.*						
Link Flow Re	eliability Factors	s (<i>R_a</i>)							
Link	D^1	D ²	n	R _a					
1	N.A.*	33.33	1	0.6666					
2	0.3333	0.0244	2	0.8211					
3 0.0244 N.A.* 1 0.9756									
Since these nodes are zones, the discontinuity cannot be computed, and the link									

Table 4-2: Example application of a heuristic approach to determining link flow reliability factors

4.4 Characteristics of Formulation

It is prudent to examine the characteristics of the proposed mathematical formulation prior to presenting the development of a solution algorithm. Since the formulation provides the basis of the methodology, its characteristics, and in particular, the effects these characteristics have on the estimated solution, should be fully understood.

Three particular characteristics of the least squares formulation need to be examined. First, the constraints do not ensure that the feasible region contains only non-negative demands. Second, it is not clear that an error function based on the squared flow differences is the most appropriate measure of error. Third, in the event that multiple solutions exist, that all result in the same error function value, how is a single solution chosen? Each of these three issues is discussed and illustrated in the following sections.

4.4.1 Negative demand estimates

As with least squares regression, the constraints proposed in Equation [4-10] do not restrict feasible demands to being non-negative. This lack of restriction can lead to negative demand estimates as illustrated using the following simple example. Consider once again, the three-link network illustrated previously in Figure 4-1.

The observed link flows, $V'_1 = 15$ vph, $V'_2 = 10$ vph, and $V'_3 = 50$ vph, do not exhibit node or path flow continuity. If all observed flows are considered to be equally reliable, then the constraints expressed in Equation [4-2] lead to the following demand estimates:

$$T_{AC} = (2 \cdot V'_3 + V'_1 - V'_2) / 3 = (2 \cdot 50 + 15 - 10) / 3 = 35 \text{ vph}$$

 $T_{AB} = T_{AC} - V'_3 + V'_2 = 35 - 50 + 10 = -5 \text{ vph}$

These demand estimates produce link flows of $V_1 = 30$ vph, $V_2 = -5$ vph, and $V_3 = 35$ vph.

The above solution minimizes the squared link flow error, however, the negative demand is not valid and does not have any consistent physical interpretation. It is potentially possible to truncate any non-positive demand estimates to zero. This technique has been used by other researchers (Hendrickson and McNeil, 1984a), but the magnitude of its effect on the optimality of the solution is not known *a priori*. Figure 4-10 illustrates, for this example network, the impact that truncation has on the optimality of the estimated demand. The unconstrained optimization results in a squared link flow error of 675 vph², and is defined as point *A*. Point *B* reflects the solution obtained when the unconstrained solution is truncated to ensure that all demands are non-negative (E=725 vph²). It is evident from Figure 4-10 that point *B* does not represent the optimal solution, even when T_{AB} is constrained. Point *C*, representing the constrained optimum solution, has an associated squared link flow error of 712.5 vph². As will be demonstrated in Section 4.5.6, the modified Jacobi iterative solution method proposed in Section 4.5.3 results in an O-D estimate for this same example network corresponding to point *C*.

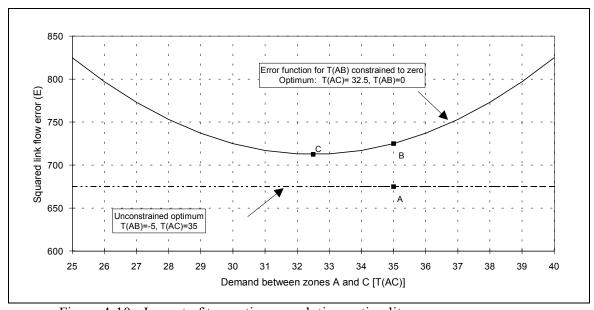


Figure 4-10: Impact of truncation on solution optimality

This example is presented to illustrate that the technique of truncation of the final model estimate can lead to the acceptance of a solution which is sub-optimal. Of course, this example cannot serve to make general conclusions regarding the magnitude of the sub-optimality, as this is likely a function of the network configuration and the observed link flows.

It is possible to add explicit non-negativity constraints to the LSE model formulation expressed in Equation [4-6] to produce Equation [4-13], however, it is significantly more difficult to carry out a closed-form algebraic solution for this modified formulation. It will be shown by way of an example, that the modified Jacobi iterative solution method proposed in Section 4.5.3 does indeed solve the constrained formulation provided in Equation [4-13].

$$\sum_{r} \left[\left(\sum_{ij} \left(T_{ij} P_{ij}^{r} \right) - V_{r}' \right) P_{ij}^{r} \right] = 0 \forall_{ij}$$

$$T_{ii} \ge 0 \forall_{ii}$$
[4-13]

where: a = unique link identifier

r = any link on a route between origin i and destination j

 V_a = estimated flow on link a (vph) V'_a = observed flow on link a (vph) P_{ij}^a = probability that demand between i and j will use link a

 T_{ii} = demand between origin *i* and destination *j* (vph)

4.4.2 Suitability of error function

The suitability of defining an error function based on the squared link flow differences is partially a philosophical debate and partially a mathematical convenience. The error function actually consists of two elements — the link flow difference, and the second order exponent. The latter element is most easily justified. Since there is no reason to believe that errors resulting from underestimated flows are more or less important than those resulting from over-estimated flows, the contribution to the error function should be the same for the same absolute error. The most convenient manner in which to achieve this equality is to square the error, negating the sign (i.e. negative or positive) of the error. Squaring the error also has the effect of weighting the large errors proportionately more heavily than smaller errors.

The philosophical preference arises from the choice of link flow differences as a measure of error. Adopting this measure of error implies that under (or over) estimating a link flow of 10 vph by 10 vph is as important as under (or over) estimating a link flow of 1000 vph by 10 vph. The link flow difference in each case is simply 10 vph. No measure of the relative error is captured. It may be argued, that it is more important to reduce the errors that constitute a large proportion of the observed flow, than it is to reduce the small proportional errors.

No answer to this philosophical debate is proposed at this point in the thesis. However, the development of an O-D estimation technique based on minimizing the proportional link flow errors is described in the next chapter. Subsequent chapters describe the application of both models to a number of hypothetical and actual networks. Comparisons between the two models are made and conclusions regarding the most appropriate error function are drawn.

Selecting a solution when multiple solutions exist 4.4.3

The LSE model, which has been presented, is based on the minimization of the squared link flow differences. In theory, it is possible that more than one solution, that minimizes this error, exists. The question that must be answered, is, how does the model select one of these feasible solutions?

To illustrate the problem, consider the simple two-link network presented in Figure 4-11 for which the observed flows on links 1 and 2 are $V'_1=16$ vph and $V'_2=18$ vph. Three O-D demands are feasible – T_{AB} , T_{AC} , and T_{BC} . As indicated in Figure 4-12, it is possible, in this simple example, to graphically illustrate the feasible solution space defined by the following three normal equations:

$$T_{AB} + T_{AC} - 16 = 0$$

 $(T_{AB} + T_{AC} - 16) + (T_{BC} + T_{AC} - 18) = 0$
 $T_{BC} + T_{AC} - 18 = 0$

The feasible solution space is defined by the line describing the intersection of two planes. The model formulation does not provide any means for selecting one solution from the many feasible solutions that lie on this line. There are two issues that must be addressed. First, the model must be consistent so that under the same circumstances, it always selects the same solution as optimal. Second, when multiple solutions exist, it is advantageous for the model to select a solution which, by some measure, is more likely than the other feasible solutions.

It will be shown that the solution algorithm used to solve the LSE formulation does provide a means of selecting one solution when multiple solutions exist. Furthermore, it will be shown by way of example, that in the absence of prior information, the selected solution approximates the most likely solution.

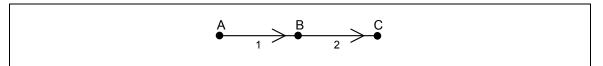


Figure 4-11: Example two-link network

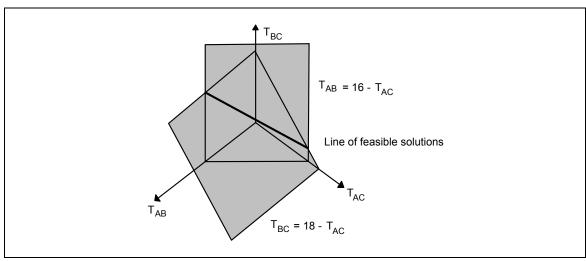


Figure 4-12: Illustration of the feasible O-D solution space for a two-link example network

4.5 Development of a Solution Algorithm

The least squares model that has been developed and illustrated in the preceding sections may be thought of as a linear regression of multiple independent variables. Regression is a well established analysis technique on which much has been written (two particularly useful references are Draper and Smith, 1980 and Box *et al.*, 1978). We draw on this literature to briefly provide the formal solution to find a multiple linear regression model that minimizes the least squared errors.

4.5.1 Formal solution

Consider an O-D estimation problem in which there exists L observed link flows, and N unknown O-D demands. Then, the set of equations of the form

$$V_a' = \sum_{a} \left(P_{ij}^a T_{ij} \right) \tag{4-14a}$$

where: V'_a = observed flow on link a (vph)

 P_{ij}^{a} = probability that demand between *i* and *j* will use link *a*

 T_{ii} = demand between origin i and destination j (vph)

may be written in matrix form as

$$\begin{bmatrix} V_1' \\ \vdots \\ V_L' \end{bmatrix} = \begin{bmatrix} P_1^1 & \cdots & P_N^1 \\ \vdots & & \vdots \\ P_1^L & \cdots & P_N^L \end{bmatrix} \times \begin{bmatrix} T_1 \\ \vdots \\ T_N \end{bmatrix}$$

$$[4-14b]$$

or,

$$\mathbf{V} = \mathbf{P} \mathbf{T} \tag{4-14c}$$

where: V = the column vector of observed link flows

P = the matrix of link use probabilities

T = the column vector of unknown O-D demands

Note, that in order to present both **P** and **T** as two-dimensional matrices, T_i refers, in this section only, to the unknown demand between the ith origin - destination pair.

Values for **T**, that minimize the squared link flow error, may be found by solving the system of equations presented in Equation [4-6]. These equations, algebraically referred to as the *normal* equations, are expressed in matrix notation as

$$\mathbf{P}^{\mathsf{t}}\left(\mathbf{V}-\mathbf{PT}\right)=0\tag{4-15}$$

where: P^t = the transpose of the matrix P

The transpose of a matrix A, having m rows and n columns, is a matrix A^t having n rows and m columns. The ith column in matrix A is transposed to become the ith row in matrix A^t . Equation [4-15] can be expanded to produce Equation [4-16].

$$\mathbf{P}^{\mathsf{t}} \mathbf{V} - \mathbf{P}^{\mathsf{t}} \mathbf{P} \mathbf{T} = 0 \tag{4-16}$$

It is desirable to express Equation [4-16] in terms of the unknown column demand matrix **T**. To achieve this goal, it is necessary to make use of the properties of the *inverse* of a matrix. If the inverse of some matrix **A** is denoted **B**, then, as indicated in Equation [4-17], the matrix multiplication of **A** and **B** in any order, results in the unit matrix **I**.

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I} \tag{4-17}$$

If the term [P^tP] from Equation [4-16] has an inverse, then the vector of unknown demands can be determined from Equation [4-18].

$$[P^{t}P]^{-1} P^{t}V - [P^{t}P]^{-1} P^{t}PT = 0$$
 [4-18a]

$$[\mathbf{P}^{t}\mathbf{P}]^{-1} \mathbf{P}^{t}\mathbf{V} - \mathbf{I} \mathbf{T} = 0$$
 [4-18b]

$$\mathbf{T} = [\mathbf{P}^{\mathsf{t}}\mathbf{P}]^{-1} \mathbf{P}^{\mathsf{t}}\mathbf{V}$$
 [4-18c]

where: V = the column vector of observed link flows

P = the matrix of link use probabilities

 $\mathbf{P}^{\mathbf{t}}$ = the transpose of the matrix of link use probabilities

T = the column vector of unknown O-D demands

4.5.2 Limitations of the formal solution

The formal solution expressed in Equation [4-18] is only valid when $[P^tP]$ has an inverse. In order for $[P^tP]$ to have an inverse, two conditions must be met. First, $[P^tP]$ must be square, and second, the determinant must be non-zero. If the determinant of a square matrix is zero, then the matrix is considered to be singular and cannot be inverted.

The first condition is always met, for if **P** is of dimensions $L \times N$, then **P**^t must have dimensions $N \times L$. Then, the dimensions of **P**^t**P** are $L \times L$, which is a square matrix.

The second condition can be shown to not always be satisfied for a general case. To illustrate this, consider the two-link network presented in Figure 4-11 for which three O-D demands are possible (T_{AB} , T_{AC} , and T_{BC}) and the observed flows are V'_{I} =16 vph and V'_{I} =18 vph. For this network, **P**, **P**^t, **T**, and **V** can be defined as follows:

$$\mathbf{P} = \begin{bmatrix} P_{AB}^{1} & P_{AC}^{1} & P_{BC}^{1} \\ P_{AB}^{2} & P_{AC}^{2} & P_{BC}^{2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}; \ \mathbf{P^{t}} = \begin{bmatrix} P_{AB}^{1} & P_{AB}^{2} \\ P_{AC}^{1} & P_{AC}^{2} \\ P_{BC}^{1} & P_{BC}^{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{T} = \begin{bmatrix} T_{AB} \\ T_{AC} \\ T_{BC} \end{bmatrix}, \mathbf{V} = \begin{bmatrix} V'_1 \\ V'_2 \end{bmatrix} = \begin{bmatrix} 16 \\ 18 \end{bmatrix}$$

The matrix multiplication of $[P^tP]$ is computed as follows:

$$\mathbf{P}^{\mathbf{t}}\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

The determinant of [**P**^t**P**] is computed as $a_{11}A_{11}$ - $a_{12}A_{12}$ + $a_{13}A_{13}$ * where A_{13} , A_{13} , and A_{13} are computed as follows:

$$A_{11} = \begin{vmatrix} 2 & 1 \\ 1 & 1 \end{vmatrix} = a_{11}a_{21} - a_{12}a_{21} = (2 \cdot 1) - (1 \cdot 1) = 1$$

$$A_{12} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1; A_{13} = \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} = 1$$

Then, the determinant of $[\mathbf{P}^t\mathbf{P}]$ is equal to $(1\cdot1)$ - $(1\cdot1)$ + $(0\cdot1)$ = 0. Since the determinant is equal to zero, the matrix $[\mathbf{P}^t\mathbf{P}]$ is singular and cannot be inverted. This singularity occurs because the O-D estimation problem is underspecified as there are three unknown demands and only two constraint equations. There exist multiple solutions and the formal solution to the least squares model cannot be solved.

4.5.3 Iterative solution algorithm

The formal solution approach discussed in the previous section is only valid when the problem is not underspecified. However, we wish to be able to solve the problem regardless of the number of feasible solutions that exist. Existing iterative methods for solving systems of linear equations include the Gauss-Seidel method (Hornbeck, 1975) and the Jacobi method (Williams, 1984). Though these methods themselves will be shown to be unsuitable for the purposes of this work, they do form the basis of the proposed iterative solution method and as such will be briefly examined below.

a. Gauss-Seidel Method:

The Gauss-Seidel iterative process consists of repeatedly cycling through the system of linear equations. During each cycle, the current equation is solved for one of the unknowns using previously determined values for each of the other unknowns. Consider as an example, a set of three linear equations:

$$c_{11}x_1 + c_{11}x_2 + c_{11}x_3 = r_1$$

$$c_{21}x_1 + c_{22}x_2 + c_{23}x_3 = r_2$$

$$c_{31}x_1 + c_{32}x_2 + c_{33}x_3 = r_3$$
[4-19]

The first equation is now solved for x_1 , the second for x_2 , and the third for x_3 to provide Equation [4-20].

^{*} Williams (1984) provides a comprehensive discussion of the calculation of determinants and their properties.

$$x_{1} = (r_{1} - c_{12}x_{2} - c_{13}x_{3})/c_{11}$$

$$x_{2} = (r_{2} - c_{21}x_{1} - c_{23}x_{3})/c_{22}$$

$$x_{3} = (r_{3} - c_{31}x_{1} - c_{32}x_{2})/c_{33}$$
[4-20]

An arbitrary initial guess is made for each unknown and designated by x_1^0, x_2^0, x_3^0 . Then, the first equation in Equation [4-20] can be used to estimate a new value for x_1 .

$$x_1^1 = \left(r_1 - c_{12}x_2^0 - c_{13}x_3^0\right)/c_{11}$$
 [4-21a]

A new estimate for x_2 is determined using the most recently computed value of x_1 and the value of x_3 is computed using the most recently computed value of x_1 and x_2 .

$$x_2^1 = \left(r_2 - c_{21}x_1^1 - c_{23}x_3^0\right)/c_{22}$$
 [4-21b]

$$x_3^1 = (r_3 - c_{31}x_1^1 - c_{32}x_2^1)/c_{33}$$
 [4-21c]

According to Hornbeck (1975) convergence of the Gauss-Seidel iterative method cannot be guaranteed unless the absolute value of the diagonal is greater than the sum of the absolute value of the remaining elements on that row (Equation [4-22]).

$$\left|c_{ii}\right| > \sum_{j=1}^{n} \left|c_{ij}\right| \tag{4-22}$$

b. Jacobi Method:

The Jacobi method of iteratively solving a system of linear equations is similar to the Gauss-Seidel method with the exception that new estimates of unknowns are made only on the basis of the values computed during the previous iteration. Computations are not made using the most recently computed values as was done in the Gauss-Seidel method. Examining the set of three linear equations presented in Equation [4-19], the Jacobi method computes x_1 , x_2 , and x_3 from the following equations:

$$x_1^1 = \left(r_1 - c_{12}x_2^0 - c_{13}x_3^0\right) / c_{11}$$
 [4-23a]

$$x_2^1 = (r_2 - c_{21}x_1^0 - c_{23}x_3^0)/c_{22}$$
 [4-23b]

$$x_3^1 = (r_3 - c_{31}x_1^0 - c_{32}x_2^0)/c_{33}$$
 [4-23c]

It can be noted that in Equation [4-23b] and [4-23c] all computations are made using values determined in the previous iteration, in this case iteration zero. In the Gauss-Seidel method, these values were the most recent values available.

The convergence criterion for the Jacobi method is the same as that for the Gauss-Seidel method. Williams (1984) indicates that, in general, the Gauss-Seidel method converges more quickly than the Jacobi method, however, this is not guaranteed.

c. Evaluation of the Gauss-Seidel and Jacobi Methods:

It would appear that both the Gauss-Seidel method and the Jacobi method would be potentially suitable for solving the system of linear equations resulting from the use of the LSE model for the estimation of O-D demands. However, it is not clear that the convergence criterion of these iterative methods, expressed in Equation [4-22], is likely to be met. To examine this, consider the now familiar two-link example network presented in Figure [4-11]. Consider the observed link flows to be equal to V'_{I} =16 vph and V'_{2} =18 vph. The system of linear constraints resulting from the LSE model is provided in Equation [4-24].

$$\begin{aligned} &1T_{AB} + 1T_{AC} + 0T_{BC} = 16 \\ &1T_{AB} + 2T_{AC} + 1T_{BC} = 34 \\ &0T_{AB} + 1T_{AC} + 1T_{BC} = 18 \end{aligned} \tag{4-24}$$

Table 4-3 provides the results obtained from the first 5 iterations of the Gauss-Seidel and Jacobi iterative methods. It can be seen that the Gauss-Seidel method quickly converges to a solution while the Jacobi method does not converge at all. An examination of the system of linear equations in Equation [4-24] indicates that the convergence criterion of Equation [4-22] is not met and explains why the Jacobi method fails to converge. Despite the convergence criterion not being met in this example, the Gauss-Seidel method converges. Hornbeck (1975) indicates that in practice, convergence can be obtained with the Gauss-Seidel method with much weaker diagonal dominance than is specified in Equation [4-22], however, under these conditions convergence cannot be guaranteed.

Table 4-3: Gauss-Seidel and Jacobi iterative method results from a simple network having multiple solutions that satisfy the system of linear equations

	1100 1 1	8		20-07-0		ory erre by	200		
Iteration	Ga	auss-Se	eidel M	ethod	Jacobi Method				
Number	T_{AB}	T_{AC}	T _{BC}	Ε	T_{AB}	T_{AC}	T_{BC}	Ε	
	(vph)	(vph)	(vph)	(vph ²)	(vph)	(vph)	(vph)	(vph ²)	
0	5.0	5.0	5.0	100.0	5.00	5.00	5.00	100.0	
1	11.0	9.0	9.0	16.0	5.00	5.00	10.00	98.0	
2	7.0	9.0	9.0	0.0	11.00	12	20.00	98.0	
3	7.0	9.0	9.0	0.0	4.00	5	16.00	98.0	
4	7.0	9.0	9.0	0.0	11.00	12	16.00	98.0	
5	7.0	9.0	9.0	0.0	4.00	5	16.00	98.0	

Since the Gauss-Seidel method utilizes the most recent O-D estimates in each computation, it is possible that the estimated solution is sensitive to the order in which the system of equations is examined. Consider the same two-link example but in this case, assume that the order in which the O-D traffic demands appear in each equation is altered.

$$\begin{aligned} &1T_{BC} + 1T_{AC} + 0T_{AB} = 18 \\ &1T_{BC} + 2T_{AC} + 1T_{AB} = 34 \\ &0T_{BC} + 1T_{AC} + 1T_{AB} = 16 \end{aligned} \tag{4-24}$$

Applying the Gauss-Seidel method to this system of linear equations results in an estimated O-D solution of $T_{AB} = 8$ vph, $T_{AC} = 8$ vph, $T_{BC} = 10$ vph. This solution satisfies the system of linear equations, but it is not the same solution as the one found earlier using the Gauss-Seidel method. Thus, in this example, the O-D solution estimated by the Gauss-Seidel method is sensitive to the order in which the system of equations is written. This sensitivity is

undesirable, as the estimated O-D demand is potentially dependent on the arbitrary order in which the equations are written.

In this section it has been shown that both the Gauss-Seidel method and the Jacobi method may not converge to a solution if diagonal dominance does not exist. It has also been shown, by way of a simple example, that diagonal dominance is not guaranteed to exist for all O-D estimation problems. Furthermore, it was shown that the O-D solution estimated by the Gauss-Seidel method, for this particular example network, was sensitive to the order in which the system of linear equations was written. As a result, the next section proposes an iterative solution method that is a modification of the standard Jacobi method.

d. Proposed Method:

The problem with the standard Jacobi method is that it is only guaranteed to converge when the diagonal dominance expressed by Equation [4-22] exists. Since O-D estimation problems exist in which this diagonal dominance criterion is not met, it is desirable to modify the standard Jacobi method to enable it to converge even when diagonal dominance does not exist.

The standard Jacobi method is modified, using a *relaxation* technique (Hornbeck, 1975), so that after an iteration is completed (i.e. values have been computed for all unknowns), these computed values are modified by some factor $(1-\alpha)$ prior to using them in the next iteration. It should be noted that when $\alpha=1.0$, then the modified Jacobi method reverts back to the standard Jacobi method.

$$x_i^{l+1^*} = \alpha x_i^{l+1} + (1 - \alpha) x_i^l$$
 [4-25]

where: $x_i^{l+1^*}$ = the modified value for unknown x_j that is estimated in iteration l+1 x_i^{l+1} = the standard Jacobi estimate of x_j in iteration l+1

 x_i^l = estimate of unknown x_j in iteration l

 α = relaxation factor

Consider the set of three general linear equations presented in Equation [4-19]. From, Equation [4-25], the modified estimate for x_I for the first iteration is expressed as Equation [4-26a]. From Equation [4-23a], the expression for x_1^1 can be substituted into Equation [4-26a]. Algebraically simplifying Equation [4-26a] results in Equation [4-26c], an expression for x_1^1 that is a function of the value of x_I determined in the previous iteration (x_1^0) and the value of the constraint equation from which x is being evaluated (C_i) . Similar expressions can be derived for the other unknowns x_2 , and x_3 .

$$x_{1}^{1^{*}} = \alpha x_{1}^{1} + (1 - \alpha)x_{1}^{0}$$

$$x_{1}^{1} = (r_{1} - c_{12}x_{2}^{0} - c_{13}x_{3}^{0})/c_{11}$$

$$x_{1}^{1^{*}} = \alpha ((r_{1} - c_{12}x_{2}^{0} - c_{13}x_{3}^{0})/c_{11}) + (1 - \alpha)x_{1}^{0}$$

$$x_{1}^{1^{*}} = \alpha ((r_{1} - c_{12}x_{2}^{0} - c_{13}x_{3}^{0})/c_{11}) + x_{1}^{0} - \alpha x_{1}^{0}$$

$$x_{1}^{1^{*}} = x_{1}^{0} - \frac{\alpha}{c_{11}}(c_{11}x_{1}^{0} - r_{1} + c_{12}x_{2}^{0} + c_{13}x_{3}^{0})$$
[4-26b]

if,

$$C_1 = c_{11}x_1^0 + c_{12}x_2^0 + c_{13}x_3^0 - r_1$$

then,

$$x_1^{1^*} = x_1^0 - \frac{\alpha}{c_{11}}(C_1)$$
 [4-26c]

$$x_2^{1^*} = x_2^0 - \frac{\alpha}{c_{22}}(C_2)$$
 [4-26d]

$$x_3^{1^*} = x_3^0 - \frac{\alpha}{c_{33}}(C_3)$$
 [4-26e]

For the more general O-D estimation case, Equation [4-27] provides the modified estimated of demand T_{ij} for the current iteration.

where:
$$n_{ij} = \sum_{r} \left(P_{ij}^{r^2} R_r^2 \right)$$

$$C_{ij} = \sum_{r} \left(\left(V_r - V_r' \right) P_{ij}^r R_r^2 \right)$$

$$V_r' = \text{observed flow on link } r \text{ (vph)}$$

$$V_r = \text{estimated flow on link } r \text{ (vph)}$$

$$P_{ij}^r = \text{probability that demand between } i \text{ and } j \text{ will use link } r$$

$$R_r = \text{relative reliability of flow observed on link } r$$

$$\alpha = \text{relaxation factor } (0 < \alpha < 1.0)$$

The format of the proposed iterative method, as expressed in Equation [4-27], is consistent with engineering intuition. First, when C_{ij} is equal to zero for each O-D pair, then the set of linear constraint equations is satisfied and the current estimate of T_{ij} must be optimal. However, if C_{ij} is not equal to zero, then the current demand estimate does not satisfy the constraint equations, implying that T_{ij} is not yet optimal and some change to T_{ij} will result in a reduction of the total squared link flow error.

= iteration number

Iterative changes are made to the current demand estimate by computing the average flow errors across all links that each demand traverses, and subtracting this average error from the current demand estimate. Since the **average** flow error across all links is computed, iterative changes to all contributing demands are allocated in a least biased manner. This approach leads to a solution that approximates the most likely solution when multiple solutions exist and prior information is not available. The relaxation factor α , is used to achieve convergence of the algorithm, even when diagonal dominance does not exist. The complete iterative algorithm is presented in Table 4-4 and is applied to a simple example network in Section 4.5.5.

The proposed modified Jacobi iterative solution method provides several advantages over the Gauss-Seidel method and the standard Jacobi method:

- 1. The convergence criterion for the modified Jacobi method is less restrictive than that of the standard Jacobi method. It will be demonstrated in Section 4.5.5, that the modified Jacobi method converges to a solution for the simple example two-link network examined earlier, for which the standard Jacobi method did not converge.
- 2. The modified Jacobi method is insensitive to the order in which the system of equations is arranged. It was shown in this section that, for the examined example, the Gauss-Seidel method exhibited sensitivity to the order of the equations.

Table 4-4:	Iterative solution algorithm for estimating O-D demands that minimize
	the squared link flow errors

Step	Description	Equation
	Initially, the iteration counter is set to zero	l = 0
1	Use the prior matrix as the initial estimate of the unknown demand matrix for iteration zero	$T_{ij} = t_{ij} \ \forall_{ij}$
2	Assign the current estimate of demand to the network using known link use probabilities	$V_a = \sum_{ij} \left(T_{ij} \cdot P_{ij}^a \right) \forall_a$
3	Compute the sum of the difference between estimated and observed link flows for all links associated with each O-D pair in turn	$C_{ij} = \sum_{r} \left(\left(V_r - V_r' \right) P_{ij}^r R_r^2 \right)$
4	Increment the iteration counter	l = l + 1
5*	Based on the estimated and observed link flows, compute new estimates of all O-Ds	$T_{ij}^{l+1} = T_{ij}^l - \frac{\alpha}{n_{ij}} \left(C_{ij} \right)$
		$n_{ij} = \sum_{r} \left(P_{ij}^{r^2} R_r^2 \right)$
6	Incorporate non-negativity constraints	$if T_{ij}^{l+1} < 0 then T_{ij}^{l+1} = 0$
7	If stopping criterion not yet met, go to Step 2	See section 4.5.4

4.5.4 Algorithm stopping criteria

Since the algorithm proposed in Section 4.5.3 is iterative in nature, criteria must be defined that can be used to determine when sufficient iterations have been carried out to warrant termination of the procedure. An obvious criterion to use is the link flow error function, E. The algorithm has been formulated to ensure that E never increases if more iterations are carried out. Unfortunately, since E is the squared difference between the estimated and observed flows, its magnitude has little physical meaning. Using Equation [4-28], it is possible to determine the average absolute link flow error based on E. The true average absolute link flow error is computed from Equation [4-29].

$$\overline{E} = \sqrt{E/N} \tag{4-28}$$

$$E^{abs} = \frac{1}{N} \sum_{a} |V_a - V_a'|$$
 [4-29]

^{*} If seed O-D demand reliability factors are to be considered, then Equation [4-34] from Section 4.6 must be used instead of the equation presented here.

where: N = number of links over which the error has been computed

 V'_a = observed flow on link a (vph) V_a = estimated flow on link a (vph)

Another stopping criterion, the average link flow error measured as a proportion of the observed flow, can be computed using Equation [4-30]. The true average relative link flow error is determined from Equation [4-31].

$$\overline{E}^{n} = \overline{E}/_{\overline{V}}, \qquad [4-30]$$

$$\overline{E}^{rel} = \frac{1}{N} \sum_{a} \left(\frac{|V_a - V_a'|}{V_a'} \right)$$
 [4-31]

where: N = number of links over which error has been computed

 \overline{V}' = average observed link flow

 V'_a = observed flow on link a (vph) V_a = estimated flow on link a (vph)

It is also advantageous to develop a stopping criterion that is based on the marginal decrease in link flow error. If each subsequent iteration of the algorithm provides a large reduction in the link flow error, then carrying out further iterations is likely to be beneficial. If the marginal reduction in link flow error is very small, then regardless of the absolute or relative measures of error, further iterations are not likely to result in significant reductions in error. An appropriate measure of the marginal reduction in link flow error is obtained by computing the difference between the average relative link flow error of the previous iteration, with that of the current iteration.

$$\Delta \overline{E}^{n} = (\overline{E}^{n})^{j-1} - (\overline{E}^{n})^{j}$$
 [4-32]

where: i = the current iteration number

For practical purposes, it is most useful to use Equation [4-32] as a stopping criterion for the iterative algorithm as the marginal improvement provided by each iteration provides a true measure of the benefit of carrying out a successive iteration. This benefit can be weighed against the cost of carrying out another iteration, and the user can then determine whether the resulting benefit - cost ratio is sufficiently high to warrant further iterations. The cost of carrying out each iteration would typically be measured by the amount of computing time required to carry out the iteration. For the simple example network used to illustrate features of the algorithms and their formulations, computing time is very short. However, as will be demonstrated in Chapter 7, estimating time-varying demands on large networks may require significant computing time to carry out each iteration.

Application of iterative algorithm to a simple network 4.5.5 having multiple solutions

The iterative solution technique is illustrated using the now familiar two-link example network presented earlier in Figure 4-11. Three O-D demands are feasible (T_{AB} , T_{AC} , and T_{BC}) and observed flows are assumed to be $V'_{I}=16$ vph and $V'_{I}=18$ vph. The relaxation factor α is initially assumed to be equal to 0.5.

The system of linear equations that are to be solved are those from Equation [4-24]. Based on Equation [4-27], the following expressions for each unknown O-D demand can be determined. These expressions can be used to generate estimates of the O-D demands in each iteration as provided in Table 4-5.

$$T_{AB}^{1^*} = T_{AB}^0 - \frac{\alpha}{1} \left(T_{AB}^0 + T_{AC}^0 - 16 \right)$$

$$T_{AC}^{1^*} = T_{AC}^0 - \frac{\alpha}{2} \left(T_{AB}^0 + 2T_{AC}^0 + T_{BC}^0 - 34 \right)$$

$$T_{BC}^{1^*} = T_{BC}^0 - \frac{\alpha}{1} \left(T_{BC}^0 + T_{AC}^0 - 18 \right)$$
[4-33]

Table 4-5: Iterative algorithm results for a simple example network

Iteration	T_{AB}	T_{AC}	T_{BC}	Flow on L	ink (vph)	Error
Number	(vph)	(vph)	(vph)	1	2	(vph²)
0	5.00	5.00	5.00	10.00	10.00	100.000
1	8.00	8.5	9.00	16.50	17.50	0.500
2	7.75	8.5	9.25	16.25	17.75	0.125
3	7.63	8.5	9.38	16.13	17.88	0.031
4	7.56	8.5	9.44	16.06	17.94	0.008
5	7.53	8.5	9.47	16.03	17.97	0.002
6	7.52	8.5	9.48	16.02	17.98	0.000
7	7.51	8.5	9.49	16.01	17.99	0.000
8	7.50	8.5	9.50	16.00	18.00	0.000
9	7.50	8.5	9.50	16.00	18.00	0.000
10	7.50	8.5	9.50	16.00	18.00	0.000

After eight iterations, the algorithm has converged to a solution. Comparing this solution to the one computed by the link-count based maximum likelihood model in Chapter 3 ($T_{AB} = 7.53 \text{ vph}$, $T_{AC} = 8.47 \text{ vph}$, $T_{BC} = 9.53 \text{ vph}$), indicates that the two solutions are approximately equal.

Unlike the trip-based maximum likelihood model examined in Chapter 3, the LSE model is insensitive to the magnitude of the seed, so that the same demand is estimated for all uniform seeds. For a non-uniform seed of $T_{AB}=10$ vph, $T_{AC}=10$ vph, and $T_{BC}=4$ vph, the LSE model estimates an O-D demand of $T_{AB}=6.0$ vph, $T_{AC}=10.0$ vph, and $T_{BC}=8.0$ vph. Again, this solution is similar to the maximum likelihood estimates provided by the trip-based entropy model and the link-count based information model in Chapter 3.

It is also instructive to examine the impact that the relaxation factor α has on the final solution. Figure 4-13 illustrates the variation in total link flow error as a function of iteration number for three different levels of α .

Figure 4-13 indicates that, for this simple example, the algorithm converges very quickly when α is equal to 0.5, and somewhat less quickly when α is equal to either 0.1 or 0.95. The relaxation factor acts to prevent the entire correction factor from being utilized when computing the new O-D demand. When α is very small, a small portion of the computed correction factor is used to modify the current O-D estimate. Since only small corrections are

made during each iteration, it would also be expected that the algorithm would converge more slowly. This intuitive expectation is supported by Figure 4-13.

Unfortunately, using this same line of reasoning, it would be expected that as α approaches 1.0, the algorithm would converge more and more rapidly. This expectation is not supported by Figure 4-13. In fact, when α =0.95 the algorithm converges more slowly than when α =0.5 or α =0.1. To provide insight into the cause for these results, we examine the demand, T_{AC} , estimated during each iteration (Figure 4-14).

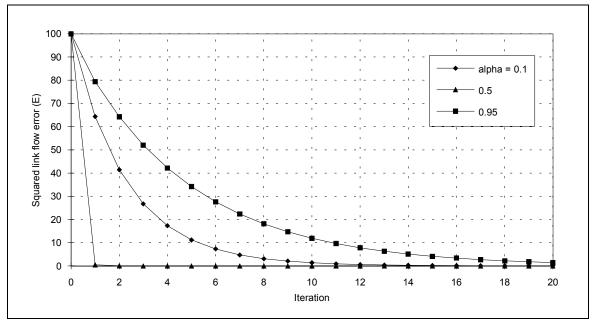


Figure 4-13: Effect of the relaxation factor and the number of iterations on the squared link flow error (Seed matrix = (5,5,5))

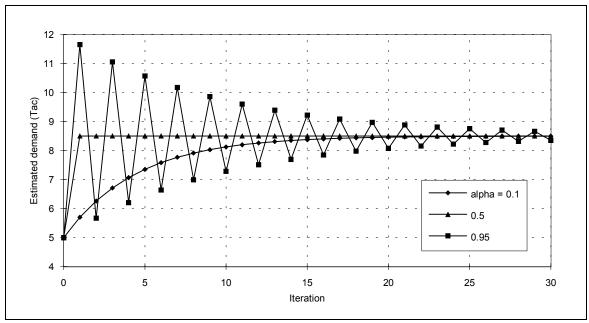


Figure 4-14: Effect of the relaxation factor and the number of iterations on a selected O-D demand - T_{AC} (Seed matrix = (5,5,5))

Figure 4-14 indicates that the algorithm over-estimates the necessary correction when α =0.95. Since the algorithm continually over-estimates the necessary correction, it also converges more slowly. However, when α is equal to either 0.1 or 0.5, a small enough portion of the computed correction is used, so that the algorithm never over-estimates the necessary correction.

This example has indicated that the iterative solution method converges most rapidly when α =0.5. However, this finding is not likely to be true in general as the effect of α on the rate of convergence is likely network dependent. Fortunately, it is a simple matter to calibrate α for each network by carrying out a simple sensitivity analysis and then utilizing the most appropriate value. As the value of α approaches 1.0, there is a danger that if the system of constraints is not diagonally dominant, the algorithm will not converge. This problem is avoided by selecting a smaller value for α such that the algorithm applies only a portion of the computed correction factors.

The automatic determination of an optimal value for α is outside of the scope of this thesis, however, it is hypothesized that a methodology could be created that selects the value of α on the basis of the number of sign reversals observed in the alterations made to each O-D demand. For example, in Figure 4-14, the oscillations that occur when α =0.95 indicate that the convergence could be accelerated by reducing the magnitude of the relaxation factor. This hypothesis is not tested in this thesis.

The LSE model is insensitive to the number of trips contained within the seed matrix. The results that have been presented here are based on a uniform seed matrix having a

magnitude of 5 vph for each O-D demand. The same final O-D estimate would have been obtained for this network regardless of the magnitude of the seed matrix, providing that the matrix was uniform and the relaxation factor was sufficiently small to ensure convergence. In general, when the magnitude of the seed matrix becomes less similar to the magnitude of a feasible solution, more iterations are required to achieve the same level of link flow error.

4.5.6 Application of iterative algorithm to a second simple network

It is instructive to re-examine the example of Section 4.4.1. This example served to illustrate how the unconstrained LSE model formulation can lead, under certain circumstances, to negative demand estimates. It was further illustrated that the simple truncation of any final O-D estimates that are negative may lead to a solution that is not optimal. In this section, the modified Jacobi iterative solution technique, that was provided in Table 4-4, is used to solve for the optimal O-D demands for the example of Section 4.4.1.

Consider the three-link network that was originally presented in Figure 4-1. The observed link flows, $V'_{I} = 15$ vph, $V'_{2} = 10$ vph, and $V'_{3} = 50$ vph, do not exhibit node or path flow continuity. All observed flows are considered to be equally reliable.

As indicated in Figure 4-15, the algorithm's final demand estimate is $T_{AB} = 0$ vph and $T_{AC} = 32.5$ vph with an associated squared link flow error of 712.5 vph². It can be seen that the estimated solution is closer to the unconstrained optimum ($T_{AB} = -5$ vph and $T_{AC} = 35$ vph; E = 675 vph²) than the solution obtained when the unconstrained solution estimate is truncated ($T_{AB} = 0$ vph and $T_{AC} = 35$ vph; E = 725 vph²). The algorithm provides the optimal solution under the constraint that $T_{AB} > 0$ vph and $T_{AC} > 0$ vph. Since link flow continuity does not exist, there does not exist a solution that results in a link flow error of zero.

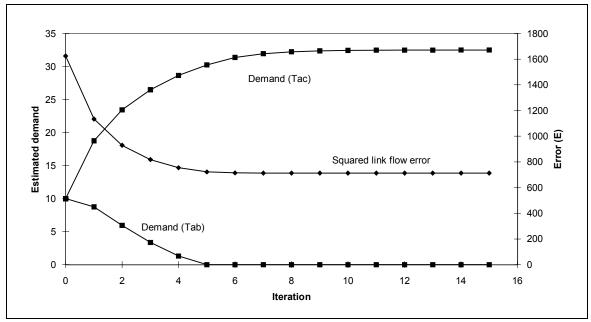


Figure 4-15: Demand estimates and associated squared link flow error as a function of the iteration for a situation in which the unconstrained optimum results in a negative O-D estimate

4.5.7 Application of iterative algorithm to a network with non-uniform link reliability factors

It is instructive to examine the performance of the algorithm when non-uniform link flow reliability factors are considered. This examination is accomplished by applying the LSE model to the three-link example network examined in the previous section. The observed link flows are again assumed to be $V'_1 = 15$ vph, $V'_2 = 10$ vph, and $V'_3 = 50$ vph. It is initially assumed that the flow observed on link three is suspect, and as such has an associated reliability factor of 0.5, while the other two links are considered totally reliable.

Figure 4-16 illustrates the estimated demands, and associated squared link flow error, for the first 15 iterations of the LSE model algorithm. As is evident, the model converges to a solution which provides a minimum squared link flow error. Since link flow continuity does not exist, no O-D solution can exactly replicate the observed link flows, and thus the squared link flow error never reaches zero. After 30 iterations of the algorithm, the O-D estimates of $T_{AB} = 2.5$ vph and $T_{AC} = 20$ vph are different from those estimated in the previous example ($T_{AB} = 0$ vph and $T_{AC} = 32.5$ vph). The associated squared link flow error is also quite different (E = 337.7 vph²; previous example E = 712.5 vph²), as the objective function is also a function of the squared link flow reliability factor.

Figure 4-16 provides results from the algorithm for the specific case when R_3 =0.5. It is instructive to examine the final O-D estimates as a function of the link flow reliability of link

three. Figure 4-17 provides the final estimates of the O-D demands, as well as the associated squared link flow error, over the complete range of flow reliability for link three. It can be seen that the squared link flow error increases as the magnitude of the link flow reliability factor increases. This occurs due to the presence of the link flow reliability factor within the objective function. It can also be seen that for a link flow reliability factor of 1.0, the O-D estimates and the associated squared link flow error, are equal to those determined in the previous section and illustrated in Figure 4-15.

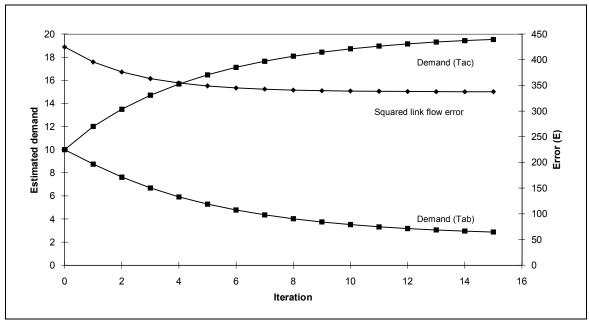


Figure 4-16: Demand estimates and associated squared link flow error as a function of the iteration for a situation in which non-uniform link flow reliability factors exist

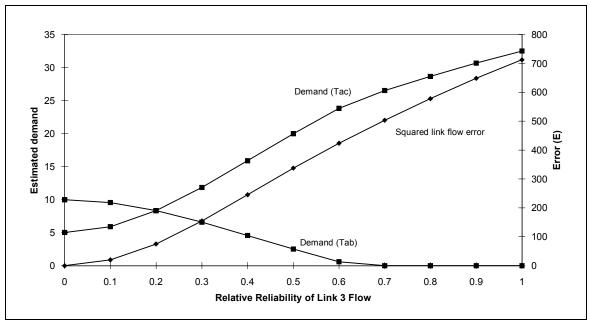


Figure 4-17: Demand estimates and associated squared link flow error as a function of the link flow reliability

4.6 Incorporation of Seed O-D Demand Reliability

4.6.1 Introduction

In practice, some prior O-D demand information is often available when a new O-D matrix needs to be estimated. These data may be somewhat outdated, or be available for only a portion of the network, but these data usually contain some information which may be relevant to the current demand pattern. As such, they could potentially be used to produce more accurate demand estimates when multiple solutions can minimize the objective function.

Until this point, it has been assumed that all the cells in the prior demand matrix are known with an equal degree of certainty. In the case when no prior information exits, all cells are set to have magnitudes which are equal to each other, implying that O-D demands are equally likely. However, the situation in which some portion of the seed matrix is known with certainty, and less is known about the remaining demands, is more difficult to reflect with the current LSE model formulation, as it does not provide a means of explicitly indicating the relative reliability of the prior information.

Before discussing how a seed O-D demand reliability can be incorporated into the LSE model, it is useful to first define what exactly a seed demand reliability factor should accomplish.

The seed O-D reliability factor (R_{ij}) is similar to the link flow reliability factor (R_a) except its subscripts refer to an origin - destination pair, rather than a link. As with the link flow reliability factor, the seed O-D reliability factor is given the range of $0 \le R_{ij} \le 1.0$.

When $R_{ij} = 0$, then the seed O-D demand t_{ij} represents an initial demand estimate which is to have virtually no bearing on the final solution. In contrast, when $R_{ij} = 1.0$, then the seed demand, t_{ij} is considered to be known with the highest level of certainty, and as such, should not be altered to reduce any associated link flow error. When $0 < R_{ij} < 1.0$, the reliability factor represents an intermediate relative reliability of the associated seed demands.

4.6.2 Modification to the solution algorithm

The seed reliability factor is at present incorporated exclusively into the iterative solution algorithm of the LSE model, rather than its formulation. In fact, even without the reliability factor, the seed matrix does not even appear in the model formulation.

The solution algorithm successively modifies an initial seed matrix. In the absence of any seed reliability factors, all O-D demands are modified, with the magnitude of the modification being proportional to the relative magnitude of the seed demand. However, with seed reliability factors, it is desired to modify the algorithm so that the seed demand is successively modified at rates which are in proportion to the relative magnitude of the seed demand and the associated seed reliability factor. Thus, an O-D cell of high reliability, which may require a large modification, will actually only undergo a small change.

To incorporate this into the algorithm, the equation in Step 5 of Table 4-4 is replaced with Equation [4-34].

where:
$$n_{ij} = \sum_{r}^{l} \left(C_{ij} \right) (1 - R_{ij})$$
 [4-34]
 $C_{ij} = \sum_{r} \left(P_{ij}^{r^2} R_r^2 \right)$ $C_{ij} = \sum_{r} \left((V_r - V_r') P_{ij}^r R_r^2 \right)$ $V_r' = \text{observed flow on link } r \text{ (vph)}$ $V_r = \text{estimated flow on link } r \text{ (vph)}$ $P_{ij}^r = \text{probability that demand between } i \text{ and } j \text{ will use link } r$ $R_r = \text{relative reliability of flow observed on link } r$ $R_{ij} = \text{relative reliability of demand between } i \text{ and } j$ $\alpha = \text{relaxation factor } (0 < \alpha < 1.0)$ $\alpha = \text{iteration number}$

4.6.3 Example application considering the relative reliability of seed O-D demands

To illustrate the impact of relative seed reliability, consider the four-link network in Figure 4-18 having four possible O-D demands (T_{AC} , T_{AD} , T_{BC} , T_{BD}), and assume the observed flows are $V'_1 = 100$ vph, $V'_2 = 100$ vph, $V'_3 = 80$ vph, and $V'_4 = 120$ vph.

Using this example network, five different scenarios will be examined. For each scenario, the LSE model will be used to estimate an O-D solution. Table 4-6 provides a summary of the scenario configurations and the resulting estimated demands.

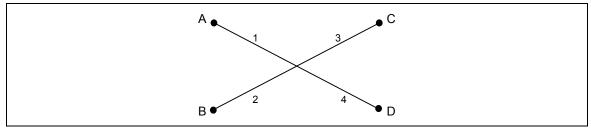


Figure 4-18: Example four-link network used to illustrate the impact of the relative reliability of seed O-D demands

Table 4-6: Scenario configurations and estimated demands used to illustrate the impact of seed O-D demand reliability factors

Scenario	Seed O-D (vph)				Seed Reliability				Estimated O-D (vph)				
	t _{AC}	t _{AD}	t _{BC}	t _{BD}	R_{AC}	R_{AD}	R _{BC}	R_{BD}	T_{AC}	T_{AD}	T _{BC}	T_{BD}	
1	10	10	10	10	0	0	0	0	40.0	60.0	40.0	60.0	
2	5	10	1	15	0	0	0	0	42.3	57.7	37.7	62.3	
3a	5	10	1	15	1	0	0	0	5.0	95.0	75.0	25.0	
3b	5	10	1	15	0	1	0	0	85.0	10.0	0.0	105.0	
3c	5	10	1	15	0	0.2	0	0	45.1	54.9	34.9	65.1	

As indicated in Table 4-6, when no prior information is known, the LSE model estimates a demand for Scenario 1 that is the same as the maximum likelihood solution. The demand estimate for Scenario 2, when prior information of uniform reliability exists, is, as expected, not equal to estimate of Scenario 1. In Scenario 1, an O-D estimate is made in the absence of any prior information. In Scenario 2, knowledge of a prior demand matrix is used as a starting estimate and modified to reflect the existing link flows.

Scenario 3a illustrates the impact of the prior O-D demand estimate, t_{AC} , being known with certainty. Consistent with the intent of seed reliability factors, the estimated demand has preserved the magnitude of the prior demand t_{AC} , since this value was known to be completely accurate. Scenario 3b indicates this same result when another seed demand is known with certainty. Scenario 3c illustrates the impact of knowing that a seed demand is somewhat more reliable than the other seed demands. It would be expected that the estimate of T_{AD} would fall somewhere between 57.7 vph, the value obtained when the seed in not known with any degree of confidence, and 10.0 vph, the value obtained when the seed demand is known with certainty. The estimate of 54.9 vph for T_{AD} does fall between these two limits and, as the seed was known with only a small degree of certainty, the estimated value is more similar to the value obtained when $R_{AD} = 0$ than when $R_{AD} = 1.0$.

Figure 4-19 illustrates the estimates of T_{AD} for five different levels of R_{AD} for each iteration. As expected, when the prior estimate of the demand between zones A and D is known to be completely accurate ($R_{AD} = 1.0$), the final estimate of the demand between

zones A and D is equal to the prior estimate ($T_{AD} = 10$ vph). When $R_{AD} = 0$, all prior information is known to the same degree of certainty, and as indicated in Scenario 2 of Table 4-6, the estimate for the demand $T_{AD} = 57.7$ vph. If the prior knowledge of the demand between zones A and D ($t_{AD} = 10$ vph) is not perfectly reliable, but is known with a high level of confidence (i.e. $R_{AD} = 0.75$), then it would be expected that the final estimate of T_{AD} would not be very different from the prior estimate of 10 vph. For this example, when $R_{AD} = 0.75$, T_{AD} is estimated to be 37.3, vph which, though different from 10 vph (when $T_{AD} = 0.75$), is also quite different from 57.7 vph, the value of T_{AD} when $T_{AD} = 0.75$

Figure 4-20 illustrates the similarity of the estimated demand T_{AD} to the seed demand t_{AD} as a function of the seed reliability factor R_{AD} . The similarity between the estimated and seed demand is measured by the relative distance that the estimated demand is from the seed demand (Equation [4-35]).

Similarity Index =
$$\frac{T_{ij}^{R_{ij}=0} - T_{ij}}{T_{ij}^{R_{ij}=0} - T_{ij}^{R_{ij}=1}}$$
 [4-35]

It is clear that the relationship between the seed reliability factor and the Similarity Index, in this example, is non-linear. However, this relationship is likely to be dependent on network conditions, so that in general, the relationship between R_{ij} and the Similarity Index is not known *a priori*. The effect of R_{ij} is a function of how quickly the other available O-D demands can be altered from their prior estimate values to converge to a solution that minimizes the objective function. If all other O-D demands are constrained, then any value of R_{ij} less than 1.0 results in the same final solution. The only impact of that R_{ij} has under these conditions is to change the number of iterations required by the algorithm to converge.

Figure 4-21 illustrates the final O-D estimates for each of the four possible demands, for the complete range of values of R_{AD} . Here again, the impact of R_{AD} on T_{AD} can be seen. As more confidence is placed on the prior estimate of T_{AD} , the closer the final estimate is to the prior estimate. In this case, for each value of R_{AD} , the four estimated demands represent a feasible solution which exactly replicates the observed link flow.

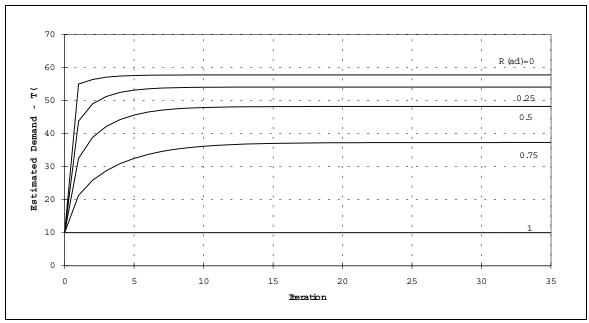


Figure 4-19: Intermediate estimates of O-D demand T_{AD} for different levels of seed O-D demand reliability

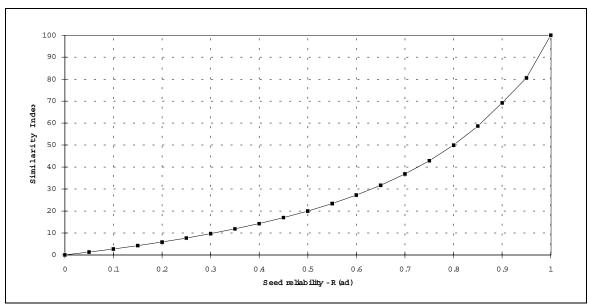


Figure 4-20: Similarity of estimated demand (T_{AD}) to seed demand (t_{AD}) as a function of seed reliability factor (R_{AD})

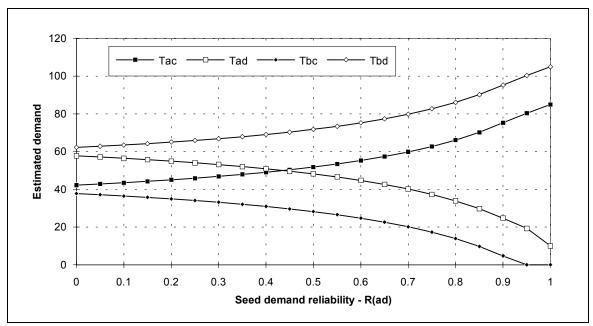


Figure 4-21: Final estimates of O-D demands for different levels of seed O-D demand reliability (t_{AD})

4.7 Impact of Redundant Link Flows

In Chapter 3, it was shown that the link-count based maximum likelihood model was sensitive to the presence of redundant link flow information, while the trip-based model was not. Utilizing the four-link network presented in Figure 4-18, the results in Table 4-3 indicate that the LSE model estimates are sensitive to the presence of redundant link flow information. However, as discussed in Section 3.6.3, since link flow continuity rarely exists in field data, link flow data that are completely redundant rarely exist, and the LSE model's sensitivity to redundant link flow data, is not of great concern.

Table 4-7: LSE model O-D estimates for a four link network with redundant link flows

O-D	Trip	LSE Model										
	Based	Links having observed flows										
	Model	all	1,2,3	1,2,4	1,3,4	2,3,4						
A-C	12.5	12.5	11.67	13.33	11.67	13.33						
A-D	12.5	12.5	11.67	13.33	13.33	11.67						
B-C	10.5	10.5	9.67	11.33	11.33	9.67						
B-D	17.5	17.5	16.67	18.33	16.67	18.33						

4.8 Summary

This chapter has described the development of a model based on the assumption that estimated O-D demands should minimize the squared link flow errors. The resulting model is comparable to an unconstrained multiple linear regression model. It has been shown that this formulation permits negative demand estimates, and that truncation of negative estimates can result in the acceptance of sub-optimal solutions.

It has also been shown that observed link flow data are often obtained from different sources and are known with different levels of confidence. A link flow reliability factor was incorporated into the model formulation that performs a function which is analogous to that of the weights in a weighted least squares regression. Two approaches for computing these factors have been suggested. The first utilizes the inverse of the variance of link flows computed from a time series of observations, while the second is heuristic and uses the magnitude of flow discontinuities at nodes to determine link flow reliability factors.

The model was illustrated analytically for four different simple network examples. These network examples considered link flow continuity and discontinuity, multipath routings, and link flow reliability.

The formal solution to the model formulation was given in matrix notation. It was noted that a solution exists only if the inverse to the $[P^tP]$ matrix could be found. In general, it cannot be stated that this inverse always exists. In particular, when multiple solutions exist, $[P^tP]$ is singular and cannot be inverted.

An iterative solution algorithm was proposed which is applicable regardless of the number of feasible solutions,. This algorithm, which is a modification of the Jacobi iterative solution technique, successively modifies an initial seed matrix based on the differences between observed and estimated link flow errors resulting from the previous iteration. The iterative technique incorporates a relaxation factor in order to achieve convergence in situations when diagonal dominance does not exist in the system of constraint equations.

This algorithm was applied to a simple example problem having multiple feasible solutions. The algorithm was shown to converge successfully and, when the seed matrix was uniform, led to a solution that closely approximated the most likely solution. It was also shown that the algorithm provides O-D estimates that are optimal under non-negativity constraints.

Since, in practice, prior information often has varying degrees of reliability associated with it, seed O-D reliability factors were introduced into the algorithm. Using a simple example network, the effect of using these seed O-D demand reliability factors was shown to be consistent with intuition.

Chapter 5 presents another model which is based on a procedure for minimizing the relative link flow error rather than the absolute link flow differences. The presentation of the development of this model parallels the presentation in this chapter. In Chapter 6, both models are extended to permit the estimation of time varying O-D demands. Subsequently, Chapter 7 will describe the application of both models to a 35 kilometer section of freeway in Toronto, Ontario, for the estimation of time varying demands.

CHAPTER 5

DEVELOPMENT OF A LEAST RELATIVE ERROR O-D ESTIMATION MODEL

5.1 Introduction

The least relative error (LRE) model described in this chapter is based on the premise that the estimated O-D should minimize some measure of the relative link flow differences. As will be shown, this premise leads to a solution algorithm that iteratively corrects an initial seed matrix through successive multiplication of correction factors. This method does not require observed link flows to exhibit either node or path flow continuity.

This chapter first describes the development of the LRE model, following which the mathematical basis of the model is presented. The associated link flow error function is defined and illustrated, and the model formulation is illustrated using the same three simple example network scenarios that were examined in Chapter 4. These scenarios serve to illustrate the model's ability to function correctly when flow continuity, flow discontinuity, and multipath routes are considered. Characteristics of the model formulation are identified and examined. It is shown that a formal solution to the model formulation cannot be solved for algebraically, but that an iterative solution algorithm can solve the model formulation.

5.2 Development of the LRE Model

In the previous chapter, the mathematical basis for the LSE model was presented. The premise for the LSE model is that the best unbiased estimate of a target O-D matrix can be defined as the one which minimizes the squared errors between a set of observed and estimated link flows. However, this premise leads to a model formulation that may lead to negative demand estimates, and which considers only the magnitude of the link flow errors,

not their size relative to the observed (or true) flow. The modified Jacobi iterative solution method used to solve the LSE model formulation prevented the estimation of negative demands through the introduction of non-negativity constraints, and required the introduction of a relaxation factor, which, if insufficiently small, prevented convergence of the algorithm. In this chapter, a model is formulated that incorporates non-negativity constraints, considers a different error function, and does not require the introduction of a relaxation factor within the solution algorithm.

5.2.1 **Mathematical basis**

The mathematical derivation of the relative link flow error model is initiated by defining a suitable relative link flow error function. This function, provided in Equation [5-1], explicitly reflects the relative size of link flow errors, and attributes the same amount of error to the objective function when under-estimating link flows as when overestimating flows.

$$E_R = \sum_{a} \left[\ln \left(\frac{V_a}{V_a'} \right) \right]^2 \tag{5-1}$$

where: a = link identifier

 E_R = total relative link flow error (dimensionless) V_a = estimated flow on link a (vph) V'_a = observed flow on link a (vph)

Figure 5-1 illustrates the change of E_R as a function of the ratio of estimated to observed flow. It can be noted that E_R is only sensitive to the ratio of the flows. For example, the contribution to the error function when the estimated flow is five times less than the observed flow (Point A), is the same as when an estimated flow is five times larger than the observed flow (Point B). Therefore, the error function does not differentiate between underestimation or overestimation of the estimated flow.

It should also be noted that when the estimated and observed flows on a given link are equal, then the contribution of that link to the error function is equal to zero. This is a desirable characteristic as it provides an appropriate scale in which the global minimum error is achieved when $E_R = 0$.

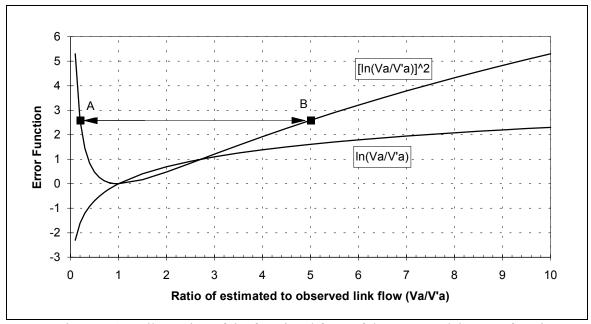


Illustration of the functional form of the LRE model's error function

Having defined a suitable error function, the objective is to identify a set of solutions that minimizes E_R . Since E_R is a continuous convex function, its minimum can be found by computing the partial derivative of E_R with respect to each unknown O-D demand (T_{ij}) in turn, and setting each resulting equation equal to zero (Equation [5-2]). On the basis of Equation [5-3], Equation [5-2] can be transformed to provide Equation [5-2].

$$\frac{\partial}{\partial} \frac{E_R}{T_{ij}} = \sum_{a} \left[2 \ln \left(\frac{V_a}{V_a'} \right) \frac{\partial \ln \left(\frac{V_a}{V_a''} \right)}{\partial T_{ij}} \right] = 0$$
 [5-2]

$$\frac{d}{dx}\ln u = \frac{1}{u}\frac{du}{dx} \tag{5-3}$$

$$\frac{\partial}{\partial} \frac{E_R}{T_{ij}} = \sum_{a} \left[2 \ln \left(\frac{V_a}{V_a'} \right) \frac{1}{V_a} \frac{\partial}{\partial} \frac{V_a}{T_{ij}} \right] = 0$$
 [5-4]

a = link identifierwhere:

 E_R = total relative link flow error (dimensionless)

 V_a = estimated flow on link a (vph) V'_a = observed flow on link a (vph)

 T_{ii} = demand between origin *i* and destination *j* (vph)

Equation [5-3] expresses the estimated link flow V_a , in terms of the as yet unknown demands. Since each O-D demand (T_{ij}) can appear in Equation [5-4] not more than once, the partial derivative of V_a with respect to some unknown demand T_{ii} is simply the proportion of that demand contributing to the flow on link a (Equation [5-4]).

$$V_a = \sum_{ij} \left(T_{ij} P_{ij}^a \right) \tag{5-5}$$

$$\frac{\partial V_a}{\partial T_{ii}} = P_{ij}^a \tag{5-6}$$

where: a = link identifier

 V_a = estimated flow on link a (vph) T_{ij} = demand between origin i and destination j (vph) P_{ij}^a = probability that demand between i and j will use link a

Substituting Equation [5-4] into Equation [5-2] results in a system of N non-linear equations in terms of N unknown O-D demand traffic demands (Equations [5-5a] and [5-5b]). One may note, however, that not all of these N equations are necessarily independent, implying that, similar to the LSE model, the existence of a unique solution is not guaranteed.

$$\frac{\partial}{\partial} \frac{E_R}{T_{ij}} = \sum_a \left[2 \frac{P_{ij}^a}{V_a} \ln \left(\frac{V_a}{V_a'} \right) \right] = 0$$
 [5-7a]

$$\frac{\partial}{\partial} \frac{E_R}{T_{ij}} = \sum_a \left[\ln \left(\frac{V_a}{V_a'} \right)^{P_{ij}^a / V_a} \right] = 0$$
 [5-7b]

Since $e^{\Sigma x} = \prod e^x$, it is possible to transform Equation [5-5b] into Equation [5-5c] by raising each side of the equation to the exponent e.

$$\frac{\partial}{\partial} \frac{E_R}{T_{ij}} = \prod_a \left(\frac{V_a}{V_a'}\right)^{P_{ij}^a/V_a} = 1$$
 [5-7c]

where: a = link identifier

 V_a = estimated flow on link a (vph) V'_a = observed flow on link a (vph) T_{ij} = demand between origin i and destination j (vph)

 P_{ii}^{a} = probability that demand between i and j will use link a

A closer inspection of Equation [5-7c] indicates that, when considering some link a, that is not part of any route between origin i and destination j (i.e., $P_{ij}^a = 0$), there is no contribution to the product. Therefore, Equation [5-7c] can be simplified such that the product need be across only those links which are on a path between origin i and destination j.

$$\frac{\partial}{\partial} \frac{E_R}{T_{ij}} = \prod_r \left(\frac{V_r}{V_r'}\right)_{j}^{P_{ij}^r/V_r} = 1$$
 [5-7d]

where: r = any link on a route between origin i and destination j

Two important characteristics of Equation [5-7d] can be identified:

- 1. Any O-D demand matrix, that results in estimated link flows that are exactly equal to the observed flows, satisfies the system of non-linear constraints.
- 2. If the observed flow on a link is zero and this link is utilized by an O-D, then either this particular O-D's demand is equal to zero, or the observed flow is erroneous, or the assumed route (and thus P_{ii}^a) is incorrect. Therefore, in the event that an observed link flow of zero is encountered, two courses of action can be taken. One can choose to ignore the link entirely, implying that little confidence is placed in the observed flow of zero. Second, one can impart some small but arbitrary non-zero, finite flow to the link. If this imparted flow is small, (i.e. 1 vph) it will tend to force the demands utilizing that link to become much smaller. These approaches, and their impacts, are discussed in more detail later in this thesis.

The LRE model formulation leads to a system of non-linear equations described by Equation [5-8]. This system of non-linear constraints does not require estimated link flows to exactly match those observed, implying that neither node flow nor path flow continuity need exist. This permits the LRE model to be applied to actual networks using field link flow data, something which was not possible always with the formulations of Van Zuylen and Willumsen (1980) that were examined earlier.

$$\prod_{r} \left(\frac{V_r}{V_r'} \right)^{P_{ij}^r/V_r} = 1 \forall_{ij}$$
 [5-8]

where: r = any link on a route between origin i and destination j

 V_r = estimated flow on link r (vph) V'_r = observed flow on link r (vph)

 P_{ii}^{a} = probability that demand between *i* and *j* will use link *r*

5.2.2 **Example with link flow continuity**

It is useful to illustrate the effect of the above system of constraints using the same examples that were used in Chapter 4 to illustrate the LSE model. For convenience an illustration of each network is again provided.

Consider the simple three-link network in Figure 5-1 having two origin - destination demands. Based on Equation [5-8], the system of non-linear constraints for this example consists of the following two non-linear expressions:

$$\left(\frac{T_{AB} + T_{AC}}{V_1'}\right)^{(T_{AB} + T_{AC})^{-1}} \times \left(\frac{T_{AB}}{V_2'}\right)^{(T_{AB})^{-1}} = 1$$

$$\left(\frac{T_{AB} + T_{AC}}{V_1'}\right)^{(T_{AB} + T_{AC})^{-1}} \times \left(\frac{T_{AC}}{V_3'}\right)^{(T_{AC})^{-1}} = 1$$

Unfortunately, unlike the linear constraints developed in Chapter 4, these non-linear constraints cannot be explicitly solved for algebraically to yield a closed form solution for T_{AB} and T_{AC} . Therefore, the appropriateness of the model formulation will first be illustrated by enumerating all possible O-D combinations within a reasonable range. Having demonstrated the relevance of the LRE model formulation for a simple network, the problem of solving the formulation in a manner that is more efficient and practical than performing a complete enumeration will be examined next.

Since the above network provides a uniquely utilized link for each of the two unknown O-D demands, it is a trivial task to determine the correct O-D matrix when node and path flow continuity exist. For example, if the observed link flows are V'_I =50 vph, V'_2 =20 vph, and V'_3 =30 vph, then intuitively it can be seen that the unknown demands must be T_{AB} =20 vph and T_{AC} =30 vph. Since it is not possible to algebraically solve for T_{AB} and T_{AC} , we compute the left hand side of the constraint equations for a number of possible O-D demands. The O-D demand that satisfies the constraints is the one that results in both constraint equations equaling 1.0. These results, provided in Table 5-1a, indicate that the only demand that simultaneously satisfies both of the constraint equations is T_{AB} =20 vph and T_{AC} =30 vph, which is the correct solution. Table 5-1b indicates that this solution also represents a minimum to the objective function of Equation 5-1.

This simple example indicates that when a single unique O-D solution exists that exactly replicates the observed link flows, the LRE model formulation identifies the correct solution. The estimation of O-D demands, when there exists only one single solution that replicates the observed link flows, is the simplest O-D estimation problem. Having shown the validity of the LRE model formulation for this simple case, we will examine in the next section the application of the LRE model to a situation in which there does not exist an O-D demand that exactly replicates the observed link flows.

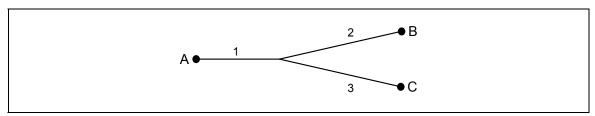


Figure 5-2: Example three-link network

	demands (vpn)												
T _{AC}					T,	AΒ		·					
		5	1	0		5	2	0	25	5			
	AB	AC	AB	AC	AB	AC	AB	AC	AB	AC			
5	0.645	0.595	0.861	0.645	0.937	0.668	0.973	0.680	0.992	0.11			
10	0.699	0.827	0.891	0.856	0.954	0.871	0.983	0.881	0.999	0.48			
15	0.724	0.912	0.908	0.929	0.964	0.939	0.990	0.945	1.003	0.70			
20	0.737	0.953	0.917	0.963	0.971	0.970	0.994	0.974	1.007	0.83			
25	0.745	0.976	0.924	0.983	0.976	0.987	0.998	0.990	1.009	0.91			
30	0.750	0.990	0.928	0.994	0.979	0.998	1.000	1.000	1.011	0.96			
35	0.754	0.999	0.931	1.002	0.981	1.004	1.002	1.006	1.012	1.00			
40	0.756	1.005	0.933	1.007	0.983	1.009	1.003	1.010	1.013	1.02			
45	0.758	1.009	0.935	1.011	0.984	1.012	1.004	1.013	1.014	1.04			
50	0.759	1.012	0.936	1.013	0.985	1.014	1.005	1.015	1.014	1.05			
55	0.760	1.014	0.937	1.015	0.986	1.016	1.005	1.017	1.015	1.06			
60	0.761	1.016	0.938	1.016	0.986	1.017	1.006	1.018	1.015	1.07			
65	0.762	1.017	0.938	1.017	0.987	1.018	1.006	1.018	1.016	1.07			
70	0.762	1.018	0.939	1.018	0.987	1.019	1.007	1.019	1.016	1.08			
75	0.762	1.018	0.939	1.019	0.987	1.019	1.007	1.019	1.016	1.08			

Table 5-1a: Left hand side of non-linear constraints computed for several O-D demands (vph)

Table 5-1b: Relative link flow error (E_R) for all potential O-D solutions in the vicinity of the analytical solution

T _{AC}				T _{AB}			
	5	10	15	20	25	30	35
5	7.7225	5.1404	4.1328	3.6909	3.5211	3.5020	3.5734
10	4.5783	2.5270	1.7702	1.4679	1.3840	1.4211	1.5312
15	3.2419	1.4414	0.8242	0.6077	0.5800	0.6560	0.7936
20	2.5667	0.9058	0.3744	0.2142	0.2253	0.3288	0.4867
25	2.2160	0.6409	0.1658	0.0443	0.0830	0.2067	0.3797
30	2.0490	0.5302	0.0939	0.0000	0.0589	0.1976	0.3820
35	1.9954	0.5153	0.1065	0.0328	0.1068	0.2570	0.4501
40	2.0157	0.5632	0.1746	0.1160	0.2014	0.3604	0.5603
45	2.0862	0.6539	0.2804	0.2332	0.3274	0.4932	0.6985
50	2.1918	0.7746	0.4125	0.3742	0.4751	0.6462	0.8557
55	2.3225	0.9167	0.5634	0.5318	0.6381	0.8134	1.0261

5.2.3 Example with link flow discontinuity

A less trivial O-D estimation task than the one in Section 5.2.2 is the determination of the most appropriate O-D matrix when path and node flow continuity do not exist. Using the same three-link network used in Section 5.2.2, one may consider the observed link flows of V'_I =50 vph, V'_2 =35 vph, and V'_3 =30 vph. Due to the lack of flow continuity, no O-D demand can be found that exactly replicates the observed link flows, and it is not immediately apparent what O-D matrix should be chosen in order to minimize the total relative link flow error. Again, the LRE model formulation is applied by computing the left hand side of the constraint equations for a number of possible O-D demands. The results of these computations, which are provided in Table 5-1a, indicate that the only demand that simultaneously satisfies both of the constraint equations is T_{AB} =32 vph and T_{AC} =28 vph. The application of these demands to the network results in estimated link flows of V_I =60 vph, V_2 =32 vph, and V_3 =28 vph, which do not replicate those observed, but do provide the lowest possible relative link flow error. The latter fact is illustrated in Table 5-1b, which provides the objective function values for each candidate O-D matrix.

	demands for a network exhibiting fink now discontinuity (vph)													
T _{AC}					T,	AΒ								
	3	0	3	1		2	3	3	3	4				
	AB	AC	AB	AC	AB	AC	AB	AC	AB	AC				
25	0.997	0.994	0.998	0.995	0.999	0.995	1.001	0.995	1.002	0.996				
26	0.997	0.997	0.998	0.997	1.000	0.997	1.001	0.997	1.002	0.998				
27	0.997	0.998	0.999	0.999	1.000	0.999	1.001	0.999	1.002	0.999				
28	0.997	1.000	0.999	1.000	1.000	1.000	1.001	1.001	1.003	1.001				
29	0.998	1.002	0.999	1.002	1.000	1.002	1.002	1.002	1.003	1.003				
30	0.998	1.003	0.999	1.003	1.001	1.003	1.002	1.004	1.003	1.004				
31	0.998	1.004	1.000	1.005	1.001	1.005	1.002	1.005	1.003	1.005				
32	0.998	1.006	1.000	1.006	1.001	1.006	1.002	1.006	1.003	1.006				
33	0.999	1.007	1.000	1.007	1.001	1.007	1.002	1.007	1.004	1.007				
34	0.999	1.008	1.000	1.008	1.001	1.008	1.003	1.008	1.004	1.008				
35	0.999	1.008	1.000	1.009	1.002	1.009	1.003	1.009	1.004	1.009				
36	0.999	1.009	1.000	1.009	1.002	1.010	1.003	1.010	1.004	1.010				
37	0.999	1.010	1.001	1.010	1.002	1.010	1.003	1.011	1.004	1.011				
38	0.999	1.011	1.001	1.011	1.002	1.011	1.003	1.011	1.004	1.011				
39	1.000	1.011	1.001	1.012	1.002	1.012	1.003	1.012	1.004	1.012				

Table 5-2a: Left hand side of non-linear constraints computed for several O-D demands for a network exhibiting link flow discontinuity (vph)

The squared natural logarithm of the relative link flow error (E_R) associated with the suggested solution is computed to be 0.0460. To illustrate that the above suggested solution minimizes the relative link flow error, an enumeration of the link flow error of all potential O-D solutions located near the suggested optimum can be carried out. The results of this enumeration, presented in Table 5-1b, indicate that the suggested solution has the smallest associated relative link flow error. Figure 5-3 graphically portrays these data to illustrate that the objective function is concave in the vicinity of the optimum solution.

Table 5-1b: Relative link flow error (E_R) for all potential O-D solutions in the vicinity of the analytical solution for a network exhibiting link flow discontinuity

T _{AC}		T _{AB}											
	30	31	32	33	34	35	36						
25	0.0661	0.0608	0.0584	0.0587	0.0615	0.0665	0.0736						
26	0.0571	0.0524	0.0505	0.0513	0.0546	0.0600	0.0675						
27	0.0520	0.0479	0.0465	0.0478	0.0515	0.0574	0.0653						
28	0.0506	0.0469	0.0460	0.0478	0.0519	0.0582	0.0665						
29	0.0523	0.0491	0.0487	0.0509	0.0554	0.0621	0.0708						
30	0.0570	0.0543	0.0543	0.0569	0.0618	0.0688	0.0779						
31	0.0644	0.0621	0.0625	0.0655	0.0708	0.0782	0.0875						
32	0.0742	0.0723	0.0731	0.0765	0.0821	0.0898	0.0995						
33	0.0863	0.0848	0.0859	0.0896	0.0956	0.1036	0.1136						
34	0.1004	0.0992	0.1008	0.1048	0.1111	0.1194	0.1297						
35	0.1164	0.1156	0.1174	0.1218	0.1283	0.1370	0.1475						

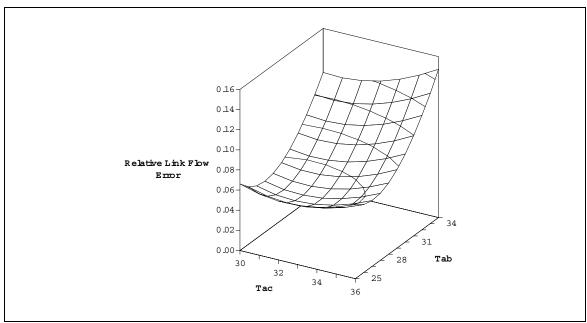


Figure 5-3: Graphical illustration of the concavity of the relative link flow error function in the vicinity of the optimal solution

Since it is not clear what the correct solution to this example should be, it is of interest to compare the demand estimated by the LRE model with the demand estimated by the LSE model (Section 4.2.3). The LSE model indicates that the optimum demand is T_{AB} =30 vph and T_{AC} =25 vph, while the LRE model indicates that the optimum demand is T_{AB} =32 vph and T_{AC} =28 vph. It is not possible to state categorically which model provides a more accurate O-D estimate, as the most accurate solution cannot be defined in absolute terms since this depends on the measure of error that is chosen. It is also interesting to note that the LRE model estimates a total of 60 trips, while the LSE only estimates a total of 55 trips. The LRE model does not, however, always estimate more trips than the LSE model, as can be readily demonstrated by a very simple example.

Consider a two-link linear network having one O-D demand, T, and observed flows of $V'_1 = 20$ and $V'_2 = 40$. The LSE model formulation expressed in Equation [4-6] leads to the following linear expression:

$$(T - V'_1) + (T - V'_2) = 0$$

 $T = (V'_1 + V'_2)/2 = (20 + 40)/2 = 30$

The LRE model formulation expressed in Equation [5-8] leads to the following non-linear expression:

$$\left(\frac{T}{V_1'}\right)^{T^{-1}} \times \left(\frac{T}{V_2'}\right)^{T^{-1}} = 1$$

$$\left(\frac{T^2}{V_1 V_2'}\right)^{T^{-1}} = 1$$

$$T = \sqrt{V_1 V_2'} = \sqrt{20 \cdot 40} = 28.3$$

In contrast to the previous example, the number of trips estimated by the LRE model (28.3) in this example, is less than the number estimated by the LSE model (30). Thus, no general statement can be made concerning the total number of trips estimated by the LRE model relative to the LSE model.

5.2.4 Example with multipath routes

The illustration of the appropriateness of the proposed system of non-linear constraints has been limited thus far to an example in which all-or-nothing routes have been assumed. It is instructive to demonstrate the appropriateness of the constraints for a simple example in which multipath routes are assumed. Consider the network in Figure 5-2 in which there exists two routes between zones A and C. Assume that the observed link flows are $V_1'=30$ vph, $V_2'=20$ vph, $V_3'=10$ vph, and $V_4'=20$ vph and that 2/3 of the traffic demand between zones A and C use link 4.

Since the LRE formulation expressed in Equation [5-8] leads to a system of non-linear constraints that cannot be solved analytically, the left hand sides of the two constraint equations are computed for a range of potential O-D demand solutions and presented in Table 5-3. These results indicate that, despite the multiple paths between zones A and C, the solution space contains only a single solution. Since the above network provides a uniquely utilized link for each of the two unknown O-D demands, and since node and path flow continuity exist in the observed link flow data, the O-D demand of T_{AB} =20 vph and T_{AC} =30 vph satisfies the constraint equations. Once again, the proposed model formulation leads to the selection of the correct O-D demand.

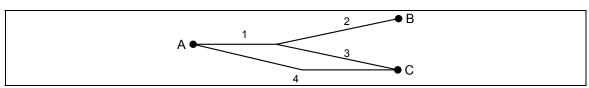


Figure 5-4: Example four-link network exhibiting multiple paths between an origin and destination zone

	demands for a network exhibiting multipath routes (vph)													
T _{AC}					Τ,	AB								
	1	8	1	9		.0	2	1	2	2				
	AB	AC	AB	AC	AB	AC	AB	AC	AB	AC				
25	0.989	0.984	0.994	0.984	0.998	0.985	1.002	0.985	1.005	0.986				
26	0.990	0.988	0.994	0.988	0.998	0.989	1.002	0.989	1.005	0.989				
27	0.990	0.991	0.995	0.991	0.999	0.992	1.002	0.992	1.005	0.993				
28	0.991	0.994	0.995	0.994	0.999	0.995	1.003	0.995	1.006	0.996				
29	0.991	0.997	0.996	0.997	1.000	0.998	1.003	0.998	1.006	0.998				
30	0.992	0.999	0.996	1.000	1.000	1.000	1.003	1.000	1.006	1.001				
31	0.992	1.001	0.997	1.002	1.000	1.002	1.004	1.003	1.007	1.003				
32	0.993	1.004	0.997	1.004	1.001	1.004	1.004	1.005	1.007	1.005				
33	0.993	1.005	0.997	1.006	1.001	1.006	1.004	1.006	1.007	1.007				
34	0.993	1.007	0.998	1.008	1.001	1.008	1.005	1.008	1.008	1.008				
35	0.994	1.009	0.998	1.009	1.002	1.009	1.005	1.010	1.008	1.010				
36	0.994	1.010	0.998	1.011	1.002	1.011	1.005	1.011	1.008	1.011				
37	0.995	1.012	0.999	1.012	1.002	1.012	1.005	1.012	1.008	1.013				
38	0.995	1.013	0.999	1.013	1.003	1.013	1.006	1.014	1.009	1.014				
39	0.995	1.014	0.999	1.014	1.003	1.015	1.006	1.015	1.009	1.015				

Table 5-3: Left hand side of non-linear constraints computed for several O-D demands for a network exhibiting multipath routes (vph)

5.3 **Incorporation of Link Flow Reliability**

It is also desirable to incorporate a relative link flow reliability factor (R_a) into the LRE model formulation. This is accomplished in much the same fashion as was done with the LSE model discussed in Chapter 4. Specifically, the relative reliability of observed link flows is incorporated into the error function expressed in Equation [5-1], to produce Equation [5-9].

Following the same development as presented in Section 5.2.1, it can be shown that minimizing the error function which is presented in Equation [5-9] results in the system of non-linear constraints expressed by Equation [5-10].

$$E_R = \sum_{a} \left[\ln \left(\frac{V_a}{V_a'} \right) R_a \right]^2$$
 [5-9]

$$\frac{\partial}{\partial} \frac{E_R}{T_{ij}} = \prod_r \left(\frac{V_r}{V_r'}\right)^{R_r^2 P_{ij}^r / V_r} = 1$$
 [5-10]

where: a = link identifier

 E_R = total relative link flow error

 V_a = estimated flow on link a (vph) V_a' = observed flow on link a (vph)

 R_a = relative link flow reliability (0 $\leq R_a \leq 1.0$)

Characteristics of Formulation 5.4

Similar to the analysis carried out in Chapter 4 for the LSE model, this section will examine some of the more interesting characteristics of the LRE model formulation. In the examination of the LSE model, it was stated that the formulation's ability to estimate negative demands, in the absence of explicit non-negativity constraints, was undesirable. It is of interest here to examine whether the LRE formulation requires similar guards against possible solutions which are actually negative demand estimates.

The fact that the LRE model does not permit the estimation of negative demands can be shown to be true by examining the general formulation presented in Equation [5-10], which can be expressed by Equation [5-11a].

$$\prod_{r} (X_r) = 1 \tag{5-11a}$$

where

$$X_r = \left(\frac{V_r}{V_r'}\right)^{R_r^2 P_y''/V_r}$$
 [5-11b]

and:

r = link identifier

 E_R = total relative link flow error V_r = estimated flow on link r (vph) V'_r = observed flow on link r (vph)

 R_r = relative link flow reliability ($0 \le R_r \le 1.0$)

It is known that the observed link flow, V'_r , must always be positive. If a demand T_{ij} , is negative, then the link flow resulting from this demand, V_r , will also be negative. Assuming that a demand is negative, let us determine if Equation [5-11b] is feasible.

$$X_{r} = \left(\frac{-V_{r}}{V_{r}'}\right)^{R_{r}^{2} P_{ij}^{r} / -V_{r}} = \left(-\frac{V_{r}'}{V_{r}}\right)^{R_{r}^{2} P_{ij}^{r} / V_{r}}$$
 [5-11c]

In Equation [5-11b], the exponent $P_{ii}^r/V_r \le 1.0$, the calculation of X_r requires that the root of a negative number be found. This is not possible to do without resorting to complex numbers, which do not have any physical meaning within this context. Therefore, it is not possible for negative demands to satisfy the system of constraints expressed by the LRE model.

5.5 **Development of a Solution Algorithm**

5.5.1 Formal solution

Unlike the LSE model, a formal closed form algebraic solution does not exist for solving the LRE model. Equation [5-10] cannot be written explicitly in terms of the estimated flow, which is itself a function of the unknown demands T_{ij} . Despite this lack of a formal solution, it is still possible to solve iteratively for the unknown demands. This iterative approach is described in the next section.

5.5.2 **Iterative solution algorithm**

This section presents an iterative solution algorithm for solving the LRE model defined by Equation [5-10]. This algorithm, which is similar to the modified Jacobi iterative method developed in Section 4.5.3, is presented without proof that it satisfies the formulated constraints or that it minimizes the objective function. However, its ability to do both is illustrated using the same simple network examined earlier. Further characteristics of the algorithm are also examined. Similar to the original mathematical formulation, the iterative solution algorithm requires that all link use probabilities be exogeneously defined. It is also assumed that a prior matrix exists, though all cell entries may be set to be equal to each other, when no prior knowledge is available.

The first step in this algorithm requires that Equation [5-12] be used to compute the left hand side of Equation [5-10] for each O-D pair, based on the initial demands and the known link use probabilities. When C_{ij} is equal to 1.0 for all i and j, then the constraints of Equation [5-10] are satisfied. However, if C_{ij} is not equal to 1.0, then T_{ij} is not yet optimal and some change to T_{ij} is possible that will reduce E_R . The existing estimate of T_{ij} is then successively modified by the magnitude of C_{ii} as expressed by Equation [5-13].

$$C_{ij} = \prod_{r} \left(\frac{V_r}{V_r'} \right)^{R_r^2 P_{ij}^r / V_r}$$
 [5-12]

$$T_{ij}^{l+1} = T_{ij}^{l} \times C_{ij}^{-1/n_{ij}}$$
 [5-13]

where:
$$n_{ij} = \sum_{r} \left(\frac{R_r^2 P_{ij}^{r^2}}{V_r} \right)$$

 V'_r = observed flow on link r (vph) V_r = estimated flow on link r (vph) P_{ij}^r = probability that demand between i and j will use link r

l' = iteration number

 R_r = relative link flow reliability ($0 \le R_r \le 1.0$)

Physically, the geometric average of the relative flow errors across all links that each demand traverses is computed, and the current demand estimate is multiplied by the inverse of this average error. Since the **average** flow error across all utilized links is computed, iterative changes to all contributing demands are allocated in a systematic manner. This approach automatically leads to the selection of a demand, that is the most likely solution, when multiple solutions exist.

The iterative solution algorithm is presented in Table 5-4 while potential stopping criteria for this algorithm are examined in the next section.

Table 3	-4: LRE iterative solution algorithm	
Step	Description	Equation
	Initially, the iteration counter is set to zero	<i>l</i> = 0
1	Use the prior matrix as the initial estimate of the unknown demand matrix	$T_{ij} = t_{ij} \ \forall_{ij}$
2	Assign the current estimate of demand to the network using known link use probabilities	$V_a = \sum_{ij} \left(T_{ij} \cdot P_{ij}^a \right) \forall_{ij}$
3	Compute the product of the ratios of estimated and observed link flows for all links associated with each O-D pair in turn	$C_{ij} = \prod_{r} \left(\frac{V_r}{V_r'} \right)^{R_r^2 P_{ij}^r / V_r}$
4*	Based on the estimated and observed link flows, compute new O-D	$T_{ij}^{l+1} = T_{ij}^{l} \times C_{ij}^{-\frac{1}{n_{nj}}}$ $n_{ij} = \sum_{r} \left(\frac{R_r^2 P_{ij}^{r^2}}{V_r} \right)$
5	Increment the iteration counter	$n_{ij} = \sum_{r} \left(\frac{V_r}{V_r} \right)$ $l = l + 1$
6	If stopping criteria not yet met, go to Step 2	See section 5.5.2

Table 5-4: LRE iterative solution algorithm

5.5.3 Algorithm stopping criteria

Since the algorithm proposed in Section 5.5.1 is iterative in nature, criteria must be defined that can be used to determine when enough iterations have been carried out. An obvious candidate criterion to use is the link flow error function, E_R . The algorithm has been formulated to ensure that E_R never increases if more iterations are carried out. Unfortunately, since E_R is the squared natural logarithm of the ratio of estimated to observed flows, its magnitude usually has little physical meaning. Using Equation [5-28], it is possible to determine the average relative link flow error based on E_R . The true average relative link flow error is computed from Equation [5-29] in which the flow to be used in the numerator and denominator is chosen to ensure that the ratio is always greater than one.

$$\overline{E}_R^n = e^{\sqrt{E_R/N}}$$
 [5-14]

$$\overline{E}_R^{rel} = \frac{1}{N} \sum_{a} \left(\frac{V_1}{V_2} \right)$$
 [5-15]

^{*} If seed O-D demand reliability factors are to be considered, then Equation [5-34] from Section 5.6 must be used instead of the equation presented here.

 V_1 = smaller of the observed or the estimated link flow (vph) V_2 = greater of the observed or the estimated link flow (vph) N = number of links over which the total error has been computed

Equation [5-15] provides an estimate of the average estimated link flow as a proportion of the observed flows. The average absolute link flow error can also be computed using Equation [5-30] while the true average absolute link flow error is determined from Equation [5-31].

$$\overline{E}_R = (\overline{E}_R^n - 1)\overline{V}'$$
 [5-16]

$$E_R^{abs} = \frac{1}{N} \sum_{a} |V_a - V_a'|$$
 [5-17]

where: $V'_a = \text{observed flow on link } a \text{ (vph)}$ $V_a = \text{estimated flow on link } a \text{ (vph)}$ $\overline{V}' = \text{average observed link flow (vph)}$

N = number of links over which error has been computed

Before continuing with the examination of additional candidate stopping criteria, it is instructive to examine the relative accuracy of using Equations [5-28] and [5-30] to approximate Equations [5-29] and [5-31], respectively.

Figure 5-5 provides a comparison between the actual relative error and the relative error estimated from E_R . As can be seen, the estimated relative error consistently underestimates the true relative error. This underestimation is particularly severe when the estimated flow is much smaller than the observed flow. When the estimated flow is reasonably similar to the observed flow (say within 50% of the observed flow) then the estimated relative error is a sufficiently accurate approximation of the actual relative error.

Figure 5-6 provides a comparison between the true absolute error and the absolute error estimated from E_R . As can be seen, the estimated absolute error underestimates the true absolute error over a significant range. Furthermore, the magnitude of underestimation is not constant. The estimated error is greater than the true error only when the flow difference is less than 65% of the observed flow. Clearly, the average absolute flow differences estimated by Equation [5-30] must be interpreted with some caution.

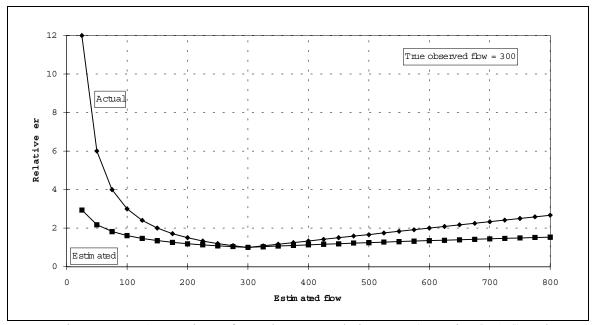


Figure 5-5: Comparison of actual average relative error (Equation [5-29]) and average relative error estimated using Equation [5-28] over a range of flow errors

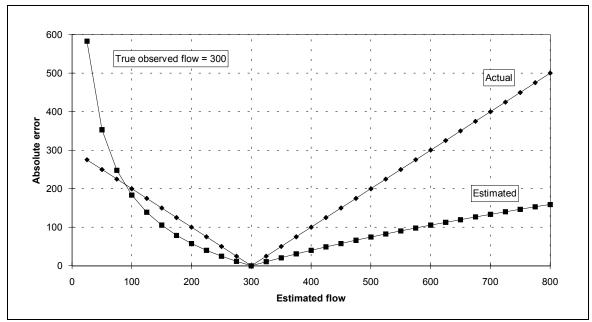


Figure 5-6: Comparison of actual absolute error (Equation [5-31]) and absolute error estimated using Equation [5-30] over a range of flow errors

Finally, it is advantageous to develop a stopping criteria that is based on the marginal decrease in link flow error. If each subsequent iteration of the algorithm provides a large reduction in the link flow error, then carrying out further iterations is likely to be beneficial. If the marginal reduction in link flow error is very small, then regardless of the absolute or relative measures of error, further iterations are not likely to result in significant reductions in error. An appropriate measure of the marginal reduction in link flow error is obtained by computing the difference between the average relative link flow error of the previous iteration, with that of the current iteration.

$$\Delta \overline{E}_{R}^{n} = \left(\overline{E}_{R}^{n}\right)^{i-1} - \left(\overline{E}_{R}^{n}\right)^{i}$$
 [5-18]

where: i = the current iteration number

As noted in Chapter 4 for the LSE model, a measure of the marginal improvement is likely the most appropriate stopping criterion.

5.5.4 Application of iterative algorithm to a simple network having multiple solutions

We illustrate the iterative solution technique by way of a variation of the now familiar example two-link network. Consider the four-link network in Figure 5-7 in which three feasible O-D demands exist (T_{AB} , T_{AC} , and T_{BC}).

To be able to compare the LRE model estimates with the maximum likelihood solutions determined in Chapter 3, it is initially assumed that links 3 and 4 are not utilized by any of the demands. The observed flows are assumed to be $V'_{I}=16$ vph, $V'_{2}=18$ vph, $V'_{3}=0$ vph, and $V'_{4}=0$ vph. Table 5-5 provides the results from each step of the iterative algorithm.

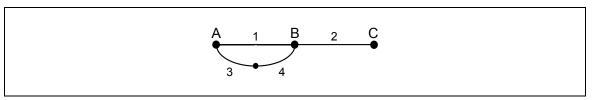


Figure 5-7: Example four-link network in which multiple feasible O-D solutions are possible

Table 5-5: Iterative algorithm results of LRE model for a simple example network with multiple solutions and all-or-nothing routings

	Estim	ated De	mand	Est	imated vr	Link Fl	ow	Corre			
Iter	T _{AB}	T _{AC}	T _{BC}	V_I	V ₂	V ₃	V_4	C_{AB}	C_{AC}	C_{BC}	E_R
0	5.00	5.00	5.00	10.00	10.00	0.00	0.00	0.954	0.900	0.943	0.5664
1	8.00	8.49	9.00	16.49	17.49	0.00	0.00	1.002	1.000	0.998	0.0017
2	7.86	8.48	9.15	16.34	17.63	0.00	0.00	1.001	1.000	0.999	0.0009
3	7.76	8.48	9.26	16.23	17.73	0.00	0.00	1.001	1.000	0.999	0.0004
4	7.69	8.47	9.34	16.16	17.81	0.00	0.00	1.001	1.000	0.999	0.0002
5	7.64				17.86	0.00	0.00	1.000	1.000	1.000	0.0001

6	7.61	8.47	9.43	16.08	17.90	0.00	0.00	1.000	1.000	1.000	0.0001
7	7.58	8.47	9.46	16.06	17.93	0.00	0.00	1.000	1.000	1.000	0.0000
8	7.57	8.47	9.48	16.04	17.95	0.00	0.00	1.000	1.000	1.000	0.0000
9	7.56	8.47	9.50	16.03	17.97	0.00	0.00	1.000	1.000	1.000	0.0000
10	7.55	8.47	9.51	16.02	17.98	0.00	0.00	1.000	1.000	1.000	0.0000
11	7.54	8.47	9.51	16.01	17.98	0.00	0.00	1.000	1.000	1.000	0.0000
12	7.54	8.47	9.52	16.01	17.99	0.00	0.00	1.000	1.000	1.000	0.0000
13	7.54	8.47	9.52	16.01	17.99	0.00	0.00	1.000	1.000	1.000	0.0000
14	7.53	8.47	9.52	16.00	17.99	0.00	0.00	1.000	1.000	1.000	0.0000
15	7.53	8.47	9.53	16.00	18.00	0.00	0.00	1.000	1.000	1.000	0.0000

The results presented in Table 5-5 permit a number of conclusions to be made. First, the LRE model estimates an O-D having an associated link flow error of zero. This indicates that the estimated demand exactly replicates the observed link flows. Second, even though multiple O-D solutions exist which replicate the observed link flows, the estimated O-D demand is equal to the maximum likelihood solution determined in Chapter 3 ($T_{AB} = 7.53$, $T_{AC} = 8.47$, $T_{BC} = 9.53$). Third, the algorithm converges rapidly to the final solution.

It is useful to provide a graphical representation of the estimated O-D demands and resulting link flows at each iteration. It is evident from Figure 5-8, which illustrates the O-D demand estimates for each of the first 15 iterations of the algorithm, that the iterative modifications made to the initial seed demand are initially large, but become progressively smaller as the estimated link flows become more similar to the observed flows. Figure 5-8 also indicates that the successive modifications to the initial seed demand are well behaved, with no erratic or oscillatory behaviour evident.

Figure 5-9 illustrates the link flows resulting from the O-D demands estimated in each of the first 15 iterations of the algorithm. Again, the transition of the link flows over successive iterations is well behaved. The link flows approach the observed link flows of 16 vph and 18 vph for links 1 and 2 respectively.

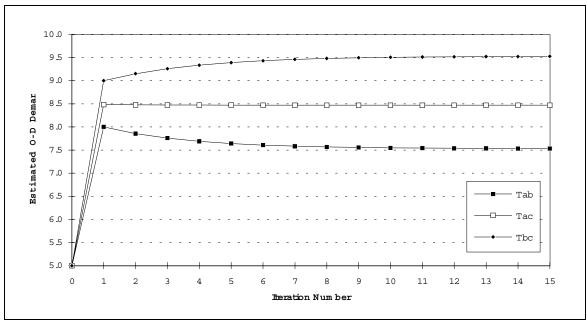


Figure 5-8: Iterative estimates of O-D demands for simple example network having all-or-nothing routes

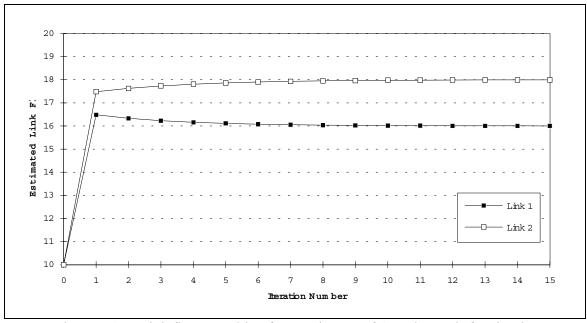


Figure 5-9: Link flows resulting from estimates of O-D demands for simple example network having all-or-nothing routes

5.5.5 Application of iterative algorithm to a simple network having a single solution with multiple paths

The previous example illustrated the ability of the proposed iterative solution to determine the most likely solution when a number of feasible solutions exist. However, as this previous example utilized all-or-nothing routings, link use probability values were restricted to either 0 or 1. It is useful to illustrate the performance characteristics of the algorithm for the same four-link network in which multipath routings are utilized. It is assumed that the link use probabilities associated with the network in Figure 5-7 are indicated in Table 5-6:

Table 5-6: Link use probabilities associated with the four-link example network in which multiple paths exist

	Link Number										
	1 2 3 4										
P(AB)	0.6	0	0.4	0.4							
P(AC) P(BC)	0.5	1	0.5	0.5							
P(BC)	0 1 0 0										

For this example network configuration, it is simple to describe algebraically the O-D solution as a function of the system of linear constraints (Equation [5-19]).

$$V'_{I} = 0.6 T_{AB} + 0.5 T_{AC} + 0.0 T_{BC}$$
 [5-19a]

$$V'_2 = 0.0 T_{AB} + 1.0 T_{AC} + 1.0 T_{BC}$$
 [5-19b]

$$V'_3 = 0.4 T_{4R} + 0.5 T_{4C} + 0.0 T_{RC}$$
 [5-19c]

$$V'_4 = 0.4 \ T_{AB} + 0.5 \ T_{AC} + 0.0 \ T_{BC}$$
 [5-19d]

If it is assumed that the observed flows are known exactly, and that node and path continuity exist, then the flows on links 3 and 4 must be equal $(V'_3 = V'_4)$. Under these assumptions, the system of four linear constraints is reduced to a system of three independent linear constraints that can be solved for the three unknown demands as functions of the observed link flows (Equation [5-20]).

$$T_{AB} = 5(V'_1 - V'_3)$$
 [5-20a]

$$T_{AC} = 2(3V'_3 - 2V'_1)$$
 [5-20b]

$$T_{BC} = V'_2 - 6V'_3 + 4V'_1$$
 [5-20c]

If the observed link flows are assumed to be V'_{I} =19 vph, V'_{2} =30 vph, V'_{3} =16 vph, and V'_{4} =16 vph, Equation [5-20] indicates that T_{AB} = 15 vph, T_{AC} = 20 vph, and T_{BC} = 10 vph.

Having analytically determined that a single unique demand solution exists that replicates the observed link flows, one can now examine the results provided by the iterative algorithm. Figure 5-10 illustrates the rate of convergence of the algorithm by depicting the O-D demands estimated in each iteration. As indicated, the model estimates of demand eventually converge to the correct solution, however, this convergence requires approximately 2000 iterations. Consistent with the results of the previous example, the magnitude of the modifications, that are made to the demands during each iteration, become smaller with each additional iteration.

Figure 5-11 provides the associated relative link flow error (E_R) for each iteration. It should be noted that the x-axis of the figure, denoting the iteration number, has been logarithmically transformed, providing a clearer illustration of the reduction in error (E_R) for changes in the order of magnitude of the number of iterations performed. If only one iteration is performed, the resulting error is approximately 0.006. If 10 iterations are performed, the error is reduced to approximately 0.001. However, performing 100 iterations, only reduces the error to 0.0007, and 1000 iterations to approximately 0.0. Clearly, though the algorithm does converge to the correct solution, there exists diminishing returns in terms of the reductions in error obtained by carrying out additional iterations.

To further explore these diminishing returns, the marginal improvement provided by each iteration is examined using Equation [5-20]. Figure 5-12 illustrates the marginal improvement in relative link flow error for each iteration. Note, that plotted values begin at iteration two, instead of one, as the value of 5.95 for iteration one would dominate all other values to such an extent that they could not be clearly seen in the figure. Figure 5-12 indicates that the marginal improvement of subsequent iterations decreases exponentially, such that after 25 iterations, the marginal improvement is less than 0.00001.

These results have several implications for the use of the LRE iterative solution algorithm:

- 1. The algorithm ultimately converges to the correct O-D solution.
- 2. The marginal reduction in relative error decreases approximately exponentially.
- 3. The rate of convergence of the algorithm is largely dependent on the network's characteristics and the quality of the prior information.

These characteristics indicate that the algorithm makes the most significant changes to the prior demand estimate (seed matrix) during the first number of iterations. Subsequent alterations are of a progressively smaller magnitude. If a limited number of iterations can be carried out, the greatest reduction in relative link flow error that can be achieved within this number of iterations will be obtained.

The knowledge that the marginal reduction in link flow error always decreases, allows a minimum marginal improvement threshold to be used as a stopping criterion. If the marginal improvement of the current iteration is insufficient, then further iterations need not be carried out as the marginal improvement of each subsequent iteration will be less than that obtained in the current iteration.

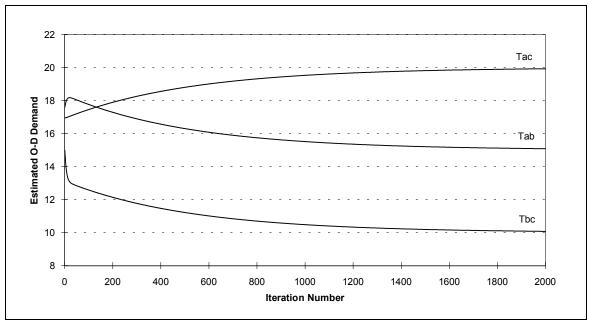


Figure 5-10: Rate of convergence of demands estimated by the LRE model for a simple network having a single solution and multiple paths

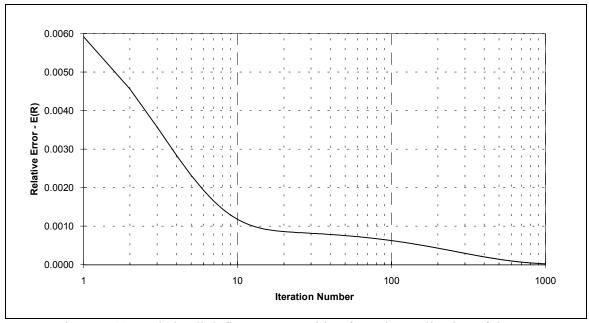


Figure 5-11: Relative link flow error resulting from the application of the LRE algorithm to a simple network having multiple path routes

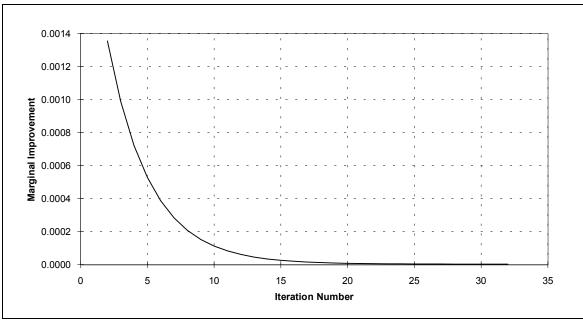


Figure 5-12: Marginal improvement in relative link flow error (E_R) resulting from the application of the LRE algorithm to a simple network having multiple paths

5.6 Incorporation of Seed O-D Demand Reliability

The need for seed O-D demand reliability factors and their characteristics were discussed earlier in Section 4-6. This section provides the modifications that are required to the LRE model algorithm to incorporate seed reliability factors, and illustrates the impacts of these factors using the same example network used in Section 4-6.

5.6.1 Modification to algorithm

To incorporate the seed demand reliability factor R_{ij} into the LRE solution algorithm, the equation in Step 4 of Table 5-4 is replaced with Equation [5-34].

$$T_{ij}^{l+1} = T_{ij}^{l} \times C_{ij}^{-\frac{1}{n_{ij}}} \vec{\mathbf{Q}}_{R_{ij}} \vec{\mathbf{E}}$$
 [5-21]

where: $T_{ij}^{\ l}$ = demand between i and j computed during the previous iteration (vph)

(vph) T_{ij}^{l+1} = demand between i and j computed during the current iteration (vph)

 R_{ij} = relative reliability of prior demand between i and j

 R_a = relative reliability of flow observed on link a

 V_r = estimated flow on link r (vph)

$$n = \sum_{r} \left(\frac{R_r^2 P_{ij}^r}{V_r} \right)$$

5.6.2 Example application considering the relative reliability of seed O-D demands

In order to examine the impact of seed reliability factors, consider the four-link network originally presented in Chapter 4 and illustrated again in Figure 5-13. Assume the observed flows are $V'_1 = 100$ vph, $V'_2 = 100$ vph, $V'_3 = 80$ vph, and $V'_4 = 120$ vph. Using this example network, the same five scenarios examined in Chapter 4 are examined using the LRE model. Table 5-6 provides a summary of the scenario configurations and the resulting estimated demands.

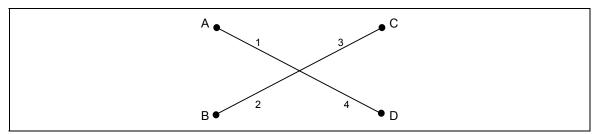


Figure 5-13: Example four-link network used to examine the relative reliability of seed O-D demands

Table 5-7: Scenario configurations and estimated demands used to illustrate the impact of seed O-D demand reliability factors

Scenario		Seed (vp	l O-D oh)		Se	Seed Reliability				Estimated O-D (vph)				
	t _{AC}	t_{AD}	t _{BC}	t_{BD}	R_{AC}	R_{AD}	R_{BC}	R_{BD}	T_{AC}	T_{AD}	T_{BC}	T_{BD}		
1	10	10	10	10	0	0	0	0	40.0	60.0	40.0	60.0		
2	5	10	1	15	0	0	0	0	63.0	37.0	17.0	83.0		
3a	5	10	1	15	1	0	0	0	5.0	95.0	75.0	25.0		
3b	5	10	1	15	0	1	0	0	84.3	10.0	0.0	104.4		
3c	5	10	1	15	0	0.2	0	0	64.9	35.1	15.1	84.9		

For Scenario 1, when no prior information is known, the LRE model estimates a demand that is the same as the maximum likelihood solution and the LSE model estimate. The demand estimate for Scenario 2, when prior information of uniform reliability exists, is quite different from the LSE model estimate determined in Section 4-6 ($T_{AC} = 42.3$ vph, $T_{AD} = 42.3$

57.7 vph, T_{BC} = 37.7 vph, T_{BD} = 62.3 vph). Examining the role of prior information is important in gaining an understanding of why this significant difference occurs.

The seed demand is used to provide information about both the relative and absolute magnitude of the true demand. The absolute magnitude of the seed matrix is altered based on the observed link flows. The results of Scenario 2 indicate that, though the seed matrix serves as an initial estimate of the demand, no weight is placed on preserving either the magnitude or the structure of this prior information. The LRE and the LSE models attempt to find an O-D matrix that minimizes their respective error functions, regardless of how different this demand might be from the seed matrix. This characteristic is also shared by the entropy and information models examined in Chapter 3.

It is possible to formulate both the LRE and LSE models so that the error function explicitly incorporates the degree to which the prior information has been altered (Equations [5-22a] and [5-22b], respectively). Each formulation combines the error associated with differences between the estimated and observed link flows, and the error associated with differences between the estimated demand matrix and the seed matrix. The relative weighting on these two components depends on the value of λ . When $\lambda=1$, then no weighting is placed on the discrepancies between the estimated and seed matrix, and Equations [5-22a] and [5-22b] revert to those developed earlier in this chapter and in Chapter 4. When $\lambda=0$, then no weighting is placed on link flow errors, and the estimated demand will remain equal to the prior matrix. The difficulty in applying Equations [5-22a] and [5-22b] is determining an appropriate value for λ . In most instances, no knowledge of the relative reliability of the prior matrix exists, while link flows are usually obtained through the use of loop detectors, and as such are known to be relatively highly reliable. Without any explicit knowledge of the reliability of the prior matrix, there is little reason to place any weight on the discrepancy between the estimated and seed demands ($\lambda=1$), and the model formulations of Equations [5-22a] and [5-22b] are not developed further.

$$E_R = \frac{\lambda}{L} \sum_{a=1}^{L} \left[\ln \left(\frac{V_a}{V_a'} \right) \right]^2 + \frac{1 - \lambda}{N} \sum_{ij=1}^{N} \left[\ln \left(\frac{t_{ij}}{T_{ij}} \right) \right]^2$$
 [5-22a]

$$E = \frac{\lambda}{L} \sum_{a=1}^{L} (V_a - V_a')^2 + \frac{1 - \lambda}{N} \sum_{ij=1}^{N} (t_{ij} - T_{ij})^2$$
 [5-22b]

where: T_{ij} = traffic demand between i and j (vph)

 t_{ij} = prior traffic demand between i and j (vph)

 V_a = estimated flow on link a (vph) V'_a = observed flow on link a (vph)

 λ = weighting placed on the link flow error

Returning to the examination of the LRE model results provided in Table 5-6, results for Scenarios 3a and 3b illustrate the impact of t_{AC} being known with certainty. In each case, the estimated demand retains the value of the seed demand that is known with certainty. These estimates are also the same as those estimated by the LSE model.

Scenario 3c illustrates the impact of knowing that the prior estimate of one of the O-D pairs is somewhat more reliable than the prior information for the other O-D pairs. The estimate of $T_{AD} = 35.1$ vph falls between 37.0 vph, the value obtained when the seed in not known with any degree of confidence, and 10.0 vph, the value obtained when the seed demand is known with certainty. Since the demand estimated when $R_{AD} = 0$ is different from that estimated by the LSE model ($T_{AD} = 57.7$ vph), it would not be expected that T_{AD} be the same for both the LSE and LRE models when $R_{AD} = 0.2$. However, it might be expected that the impact of $R_{AD} = 0.2$ relative to the estimate when $R_{AD} = 0$ would be similar for both the LSE and LRE models. This expectation can be tested by computing the Similarity Index using the equation presented in Section 4-6, resulting in a Similarity Index of 6.9 for the LRE model and 5.9 for the LSE model. In the absence of any context in which to interpret these results, no conclusions can be made regarding the similarity of the two values. To provide such context, Figure 5-20 illustrates the Similarity Index as a function of R_{AD} for both the LRE model and the LSE model. It is evident from this figure, that the Similarity Indices follow the same general non-linear trend for both models.

Figure 5-15 which illustrates the intermediate O-D demand estimates for the first 30 iterations of the LRE model, indicates that the LRE model tends to over-correct the seed demand for demands T_{AD} and T_{BD} in the first iteration. With a seed demand of 10 vph, the model estimates T_{AD} to be approximately 59 vph in the first iteration, when in fact, the final estimate is only 37 vph.

Figure 5-16 provides a similar illustration for Scenario 3a in which t_{AC} is assumed to be completely reliable. The final estimate under these conditions retains this prior information as $T_{AD}=5$. This prior information acts as an additional constraint on the feasible solution space. The reduced solution space still contains at least one feasible solution as the estimated O-D can be seen to exactly replicate the observed link flows. Like Figure 5-15, Figure 5-16 illustrates the over-reaction of the algorithm in predicting intermediate demand values. However, in Figure 5-16, though the over-compensation appears to be significantly greater than in Figure 5-15, this trend occurs primarily due to the constraint on T_{AD} . In the first iteration, the algorithm computes correction factors under the assumption that all demands are able to be corrected. Only when correction factors are applied, does the impact of R_{ij} become evident. Thus, it is not until iteration 2 that the model begins to account for the additional constraint on T_{AD} .

To prove that throughout these successive iterative modifications, the algorithm never diverges, Figure 5-17 illustrates that the relative link flow error (E_R) never increases.

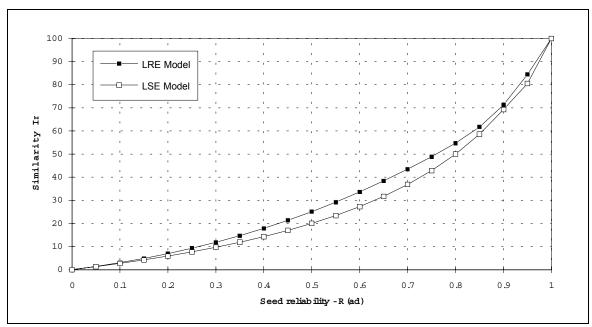


Figure 5-14: Similarity of estimated demand (T_{AD}) to seed demand (t_{AD}) as a function of seed reliability factor (R_{AD})

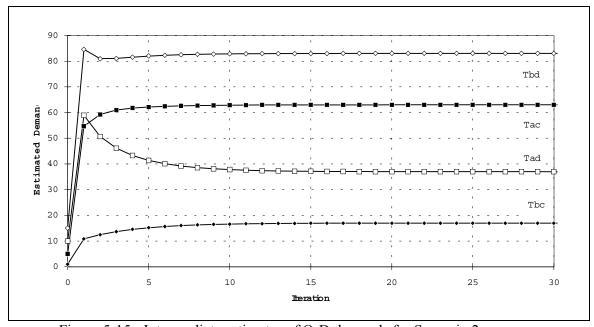


Figure 5-15: Intermediate estimates of O-D demands for Scenario 2

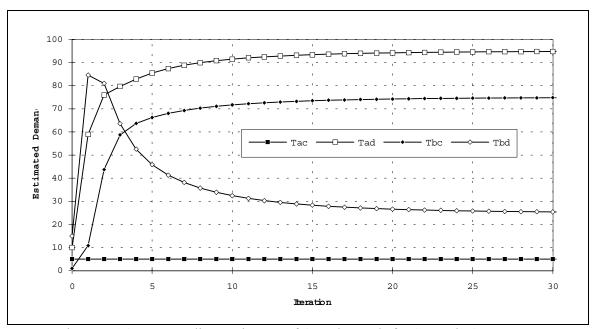


Figure 5-16: Intermediate estimates of O-D demands for Scenario 3a

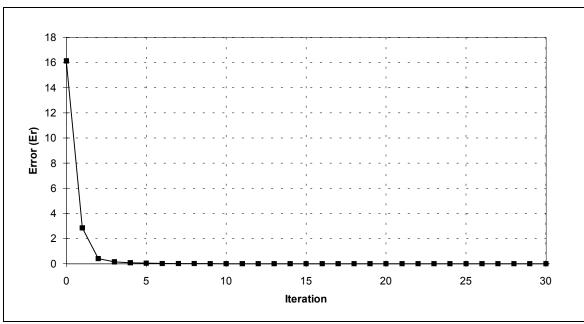


Figure 5-17: Relative error (E_R) of intermediate O-D demand estimates for Scenario 3a

It has been shown by way of a simple example that the LRE solution algorithm is well behaved when R_{ij} takes on a binary value. Of greater interest is the behaviour of the algorithm when R_{ij} assumes some fractional value less than 1.0. This is illustrated using the

same four-link network, except that the reliability of the prior information for demand between zones A and D is varied.

Figure 5-19 illustrates the intermediate estimates of T_{AD} for five different levels of R_{AD} . As expected, when $R_{AD} = 1.0$, T_{AD} does not change from the seed value ($T_{AD} = 10$). When $R_{AD} = 0$, all prior information is known to the same degree of certainty and, as indicated by the results from Scenario 2 in Table 5-6, $T_{AD} = 37$. If the prior estimate that $t_{AD} = 10$ is not completely reliable, but is known with a high level of confidence, (i.e. $R_{AD} = 0.75$) then it is expected that the final estimate of T_{AD} will not be very different from the prior estimate ($t_{AD} = 10$). For this example, when $R_{AD} = 0.75$, T_{AD} is estimated to be 24, which, though different from the prior estimate of 10, is also quite different from 37, the value of T_{AD} when $T_{AD} = 0$.

In general, the relationship between R_{ij} and the degree of similarity between T_{ij} and t_{ij} is not known *a priori*. The effect of R_{ij} is a function of how quickly the other available O-D demands can be altered from their prior estimate values to converge to a solution. If all other O-D demands are constrained, then any value of R_{ij} less than 1.0 results in the same final solution. The only impact that R_{ij} has under these conditions is to change the number of iterations that are required to converge.

Figure 5-21 illustrates the final O-D estimates for each of the four possible demands, for the complete range of values of R_{AD} . As more confidence is placed on the prior estimate of T_{AD} , the closer the final estimate of T_{AD} is to the prior estimate. In this example, for each value of R_{AD} , the four estimated demands represent a feasible solution which exactly replicates the observed link flow.

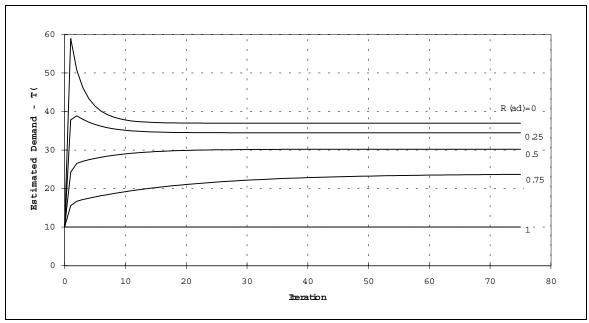


Figure 5-18: Intermediate estimates of O-D demand T_{AD} for different levels of seed O-D demand reliability

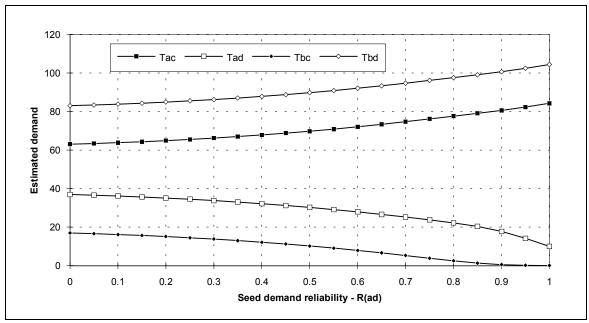


Figure 5-19: Final estimates of O-D demands for different levels of seed O-D demand reliability (t_{AD})

5.7 Summary

In this chapter a model based on the assumption that the estimated demand should minimize some function of the relative magnitude of the link flow errors has been developed. A function that implements this assumption is the squared natural logarithm of the link flow errors. It has been shown that this formulation is sensitive to the relative link flow error, regardless of whether this error results from an under-estimation or an overestimation of the flows. It is noted that this error function is continuous and differentiable. The solution that minimizes the error function is one that satisfies a system of non-linear constraints. Unlike the LSE formulation, this system of constraints does not lead to negative demand estimates. Unfortunately, these non-linear constraints cannot be explicitly solved for the unknown O-D demand, and hence requires an iterative solution.

The model formulation was applied to a number of simple network examples. These networks illustrated the model's estimation capabilities in response to link flow continuity and discontinuity, single and multiple solutions, single and multiple path routes, uniform and non-uniform link flow reliability, and finally, uniform and non-uniform seed demand reliability.

Since the system of constraints cannot be explicitly solved for the unknown O-D demands, an explicit closed form algebraic solution to the formulated model is not available. However, an iteration solution algorithm was proposed and demonstrated. The solutions resulting from this algorithm satisfy the formulated model's constraints. Similar to the

iterative solution method developed in Chapter 4, the algorithm successively modifies an initial seed matrix based on the relative differences between observed and estimated link flows. No mathematical proof is given that the algorithm converges, however, it is shown by way of several example applications, that the solution algorithm is usually well behaved and, for the given examples, does converge to a unique global optimum. Furthermore, it is shown that when the seed matrix is uniform, and multiple solutions exist, the algorithm estimates an O-D demand that closely approximates the most likely solution. Neither the formulation, nor the solution algorithm are sensitive to the magnitude of the uniform seed matrix.

Several algorithm stopping criteria have been defined and examined. It was determined that the most appropriate criteria, under most circumstances, is the marginal reduction in relative link flow error.

Chapter 6 extends both the LSE and LRE static model formulations to permit the estimation of dynamic demands. Alterations to the static formulations described in Chapter 4 and in this chapter are described. The abilities and performance characteristics of these extended models are demonstrated through their application to two hypothetical networks.

CHAPTER 6

Dynamic Extension of Static LSE and LRE Models

6.1 Introduction

In Chapters 4 and 5, the development of two similar O-D estimation approaches, each based on a different error function, was described. These models could be used to estimate a single static demand that, when reapplied to the network, replicated the observed link flows as closely as possible. In reality, O-D demands exhibit substantial peaking characteristics throughout the day. It may be desirable therefore to extend the LSE and LRE models to provide the capability for estimating a dynamic demand, which may vary both temporally and spatially.

This chapter first demonstrates, using field data, the need for estimating time varying O-D demands. Next, the additional complexities that are introduced when considering dynamic demands are identified and discussed. Modifications to the LSE and LRE models, which are required to address the identified complexities, are presented. These extensions are illustrated using two small hypothetical networks for which dynamic demands are estimated. Conclusions are made concerning the remaining limitations of the LSE and LRE models. To illustrate further the abilities of the models, and to identify several practical difficulties that are typically encountered when using actual field data, Chapter 7 describes the applications of these dynamic models to a 35 km section of freeway in Toronto, Ontario.

6.2 Establishing that a Need for Dynamic Demands Exists

The need for dynamic demand estimation capabilities can be established using two separate, but related arguments. The first argument which is presented, shows by way of field data, that O-D demands typically exhibit significant temporal and spatial variations. Building on the conclusions of this first debate, the second argues that the assumptions which are made in the estimation of static demands are increasingly violated as the time dimension becomes more disaggregated. These violations lead to inaccuracies within the estimated static demands. It is shown that dynamic O-D demand estimation does not need to make these same assumptions, and thus can be more accurate under these same conditions.

6.2.1 Spatial and temporal variations evident in field O-D patterns

The analysis of the characteristics of actual O-D demand patterns can rarely be carried out as the true demands are usually unknown. This is particularly true when this examination is to be carried out within the context of traffic management and control, rather than urban planning. In the latter case, the temporal and spatial variations that are of interest are those that occur over much longer periods of time and over greater disntance than within the context of traffic management and control. The obvious question then, is how can the importance of dynamic demands be justified through the use of field data, when these data are, in fact, not available?

A possible answer to this question lies in the analysis of available link count data to identify temporal and spatial patterns that result from the underlying O-D demands. Without knowing the true demands, it is possible to show, from observed link flows, that demands that are active in the field, do exhibit such significant variations.

Consider Figure 6-1, which depicts typical 15 minute link flows by time of day, observed on a three lane section of a major freeway in Toronto, Canada. These flows exhibit significant variation over time of day. The minimum flow (500 vph) is observed at 4 AM, while the maximum flow (5500 vph) is observed at approximately 7:30 AM. The greatest rate of change in traffic flow occurs between 5:30 AM and 6:30 AM, when the traffic flow increases rapidly from approximately 1000 vph to 5000 vph. Over this one hour period, there exists an average rate of flow increase of 67 vph/minute. Clearly, O-D demands that do not vary temporally would not be able to represent adequately these types of observed temporal characteristics.

It might be argued that observed flows, that result from the combined interaction of many O-D demands, do not possess significant correlation to the pattern of any single demand. Since this may be true, two additional figures (Figures 6-2 and 6-3) are provided. These figures illustrate total origin demand productions and total destination demand attractions, respectively, observed at several on and off-ramps to Highway 401 in Toronto, Canada. The flows illustrated in these figures reflect the effects of O-D matrix column and row totals,

respectively, instead of the entire matrix. In each figure, data are provided for two locations that are separated by approximately 6 km. These figures indicate that the flow characteristics observed in the two locations exhibit significant fluctuations over time and that these temporal variations can differ significantly for spatially separated locations. These data serve to reinforce the notion that field O-D patterns exhibit significant temporal and spatial variations.

It is possible to estimate a time series of independent static demands in order to capture temporal variations. This approach is illustrated in Figure 6-4, in which a portion of the flows depicted in Figure 6-1 are reproduced, along with a hypothetical time series of link flows resulting from a time series of one hour static O-D demands. It is clear from this figure that, though the one hour aggregate demands lead to flow discrepancies (particularly between 5 and 7 AM), the time series of static demands results in flows that can capture the more general temporal trends in the observed flows. One might conclude then, that the extension from a single demand estimate, to a time series of static demands, is sufficient to capture observed temporal fluctuations. However, static demand estimation makes no distinction as to when vehicles begin their trips, or when these trips arrive at their destinations. Vehicles are assumed to depart from their origins at a constant rate throughout the entire period and are assumed to complete their trips within the analysis period.

A time series of static demand provides a means of disaggregating the departure time, so that temporal variations in departure time can be reflected. However, demands are still assumed to complete their trips within each time slice. As the duration of the departure time slice becomes shorter, it is possible to reflect more accurately temporal variations in O-D departures. Conversely, as the departure time slice duration becomes shorter, the assumption that all vehicles complete their trips within this time slice, becomes increasingly less valid. It is the lack of validity of this assumption that deteriorates the accuracy of the time series of estimated static demands. The next section will examine, by way of an example, the relative impact of violating these static demand assumptions.

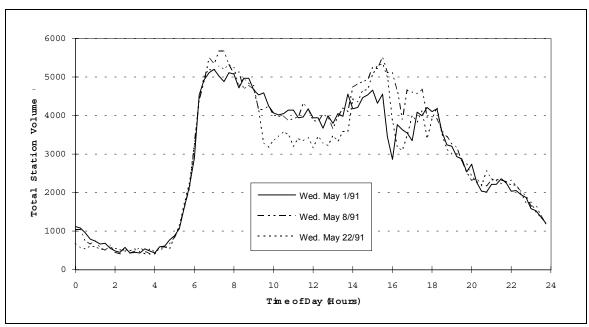


Figure 6-1: Typical observed aggregated 15 minute link volumes by time of day (Source: Hellinga *et al.*, 1994c)

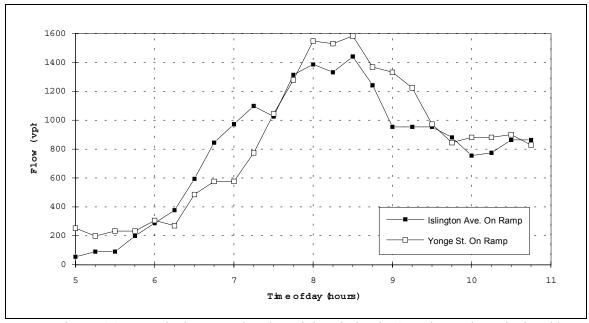


Figure 6-2: Typical temporal and spatial variation in O-D demands as depicted by total observed flow from specific origin zones (FTMS detector data, Highway 401 eastbound, June 8, 1992)

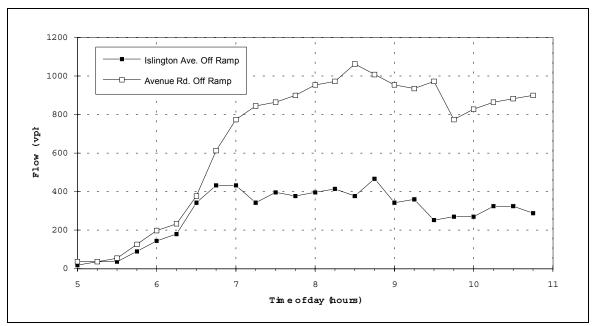


Figure 6-3: Typical temporal and spatial variation in O-D demands as depicted by total observed flow to specific destination zones (FTMS detector data, Highway 401 eastbound, June 8, 1992)

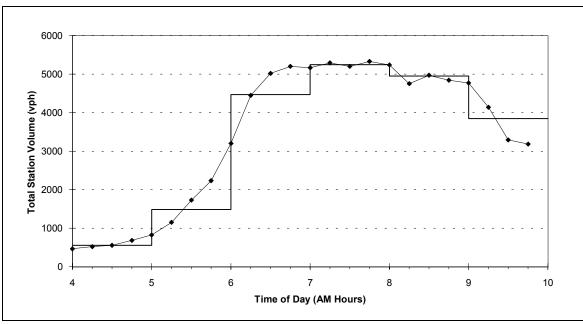


Figure 6-4: Approximating temporal variations in link flows through the use of a time series of static O-D demands (Observed data from May 22/91)

6.2.2 Effects of violating static estimation assumptions

To illustrate the potential effects of the static estimation assumption that all trips can be completed within a single 10 minute time slice, consider the simple example network illustrated in Figure 6-5, in which one demand between one O-D pair traverses five links.

It is assumed that time is discretized into time slices of equal duration. It is further assumed that the true demand pattern consists of 10 vehicles which depart during the first 10 minute time slice, and no other vehicles depart prior to, or after, this time slice. If the travel time on each link is assumed to be exactly 10 minutes, or one time slice, and loop detectors are located at the end of each link, then the link counts that would be observed during each time slice are presented in Table 6-1.

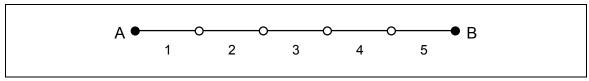


Figure 6-5: Example linear five-link network used to illustrate the relative accuracy of a static demand, a time-series of static demands, and a dynamic demand

1 abic 0-1.	Link counts	Link counts observed on example inlear rive-link network						
Time			Link					
Slice	1	2	3	4	5			
1	0	0	0	0	0			
2	10	0	0	0	0			
3	0	10	0	0	0			
4	0	0	10	0	0			
5	0	0	0	10	0			
6	0	0	0	0	10			
Total	10	10	10	10	10			

Table 6-1: Link counts observed on example linear five-link network

The estimation of a static O-D, with the period of analysis encompassing all 6 observation time slices, would assume that a link count of 10 vehicles was observed on all links during the 60 minute period of analysis and as such, the static O-D demand leading to these observed link counts must be one with a departure rate equal to 10 vehicles/hour for the entire 60 minute period.

The estimation of a time series of static demands would result in the problem of estimating, for each departure time slice, except the first time slice, a demand based on four link counts of zero, and one link count of 10 (the first time slice has no observed flows so the demand must be zero). A reasonable estimate for the O-D that would depart the origin in each time slice would be the mean flow of $(4\times0 + 1\times10)/5 = 2$ vehicles per 10 minute time slice, or 12 vph.

The estimation of a dynamic demand explicitly recognizes the relationship between departure time and the time to travel each link. The only solution that satisfies the observed link count is one in which 10 vehicles depart during the first 10 minute time slice at a rate of

60 vph, and no vehicles depart at any other time. Each of the three estimation methods are illustrated in Figure 6-6.

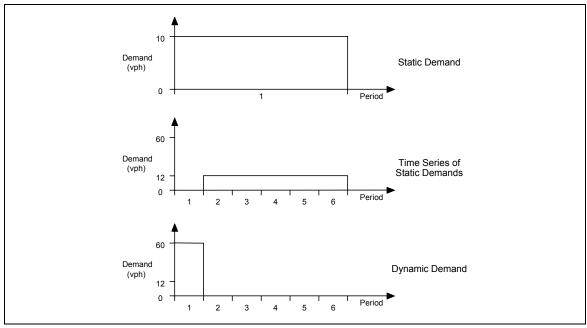


Figure 6-6: O-D estimation results as a function of the estimation process utilized

6.3 Extensions to LSE and LRE Static Models

The most obvious extension required to the static models, is the inclusion of some departure time index. In order to estimate dynamic demands, there must exist some means of tracking when demands depart their origins. Following from this need is the need for differentiating between link flows observed at different times. This section will first examine how the static versions of the LSE and LRE models can be expanded to accommodate both of these time based indices.

Since the dynamic model formulations require the addition of a number of indices to the previously defined variables, all relevant variables are defined again below:

i = origin zone

j = destination zone

a = link number

= time slice in which a given link flow is observed

d = time slice during which a given traffic demand departs an origin

 V_{ao} = estimated flow on link a during slice o

 V'_{ao} = observed flow on link a during slice o

 T_{iid} = demand between origin i, destination j, departing at time d

 P_{ijd}^{ao} = probability of demand departing origin *i* at time *d*, *en route* to

destination j, utilizing link a during time slice o

 R_{iid} = reliability of seed demand for origin i, destination j, time slice d

 R_{ao}^{s} = link flow reliability factor for link a, during time slice o

6.3.1 Formulation of dynamic LSE and LRE models

The derivation of the dynamic LSE model formulation is commenced with the definition of the error function. As presented in Chapter 4, the LSE model's error function is the sum of the squared link flow differences. For the dynamic case, this function is extended to consider a time series of observed and estimated link flows by summing across an index for both links and time slices.

$$E = \sum_{o} \sum_{a} (V_{ao} - V'_{ao})^{2}$$
 [6-1]

Having defined the objective function for the dynamic LSE model, the derivation presented in Chapter 4 is followed to determine the set of equations that define the minimum of this objective function. The partial derivatives of the error function with respect to each unknown demand T_{ijd} are expressed by Equation [6-2a]. Equation [6-2b] provides the expression for the estimated link flows, while Equation [6-2c] provides the derivative of the estimated flow with respect with the unknown demand T_{ijd} . Substituting Equations [6-2b] and [6-2c] into Equation [6-2a] produces Equation [6-2d], the system of equations that results in the minimum E.

$$\frac{\partial}{\partial} \frac{E}{T_{ijd}} = \sum_{o} \sum_{a} \left[2(V_{ao} - V'_{ao}) \frac{\partial V_{ao}}{\partial T_{ijd}} \right] = 0$$
 [6-2a]

$$V_{ao} = \sum_{d} \sum_{ij} \left(T_{ijd} P_{ijd}^{ao} \right)$$
 [6-2b]

$$\frac{\partial V_{ao}}{\partial T_{iid}} = P_{ijd}^{ao}$$
 [6-2c]

$$\frac{\partial}{\partial} \frac{E}{T_{ijd}} = \sum_{o} \sum_{a} \left[2 \left(\sum_{d} \sum_{ij} \left(T_{ijd} P_{ijd}^{ao} \right) - V'_{ao} \right) P_{ijd}^{ao} \right] = 0 \forall_{ijd}$$
 [6-2d]

An iterative algorithm can be used to solve Equation [6-2d]. Equations [6-3a], [6-3b], and [6-3c] define this algorithm.

$$T_{ijd}^{l+1} = T_{ijd}^{l} - \frac{\alpha}{n} C_{ijd} \left(1 - R_{ijd} \right)$$
 [6-3a]

$$C_{ijd} = \sum_{o} \sum_{r} \left((V_{ro} - V'_{ro}) P_{ijd}^{ro} \right)$$
 [6-3b]

$$n = \sum_{o} \sum_{r} \left(P_{ijd}^{ro} R_{ro}^2 \right)$$
 [6-3c]

It can be similarly shown that the dynamic extensions of the static LRE model lead to an iterative algorithm as defined in Equations [6-4a], [6-4b], and [6-4c].

$$T_{ijd}^{l+1} = T_{ijd}^{l} \times C_{ijd}^{-\frac{1}{n}(1-R_{ijd})}$$
 [6-4a]

$$C_{ijd} = \prod_{o} \prod_{r} \left(\frac{V_{ro}}{V'_{ro}} \right)^{R_{ro}^2 P'_{ijd} / V_{ro}}$$
 [6-4b]

$$n = \sum_{o} \sum_{r} \left(\frac{P_{ijd}^{ro} R_{ro}^2}{V_{ro}} \right)$$
 [6-4c]

The LSE and LRE dynamic models have been incorporated into a computer model called QUEENSOD. This model is used to carry out the estimation of dynamic and static demands for the two example applications described later in this chapter. A user's guide, describing the necessary data inputs into the model and the available outputs from the model, appears in Appendix A.

6.3.2 Characteristics of dynamic estimation

The first characteristic of the dynamic LSE and LRE formulation to note is the existence of a five-dimensional link use probability matrix (P_{ijd}^{ro}) . As a practical consideration, it is worthwhile to quantify the computer memory requirements that this matrix imposes for a medium to large sized network.

Consider a network having 50 origin zones, 50 destination zones, and 2000 links. Assume that demands are to be estimated for 16 time slices and that flows are observed for 20 of such time slices. Based on these network characteristics, the link use probability matrix would contain $50 \times 50 \times 16 \times 2000 \times 20 = 1.6 \times 10^9$ cells. Since the magnitude of the link use probability may be any real value between 0 and 1.0, the matrix must be stored as a *real* variable, which requires 4 bytes per value. For the above network, the link use probability matrix requires 1530 megabytes of memory. Clearly, this requirement presents a significant practical problem, and unless it is circumvented, limits the usefulness of the dynamic LSE and LRE models in terms of most practical applications.

Two approaches might be taken to address the memory requirement constraints of storing the entire link use probability matrix. First, it should be recognized that the matrix is very sparse, since flows resulting from a demand departing at a given time, are only observed on very few links during a given time slice. The sparseness of the matrix may be used to advantage as the cells containing zeros might be collapsed, reducing the overall dimensions of the matrix.

A second approach, relies on the fact that the link use probability matrix can be generated from a time series of route choice trees, and a time series of link travel times. These trees require much less storage space as their dimensions are (links × destinations × departure time slices). For the previous sample network scenario, the memory requirement of a route tree is only $2000 \times 50 \times 16 \times 4$ bytes ≈ 6 megabytes. The link use probabilities can, therefore, be generated from the stored trees and link travel times each time they are required. This latter method has been adopted within the QUEENSOD model for overcoming the memory constraint of storing the entire link use probability matrix.

6.3.3 Generating link use probabilities from known route trees and link travel times

Previously, it has been assumed that the link use probability matrix has been exogeneously computed and as such, its calculation required no further discussion. Now, however, the link use probability matrix is computed from stored routes and link travel times, both of which are assumed known. Since the link use probability matrix is computed within the proposed models, these calculations need to be presented first. It is prudent, therefore, to describe first the format and nature of the input data; route choice trees and link travel times.

Route choice trees represent, in a compact format, the sequence of links that must be traversed in order to travel from any current location, to any valid destination zone. These trees provide spatial information in terms of the links that are to be utilized, but they do not contain any temporal information indicating how long it will take to traverse any link. Thus, route trees alone are not sufficient to produce a link use probability matrix as there is no way to compute **when** a demand will use a particular link, only that it will, or will not use a link at some time.

Route choice trees may vary temporally, so that later departures choose different routes from earlier departures. Trees may also be multipath, so that the total demand from a specific origin to a specific destination, is apportioned to more than one route.

Temporal information is obtained from sources of link travel times. If a time series of link travel times are known, it is possible, given the route tree and departure time, to determine the demand that will be observed on any link. Travel times are assumed to be determined exogeneously from the O-D estimation methodology presented in this thesis. The travel times may be observed directly from probe vehicles having the communication capability of transmitting link travel time information to a central computer, and/or they may be estimated from spot detector speeds, traffic flows and signal plans. Regardless of the data source, the travel time data must represent the total time required for a vehicle to traverse the link during that time interval. This total must include time spent traveling and time spent in a queue due to congestion or a traffic signal. In this way, congestion effects on travel time are accounted for.

As illustrated in Figure 6-7, time is discretized into time slices of uniform duration for both the departure of demands and the observation of flows. In this figure, the travel time on each link is assumed to be equal to the duration of one time slice, though in general, this need not be the case, nor is this typically encountered with actual field data. Demand is also assumed to depart from its origin with uniform headway throughout the departure time slice. In Figure 6-7, a platoon of traffic demand departs at a uniform rate during the first departure time slice (time 0 to time t_I). As platoon dispersion is not considered, the band-width of the platoon remains constant. It is also assumed that loop detectors are located at the end of each link, so that vehicles are not observed until they have actually traversed the entire link.

Figure 6-7 depicts the uncommon situation in which all link travel times are conveniently, not only equal to each other, but they are also equal to the duration of the observation time slice. This non-typical situation results in the entire platoon crossing the detector at the end of each individual link during a single time slice. In practice, the situation is rarely so convenient.

Consider therefore, the linear four-link network depicted in Figure 6-8, in which the four links are not all of equal length, and travel times on each link are not equal to each other. The depicted proportions of demand observed on each link during each time slice are determined by computing the arrival time of the front of the platoon at each detector station. Since platoon dispersion is not modeled, the tail of the platoon always arrives one departure time slice duration after the head of the platoon. If the head and tail of the platoon arrive during two different time slices, then the demand is apportioned to each time slice in proportion to the fraction of the demand arriving in each time slice. This method also permits link travel times to change temporally, as the travel time used to compute the platoon's arrival is the one that is in effect during the time slice that the platoon enters the link.

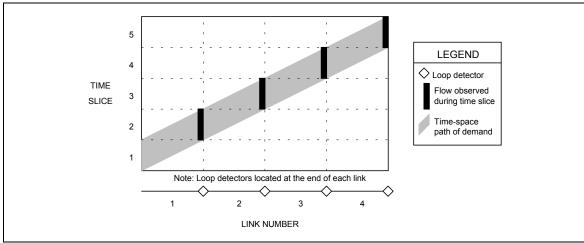


Figure 6-7: Discretized representation of demand propagation over time and space

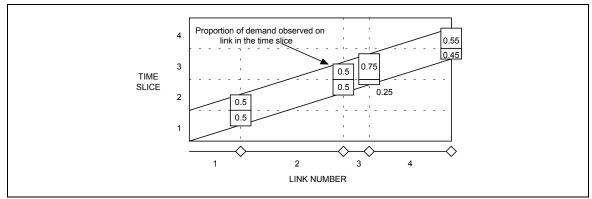


Figure 6-8: Computation of observed link flows based on demand departure time and link travel times

6.3.4 Identifying extraneous link flows

One difficulty that arises, when considering a time series of link flows, is identifying those flows that result from demands departing their origins prior to the analysis time period. Clearly, flows that do not result from demands being estimated, should not be considered within the analysis, as they will bias the results. This problem is best illustrated by way of a simple example. Consider Figure 6-9, which depicts the time-space path of two demands. The first demand path, denoted A, results from demand departing the origin during time slice t. The second demand path, B, represents demand departing during time slice t+1. In this figure, the travel time on each link is assumed to be one time slice, though in general, this need not be the case. If demands are to be estimated for time slices t+1 to t+5, then any demands departing prior to time slice t+1, are not considered. Accordingly, flows resulting from demands departing prior to time slice t+1, must also be ignored. Some systematic method is therefore required to determine if an observed flow results from a demand that departed prior to the start of the analysis period.

In this simple example, it is evident that only flows in demand path B should be considered, and those in path A result from demands departing prior to the start of the analysis period. For networks having more than one O-D demand, the determination of irrelevant, or extraneous flows is somewhat more involved. Consider for example, the simple network illustrated in Figure 6-10. One may wish to define a method for determining if flows observed on link A, are extraneous. Since the network is not linear, and many demands contribute to the flow on link A, the problem is more complex than that illustrated in Figure 6-9. One needs to first determine the minimum time required for the demand departing each origin to reach link A. Any flows observed on link A prior to this minimum time must be extraneous flows. Unfortunately, as illustrated in Figure 6-11, the flow on link A is likely to result from a number of demands, all departing at different times. A flow that is not completely extraneous, may result in part from demands departing prior to the start of the analysis period, but since the true demands are unknown, it is not possible to determine a priori, the proportion of the flow that results from demands departing prior to the analysis period.

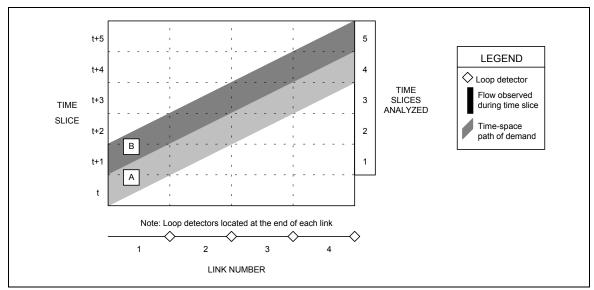


Figure 6-9: Illustration of occurrence of flows that do not result from demands considered within analysis period

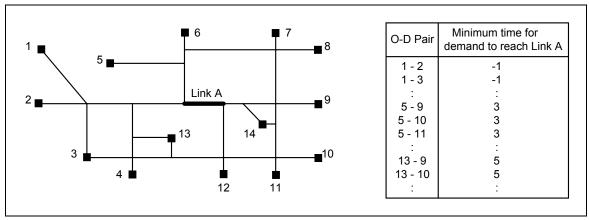


Figure 6-10: Illustration of a more complex situation for the determination of extraneous link flows

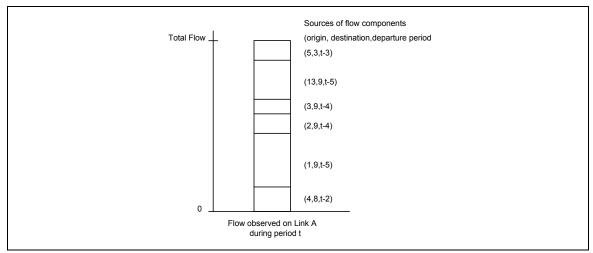


Figure 6-11: Illustration of the occurrence of partially extraneous flows

6.4 Application of Dynamic Models to a Small Hypothetical Network

6.4.1 Network characteristics and scenario descriptions

The characteristics and performance of the dynamic LSE and LRE models are illustrated using the network in Figure 6-12. The network is composed of five links and six nodes, of which two are origins (nodes A and B), and two are destinations (nodes C and D). It is assumed that links 1, 2, 4, and 5, have a fixed travel time of five minutes, while the travel time on link 3 is exactly 10 minutes. Four O-D demands (T_{AC} , T_{AD} , T_{BC} , and T_{BD}) are active during each of three consecutive five minute departure time slices. The magnitude of these demands are provided in Table 6-2. The traffic flows resulting from these O-D demands, are observed during five consecutive five minute observation time slices. It is assumed that these flows are measured exactly, such that node and path flow continuity exist. These observed flows are also provided in Table 6-2. It should be noted that as there exists only a single path from each origin to each destination zone, the problem of determining appropriate routes and route weights, is not encountered in this example.

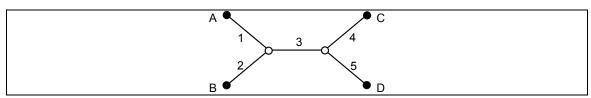


Figure 6-12: Example five-link network used to illustrate the performance characteristics of the dynamic LSE and LRE models

	Departure Time Slice					Observ	ation Tim	e Slice	
O-D	1	2	3	Link	1	2	3	4	5
A-B	50	85	60	1	0	175	285	235	0
A-C	125	200	175	2	0	125	250	200	0
B-C	75	100	100	3	0	0	300	535	435
B-D	50	150	100	4	0	0	0	125	185
				5	0	0	0	175	350

Table 6-2: Actual network O-D demands and observed link flows (vph)

Table 6-3: Description of scenarios evaluated to illustrate the dynamic estimation abilities of the LSE and the LRE models

Ī	Scenario	Seed demand	Estimation Method	Number of Iterations
Γ	1a	Actual	Dynamic	60
	1b	Uniform	Dynamic	60
Γ	2a	Actual	Static	0
	2b	Actual	Static	20

Table 6-3 provides the scenario configurations that are examined on the test network. Scenario 1a provides a reality check on the estimation model, since the actual demand is provided to the model as the seed. If the model attempts to alter the seed, or indicates that any link flow error exists, then an error is likely to exist in the model formulation or implementation. Scenario 1b represents the situation likely to be encountered in the field, as the uniform seed implies no prior knowledge of the true demand. Scenarios 2a and 2b provide an opportunity to examine the appropriateness of approximating dynamic demands with a time series of static demands.

6.4.2 Measures for comparison

In order to facilitate the quantitative evaluation of the LSE and LRE dynamic models, several Measures Of Performance (MOP) are defined. The first MOP is average link flow error, represented as a proportion of the observed link flow (\overline{E}^n for the LSE model, and \overline{E}_R^n for the LRE model). When the estimated link flows are equal to the observed flows, then, for the LSE model, this value must be equal to zero. For the LRE model, the normalized link flow error results from a natural logarithm transformed ratio of the observed and estimated flows. As a result, when the estimated flows are equal to the observed flows, this measure equals 1.0 instead of 0.0.

The second MOP is the sample correlation coefficient computed between the estimated and observed flows (r_v) , and between the estimated and true O-D demand (r_{od}) . When the two quantities being compared are equal, the correlation coefficient should equal 1.0. The correlation coefficient provides a quantitative measure of the degree of linear association between the two quantities. As the values of the two quantities become less associated, the magnitude of the correlation coefficient also decreases. The correlation coefficient may take on values from -1.0 to +1.0, where negative values indicate that as one of the two quantities increases, the other quantity decreases. The correlation coefficient is computed using Equation 6-5.

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{(n - 1)S_x S_y}$$
 [6-5]

where: $r = \text{sample correlation coefficient } (0.0 \le r \le 1.0)$

x,y = quantities being correlated

 S_x = sample standard deviation of data series x S_y = sample standard deviation of data series y n = number of observations compared

The last MOP is the root-mean-square error computed between the estimated and true O-D demand. When the estimated demand is equal to the true demand, then the RMS error is zero. As the matrices become increasingly less similar, the RMS error increases. The RMS error is computed using Equation 6-6.

$$RMS = \sqrt{\frac{\sum (obs - est)^2}{n}}$$
 [6-6]

It is of interest to examine the relative sensitivity of the normalized link flow error and the correlation coefficient to discrepancies between the observed and estimated link flows. To illustrate such sensitivity, consider 10 links, on which flows of 1, 2, 3, ..., and 10 are observed, respectively. Now consider that the estimated flows for all links, except link 6, are equal to the observed flows. For ten different situations, the flow on link 6 is assumed to be 1, 2, 3, ... and 10, respectively. The associated relative link flow error and the correlation coefficient for each of these ten situations are presented in Figure 6-13. When the estimated flow on link 6 is equal to 6, the observed flow, then the normalized link flow error is equal to 0.0 and the correlation coefficient is equal to 1.0. As the flow estimate for link 6 becomes increasingly different from the observed flow, the normalized link error increases and the correlation coefficient decreases. It is of interest to note that the magnitude of the coefficient of correlation is less sensitive to the presence of link flow error, than is the normalized link flow error. As such, the measure of normalized link flow error is considered to be a better indicator for comparing different estimation scenarios, than is the correlation coefficient.

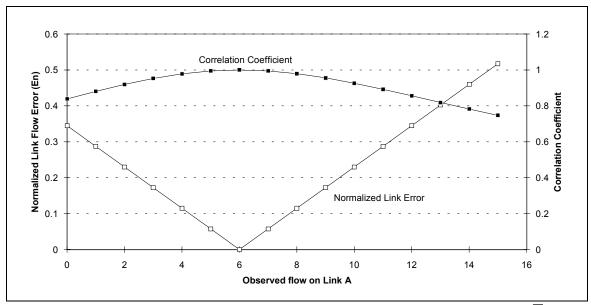


Figure 6-13: Illustration of the sensitivity of the normalized link flow error (\overline{E}^n) and the correlation coefficient to the magnitude of the link flow error

6.4.3 Analysis of model results

A number of conclusions can be drawn based on the O-D estimation results presented in Table 6-4. First, the results for Scenario 1a, indicate that both the LSE and LRE models correctly recognize that the estimated and observed link flows are identical. The correlations between estimated and observed flows, and observed and estimated demands, are perfect. Since the seed demand is the actual demand, the models do not alter the seed demand, and the resulting demand estimate is identical to the true demand.

For Scenario 1b, a uniform seed demand is provided to each estimation model. Both the LSE and LRE models predict a final O-D demand that exactly replicates the observed link flows. However, neither model predicted a demand that is identical to the true demand, nor is the demand estimated by the LSE model identical to the demand estimated by the LRE model. Since the link flow resulting from each estimated demand replicates exactly the observed link flows for each case, it is clear that multiple feasible solutions exist and the measure of how similar the estimated demand is to the actual demand (correlation coefficient) does not provide a meaningful absolute measure. As has been demonstrated earlier, when multiple solutions exist, both the LSE and LRE models choose the solution that approximates the most likely solution. The O-D correlation coefficient provides a measure of how similar this most likely feasible demand is to the true demand. However, as this true demand was arbitrarily chosen, many correlation coefficient values could have been achieved by selecting the appropriate feasible demand as the *true* demand. It is worthwhile to note, however, that both models estimate demands that have approximately the same degree of similarity to the true matrix.

The results for Scenario 2a indicate that the best time series of static seed demands does not result in a time series of flows that replicates those flows that were observed. The static estimation model assumes that all trips are completed within each departure time slice. Thus, even though, the true demand is provided to the model, the estimated link flows that result are quite different from those observed. Even when the algorithm is permitted to alter the seed demand (Scenario 2b), it is not possible to estimate a time series of static demands that results in link flows that exactly replicate the observed link flows.

Table 6-4: Results from the application of the dynamic LSE and LRE models to a simple five-link network

simple five mix network							
Scenario	Link Flow	Correlation	RMS O-D				
	Error	Flow	O-D				
LSE Model	\overline{E}^{n}						
1a	0.0000	1.0000	1.0000	0.00			
1b	0.0000	1.0000	0.7783	29.11			
2a	1.5166	-0.3872	N.A.	114.63			
2b ⁽¹⁾	1.4805	-0.0349	0.4400	67.57			
LRE Model	LRE Model \overline{E}_R^{n}						
1a	1.0000	1.0000	1.0000	0.00			
1b	1.0000	1.0000	0.8008	27.79			
2a	15.2466	-0.3872	N.A.	114.63			
2b ⁽²⁾	8.4994	0.0886	0.1594	110.34			
	(1) Convergence tolerance met after 16 iterations						
(2) Convergence tolerance met after 7 iterations							

6.5 Limitations of Dynamic Models

The previous section described the application of both the LSE and LRE dynamic models to a simple example network. However, a number of limitations still exist with the current model formulations. Two limitations, which are neither unique to dynamic models, nor to the formulations proposed within this thesis, are worth noting.

First, the estimation of dynamic demands using either the LSE or the LRE model still requires that the routes that drivers utilize are known *a priori* and that the associated weights for these routes are also known. Furthermore, dynamic estimation requires that the time needed for vehicles to traverse each link in the network is known.

Second, as illustrated in Figure 6-14, it is not possible, with the current model formulations, to estimate demands that exceed the capacity of the network. In fact, it is not possible to directly observe link flows that exceed network capacity. The only evidence that excess demand exists, is the congestion at the bottleneck. Unfortunately, it is usually not possible to determine the actual demand pattern that is in excess of the network capacity. This may initially not appear to be a problem, until one considers evaluating a number of alternative congestion reduction scenarios. The effectiveness of each scenario may depend very much on the characteristics of the true demand pattern, and not on the clipped pattern that is estimated.

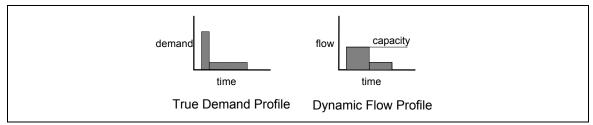


Figure 6-14: Illustration of possible discrepancies between true demand profile and observed capacity constrained flows

6.6 Application of Dynamic Models to a Larger Hypothetical Network

The previous applications of the LSE and LRE dynamic models were limited to a simple five-link network. This section examine a somewhat more extensive network, in which neither node nor flow continuity is expected to exist. Since the network, that is examined in this section, is larger than previous examples, the generation of observed link flows and appropriate routes is no longer a trivial task. To facilitate these computations, the network traffic simulation model INTEGRATION is utilized. The process of first generating appropriate observed link flows, and subsequently estimating a time varying O-D demand, is illustrated in Figure 6-15.

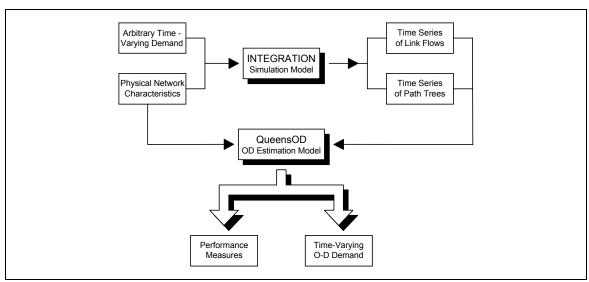


Figure 6-15: Process used to generate an appropriate time-series of observed link flows and routes using the INTEGRATION simulation model

6.6.1 Network characteristics

The physical characteristics of the network were assigned such that the network consisted of an integrated freeway and parallel arterial configuration. The network, illustrated in Figure 6-16, consists of 7 origin/destination zones, 9 additional nodes and 40 uni-directional links.

An arbitrary time-varying demand, as presented in Table 6-5, was created. This demand was considered to be the true demand, and from it the observed link flows were generated. To facilitate the generation of these observed flows, the true demand and the network characteristics were provided to the network traffic simulation model INTEGRATION. The INTEGRATION model was then used to provide estimates of appropriate link flows and routes resulting from the application of the true demand on the network. A brief description of the INTEGRATION model's relevant characteristics is provided in the next section.

The input data files required to carry out the O-D estimation using the QUEENSOD model for each of the scenarios described in the following section, are provided in Appendix B.

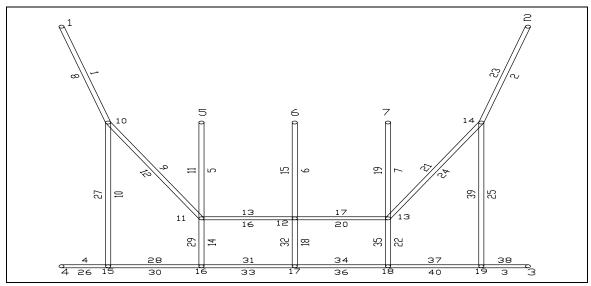


Figure 6-16: Hypothetical network used to illustrate dynamic estimation abilities of the LSE and LRE models

1 4010 0 5.			11014	ar tilliv	c varyi	115 401	mana t	ibea te	acteri	111110	OBCITO	7 u 110 v	15 (1PI
			Destination Zone										
	Origin	Departure Time Slice 1 Departure Time Slice 2											
	Zone	1	2	3	4	5	6	1	2	3	4	5	6
	1		150	200	50	25	30		125	100	25	125	50
	2	100		100	200	100	50	100		75	200	50	50
	3	125	75		650	125	75	125	75		850	75	75
	4	25	275	750		150	125	25	125	1150		50	25

Table 6-5: Actual time varying demand used to determine observed flows (vph)

6.6.2 Description of the INTEGRATION model

During the past three decades a number of different simulation models have been developed and applied to freeway and arterial networks. To circumvent some of the limitations of these earlier models, the INTEGRATION model was developed. This model, along with input data requirements and model outputs, are briefly described below. Detailed model descriptions, input requirements and other applications of the model can be found in the literature (Van Aerde, 1994; Rakha *et al.*, 1989; Van Aerde and Yagar, 1988; Van Aerde, 1985; Hellinga and Van Aerde, 1993).

The INTEGRATION model is a microscopic routing-oriented simulation model of integrated freeway and surface street networks. The model is microscopic in the sense that individual vehicle movements through the network are traced to monitor and control the unique behavior of vehicles that belong to a certain sub-population. For example, each vehicle's identity is retained, as it travels through the network, in order to route each such vehicle based on its unique trip origin, destination, departure time and vehicle type. On each link, the model determines each vehicle's speed based on the available headway.

The model is routing-based in that only a vehicle's trip origin, destination and departure times are specified external to the model, leaving the actual trip path and the arrival times at each link along the path to be derived within the simulation based on the modeled interactions with any other vehicles. The actual path taken by a vehicle depends on the vehicle type, which in turn will control what routing objective a given vehicle will attempt to achieve. For example, vehicles belonging to one group will attempt to respond in real-time to any on-line traffic information that may be available, while vehicles belonging to another group will simply consider average day-to-day traffic conditions, without concern for any characteristics that are unique to the given day. As a result, different vehicle (driver) type mixes (i.e., commuter vs. tourist, or guided vs. non-guided vehicles) can be considered concurrently. In each case, the actual arrival time at the start of a link, along the path that is finally taken will be a function of any delays that may have been incurred along the links taken so far.

6.6.3 Scenarios being examined

Having generated a time-series of link flows, routes, and route weights, and knowing the true demand, a number of O-D estimation scenarios can be undertaken. Table 6-6 provides a description of the scenarios that are examined using this network. Scenario 1a acts as a base case condition, as the true demand is provided to the model. Since zero iterations are carried out, the true demand is simply evaluated in terms of the observed link flows. As neither node, nor flow continuity are guaranteed to exist, it is expected that even the true demand will not exactly replicate the observed link flows. However, the knowledge of the magnitude of this intrinsic error provides an opportunity to interpret more accurately the results from the remaining scenarios.

Scenario 1b also assumes that the true demand is known *a priori*. However, in this case, the models are permitted to modify this prior demand in order to reduce the associated link flow error. Scenario 2a examines the models' estimation capabilities when no prior

information is provided. Scenario 2b examines the models' capabilities when it is assumed that, the O-D cells in the true matrix which have no zero demand are known, and are explicitly identified to the model.

	tollities of ESE and ERE models	
Scenario	Seed demand	Iterations
1a	Actual	0
1b	Actual	100
2a	Uniform all feasible O-D's considered	100
2b	Uniform all zero demands in actual O-D considered as structurally zero in seed	100

Table 6-6: Description of scenarios evaluated to illustrate the dynamic estimation abilities of LSE and LRE models

6.6.4 Analysis of model results

Each of the four scenarios were analyzed using both the dynamic LSE and LRE models. For each model's application, four measures of performance have been provided in Table 6-7.

The analysis begins by examining the results for Scenario 1, in which the true demand is assumed as the seed demand. No attempt is made to alter the seed matrix. The observed link flows are known without error, the routes and route weights are known exactly, and the true demand is also known, yet the results in Table 6-7 indicate that the resulting link flow error is not equal to zero. The obvious question then, is, "Why do the results indicate a non-zero link flow error?".

The answer to this question lies in the manner in which the observe link flows were obtained. The INTEGRATION model is a microscopic simulation model that reflects the movements of individual vehicles within a traffic network. For this application, the O-D demands depart their origins with uniform headways, but each vehicle's speed, and thus travel time on a link, is a function of the link's speed-flow relationship and the available headway. Link flows and travel times are output at five minute intervals, and it is these values that are used as *observed* flows and travel times. However, these values represent the average flow experienced on the link during the previous time slice, while the travel time is the travel time that is experienced at that instant in time. This travel time is not necessarily the travel time experienced by all of the vehicles that traveled the link in the previous time slice. Vehicles that traversed the link just prior to or just after the time when the link travel time was recorded, are likely to have experienced travel times that were somewhat longer or shorter than the recorded time. There exists some variation in the travel times experienced by the vehicles. Thus, the assignment of the true demand to the network using the observed routes and reported link travel times does not exactly replicate the observed flows.

The link flow errors that result from the assignment of the true demand are rather small. Figure 6-17 illustrates the correlation between the observed flows and those estimated by the dynamic LRE model for Scenario 1a. The scatter about the line of perfect correlation is small, with no obvious systematic bias. In fact, the correlation coefficient is equal to 0.9984

for both the LSE and LRE models. As expected, since no modifications are made to the seed matrix, both models provide the same estimated link flows.

In Scenario 1b, we determine whether the LSE or LRE models can reduce the link flow error found in Scenario 1a. If the error can be reduced, it is of interest to determine how different from the true demand matrix, the estimated O-D demands must become. The results in Table 6-7 indicate that both models are able to reduce the link flow error when this error is measured by the normalized link flow error. When measured by the correlation coefficient, the LSE model reduces the error, but the LRE model slightly increases the error. This apparent discrepancy occurs because the LRE model reduces the relative error, not the absolute error which is the basis upon which the correlation coefficient is determined.

Of greater interest is the change in the error associated with the estimated O-D demand. In Scenario 1a, the seed demand was not altered, so the estimated demand was exactly equal to the true demand. In Scenario 1b, both models modified the seed matrix in order to reduce the link flow errors, but in doing so, increased the O-D error. This error acts as a benchmark against which the results from Scenario 2 can be compared. One would expect that the minimum RMS O-D error that could be achieved, regardless of the quality of the prior information, would be approximately equal to 34 for the LSE model and 48 for the LRE model. If, as in Scenario 2, O-D demands are estimated without the benefit of high quality prior information, then any resulting RMS O-D errors in excess of these values could be assumed to result from the model's inability to determine the true O-D demand pattern without high quality prior information.

In Scenario 2a, it is assumed that no prior information is available. After 100 iterations, the LSE model's O-D estimate results in an average link flow error that is approximately 12% of the average observed flow. The associated RMS O-D error is almost twice that found in Scenario 1b. Clearly, the lack of prior information adversely affects the model's ability to reproduce the true demand. However, the magnitude of the link flow error indicates that further iterations would likely improve this estimate.

After 100 iterations of the LRE model, the estimated demand has a larger associated link flow error than was found in either Scenario 1a or 1b, however, the resulting RMS O-D error is slightly less than that determined in Scenario 1b. Clearly, despite the lack of prior information, the LRE model is able, in this example, to estimate an O-D demand that is as similar to the true demand as that estimated when very high quality prior information is available.

In Scenario 2b, it is assumed that it is known *a priori* which O-D matrix cells contain zeros. However, no knowledge exists regarding the relative magnitude of the non-zero cells. This prior knowledge reduces the number of unknowns and allows both models to reduce the link flow error more quickly. As illustrated in Table 6-7, after 100 iterations the normalized link flow errors for both the LSE and LRE models are lower than those found in Scenario 2a. The associated RMS O-D error is comparable to that determined in Scenario 1b. Thus, despite the little available prior information, both models were able to estimate O-D demands having the same level of error as when the true O-D demand was used as the prior information.

Figure 6-18 illustrates, for each iteration, the LRE model estimates of four selected O-D pairs. It can be seen that initial modifications to the uniform seed are large, but become successively smaller.

Figure 6-19 illustrates the normalized link flow error resulting from each iteration of the LRE model for Scenarios 1b, 2a, and 2b. Scenario 2a, in which no previous information is available, has the largest initial, and final, link flow error. Scenario 2b, in which some prior information is available, has an initial error that is larger than that experienced in Scenario 1b. However, this error is quickly reduced so that by the sixth iteration, the normalized link flow error for Scenario 2b is approximately equal to that of Scenario 1b. Figure 6-20 illustrates the same trends for the LSE model results, except that the LSE model requires more iterations to respond.

Table 6-7: LSE and LRE dynamic model results for a hypothetical integrated freeway/arterial network

neeway areena neework							
Scenario	Link Flow	Correlation	RMS O-D				
	Error	Flow	O-D				
LSE Model	\overline{E}^{n}						
1a	0.0909	0.9984	1.0000	0.00			
1b	0.0516	0.9994	0.9942	34.28			
2a	0.1162	0.9972	0.9675	62.22			
2b	0.0546	0.9994	0.9928	38.81			
LRE Model	$\overline{E}_{\scriptscriptstyle R}^{n}$						
1a	1.2808	0.9984	1.0000	0.00			
1b	1.2207	0.9983	0.9909	47.94			
2a	1.4676	0.9983	0.9910	47.66			
2b	1.2207	0.9983	0.9909	47.75			

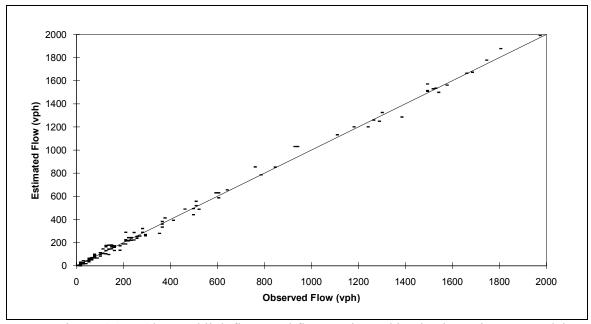


Figure 6-17: Observed link flows and flows estimated by the dynamic LRE model from the true dynamic demand (Scenario 1a)

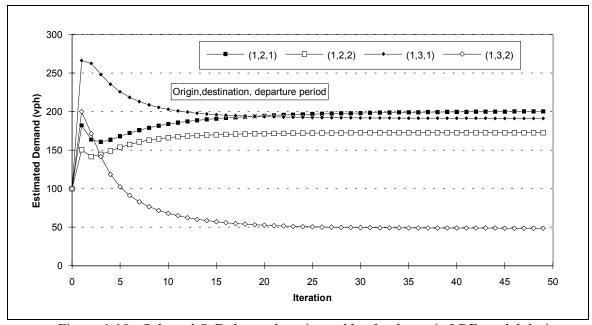


Figure 6-18: Selected O-D demands estimated by the dynamic LRE model during each iteration (Scenario 2b)

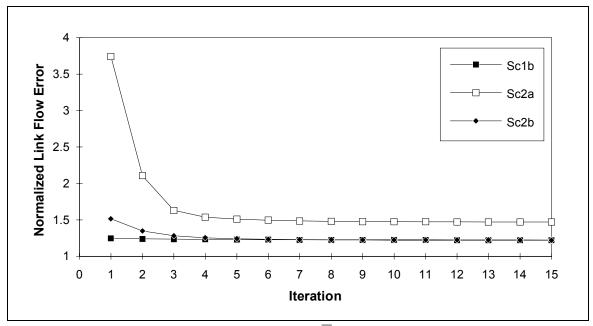


Figure 6-19: Normalized link flow error (\overline{E}_R^n) resulting from the LRE model as a function of the iteration number

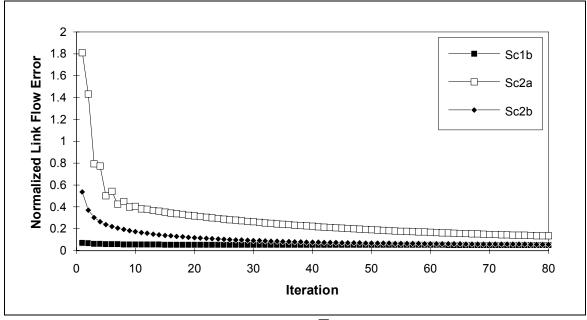


Figure 6-20: Normalized link flow error (\overline{E}^n) resulting from the LSE model as a function of the iteration number

6.7 Summary

Chapter 6 has shown that there exists a need for the estimation of dynamic, as opposed to, static O-D demands. This need was demonstrated by illustrating, using field data, that significant temporal and spatial variations exist in observed link flows as well as in total origin productions and total destination attractions.

Next, it was demonstrated that static demands become increasingly less accurate as the duration of the time slice, for which the demands are in effect, becomes shorter. The accuracy of static O-D demand estimation, of the estimation of a time series of static demands, and of dynamic estimation was illustrated using a simple linear network. It was shown that static O-D estimation assumes that all trips are completed during their departure time slice.

Having established the need for dynamic O-D estimation, dynamic extensions to the static LSE and LRE models were formulated. Revised iterative solution algorithms were also developed. The dynamic extension of the static models required the inclusion of two additional indices: O-D departure time, d, and link flow observation time, o. The inclusion of these two additional indices caused the link flow probability matrix to become five dimensional. It was illustrated that the computer memory storage requirements for this matrix become excessive, even for a medium sized network. A practical alternative is to store the route trees, tree weights, and link travel times, and compute the associated link use probabilities when necessary.

A further practical difficulty was found to exist in the identification of extraneous flows – those flows that result in part or in whole, from trips that depart prior to the analysis period. A means of identifying wholly extraneous flows was proposed.

The dynamic LSE and LRE models were applied to two hypothetical networks. The superior abilities of dynamic O-D demands to capture temporal variations were illustrated. Several quantities that reflected the performance of the estimation models were defined. It was found that the LRE model tended to converge more quickly than the LSE model for these examples. The link flow errors and the O-D errors associated with the final model estimates were comparable.

CHAPTER 7

ESTIMATING O-D DEMANDS FOR A FREEWAY CORRIDOR IN TORONTO, CANADA

7.1 Introduction

The previous chapters have described the development and formulation of two dynamic O-D demand estimation models. The performance characteristics of these models have been mathematically examined and illustrated using several small hypothetical networks. Though these networks have each served to illustrate various specific characteristics of the models, they are hypothetical with known characteristics, and many of the difficulties encountered in field applications have not been illustrated. This chapter describes the derivation of time varying O-D demands from field data for a 35-kilometre section of Highway 401, a multilane freeway in Toronto, Canada.

Unlike the hypothetical examples examined previously, neither the true demands nor the routes or route weights are known *a priori*. Furthermore, since the link flow data contain inconsistencies, neither node nor path flow continuity exist. As these conditions are likely to be the norm rather than the exception for field applications, it is desirable to gain an understanding of how well the LSE and LRE models perform under these challenging conditions.

This chapter is organized as follows. The 35 km section of Highway 401 that is analyzed, is described first in Section 7.1.1. The FTMS that is operating on a portion of this section, and from which the necessary data are obtained, is then described next in Section 7.1.2. The process of extracting the data necessary for the estimation of time varying demands from the characteristics of the physical network and the available FTMS data is subsequently

discussed. The extraction process involves abstracting the physical network into a series of nodes and directional links, determining a time series of observed link flows and link travel times, and identifying a time series of routes and route weights. Each of these three components is examined and described in turn in Section 7.2. The demands estimated by the LSE and LRE models are evaluated by examining the aggregate link flow errors and by comparing estimated and observed flows for two individual links. Demand estimates are also compared to observed origin and destination zone productions and attractions. Lastly, conclusions are made about the performance of the dynamic LSE and LRE models.

7.1.1 Description of Highway 401

The freeway site studied consists of a 35-km section of Highway 401 in Metropolitan Toronto, Canada (Figure 7-1). Carrying an average annual daily traffic (AADT) of approximately 320 000 vehicles and having a cross-section consisting of up to 16 directional lanes, this section of Highway 401 is the most heavily traveled section of freeway within Canada and ranks among the busiest in the world. It attracts this volume of traffic as it is a major commuter corridor for east-west travel within Metro Toronto, and as it is the primary route for through traffic. Unfortunately, the very success of the highway has in the recent past also become its greatest liability, as the continuing growth in traffic demand has pushed this section of freeway far beyond its original design capacity. This results in recurring congestion during large portions of the a.m. and p.m. commuting periods. Non-recurring congestion is also common place throughout these same peak periods, causing considerable efficiency and safety repercussions for both private automobiles and commercial vehicles.

One unique feature of this freeway is that each direction of travel is provided with an express and collector facility. The collector facility provides primary access from and to the major arterial roads and other freeways connecting to Highway 401. The express facility has limited access to and from the collector facility via high speed transfer roadways. The system of separated facilities also provides drivers with a number of binary routing choices. At the time of this study, the system of express/collector facilities extended from Neilson Road in the east to Islington Avenue in the west.

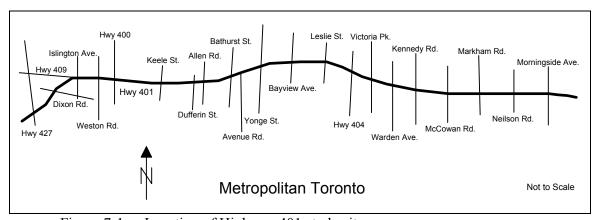


Figure 7-1: Location of Highway 401 study site

7.1.2 Description of the freeway traffic management system

Early in 1991, the Ontario Ministry of Transportation (MTO) brought on-line, on Highway 401 through Toronto, a freeway traffic management system entitled COMPASS (Korpal, 1992). This system initially encompassed 16 kilometers of express and collector freeway. The system incorporated induction loop detectors embedded in the roadway approximately every 600 meters. This spacing provided detector coverage for each unique roadway segment and for all freeway access and egress roadways.

The loop detectors are connected via fibre optic cables to the operations control centre located on the north side of Highway 401 at Keele Street. In addition, closed circuit television (CCTV) cameras with a viewing range of approximately 700 meters, are located every 1000 meters. Operators in the control centre can control the camera's view through pan and zoom capabilities to observe traffic conditions along the freeway.

Initially, the loop detector stations were operational from Yonge Street, in the middle of the study area, west to the end of the study area. The detection system was expanded eastward to Warden Avenue, but this section was not operational during the course of this study.

7.2 Acquisition of Data Necessary for O-D Estimation

7.2.1 Network representation

For this study, the Highway 401 network was examined from Morningside Avenue in the east to Highway 427, in the west. This portion encompasses the entire express/collector facility as of the summer of 1992. This section of Highway 401 includes 22 interchanges with connecting arterial roads and other freeways.

For the purposes of this study, Highway 401 was abstracted into a series of directed links and nodes. Links were defined such that a single link existed for each FTMS detector station. Link boundaries were defined such that detector stations were positioned near the midpoint of the link. This permitted the spot FTMS speed measurements to be used to estimate the link travel times. A summary of the number of links and nodes required to represent the eastbound direction of Highway 401 is provided in Table 7-1.

Figure 7-2 illustrates the abstract link and node representation of the eastbound direction of Highway 401. In order to provide a better aspect ratio of the network, the eastern half of the network is plotted below the western half. For flow progression purposes these two halves are connected and appear to the model as one continuous segment.

Table 7-1: Number of links and nodes used to represent the eastbound direction of the Highway 401 network

Node/Zone Type	Number	Link Type	Number
Origins	43	Single	10
Destinations	35	Express	48
Intermediate Nodes	147	Collector	57

Tatal	005	Ramps Transfer	115 9
Total	225	Total	239

Figure 7-2: Link and node representation of the Highway 401 eastbound network (not to scale)

7.2.2 Link flows

The COMPASS FTMS detector stations measure and transmit volume, as well as occupancy and speed data, to the central computer every 20 seconds. Twenty second detector data for the two 24-hour periods of May 1, 1991 and June 8, 1992 were obtained from the MTO. An initial examination was carried out on the data from May 1. This examination revealed several potential sources of errors arising from the manner in which the data are stored.

The total station flow is reported as the total number of vehicles detected during the previous 20 seconds for all lanes. If one of the lane detectors is inoperative, then it reports -9999 and does not contribute to the computed total. If the lane station reports data that are just considered suspect, then the station will report the data as negative. If any lane detector reports negative data, then the station total is also reported as negative, as the entire data sample may be suspect. For example, consider a three-lane section, in which one of the lane detector's is inoperative. The recorded lane flows might be 3, 5, and -9999 vehicles per 20 seconds for each respective lane detector. The station total is computed as (3+5) = -8. Clearly, this reported total is less than the actual, as the input from one of the lane detectors is not considered. Though the station data are marked by making them negative, it is not possible to distinguish between station data resulting from a suspect lane detector and a station that is inoperative.

Despite these imbedded data errors, these data represent the only available data source that quantifies existing traffic conditions. To gain an understanding of the character of the available data, it is useful to view a graphical representation of typical system data. Figure 7-3 illustrates a 22 hour time series of typical 20 second and aggregated 15 minute volume data for a three lane section in the eastbound collector lanes. Significant variations over time are evident in the 20 second volume data. To reduce this variation within the data, to reduce the storage requirements, and to provide a practical time slice duration over which time varying demands could be estimated, all data were aggregated into 15 minute average values. As illustrated in Figure 7-3 this aggregation significantly reduces the remaining variation in the data.

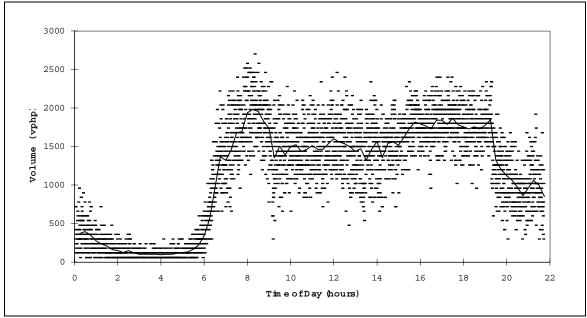


Figure 7-3: Typical 20 second and aggregated 15 minute FTMS detector volume data reflecting conditions on a 3 lane section in the eastbound collector at Avenue Road on May 1, 1991 (Station # 401DE0100DEC)

7.2.3 Route determination

Since the O-D demands are unknown, the traditional approach of using traffic assignment techniques for determining appropriate paths is not possible. For the estimation of O-D demands for the Highway 401 network, neither the utilized routes nor the O-D demands are known.

Since the true routes and route weights are unknown, some estimate of these quantities must be made. The simplest routing assumption is to route all traffic along the single shortest path. This approach is clearly unrealistic, as virtually all traffic demand would either use the collector lanes or the express lanes, exclusively. A more realistic approach is to assume that the two alternative paths, that are available, are both utilized in some proportion. For the first path (express route), a driver will attempt to access the express lanes at the earliest opportunity, and exit the express lanes at the latest possible point. For the second path (collector route), the collector lanes are utilized for the entire trip. In accepting these two paths as the only route alternatives, it is assumed that drivers do not arbitrarily alternate

between express and collector lanes without having a distinct preference for either one facility or the other.

Though the routes have been identified, the proportion of demand using each route must still be determined. To simplify the problem, it was assumed that the proportion of demand utilizing each route remained constant through both time and space for each direction. Thus, the proportion of demand choosing the express route is the same at the diversion point at Dixon Rd. during the a.m. period as it is at the transfer roadway at Markham Road during the p.m. period.

A previous simulation modeling study of Highway 401 determined, that for these same routing assumptions, the most appropriate split is for 80% of the demand to use the express route and 20% to use the collector route (Hellinga and Van Aerde, 1994c). This same split is assumed to be appropriate for this analysis. As noted by Hellinga and Van Aerde (1994b), this assumption is limiting since it is possible that this split changes over time of day as well as over distance. However, it is adequate for the purposes of illustrating the effectiveness of the LSE and LRE models when applied to field data.

7.2.4 Link travel times

The COMPASS FTMS detector stations measure and transmit volume, occupancy, and in the case of dual loop stations, speed data, to the central computer every 20 seconds. Single loop stations cannot measure speed directly, so for each 20 second interval, an average vehicle length that is computed from some nearby dual loop station, is used to compute an estimate of speed. Approximately 30% of the 175 detector stations that were operational in 1992 were dual loop stations.

Twenty second detector data for the two 24 hour periods of May 1, 1991 and June 8, 1992 were obtained from the MTO. An initial examination was carried out on the data from May 1. This examination revealed that the accuracy of the speeds estimated by single loop stations was at that time rather poor, particularly under low occupancy conditions. Two factors contributed to these inaccuracies:

1. The first is the inaccuracies in the recording of measured station occupancy. At that time, lane occupancy was measured in the field as a real value, but was truncated into integer format before being transmitted to the central computer. This truncated lane occupancy was used to compute the average station occupancy, which was also truncated into integer format before being stored on tape. The impact of this method is best illustrated using a simple example.

Consider a three-lane section of freeway from which lane occupancy data are recorded. As illustrated in Table 7-2, the effect of dual truncation can lead to recorded station occupancies of zero, while in reality the occupancy was 1.6%. The amount of error introduced by this method of recording the station occupancy becomes less significant as the occupancy increases.

Lane Occupancy (%)		Station Occupancy (%)		
True	Recorded	True	Recorded	
0.9	0	(0.9+1.9+1.9)/3	(0+1+1)/3	
1.9	1	= 1.57	= 0	
1.9	1			

Table 7-2: Truncation effect on recorded station occupancies

2. The second factor contributing to speed estimate inaccuracies is the need to have an estimate of average vehicle length for each single loop station. In 1992 each single loop station received an estimate of average vehicle length, every 20 seconds, from a nearby dual loop station. Unfortunately, the average vehicle length at two detector stations, during a 20 second interval, is not strongly correlated, particularly at low occupancies. If a single truck is recorded during a 20 second interval at a dual loop station, then the average 20 second vehicle length at this station might be 20 meters. However, unless a single truck also happens to traverse, during the same time period, the single loop station to which the average vehicle length is being provided, the estimate of average vehicle length will be quite different from the actual average vehicle length. Under heavy flow conditions, more vehicles contribute to the computed average and the probability is lower that average vehicle lengths between stations will be very different.

Dual loop stations, which can directly measure speed, are considered to provide more reliable speed data. However, these data may also contain significant systematic error. In November 1992, it was discovered that, due to incorrect loop spacing, at least three dual loop detector stations in the westbound express lanes were reporting speed data with significant error¹. Since that time, this error has been rectified, however, the data used in this study predate this time and contain these errors.

The existence of COMPASS on Highway 401, provides an efficient means of obtaining spot speeds and estimating link travel times. It is not clear, however, how accurate the estimation of travel time from spot speeds typically is, particularly in view of the previously described data inaccuracies.

A traditional floating car study was carried out to collect data describing trip duration as a function of departure time. These data provide a measure against which travel times estimated from spot speed data could be compared. A brief description of the study results is provided here, while a detailed description is available in the literature (Hellinga and Van Aerde, 1994a).

The study was carried out on three days; Monday, June 8, Thursday, August 6, and Friday, August 7, 1992. In total, seventy-seven trips were made. Each trip traversed the length of the study network in a single direction. Comparable travel times were estimated from the aggregated FTMS data. A correlation analysis of these two data sources indicated that the coefficient of correlation was 0.423, suggesting that the ability to estimate travel time from 1992 FTMS detector data was rather poor. Despite the poor correlation found in this

¹Personal communication with D. Tsui of the Freeway Traffic Management Section of the Ontario Ministry of Transportation. September 1993. Detector station I.D. numbers: 401DW0020DWE, 401DW0040DWE, 401DW0090DWE

study, no other link travel time data were available, and since some estimate of link travel time is required in order to estimate time-varying demands, the FTMS travel times were utilized.

As indicated earlier, FTMS data were only available for approximately 16 of the total 35 km network. For those regions of the network for which FTMS data were not available, link travel times were assumed to be the free speed travel time. This assumption is not accurate, unless link flows are quite small. A better approach may have been to calibrate a macroscopic speed-flow relationship, and estimate speeds based on the estimated flows. However, this approach is also not without difficulties, as it is not possible to determine from flow alone, whether a link is operating in the uncongested, or congested regions of the speed-flow relationship.

Potential sources of error within the estimation of link travel times must be evaluated in relation to the discretized time slice duration, and the average trip duration. If the time slice duration is much shorter than the average trip duration, then even small errors in the estimation of the link travel times can result in the association of a flow resulting from a particular demand to the incorrect time slice. When the time slice duration is longer, then larger link travel time errors can be tolerated before flows are associated with incorrect time slices.

It has already been noted that the available link travel time data for the Highway 401 network is not perfect. Before any projections of the impacts of these inaccuracies can be made, it is necessary to examine the characteristics of the duration of trips made on this network. Table 7-3 provides data describing trip duration as a function of departure time and route for the eastbound direction of Highway 401.

It is evident from Table 7-3, that except for trips made during the Friday p.m. peak period, trips are no longer than 34 minutes in duration. Since the duration of the observation time slice, for the estimation of the time-varying O-D demands, is 15 minutes, the longest trips will be completed in approximately two time slices. Furthermore, these trip durations are representative of trips that traverse the entire 35 km section of freeway. It would be expected that most trips do not traverse the entire network and, as a result, would also have a shorter duration. However, the true trip length distribution is not known, with the result that it is not possible to verify this expectation using field data. It will be shown in Section 7.3.5, however, that both the LSE and the LRE models predict O-D demands that result in an average trip length of approximately 10 km, suggesting that many trips will be able to be completed within a single 15 minute time slice. Due to the short trip duration, relative to the time slice length, it is possible that the errors within the estimates of link travel times may ultimately have relatively little impact on the estimated O-D demands. It would be expected that these link travel time inaccuracies would have a more significant impact if the discretized time slice were to be made shorter, to say 5 minutes. An examination of these impacts is outside of the scope of this thesis.

35.4

7

18

37

Date	Expr	ess	Collector		
	Departure Time	Duration (min)	Departure Time	Duration (min)	
Monday	7:20 a.m.	30.2	7:05 a.m.	24.7	
June 8	8:35 a.m.	27.2	9:20 a.m.	21.8	
1992	2:35 p.m.	20.3	9:50 a.m.	23.6	
	5:00 p.m.	34.4	3:30 p.m.	27.1	
			3:35 p.m.	32.2	
			6:20 p.m.	20.3	
Numb	er of trips	4		6	
Thursday	7:50 a.m.	27.8	6:20 a.m.	20.4	
Aug. 6	8:00 a.m.	26.7	7:55 a.m.	32.0	
1992	9:30 a.m.	22.0	9:35 a.m.	21.9	
	9:35 a.m.	22.1	2:40 p.m.	22.8	
	2:35 p.m.	23.5	4:20 p.m.	33.9	
	4:20 p.m.	32.2			
	4:25 p.m.	33.5			
Number of trips		7		5	
Friday	6:25 a.m.	19.4	6:30 a.m.	22.3	
Aug. 7	6:30 a.m.	20.5	7:45 a.m.	27.3	
1992	7:45 a.m.	27.9	7:55 a.m.	32.2	
	9:20 a.m.	18.7	9:20 a.m.	21.1	
	9:30 a.m.	20.9	3:00 p.m.	34.7	
	3:05 p.m.	38.8	3:10 p.m.	33.6	

41.6

41.5

8

19

5:20 p.m.

Table 7-3: Measured trip duration as a function of day, departure time, and route taken, for the eastbound direction of Highway 401

7.3 Determination of Time Varying Demands

5:20 p.m.

5:30 p.m.

Number of trips

Sub Total

Total number of trips

7.3.1 Introduction

The formulation of the dynamic LSE and LRE models has been described previously in Chapter 6. In this chapter, these formulated models are applied to the previously described Highway 401 network. The previous sections of this chapter have described the physical network characteristics, the quantity and quality of the available FTMS link flow data, and route and route weight assumptions. Before examining any results, it is necessary to describe the parameter settings of the O-D estimation process.

FTMS link flow data, reflecting traffic conditions on Highway 401 between 5 a.m. and 11 a.m. on June 8, 1992 were utilized. These data were aggregated into 24 periods, each of a 15 minute duration. Neither path, nor node continuity of flow existed. Observed flows were available for only 68 of the 239 links within the network. To provide a constraint to the model for the links for which observed flows did not exist, the link's capacity was considered to be the *observed* flow whenever the estimated flow exceeded the link's capacity.

Since no prior information was available describing the magnitude or structure of the true O-D demand, the seed demand was defined to consist of a constant value in all inter-zonal O-D cells. To limit the amount of computing time required, a maximum of 30 iterations of the algorithm were carried out.

Based on these conditions, the LSE and LRE models were applied to the Highway 401 network. After 10 iterations, the LRE model's marginal improvement was less than 0.5%, so no more iterations were carried out. All 30 iterations were carried out for the LSE model. The results of these applications are discussed in the next sections. The data used to estimate the time-varying demands for the Highway 401 eastbound network are provided in Appendix C.

7.3.2 Evaluation of estimated demands with respect to aggregated observed link flows

A primary measure of the effectiveness of an O-D estimation method, is an aggregate measure of the link flow errors. As indicated in earlier chapters, one such convenient measure is the normalized link flow error (\overline{E}^n for the LSE mode, and \overline{E}^n_r for the LRE model) in which a measure of the link flow error is represented as a proportion of the mean observed flow. A perfect correlation between observed and estimated flows results in a normalized link flow error of 0.0 for the LSE model, and 1.0 for the LRE model.

The change in the normalized link flow error as a function of the iteration number is illustrated in Figure 7-4a for the LSE model, and in Figure 7-4b for the LRE model. After 30 iterations, the demand estimated by the LSE model resulted in a normalized link flow error of 13.9% of the average observed link flow. The average observed link flow, computed from 1705 link flow observations, was found to be equal to 2755 vph, while the average estimated link flow, computed over the same links, was equal to 2766 vph. As illustrated in Figure 7-4a, the greatest reductions in the normalized error occurred during the initial number of iterations. Consistent with observations made in earlier chapters, the marginal improvement decreases with increasing numbers of iterations.

Figure 7-4b illustrates a similar pattern for the LRE model results. The final normalized link flow error after 10 iterations is 1.183, or approximately 18.3% of the average observed link flow. In fact, if \overline{E}^n (instead of \overline{E}_R^n) is computed for the LRE model results, the normalized link flow error is 16.35%. The average estimated link flow, computed for only those links for which observed data were also available, is equal to 2724 vph.

Both Figure 7-4a and Figure 7-4b indicate that carrying out more iterations would reduce the normalized link flow errors further. However, these reductions would likely be quite small, as the marginal improvements (slope of the curve) that are evident at the current maximum iteration number, are already rather small. Even for the relatively few iterations that were carried out, the magnitude of the normalized link flow error is quite small. Given the previously discussed errors in the flow data, the lack of observed data for a large portion of the system, the questionable reliability of the link travel time data, and the limiting assumptions made in selecting the routes, this level of accuracy appears quite high.

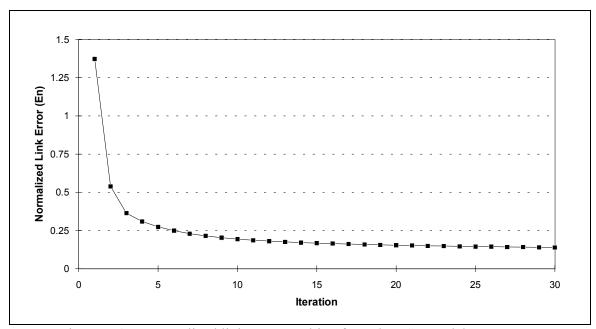


Figure 7-4a: Normalized link error resulting from the LSE model

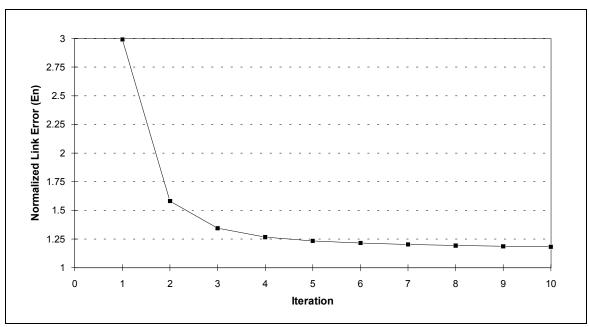


Figure 7-4b: Normalized link error resulting from the LRE model

7.3.3 Evaluation of estimated demands with respect to individual observed link flows

The previous section examined the performance of the LSE and LRE models by analyzing the link flow error aggregated across all links and all time slices. It is also of interest to examine the correlation between observed and estimated link flows by time of day for several individual links. Since observed link flow data are available for 68 of the 239 links in the network, it is not practical to individually examine all 68 links. Rather, just two samples of the total number of links are selected. The first link, link 8, consists of five lanes and is located at the west end of the network, just prior to the diverge at the beginning of the express-collector system. The second link, link 12, consists of three lanes and is located in the express lanes, between the Highway 409 on-ramp and the off-ramp to Highway 400.

Figures 7-5a and 7-5b illustrate the observed and estimated link flows by time of day, for links 8 and 12, respectively. As indicated by both figures, both the LSE and the LRE models predict flows that closely follow the trends in the observed data. Naturally, it is not possible to make any general conclusions based on these data, as it could be argued that these two links are not representative of the entire system.

In order to make more general conclusions, the correlation between observed and estimated link flows can be determined for all of the 1705 link flow observations that are available. Figures 7-6a and 7-6b illustrate this correlation for the LSE and the LRE models, respectively. The computed correlation coefficient between the observed and estimated link flows is 0.9871 for the LSE model and 0.9821 for the LRE model. In both cases, the high level of correlation indicates that there is a strong association between the estimated and observed link flows. However, these values must be considered with some caution as the observed data are not all strictly independent.

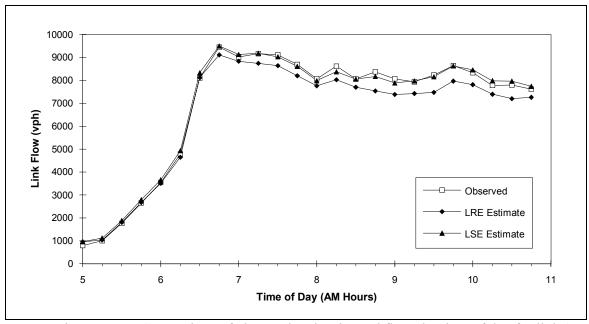


Figure 7-5a: Comparison of observed and estimated flows by time of day for link 8

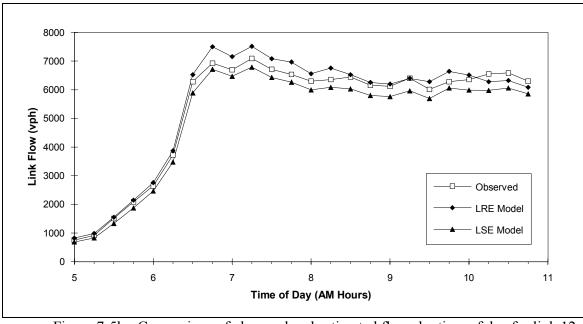


Figure 7-5b: Comparison of observed and estimated flows by time of day for link 12

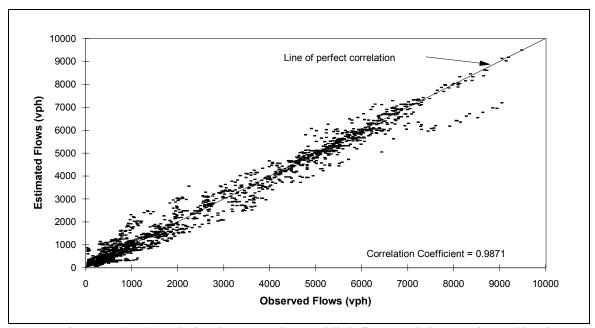


Figure 7-6a: Correlation between observed link flows and those estimated by the dynamic LSE model

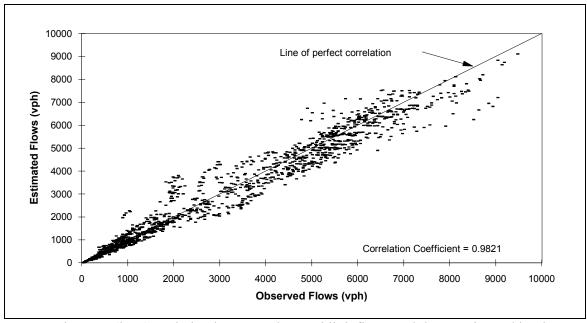


Figure 7-6b: Correlation between observed link flows and those estimated by the dynamic LRE model

7.3.4 Evaluation of estimated demands with respect to observed origin productions and destination attractions

The measures of performance examined thus far have been limited to comparisons between estimated and observed link flows. The actual goal, however, is to estimate O-D demands that closely replicate the true demands. Since the true demands are unknown, it is not possible to make any direct comparisons between the estimated and true demands. It is possible, however, to observe the total number of origin productions and the total number of destination attractions. These values represent the flows observed on the links connected to each origin and to each destination zone. In this manner, comparisons can be made between observed and estimated row and column totals of the demand matrices.

Each interchange on the freeway was considered as both an origin and a destination zone. Based on this definition, it was possible for two or more links to be associated with each origin or destination zone. This occurs when two on-ramps, one carrying southbound flow, and the other carrying northbound flow, exist within the same junction. Table 7-4 provides a listing of the links that were aggregated to form the total origin and destination flows.

It has been noted earlier that observed link flows are available for only 68 of the 239 links within the network. Observed data were not available for many of the defined origin and destination zones. In order to carry out a comparison, observed flows were required for all of the links that were associated with a zone. As a result, comparisons could only be made for four origin zones (Highway 409, Islington Ave., Weston Rd., and Yonge St.) and six destination zones (Dixon Rd., Islington Ave., Weston Rd., Keele St., Dufferin Ave., and Avenue Rd.). For each of these zones, the RMS error between the observed and estimated flows was computed as a proportion of the total observed flow. Figures 7-7a and 7-7b illustrate the results as a function of time of day for origins and destinations, respectively.

Several observations can be made based on Figures 7-7a and 7-7b. First, the LRE model provides much better estimates of the zonal flows than does the LSE model. Second, the quality of the LRE model estimates have much smaller variations across time of day, than do the LSE model estimates. This is particularly evident in Figure 7-7b. This poor performance by the LSE model seems contradictory to the other performance measures examined thus far which indicated that the LSE model performed marginally better than the LRE model.

To provide some understanding for this apparent contradiction, one must first review the objective functions of the two models. The LSE model attempts to minimize the squared absolute link flow error aggregated across all links and all time slices. Since only the absolute error is considered, the algorithm first contends with links having large absolute error. Negative flows are not permitted, so the links with a large absolute error will tend to be those links which have large observed flows. Conversely, links with low observed flows are not likely to have a significant impact on the estimated O-D. Only 30 of the 68 links, for which observed data are available, are ramp links. The remaining 38 links are either express or collector links, having significantly higher flows than the ramp, or zonal, links. Thus, the LSE model places more weight on replicating the higher flow mainline links than the lower flow zonal links.

Consider Figures 7-8a and 7-8b which depict the correlation between the observed destination flows and those estimated by the LSE model and the LRE model, respectively. It is immediately apparent that there exists a much higher correlation between the LRE model estimates and the observed destination flows (r = 0.95), than between the LSE model estimates and the observed flows (r = 0.75). The LSE model tends to over-estimate the destination flows, particularly at low flows. This trend is consistent with the earlier observation that the LSE model tends to place more weight on links having high observed flows. Since flows cannot be negative, errors on links with low flow tend to result from an over-estimation of the flows.

Since the LRE model considers the link flow error relative to the observed flow, it would be expected that the flows estimated by the LRE model would have an approximately consistent relative error. Figure 7-8b indicates that there is a high degree of correlation between the observed destination flows and the LRE model estimates. Again, this is consistent with the model formulation, as a link flow error of 10 on a link with an observed flow of 100 contributes equally to the objective function as a link error of 1000 for a link with an observed flow of 10,000.

Table 7-4: Definition of zones and the links that constitute the zonal flow

Zone Name	Node Type	Number of Links	Link N	umbers
H401 EB Thru	Origin	1	1	
Hwy 427 On	Origin	2	125	126
Dixon Rd. Off	Destination	1	127	
Dixon Rd. On	Origin	2	130	131
Hwy 409 On	Origin	2	139	138
Islington Off	Destination	1	132	
Islington On	Origin	1	142	
Weston Rd. Off	Destination	1	143	
Weston Rd. On	Origin	2	146	147
Hwy 400 Off	Destination	2	148	149
Hwy 400 On	Origin	1	155	
Keele St. Off	Destination	1	156	
Keele St. On	Origin	2	159	160
Dufferin St. Off	Destination	1	161	
Allen Expwy Off	Destination	1	164	
Allen Expwy On	Origin	2	168	173
Bathurst St. On	Origin	1	174	
Avenue Rd. Off	Destination	1	175	
Avenue Rd. On	Origin	2	178	179
Yonge St. Off	Destination	1	180	
Yonge St. On	Origin	1	185	
Bayview Off	Destination	1	186	
Bayview On	Origin	2	189	190
Leslie St. Off	Destination	1	191	
Leslie St. On	Origin	2	194	195
DVP/404 Off	Destination	2	196	197
DVP/404 On	Origin	2	204	205

Victoria Prk Off	Destination	1	208	
Victoria Prk On	Origin	2	211	212
Warden Off	Destination	1	213	
Warden On	Origin	1	218	
Kennedy Rd. Off	Destination	1	219	
Kennedy Rd. On	Origin	2	222	223
McCowan Rd. Off	Destination	1	224	
McCowan Rd. On	Origin	2	227	228
Markam Rd. Off	Destination	1	229	
Markam Rd. On	Origin	2	232	233
Neilson Rd. Off	Destination	1	234	
Neilson Rd. On	Origin	1	239	
H401 EB Thru	Destination	1	106	

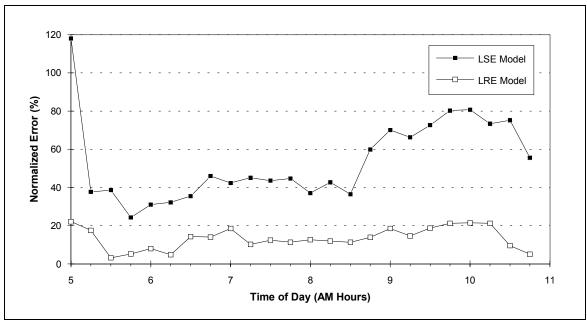


Figure 7-7a: RMS error between estimated and observed total origin productions presented as a percent of the average observed total

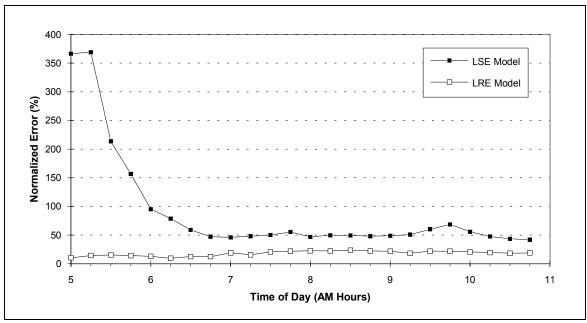


Figure 7-7b: RMS error between estimated and observed total destination attractions presented as a percent of the average observed total

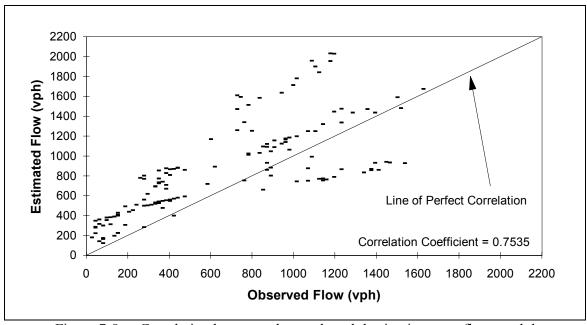


Figure 7-8a: Correlation between observed total destination zone flows and those estimated by the LSE model

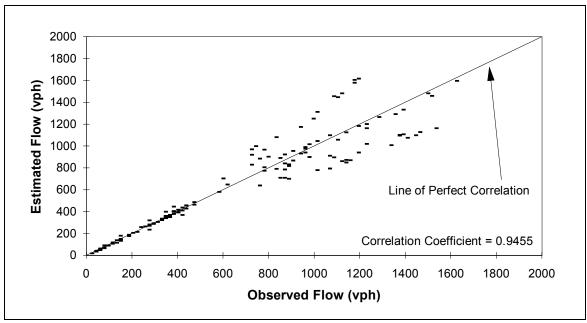


Figure 7-8b: Correlation between observed total destination zone flows and those estimated by the LRE model

7.3.5 Further characteristics of the estimated O-D demands

In addition to the performance measures examined thus far, the estimation of the O-D demand permits the average trip length and the total number of trips to be computed. These characteristics cannot be computed without either the knowledge of the O-D demands, or alternatively several assumptions, such as the average number of links traversed by a vehicle during a trip. The estimation of these characteristics provides for an opportunity to perform a reality check on the estimated O-D. Operators of FTMS and traffic engineers employed by departments of transportation often have some notion of the total number of trips that are made within a network during a specified time period. They may also have some estimate of the average trip length. A favourable comparison of these two quantities against the values derived from the estimated demands serves to place more confidence within the estimated demand.

Figure 7-9 illustrates the average trip length derived from the demand estimated by the LSE and LRE models after each iteration. It is interesting to note, that both models estimate approximately the same average trip length. The LSE model's final estimate of average trip length is 9.9 km while the LRE model's final estimate is 9.7 km. Unfortunately, since no average trip length data were available for the Highway 401 network, it was not possible to compare the model estimates against field data.

Figure 7-10 illustrates the total number of trips represented by the O-D demand estimated by the LSE and LRE model after each iteration. Again, both models estimates are approximately equal. The total number of trip departures during the six hour period from 5

a.m. to 11 a.m. on the eastbound direction of Highway 401 was estimated to be 160 400 trips by the LSE model and 171 000 by the LRE model. Again, no data were available for Highway 401 against which these estimates could be compared.

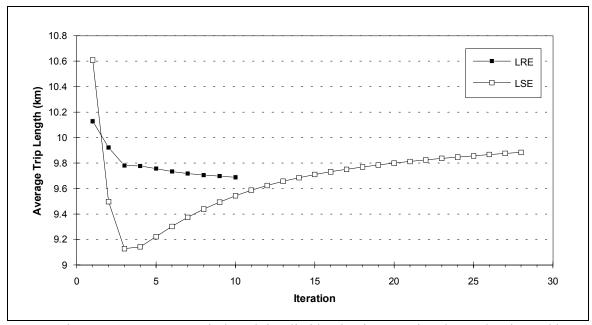


Figure 7-9: Average trip length implied by the time varying demand estimated by the LSE and LRE models as a function of iteration number

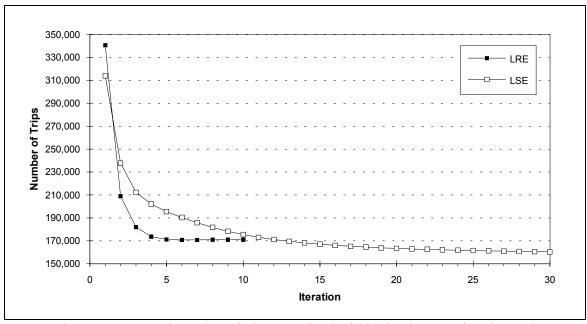


Figure 7-10: Total number of trips contained within the time varying demand estimated by LSE and LRE models as a function of iteration number

7.4 Summary

Highway 401 is a challenging urban freeway corridor to model. The problem of estimating static or dynamic demands are complicated by the existence of significant binary route choices due to the presence of the parallel express and collector facilities, and by the availability of FTMS data for only 45% of the entire system. Furthermore, there existed substantial systematic and random errors within the existing flow and speed data. All of these conditions produced an environment that was not ideal for estimating O-D demands, but was perhaps typical.

The choice of aggregation time slice duration (15 minutes) may be too long in view of the fact that few trips required more than 35 minutes to complete. However, a shorter time slice duration would significantly increase the impact of the errors within the travel time estimates.

Despite limited data, and the presence of systematic and non-systematic errors, both the LSE and LRE models were able to estimate a time-varying O-D demand. Based on the available measures of performance, both models were able to estimate a time-varying demand that successfully reflected the observed link flows.

It is recommended that dynamic O-D demands be re-estimated when the FTMS coverage has been extended to the full system. Furthermore, as the MTO has been made aware of the systematic errors within the FTMS data collection process, efforts have been made to

eliminate, or at least reduce, the impacts of these errors. As such, more recently collected data should be of higher quality.

CHAPTER 8

RGS VEHICLE PROBES AS ESTIMATORS OF DYNAMIC O-D DEMANDS

8.1 Introduction

With the current field trials and anticipated implementation of Route Guidance Systems (RGS) much effort has been expended considering system and user benefits that can be obtained when RGS equipped vehicles receive traffic information in near real-time. However, much less research has been conducted investigating how these near real-time data might be acquired and how reliable these acquired data must be in order to be of consistent marginal benefit. The provision of two-way communications between the Traffic Management Center (TMC) and RGS equipped probe vehicles makes it possible to very efficiently assemble data describing the departure rates of probe vehicles in near real-time. This chapter examines the use of RGS equipped probe vehicles as estimators of real-time O-D departure rate information for all vehicles.

8.1.1 Vehicle probe data

Two types of data can be obtained from probe vehicles, namely link travel time data and origin-destination (O-D) data. As each probe vehicle traverses a link, its travel time can be transmitted back to the TMC. This travel time provides a sample of the current travel time experienced by all drivers utilizing that link at that time. When drivers of RGS equipped vehicles initiate a trip, the intended trip destination is usually entered into the on-board computer system. The vehicle's current location (trip origin) and intended destination can then be transmitted back to the TMC, providing a centralized sample of the population O-D demand departure rates for that time period.

Little research has been conducted into investigating the expected reliability of using probe data to estimate O-D demands. Some earlier research described in the literature has examined the reliability of probe data from a planning, rather than an operational perspective [Van Aerde *et al.*, 1991a]. This work has focused on determining the number of probe vehicles required to obtain some minimum level of confidence in the O-D estimates. Van Aerde *et al.* [1993] have also examined the use of vehicle probes to estimate both travel time and O-D probe data using the same test network used in this chapter. This work, being more general in scope, does not examine the associated theories in the same level of detail, and does not explore the level of surveillance and trip length bias issues that are examined herein. Other research has examined the potential of using probe data as a source for obtaining link travel times [Boyce *et al.*, 1991]. However, this latter research has not comprehensively developed statistically based expressions that quantify the reliability of an O-D estimate made from probe data.

8.1.2 Focus of this chapter

Two critical issues must be addressed when using probe vehicles as sources of traffic information. First, it needs to be established if there is a sufficient quantity of data to provide for adequate coverage of the network. Second, it must be determined if the available coverage is of sufficient quality to generate data which possess sufficient statistical significance to be of practical benefit to the TMC or other RGS equipped vehicles. This chapter describes an analysis approach that assesses this latter issue using statistical sampling theory. The approach is developed for application to general networks, and is subsequently briefly illustrated using a simple hypothetical network using a microscopic traffic simulation model capable of modeling probe vehicles.

This chapter has two objectives. First, the mathematics associated with the relevant statistical sampling equations are developed in order to provide theoretical estimates of the anticipated reliability of O-D estimates that are made from vehicle probe data. Second, a traffic simulation model is used to provide samples of typical vehicle probe data for an example network in order to demonstrate empirically the developed equations. The analytical estimates of the mean O-D rates and the associated confidence limits are then compared with the O-D estimates from the simulation model.

The remainder of this section provides a list of nomenclature that is used strictly within this chapter. Section 8.2 describes the analytical development of equations which will provide statistically unbiased estimates of the mean O-D demands for the entire population from the probe data sample. Section 8.3 describes the development of equations that provide estimates of the reliability of these O-D demands. Section 8.4 extends the development of Section 8.3 by examining the impact of population size and level of network surveillance on estimates of O-D reliability. The developed equations are illustrated by way of a simple example. Section 8.5 then examines a method for determining if the links under surveillance provide an unbiased sample of all of the trips within the network. Section 8.6 of this chapter describes the simple example traffic network that is employed to illustrate the analysis. Results from the simulation and the analytical approach are presented and compared. Conclusions are made in Section 8.7.

8.1.3 Nomenclature

It is useful to first define several important quantities before deriving expressions for the mean and reliability of O-D demand estimates. Each of the following quantities are applicable to only one single time period out of an entire time series of data for an entire peak period. However, each can be expanded to describe a complete time series of observations or estimates by adding a subscript representing time. Since all derivations that will be presented can be applied independently for any single period, the presentation is initially simplified by omitting the time period subscript in the following variables.

 P_{ij} = the population probability of any given trip being between origin i and destination j

 p_{ij}^{y} an estimate of P_{ij} made from some sample of the total population (the probes)

 \vec{D} = the true total number of trip departures initiated in the entire network during the given period

d =an estimate of D based, in part, on probe and standard loop detector data

 D_{ij} = the true number of trip departures between origin i and destination j in the given time period

 d_{ij} = an estimate of the number of trip departures between origin i and destination j

all trips, departing at some time t, that are RGS-equipped

m = an estimate of M based on a sample of the RGS equipped and non-equipped vehicles observed on selected links on the network

 M_{ij} = the population level of market penetration for origin i and destination j

 c_{ij} = the number of probe departure calls received for origin i and destination j during the given period

 n_p = the total number of probe vehicles initiating trips in the network during the current period

In the above definitions, upper case variables represent true (though usually unknown) population values while the lower case variables represent values associated with the probe vehicles, which represent a sample of the entire vehicle population. Lower case variables associated with this sample are typically used as estimates of the population values. The term population refers to the combined fleet of RGS equipped and non-equipped vehicles. The probes are initially considered to represent an unbiased sample from this population.

In practice, only two of the above quantities are known; namely the number of probe calls associated with departures from origin i towards destination j (c_{ij}), and the total number of probe vehicles initiating trips in the network during the current period (n_p). It is desired to determine d_{ij} , an estimate of the true demand between origin i and destination j, and to be able to quantify how close this estimate may be to the true value D_{ij} . It should also be noted that it may be known how many total probe vehicles are registered in an urban area relative to the total number of registered vehicles. The ratio of these two values is different from the M and M_{ij} values that are actually existing on a network during a particular time period as the former ratio is not time dependent. The relationship between several of the above variables is illustrated in Figure 8-1.

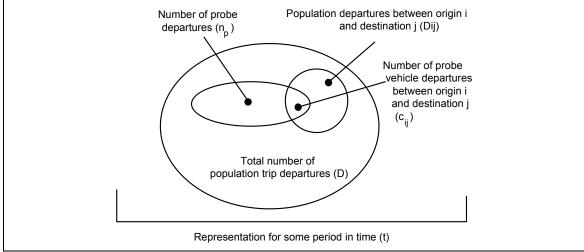


Figure 8-1: Schematic representation of relationships between several key variables which were defined

8.2 Derivation of Estimate of Population O-D Demands

The analysis method which is developed in this chapter, is based on the premise that an estimate of the prevailing O-D demand can be made by scaling the number of probe calls received by the level of market penetration. Therefore, an expression is first developed to provide an estimate of the O-D demand and the variability that is expected to exist within each of the inputs to this expression. Subsequently, expressions are developed that quantify the variability in the estimate of the O-D demand and the impact of considering that the vehicle probes are sampled from finite or infinite populations.

8.2.1 Estimating the population O-D demand D_{ij} - Approach 1

If the level of market penetration for a given time period is defined as the ratio of probe departures to total population departures, Equation [8-1a] would yield an exact estimate of the O-D specific departure rates (D_{ij}) . Unfortunately, as M_{ij} can only be computed exactly by also utilizing Equation [8-1a], Equation [8-1a] is usually of little practical value. Equation [8-1a] could be simplified by dropping the O-D specific subscripts from M_{ij} , an approach that is subsequently shown to be practical for generating an estimate of the mean of D_{ij} . Unfortunately, the generation of an estimate of the variance of D_{ij} would be complicated by the need to estimate both the sampling error in estimating M by m, and the additional error in approximating M_{ij} by M. The interested reader is directed to Appendix D where this approach is pursued and shown to yield the same results as the alternative approach which is presented below.

$$D_{ij} = \frac{c_{ij}}{M_{ij}}$$
 [8-1a]

8.2.2 Estimating the population O-D demand D_{ii} - Approach 2

The alternative approach considers that the overall O-D specific departure rate can be estimated by apportioning an overall estimate of the total fleet departures within a certain time period (D) to all the individual O-D cells which make up the complete O-D matrix, as shown in Equation [8-1b].

$$D_{ii} = P_{ii} \cdot D \tag{8-1b}$$

In this case, the population probability that a trip will be associated with a particular O-D pair (P_{ij}) is considered to be unknown, as is the total number of trip departures in the network D. If the population probability P_{ij} can be approximated by the sample probability P_{ij} , and the total number of vehicle departures D, can be approximated by d, then Equation [8-1c] provides an estimate of the population departures between origin i and destination j.

$$D_{ii} \approx d_{ii} = p_{ii} \cdot d$$
 [8-1c]

In order to use Equation [8-1c] in practice, one must be able to compute p_{ij} and d. The following subsections provide the development of these estimates. Subsequently, the reliability of d_{ij} , as an estimate of the population O-D demand, is examined as a function of the variance of p_{ij} and d.

8.2.3 Estimating the population probability P_{ij} and total departures D

The population probability that a given trip will be associated with origin *i* and destination *j* can be estimated from the number of probe vehicle departures using Equation [8-2].

$$P_{ij} \approx p_{ij} = \frac{c_{ij}}{n_p} \tag{8-2}$$

As indicated in Equation [8-3], the total number of trip departures in the network can be estimated as a function of the number of probe vehicle departures and the prevailing network level of market penetration.

$$D = \frac{n_p}{M}$$
 [8-3]

The network level of market penetration is generally unknown but can be estimated as is shown next.

8.2.4 Estimating the exact level of market penetration from route flows

The exact level of market penetration could ideally be estimated using Equation [8-4a].

$$M = \frac{n_p}{D} = \frac{\sum_{ij} \sum_{ij} a_{r,ij}}{\sum_{ij} b_{r,ij}}$$
 [8-4a]

where: $a_{r,ij} = \text{number of probe vehicles going from origin } i \text{ to destination } j \text{ using route } r \text{ that initiated their trip during the current time period}$ $b_{r,ij} = \text{number of vehicles going from origin } i \text{ to destination } j \text{ using route } r \text{ that initiated their trip during the current time period}$

In practice, there exist three difficulties that prevent the practical use of Equation [8-4a]. First, the level of market penetration, M, is strictly defined as the proportion of the vehicles departing their trip origin, at some time t, that are RGS-equipped. While it is possible to observe the number of vehicles passing an arbitrary point within the network during some time period, it is not possible to measure exclusively those vehicles that are just beginning their trip.

Second, the routes that are utilized by both equipped and non-equipped vehicles, between each O-D pair, are generally unknown. Such routes could potentially be estimated through some form of traffic assignment, but a traffic assignment requires that the O-D demands be known *a priori*. Alternatively, a heuristic route identification algorithm might be employed to satisfactorily identify the routes. Unfortunately, even if some method were to identify the utilized routes accurately, it would not be possible to quantify how close the identified routes were to the unknown true routes. Thus it would not be possible to subsequently quantify the reliability of the O-D estimate.

The third difficulty in applying Equation 8-4a arises in determining the component of the observed link flow that is associated with each O-D. Clearly, if a link is utilized by only a single O-D pair, then all of the observed flow can be attributed to that O-D. However, if a link is used by more than one O-D pair, it is not possible to exactly allocate the portion of the observed flow that is associated with each of the O-D pairs that utilize that link. Ultimately, if it were possible to determine the correct route flows for each O-D pair, and to ascertain when these vehicles began their trips, one would have already determined the O-D demands that are ultimately being sought.

8.2.5 Estimating the approximate level of market penetration from link flows

Since it is not possible to compute M exactly, one may estimate the level of market penetration for time t, based on the proportion of vehicles observed on the network, that are RGS-equipped. This is accomplished by sampling a number of links within the network, as illustrated in Equation [8-4b]. From these sampled links, the ratio of the number of probe

vehicles to the total number of vehicles can be computed as an estimate of the network wide level of market penetration.

$$M \approx m = \frac{\sum_{L} a_{L}}{\sum_{L} b_{L}}$$
 [8-4b]

where: $a_L = \text{number of probe vehicles using link } L$ $b_L = \text{total number of vehicles using link } L$

In order for the above expression to be valid it needs to be assumed that the trip length distributions of RGS-equipped and non-equipped vehicles are approximately the same. If this assumption is not valid, and RGS-equipped vehicles tend to have longer trips, then one would expect them to be observed, by way of detector counts, in a disproportionately higher number than the level of market penetration would warrant. It is further assumed, that the set of links under surveillance provides an unbiased sample of the total number of trips in the network. A method to determine if the links under surveillance do indeed provide an unbiased sample is examined later in Section 8.5.

It should also be noted that an estimate of M, based on Equation [8-4b], may lag in responding to temporal variations in M. Equation [8-4b] relies on the ratio of RGS-equipped to non-equipped vehicles observed on the network, in order to estimate M. As long as a vehicle is still en route, its presence on the network impacts the estimate of M, even though the true M may be quite different from what it was when this vehicle initiated its trip. This effect is amplified in networks which have long average trip durations.

8.2.6 Summary

Despite the above limitations, an estimate of the total number of trip departures can be made by approximating the population level of market penetration using Equation [8-4b] and substituting this value into Equation [8-3], resulting in Equation [8-5a].

$$D \approx d = \frac{n_p}{m}$$
 [8-5a]

On the basis of the previous discussions, it is possible to compute an estimate of the population O-D departures from the number of probe vehicles, the number of probe calls, and an estimate of the network level of market penetration.

$$D_{ij} \approx d_{ij} = p_{ij} \cdot d = \frac{c_{ij}}{n_p} \cdot \frac{n_p}{m} = \frac{c_{ij}}{m}$$
[8-5b]

This equation is consistent with engineering intuition, as it indicates that a best estimate of the total number of departures between any O-D pair, is the number of probe departures between this O-D pair, scaled by the estimate of the level of market penetration. The rather convoluted manner in which this simple equation is derived is necessary in order to provide a mechanism by which the associated variance of D_{ij} can be computed, as is shown next.

8.3 Derivation of Estimate of Reliability of O-D Demand Estimates

The expressions derived in Section 8.2 provide an estimate of the population O-D departure rates based on sample quantities, but these sample quantities are prone to standard sampling errors. This section will develop expressions to quantify these component errors such that the reliability of the overall O-D estimates can be estimated. First, a general expression is developed for the variance of d_{ij} . Since d_{ij} is a function of the estimate of the population probability (p_{ij}) and the total number of departures in the network (d), subsequent sections determine expressions for the reliability of these two component estimates.

8.3.1 Estimating the reliability of the O-D demand estimate

Since the O-D demand estimate d_{ij} in Equation [8-1c] is a function of two random variables $(p_{ij} \text{ and } d)$, the confidence limits of d_{ij} may be derived from the confidence limits of p_{ij} and d as is shown below:

Ang and Tang [1975] indicate that it is possible to use the first order Taylor series expansion to approximate the mean and variance of d_{ij} as shown in Equations [8-6a] and [8-6b].

$$E(d_{ii}) \approx E(p_{ii}) \times E(d)$$
 [8-6a]

$$VAR(d_{ij}) \approx \left(\frac{\partial d_{ij}}{\partial p_{ij}}\right)^{2} VAR(p_{ij}) + \left(\frac{\partial d_{ij}}{\partial d}\right)^{2} VAR(d) + \frac{\partial d_{ij}}{\partial p_{ij}} \frac{\partial d_{ij}}{\partial d} Cov(p_{ij}, d)$$
 [8-6b]

Before continuing, one may examine the third term of Equation [8-6b] which expresses the impact of the covariance of p_{ij} and d. From Equation [8-2], it can be noted that p_{ij} is a function of c_{ij} , the number of probe calls received, and n_p , the total number of probe vehicles. From Equation [8-5a], one cannote that d is a function of n_p and m, an estimate of the level of market penetration. For a given network, at some time t, the total number of probe vehicles is the sample size, and as such is constant. Therefore, any covariance between p_{ij} and d can only arise due to covariance between c_{ij} and m.

If a number of samples are taken, each consisting of n_p probe vehicles, then c_{ij} will vary due to sampling errors. However, the variance in m results from sampling link counts of probe and non-probe vehicles. Thus, sampling error for c_{ij} and m are likely to have a low level of correlation and therefore, the level of correlation between p_{ij} and d is approximated to be negligible. Then, the third term's contribution in Equation [8-6b] is negligible and can be set to zero.

In view of Equations [8-6a] and [8-6b], Equation [8-6c] can be arrived at.

$$VAR(d_{ij}) \approx d^2 VAR(p_{ij}) + p_{ij}^2 VAR(d)$$
 [8-6c]

In order to further develop Equation [8-6c], it is necessary to determine expressions for the variance of p_{ij} and d, as is carried out in the following two subsections.

8.3.2 Estimating the reliability of the population probability

The number of probe calls that are received for a given O-D pair can be considered to follow a binomial distribution where each probe vehicle is either associated with a particular O-D, or it is not. Thus, the number of probe vehicles (n_p) constitute a Bernoulli series with an unknown population probability of *success* of P_{ij} . For a single trial, the variance of the population probability can be expressed as $P_{ij}(1-P_{ij})$. For a number of trials, Equation [8-7a] provides the variance of P_{ij} .

$$\sigma_{P_{ij}}^2 = \frac{P_{ij}(1 - P_{ij})}{n_p}$$
 [8-7a]

Since P_{ij} is unknown, the population probability is approximated using the sample probability p_{ij} and the variance of the sample probability is as defined in Equation [8-7b].

$$\sigma_{P_{ij}}^2 \approx S_{p_{ij}}^2 = \frac{p_{ij}(1-p_{ij})}{n_p}$$
 [8-7b]

8.3.3 Estimating the reliability of the total number of network departures

Having developed an expression for the reliability of p_{ij} , it is now necessary to develop similar expressions for the reliability of d, the total number of network departures. First, it must be recognized that, as illustrated earlier in Equation [8-5a], d is a function of the total number of active probe vehicles and the level of market penetration.

It is again possible to use the first order Taylor series expansion as first introduced in Equations [8-6a] and [8-6b] to approximate the mean and variance of d. However, for the expansion of d, the number of vehicle probes, n_p , is a constant, and therefore the variance of n_p is zero. Thus in this case, the approximate variance of d is as shown in Equation [8-8a]. The derivative of the total number of network departures with respect to the level of market penetration is expressed in Equation [8-8b].

$$VAR(d) \approx VAR(m) \times \left(\frac{\partial}{\partial} \frac{d}{m}\right)^2$$
 [8-8a]

$$\frac{\partial}{\partial} \frac{d}{m} = -n_p m^{-2}$$
 [8-8b]

Substituting Equation 8b into Equation 8a produces Equation [8-8c].

$$VAR(d) \approx VAR(m) \times \left(\frac{-n_p}{m^2}\right)^2 = \frac{n_p^2}{m^4} VAR(m)$$
 [8-8c]

Equation [8-8c] provides an approximation of the variation expected within d, the total number of O-D departures throughout the network. Equation [8-8c] defines this variance in

terms of the variance of the level of market penetration, however, this quantity has yet to be defined.

To quantify the variance of the level of market penetration it can be stated that when a trip is sampled on a particular link, it is either an RGS-equipped vehicle, or it is not. Again, this leads to a Bernoulli sequence having a population probability of being RGS-equipped equal to M. The number of probe vehicles that exist within a finite sample of vehicles (q) from the network can then be considered to follow the binomial distribution. Using a line of reasoning similar to that used to develop Equations [8-7a] and [8-7b], it can be shown that the variance of the level of market penetration is as defined in Equation [8-9].

$$\sigma_M^2 \approx S_m^2 = \frac{m(1-m)}{q} \tag{8-9}$$

8.3.4 Summary

Section 8.3 has provided the development of a number of equations that can be used to approximate the reliability of the O-D demand estimates made based on probe information. Initially, a general expression for this reliability was presented in Section 8.3.1. This expression was a function of the variance of p_{ij} and d. Expressions for the reliability of p_{ij} and d were developed in Sections 8.3.2 and 8.3.3, respectively. These three elements can be aggregated in order to provide a more meaningful expression for the reliability of d_{ij} .

Substituting Equations [8-5a] and [8-8c] into Equation [8-6c] produces Equation [8-10a], a general form for expressing the variance of d_{ii} in terms of p_{ij} and m.

$$VAR(d_{ij}) \approx \frac{n_p^2}{m^2} S_{p_{ij}}^2 + \frac{p_{ij}^2 n_p^2}{m^4} S_m^2$$
 [8-10a]

Substituting Equations [8-7b] and [8-9], expressions for the variance of p_{ij} and m respectively, and Equation [8-2] into Equation [8-10a] results in Equation [8-10b].

$$S_{d_{ij}}^{2} = \frac{c_{ij}(n_{p} - c_{ij})}{m^{2}n_{p}} + \frac{c_{ij}^{2}(1 - m)}{m^{3}q}$$
 [8-10b]

Equation [8-10b] can be used to determine the variance of the estimated O-D demand between origin i and destination j. This equation has been developed assuming that sampling has been done from an infinite population. Section 8.4 discusses the impact of this assumption and provides a variation of Equation [8-10b] that can be used when considering a finite population. Section 8.4 also provides some discussion on how to determine the appropriate sample size, q, that should be used.

8.4 Practical Considerations

The above derivations considered that the level of market penetration and the probabilities for associating a particular trip with an O-D were estimated based on samples taken from an infinite population. The approximation is appropriate for low sampling rates, but not for

scenarios where a significant proportion of the population is sampled without replacement, as shown next.

8.4.1 **Population size**

In the equations developed thus far, it has been assumed that the variance of both p_{ii} and mcan be determined based on the binomial distribution, which assumes sampling from an infinite population. However, it is clear, that if one considers only a specific day, then there exist a finite number of trips that are initiated within the network during that day. As an extreme, if the level of market penetration is 100% then the entire population is RGSequipped and the O-D demands for the given time period should be known with certainty.

For the analysis of finite populations, it is appropriate to utilize the hypergeometric distribution instead of the binomial distribution. The mean values for the binomial and the hypergeometric distributions are the same, but the variance of the hypergeometric distribution requires that the variance of the binomial distribution be multiplied by an additional term (the finite population correction factor), as presented in Equation [8-11a].

$$F = \frac{q_p - q}{q_p - 1} \tag{8-11a}$$

where: F = finite population correction factor

 q_p = population size from which samples are taken without replacement q = sample size that is taken from the above population

Thus, for a finite population, the variance of the estimated population probability as expressed by Equation [8-7b] is modified as indicated in Equation [8-11b]. Similarly, the variance of the level of market penetration as expressed in Equation [8-9] is modified as indicated in Equation [8-11c].

$$S_{p_{ij}}^{2} = \frac{p_{ij}(1 - p_{ij})}{n_{p}} \left(\frac{d - n_{p}}{d - 1}\right)$$
 [8-11b]

$$S_m^2 = \frac{m(1-m)}{q} \left(\frac{d-q}{d-1}\right)$$
 [8-11c]

The determination of the sample size, q, in Equation [8-11c], is examined in the following section.

8.4.2 Level of surveillance

The sample used for estimating the network level of market penetration should ideally consist of individual trips which are sampled from the entire population of trips. However, since it is not possible to query non-RGS vehicles as they initiate trips, link counts are utilized as surrogates, as presented in Equation [8-4b]. If very many of the links in the network are under surveillance, then almost all of the trips will be counted at some point. Unfortunately, as each trip likely uses more than one link, the sample, consisting of the total

number of link counts, will exceed the population – the total number of trips. Thus, if the level of surveillance is very high, then the sample size should be equal to the population, and the variance of the estimated level of market penetration for a finite population is equal to zero.

For very low levels of surveillance, most trips are only counted once by loop detectors. As a result, link counts approximate independent observations of trips. In this case, the sample size is the number of link counts and the variance of the level of market penetration for a finite population is expressed by Equation [8-12].

$$S_m^2 \text{(low surveillance)} = \frac{m(1-m)}{B} \left(\frac{d-B}{d-1}\right)$$
where: $B = \sum b$.

where: $B = \sum_{L} b_{L}$

 b_L = total number of vehicles using link L.

8.4.3 **Impact on O-D reliability**

An expression for the reliability of O-D demand estimates was made in Section 8.3. However, this expression was based on the assumption that the population was infinite. Furthermore, this expression did not provide any insight into the appropriate magnitude of the sample size, q, used for the estimation of the level of market penetration. Section 8.4 has addressed each of these two issues. On the basis of the discussions presented in Section 8.4.1 and Section 8.4.2 it is possible to develop four different expressions for the variance of the O-D estimate.

$$S_{d_{ij}}^{2} = \frac{c_{ij}}{m^{2}} \left(1 - \frac{c_{ij}m}{n_{p}} \right)$$

$$\begin{cases} Infinite population \\ High level of surveillance \end{cases}$$

$$S_{d_{ij}}^{2} = \frac{c_{ij}}{m^{2}} \left(1 - \frac{c_{ij}}{n_{p}} + \frac{c_{ij}}{mB} - \frac{c_{ij}}{B} \right)$$

$$\begin{cases} Infinite population \\ Low level of surveillance \end{cases}$$

$$\begin{cases} S_{d_{ij}}^{2} = \frac{c_{ij}(n_{p} - c_{ij})(1 - m)}{m^{2}(n_{p} - m)}$$

$$\begin{cases} Finite population \\ High level of surveillance \end{cases}$$

$$S_{d_{ij}}^{2} = \frac{c_{ij}(1 - m)}{m^{3}} \left[\frac{Bm(n_{p} - c_{ij}) + c_{ij}(n_{p} - mB)}{B(n_{p} - m)} \right]$$

$$\begin{cases} Finite population \\ High level of surveillance \end{cases}$$

$$\begin{cases} Finite population \\ High level of surveillance \end{cases}$$

$$\begin{cases} S_{d_{ij}} = \frac{c_{ij}(1 - m)}{m^{3}} \left[\frac{Bm(n_{p} - c_{ij}) + c_{ij}(n_{p} - mB)}{B(n_{p} - m)} \right]$$

$$\begin{cases} Finite population \\ Low level of surveillance \end{cases}$$

$$\begin{cases} Finite population \\ Low level of surveillance \end{cases}$$

$$\begin{cases} S_{d_{ij}} = \frac{c_{ij}(1 - m)}{m^{3}} \left[\frac{Bm(n_{p} - c_{ij}) + c_{ij}(n_{p} - mB)}{B(n_{p} - m)} \right]$$

Having quantified the variance of the O-D estimate, confidence limits can be determined. If it is assumed that the normal distribution is a suitable approximation for the distribution of the variance of the O-D estimate, the 95% confidence limits are obtained by defining boundaries at plus and minus two standard deviations from the estimate of d_{ij} .

8.4.4 Example illustration

To illustrate the nature of the relationship between the variance, as computed from the appropriate equation above, and network conditions, consider a network consisting of 7 zones in which 50,000 trips are initiated during a given time period. If all O-D pairs are approximately equally likely, then the population probability that a trip will be associated with origin *i* and destination *j* may be considered to be equal to 0.02. It is assumed that a high level of surveillance exists within this network such that in the estimation of the level of market penetration, all trips are included within the sample. Then, Figure 8-2 illustrates the coefficient of variation (the standard deviation as a portion of the mean) of the O-D demand estimate as a function of the level of market penetration. For levels of market penetration less than 10%, the COV exceeds 10%. When it is assumed that sampling is taken from a finite population, then the COV reduces to zero at a level of market penetration of 100%.

It is also instructive to examine how the COV varies with total number of trips. Figure 8-3 illustrates COV as a function of the total number of trips for 10% and 25% levels of market penetration. Again, the population probability that a trip is associated with an O-D pair is 0.02. Figure 8-3 indicates, that for this example, in order to ensure that the COV does not exceed 0.5, a total of at least 2000 trips must be made at a level of market penetration of 10%. This implies that at least 200 probe vehicles are required to be active during this time period.

Though 200 probe vehicles may not initially seem to be a demanding requirement, it should be noted that these 200 probes must be initiating trips during the current time period. If O-D rates are desired for each 15 minute period, then 2400 active probes are required for a 3 hour commuting period.

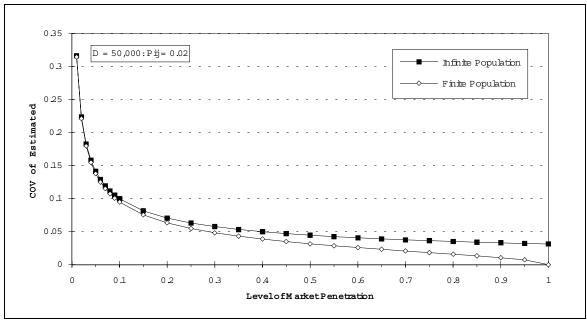


Figure 8-2: Illustration of the coefficient of variation of estimated O-D demand as a function of level of market penetration assuming all links in network are under surveillance

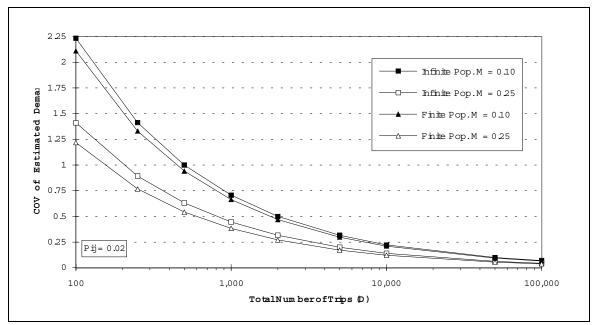


Figure 8-3: Illustration of the coefficient of variation of estimated O-D demand as a function of total number of trips assuming all links in network are under surveillance

Evaluation of Network Level of Surveillance 8.5

It has been shown that an estimate for the mean and variance of the O-D demands can be determined. However, these equations were derived under the assumption that an unbiased estimate of the level of market penetration can be made based on link flows. If the links that are under detector surveillance do not provide an unbiased sample of all of the O-D trips existing within the network, then estimates of the mean and variance of the O-D demands are not strictly correct. This section develops a method by which the existing links under surveillance can be tested to determine if they provide an unbiased sample of the existing O-D demands.

The proposed methodology is carried out by comparing the probability of a trip belonging to a specific O-D pair, as computed from the probes traversing links under surveillance, with the probabilities computed using probe calls. If the probability estimated based on the link information falls within the confidence limits of the probabilities estimated from the probe calls, then it is surmised that the links provide an unbiased sample. If, however, the probability estimated from the link information falls outside the confidence limits, then it must be concluded that the links do not provide an unbiased sample.

Equation [8-2] is used to determine an estimate of P_{ij} for all origins and destinations. Equation [8-7b] provides an estimate of the variance of the sample probability. It is desirable to be able to define confidence limits about the estimate of the mean probability. To do this, it is advantageous to use the Normal distribution as it is symmetrical. According to the literature [Lapin, 1983], the normal distribution can be used to approximate the binomial distribution when the conditions defined in Equations [8-14a] and [8-14b] are met.

$$q \cdot \pi > 5 \tag{8-14a}$$

$$q(1-\pi) > 5$$
 [8-14b]

where: q = sample size $\pi = \text{population probability of success}$

The 95% confidence limits for the probability of any trip being associated with origin i and destination j, as shown in Equation [8-15], can then be created.

$$p_{ij}(95\% \text{ C.L.}) = p_{ij} \pm 1.96S_{p_{ij}}$$
 [8-15]

It is also possible to estimate the probability that a trip will be associated with origin i and destination j from link information. Each time an RGS-equipped vehicle traverses a link under surveillance, that vehicle's origin and destination can be recorded. Then, for all links under surveillance, an estimate of p_{ij} can be made based on the proportion of the probe vehicles traveling on the links under surveillance that were associated with each O-D pair (Equation [8-16a].

$$p_{ij}^{E} = \frac{\sum_{L} a_{L,ij}}{\sum_{L} a_{L}}$$
 [8-16a]

where: $a_{L,ij}$ = number of probe vehicles associated with origin i and destination j counted on link L which is a link under surveillance a_L = number of probe vehicles traversing link L, which is a link under surveillance

If the trip length distribution for all O-D pairs is not the same, then Equation [8-16a] should be adjusted to normalize for trip length as shown in Equation [8-16b]. It is assumed that trip length provides an adequate measure of the number of links traversed.

$$p_{ij}^{E} = \left(\frac{\sum_{L} a_{L,ij}}{\sum_{L} a_{L}}\right) \left(\frac{\overline{L}}{\overline{L}_{ij}}\right)$$
[8-16b]

where: \overline{L} = average trip length across all O-D pairs \overline{L}_{ij} = average trip length for origin i and destination j

If p_{ij}^E falls outside the confidence limits calculated using Equation [8-15], then the links under surveillance are assumed to provide a biased sample of the trips within the network. If the links under surveillance do not provide an unbiased sample, then Equations [8-13a] through [8-13d] must be used with caution.

8.6 Example Network Description and Results

In order to illustrate the potential of vehicle probes to provide O-D information, an example network is required. To permit an analytical analysis to also be performed for the same network, the hypothetical network first introduced in Section 6.6.1 is once again utilized.

8.6.1 Network structure

Figure 8-4 illustrates the example network configuration. The network consists of 7 zones at the periphery of the network, 5 traffic signals, and 40 directional links. Fourteen of the 40 links are zone connectors, 8 are single lane arterials, 10 are dual lane ramps, and the remaining 8 links are two lane freeways.

A time series of 15 minute O-D demands, representing typical conditions during an 8 hour period commencing at 4 A.M., was produced. These demands, when simulated on the network, resulted in freeway link flow peaking characteristics typically observed on an urban multi-lane freeway in Toronto, Canada [Hellinga *et al.*, 1993]. As indicated in Figure 8-5, a time series of user-equilibrium multi-path routes was derived based on the time series of static O-D demands. The network structure, time series of O-D demands, and the time series of multi-path user-equilibrium paths were all used as input to the network traffic simulation model INTEGRATION. Detailed descriptions of the INTEGRATION model are available in the literature [Van Aerde *et al.*, 1991b; Van Aerde *et al.*, 1988].

The INTEGRATION simulation model has the capabilities of modeling the two way communication capabilities of probe vehicles. When each probe vehicle initiates its trip, its

origin and ultimate destination are transmitted and recorded. While the probe is en route, the time taken to traverse each link is transmitted and recorded each time a node is reached. When a probe vehicle arrives at its destination, the trip duration is reported. These data are recorded in a probe information log, a portion of which is illustrated in Figure 8-6.

The network was simulated considering all vehicles as RGS equipped. However, though the probe vehicles reported information, they did not receive information with which to alter their routes. All routes were pre-defined. To evaluate the effect of level of market penetration, the probe reports, recording when probes initiated their trips, were randomly sampled to provide levels of market penetration from 5% to 100% in 5% increments. It should be noted that in this analysis, an aggregate level of market penetration was applied. The level of market penetration did not systematically vary temporally or spatially.

On the basis of these random probe samples, an analysis was conducted to determine how closely estimates of O-D demands matched the *true* data. Within the simulation analysis, the *truth* was defined as those conditions described when 100% of the vehicles were considered to be probe vehicles.

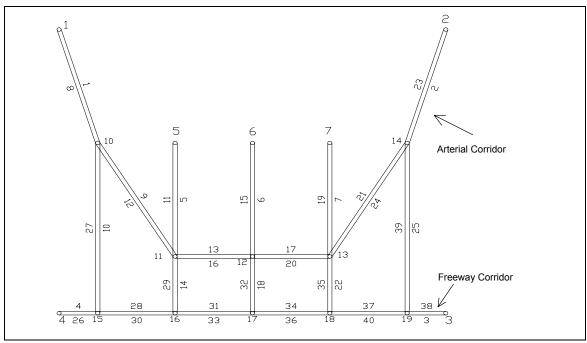


Figure 8-4: Configuration of example traffic network

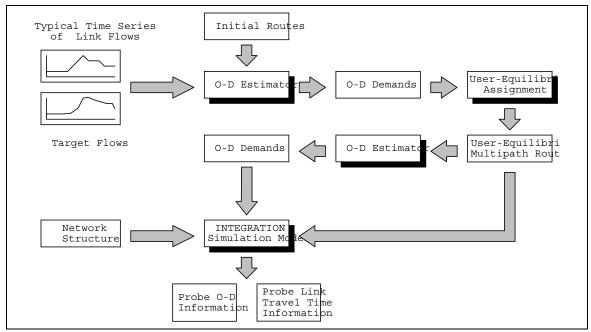


Figure 8-5: Process used to produce realistic traffic conditions on example network

pro	be data log	file :			
1	4 2 3	4	2	Data Legend:	
				Column	Description
11	4 38 38	36 (0.00	1	1 = departure record 2 = arrival record 11 = link completion record
11	. 11 4 103 37 65	5 0.00	2	vehicle number	
•	3	clock time (seconds)			
11	4 169 3	4 65	0.00	4	record type 1 and 2: origin record type 11: link number
	· ·			5	record type 1 and 2: destination record type 11: link travel time
2 4 426 4 3 1 425 1		6	record 1: expected entrance time record 11: stop fraction		

Figure 8-6: Sample of typical probe information provided by simulation model

8.6.2 Results for example network

The simulation model provides the number of probe calls (c_{ij}) received from the probe RGS equipped vehicles during each time period for each O-D pair. In addition, the number of

probe vehicles on each link and the total number of vehicles on each link are available each period. In practice, the number of probe vehicles on a link can be determined from the probe logs. The total number of vehicles on a link would be obtained from loop detector counts.

Figure 8-7 illustrates the random fluctuations in the level of market penetration that are observed on the network. It is evident, that despite the global level of market penetration being held constant throughout the simulation, significant fluctuations occur in the level of market penetration that is experienced on the network. The occurrence of these fluctuations are due only to randomness in the generation of probe and non-probe vehicles at each origin zone. When sample sizes are small (i.e. the O-D demand is small) these fluctuations are particularly evident and there is greater opportunity for extreme values of the level of market penetration to occur.

Figure 8-8 illustrates the estimated and actual O-D demand, as well as the confidence limits about the estimate, for origin zone 1 and destination zone 3 by time of day. The overall level of market penetration (*M*) is 20%. Equation [8-5b] is used to compute an estimate of the demand. Since all links in the example network are assumed to be under surveillance, Equation [8-13a] is used to provide an estimate of the confidence limits for an infinite population, and Equation [8-13c] for a finite population. The results indicate that though the O-D estimates fall within the confidence limits, the estimates can be highly inaccurate even for a level of market penetration as high as 20%.

It is instructive to examine the change in the estimated O-D demand with changes in level of market penetration. Figure 8-9 illustrates the estimated O-D demand between origin zone 1 and destination zone 3 during the time interval from 7:15-7:30 AM. Estimated demands approach the true demand with increasing level of market penetration, such that for a level of market penetration of 100% the estimate is exact. However, though the estimated values converge to the true demand, the convergence trend is only maintained on average. For example, the estimate of a particular O-D demand is not necessarily more accurate at a 20% level of market penetration than it would be for a 5% level of market penetration. This is illustrated in Figure 8-9 as, in this example, the estimate at 40% level of market penetration is more accurate than the estimate made for a 70% level of market penetration.

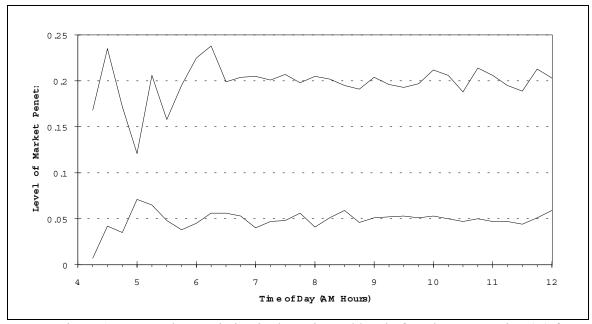


Figure 8-7: Random variation in the estimated level of market penetration (*m*) for a global level of market penetration (*M*) of 5% and 20%

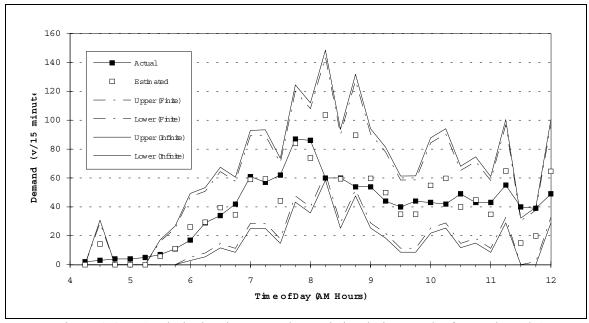


Figure 8-8: Analytical estimates and actual simulation results for a selected O-D demand over time of day for an average level of market penetration of 20% (O-D pair 1-3)

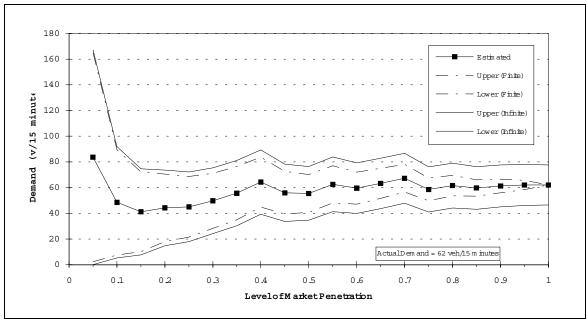


Figure 8-9: Analytical estimates and simulation results for a selected O-D demand for various levels of market penetration (O-D pair 1-3 for time period from 7:15 - 7:30 AM)

The previous discussion has been limited to the examination of a specific O-D pair. Aggregate results for all O-D pairs in the network and for all time periods can also be examined. Figure 8-10 illustrates the aggregate accuracy of O-D estimates, made based solely on probe information, as a function of the average level of market penetration. The accuracy of the O-D estimates is measured by the root-mean squared (RMS) O-D error as computed by Equation [8-6]. To provide some relative scale, the RMS error is normalized by dividing the result of Equation [8-6] by the average actual O-D demand.

RMS ERROR =
$$\sqrt{\frac{\sum_{i=1}^{N} \sum_{j=1}^{N} (D_{ij} - d_{ij})^{2}}{N^{2}}}$$
 [8-17]

where: N = number of zones in network

For an average level of market penetration of 5%, the average RMS O-D error is approximately 78% of the average true O-D demand. As the level of market penetration increases, the O-D error decreases non-linearly. At a level of market penetration of 100%, the O-D estimate is exact. If it is desired to estimate O-D demands to such a level of accuracy that the RMS error does not exceed 20% of the average observed O-D demand, Figure 8-10 indicates that, for this network example, an average level of market penetration of 40% is required.

It should be noted that the results presented in Figure 8-10 must be interpreted in conjunction with the conditions of the example network. Specifically, the level of market

penetration used within the simulation did not systematically vary either temporally or spatially. As well, loop detector data were assumed to be available for all links in the network, and to be free of measurement errors. Thus, the results presented within this chapter could be considered to be a *best case* scenario. It is likely that error rates that would be experienced in the field would be somewhat higher than those estimated here.

8.7 Summary

The above chapter has raised the issue that dynamic traffic assignment models, which are executed in real-time, will need to rely on imperfect O-D data. It also has provided an initial effort, based on statistical sampling theory, to estimate how this level of imperfection varies as a function of the level of market penetration of RGS-equipped probe vehicles.

The application of statistical sampling theory required a number of approximations to be made related to the covariance between various terms. While the implications of these approximations were not explicitly quantified, the subsequent high level of agreement between the analytical and simulation results provides a considerable level of confidence in the appropriateness of approximations that were made.

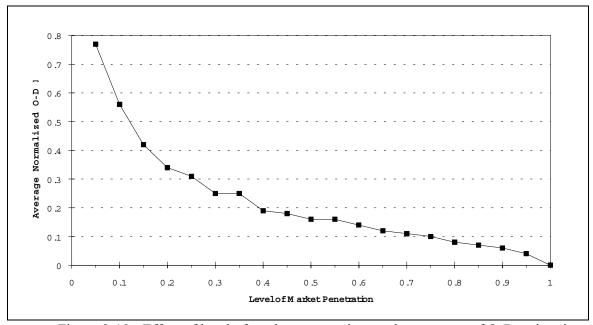


Figure 8-10: Effect of level of market penetration on the accuracy of O-D estimation

The simulation results provided a very graphic representation of the nature of the underlying sampling and estimation processes. The agreement between the simulation and analytical results also provides considerable support for the application of the simulation model for scenarios having a level of complexity that precludes the use of the analytical estimates. The high degree to which the confidence limits depend upon the specific

characteristics of the network and associated traffic flows will necessitate a unique execution of these analyses for each network site.

The lack of highly reliable O-D estimates for levels of market penetration less than 30 to 50% indicates that synthetic O-D generators are certainly not yet obsolete. Furthermore, the explicit estimation of confidence limits about probe based O-D estimates will permit a direct quantitative assessment of whether probe based O-D estimates are superior to synthetic estimates and what weightings should be applied when the probe based and synthetic estimates need to be combined.

CHAPTER 9

CONCLUSIONS AND RECOMMENDATIONS

9.1 Overview

In this thesis, two types of origin-destination demand estimation models were formulated and for them, corresponding iterative solution algorithms were developed. These models are capable of estimating static and/or dynamic O-D demands on the basis of observed link flow data for a subset of the network. These models were derived from a sound theoretical and mathematical basis, yet are sufficiently robust to permit their application to many practical real world traffic networks. To carry out the necessary computations, the LSE and LRE iterative solution algorithms were incorporated into a program called the QUEENSOD model.

Specifically, the LSE and LRE models incorporate the ability to reflect the relative reliability of the observed link traffic flows and of the prior O-D information for each O-D cell. The performance characteristics of both models were illustrated for several hypothetical networks for which true demands were known and for which analytical solutions could be obtained. The models were also applied to a 35 km section of a multi-lane urban freeway in Toronto, Canada.

The thesis also examined the expected statistical reliability of O-D estimates that are made solely on the basis of RGS equipped vehicle probe data. This examination led to the development of several analytical expressions quantifying this statistical reliability, where the expressions were compared with simulated results for a hypothetical example network.

9.2 Conclusions

The main conclusions, that can be drawn on the basis of the research reported on in this thesis, are presented in the following three sections. The first section provides conclusions that are made regarding the proposed LSE and LRE model formulations. In the second section, conclusions are made on the basis of the results obtained from the application of the LSE and LRE models to two hypothetical networks and the Highway 401 network. Finally, the third section provides conclusions regarding the potential of using RGS probe vehicle data to estimate O-D demands for all vehicles.

9.2.1 Proposed models

- a) Minimizing the squared difference between observed and estimated link flows does not ensure that O-D estimates are non-negative. The inclusion of inequality constraints, that explicitly constrain estimated demands to the non-negative region, significantly complicates the analytical solution to the optimization of the error function. The use of truncation and projection methods for ensuring non-negative demands, can result in the selection of a solution that is sub-optimal. Furthermore, in these cases, the magnitude of this sub-optimality is usually not known.
- b) In the presence of multiple optima, a system of linear equations, which describes the O-D estimation problem, is under-specified. Consequently, the traditional solution approaches which rely on matrix inversion in order to solve this system are not applicable. Furthermore, when the system is not under-specified, this matrix approach imposes excessive computer memory requirements, making it impractical for most realistically sized networks.
- c) An iterative solution approach was proposed that ensures that the estimated demands are non-negative. This approach is also applicable when multiple optima exist, or when only a single optimum exists. Furthermore, this iterative approach imposes more modest computer memory requirements that can usually be met using current personal computer capabilities. The iterative method is a modification of the Jacobi iterative technique. For the LSE model, it was found that a relaxation factor could be incorporated to ensure convergence, even when diagonal dominance criteria were not satisfied.
- d) The LRE model was formulated on the basis of minimizing the squared natural logarithm of the relative link flow error. Instead of considering the absolute magnitude of the error, this formulation considers the magnitude of the error relative to the observed flow. It was shown that this formulation leads to a system of non-linear constraints which does not have an explicit closed-form algebraic solution. It was concluded that a solution, that satisfies the system of non-linear constraints, always exists but the uniqueness of the solution was not proven. Neither node, nor path flow continuity is required, but since the system of non-linear constraints does not have a closed-form algebraic solution, an iterative solution algorithm needed to be developed that successively modifies an initial seed until some desired level of convergence is reached.

9.2.2 Application of proposed models

- a) Fully dynamic O-D demands are superior to piece-wise static demands in that they reflect the temporal relationships between trip departure times and arrival times of links. The importance of these temporal inter-relationships to the accuracy of the estimated demands increases as the trip duration becomes an increasingly larger fraction of the period duration.
- b) The LRE model was found to converge more rapidly than the LSE model for the two hypothetical network applications that were considered in this thesis. It was not possible, however, to categorically state that the LRE model always converges more rapidly than the LSE model, as each model's convergence properties are likely a function of the network's characteristics.
- c) Despite the different convergence rates observed for the LSE and LRE models, the accuracy of the final model O-D estimates were comparable.
- d) It was concluded that the Highway 401 network represented a challenging system for which to estimate O-D demands. The task of estimating demands is complicated by the existence of two parallel route choices, by the availability of FTMS data for only 45% of the network, by the lack of any prior O-D information, and by the lack of any data that would reflect the true O-D demands.
- e) It was concluded that the LSE and LRE models were able to estimate dynamic demands for the Highway 401 network that accurately reflected the observed link flows, as estimates from both models resulted in correlation coefficients between estimated and observed link flows of approximately 98%. It was also observed that the LRE model again converged more rapidly than did the LSE model. Furthermore, a comparison between the estimated and observed zonal productions and attractions indicated that, as expected, the LRE model was significantly more accurate in replicating these flows. It was not possible, however, to make any conclusions regarding which model is most appropriate in general, as each model's performance is again likely to be a function of the network's characteristics.

9.2.3 Potential of probe data

a) The anticipated implementation of operational route guidance systems (RGS) has led some to conclude that near real-time O-D information from RGS probe vehicles make the development of synthetic O-D estimation models unnecessary. To explore this hypothesis, statistical sampling theory was employed in Chapter 8 to develop expressions, which quantify the expected reliability of O-D demands that were estimated solely on the basis of RGS probe data. This development required a number of approximations to be made related to the covariance between various terms and the distribution of RGS probes throughout the general vehicle population. However, while the implications of these approximations were noted to yield conservative estimates and were not quantified further, it was concluded that the high level of correlation between

- the analytical and simulation results provides a measure of confidence in the appropriateness of these approximations.
- b) On the basis of the application of the developed expressions to a hypothetical network, it was concluded that reliable O-D estimates, made solely on the basis of probe data, are not likely to be sufficiently accurate to be of practical benefit until the level of market penetration reaches the 30 to 50% level. It was also concluded that probe-based estimates could be used effectively as prior information within an appropriate synthetic O-D estimation model, such as the LSE or LRE models that were developed in this thesis. Furthermore, the knowledge of the reliability of these probe-based estimates may also be incorporated into the synthetic O-D model.

9.3 Limitations of this Research

In this thesis, a number of assumptions and approximations needed to be which place some limitations on the applicability of the resulting models and equations. The following two sections explicitly state these limitations, first for the LSE and LRE models, and then for the statistical analysis of RGS probe data.

9.3.1 Limitations of the LSE and LRE models

- a) Similar to most existing practical dynamic O-D estimation methods, the LSE and LRE models require the *a priori* knowledge of the prevailing routes and route weights. These routes need not be completely accurate, however, inaccuracies in the chosen routes or route weights will negatively impact the quality of the estimated demands.
- b) The travel time on each link must be known for each period for which observed flows are available. These travel times may be estimated on the basis of spot speeds recorded by detectors, or by estimates from a speed-flow relationship.
- c) Due to the multiplicative nature of the LRE solution algorithm, zero cells in the seed demand matrix cannot be altered regardless of the observed flows.
- d) The modified Jacobi iterative method, which was developed for solving the system of linear equations that result from the LSE model, requires diagonal dominance in order for convergence to be guaranteed. As this dominance may not always exist in the system of normal equations, the LSE iterative solution uses a relaxation factor α to increase the likelihood of convergence. It is possible that if α is insufficiently small, the solution algorithm may not converge.
- e) Both the LSE and LRE models require the specification of a seed matrix, reflecting prior knowledge of the structure and magnitude of the O-D demand. The models permit the specification of the relative accuracy of individual seed O-D cell values, but no explicit consideration is given to maintaining the overall structure of the seed matrix.

9.3.2 Limitations of the statistical evaluation of probe data

The expressions, that quantify the expected quality of O-D demand estimates made solely on the basis of RGS probe vehicle data, required that several assumptions be made, resulting in the following limitations:

- a) It was assumed that the level of correlation between the population probability that a trip will be between origin i and destination j, and the total number of trip departures, is negligible.
- b) It was assumed that no systematic bias existed in the ownership of RGS equipped vehicle by O-D pair. A violation of this assumption would imply that probe vehicle trips no longer provide an unbiased sample of all trips.

9.4 Significance of the Thesis Contributions

9.4.1 Significance of the LSE and LRE models

The proposed LSE and LRE models maintain a sufficiently broad mathematical basis which permits generalized inferences to be drawn about these two models. The model formulations explicitly recognize that field data rarely exhibit node or path flow continuity. The formulations and solution algorithms provide results that are consistent with each other when a single solution exists, when multiple solutions exist, and when no solutions exist that exactly replicate the observed link flows.

The significance of the LSE and LRE models is their ability to perform the following:

- a) The models are able to estimate both static and dynamic O-D demands from link flows that do not need to exhibit node or path flow continuity.
- b) The model formulations explicitly incorporate the relative reliability of both observed link flows and individual seed O-D demand matrix cells.
- c) The models are capable of computing the relative reliability of an observed link flow on the basis of the magnitude of the flow discontinuity at the two nodes to which the link is joined.
- d) The models can determine if observed flows were likely to have resulted from trips departing their origins prior to the period of analysis.
- e) The QUEENSOD model, which incorporates the LSE and LRE model algorithms, is able to utilize the data file structure of the network traffic simulation model INTEGRATION. Furthermore, the demands estimated by the QUEENSOD model are output in a format that is compatible with the O-D demand data file structure required by the INTEGRATION model.

9.4.2 Significance of the evaluation of vehicle probe data

Technological advancements in the areas of communications and computer hardware have recently often outpaced the traffic engineering community's collective expertise with respect to operating these complex systems to their full potential. Much speculation has occurred over the potential of using RGS probe vehicle data to estimate dynamic O-D demands. Unfortunately, very little rigorous research has been conducted to support or disprove these speculative claims.

The statistically based development of techniques for using RGS probe vehicle data to estimate dynamic O-D demands is significant for the following specific reasons:

- a) This quantitative assessment of the expected reliability of dynamic O-D estimates made solely on the basis of RGS probe vehicle data has not been previously conducted.
- b) The expressions developed objectively quantify the expected reliability of O-D demand estimates that are made solely on the basis of RGS vehicle probe data.
- c) Despite the limitations, that result from the assumptions made in the development of these expressions, they can be used to estimate confidence limits about RGS probe vehicle based O-D estimates.
- d) Even though probe based O-D estimates are expected to be initially insufficiently accurate to be of significant practical benefit, they may be utilized as prior information in synthetic O-D estimation models that explicitly consider the reliability of individual O-D demands, such as the LSE and LRE models developed in this thesis.

9.5 Recommendations for Further Work

9.5.1 Further development of the LSE and LRE models

- a) There remains a significant need for either a practical method of simultaneously estimating O-D demands, routes, and route weights, or some means of determining appropriate routes and route weights, without the availability of prior knowledge of the O-D demands.
- b) It is necessary to assess the sensitivity of the accuracy of estimated demands, to the accuracy of routes, route weights, and link travel times. Knowledge of these sensitivities would indicate how accurate each of these data sources needs to be, and would permit a better selection of the relative priorities in data collection effort for each data item.
- c) Dynamic O-D demands for the Highway 401 network should be re-estimated when FTMS detector data are available for a greater portion of the network, such that greater confidence could be placed in the demands estimated for this portion of the network. The availability of data for a greater portion of the network would also result in more accurate measures of performance, as these measures could reflect conditions over a greater portion of the network.

- d) It is necessary to establish how sensitive dynamic traffic assignment solutions are to the accuracy of the dynamic O-D demands. The extent to which current dynamic traffic assignment methods are sensitive to imperfect real-time O-D data is especially important when such methods play a critical role in any IVHS architectures.
- e) Quantitative comparisons should be made between the most promising dynamic O-D estimation methods. These comparisons should be carried out for several hypothetical networks of varying complexity, as well as for an actual network using field data.

9.5.2 Further development of the use of probe data in estimating dynamic demands

- a) It is recommended that the approximations, that were made in deriving the analytical solutions, be investigated to establish either their associated error, or to determine methods by which the impacts of such approximations could be reduced or eliminated. The investigation of other networks with several loading scenarios needs to be carried out to determine if the observed level of market penetration trends exhibit systematic dependencies on factors which were not studied in this thesis.
- b) On a practical level, a comparison of the combined analytical and simulation results to actual data from field tests should be made. The relative benefits of demands, estimated on the basis of probe data, versus those made on the basis of synthetic estimators, should also be examined using field data. In particular, methods by which probe data could be utilized to supplement synthetic O-D analyses should be developed and explored.
- c) The use of probe-based O-D demands as prior information in a synthetic estimation approach should be illustrated. In particular, the relationship between the average network wide level of market penetration, the reliability of the associated probe-based O-D estimated, and the reliability of the population O-D estimated by the synthetic estimation model, should be examined.

REFERENCES

- Ang, A. H-S., and Tang, W. H. (1975) Probability Concepts in Engineering Planning and Design Volume 1 Basic Principles. John Wiley and Sons, Inc. pp. 196 200.
- Atkins, S.T. (1986) Transportation planning models What the papers say. *Traffic Engineering and Control* Vol. 27. pp. 460-467.
- Barbour and Fricken (1991) Estimating an O-D table using a method based on shortest augmenting paths. Presented at the Transportation Research Board's 70th Annual Meeting in Washington, D.C. January 13-17.
- Bell, Michael G.H. (1991a) The estimation of origin-destination matrices by constrained generalized least squares. *Transportation Research* Vol. 25B No. 1. pp. 13-22.
- Bell, Michael G.H. (1991b) The real time estimation of origin-destination flows in the presence of platoon dispersion. *Transportation Research* Vol. 25B No. 2/3. pp. 115-125.
- Bell, Michael G.H. (1985) Variances and covariances for origin-destination flows when estimated by log-linear models. *Transportation Research* Vol. 19B No. 6. pp. 497-507.
- Bell, Michael G.H. (1984a) Log-linear models for the estimation of origin-destination matrices from traffic counts: an approximation. *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*. VNU Science Press. pp. 451-469.
- Bell, Michael G.H. (1984b) The estimation of junction turning volumes. *Traffic Engineering and Control* Vol. 25 #5. pp. 279-281.
- Bell, Michael G.H. (1983a) The estimation of an origin-destination matrix from traffic counts. *Transportation Science* Vol. 17, No. 2. pp. 198-217.
- Bell, Michael G.H. (1983b) The estimation of origin-destination flows and their confidence intervals from measurements of link volumes: A computer program. *Traffic Engineering and Control* Vol. 24 #4.
- Brenninger-Gothe, M., Jornsten, K.O., and Lundgren, J.T. (1989) Estimation of origin-destination matrices from traffic counts using multi-objective programming formulations. *Transportation Research* Vol. 23B No. 4. pp. 257-269.
- Ben-Akiva, Moshe. (1987) Methods to combine different data sources and estimate origindestination matrices. *Transportation and Traffic Theory*. Edited by N.H. Gartner and N.H.M. Wilson. Elsevier Science Publishing Co., Inc. pp. 459-481.
- Box, G.E.P., Hunter, W.G., and Hunter, J.S. (1978) *Statistics for Experimenters*. John Wiley and Sons.

Boyce, D.E., Hicks, J. and Sen, A. (1991) *In-Vehicle Navigation Requirements for Monitoring Link Travel Times in a Dynamic Route Guidance System.* Presented at the 70th Annual Meeting, Transportation Research Board, Washington, D.C. January, 1991

- Brillouin, Leon. (1962) Science and Information Theory 2nd Edition. Academic Press Inc.
- Brillouin, Leon. (1964) Scientific Uncertainty and Information. Academic Press Inc.
- Bureau of Public Roads (1965) *Calibrating and testing a gravity model for any sized urban area.* U.S. Dept. of Commerce, Washington, D.C.
- Carey, M., Hedrickson, C., and Siddharthan, K. (1981) A method for direct estimation of origin / destination trip matrices. *Transportation Science* Vol. 15, No. 1. pp. 32-49.
- Cascetta, E., Inaudi, D., and Marguis, G. (1993) Dynamic Estimators of origin destination matrices using traffic counts. *Transportation Science* Vol. 27 No. 4. pp. 363-373.
- Cascetta, Ennio. (1984) Estimation of trip matrices from traffic counts and survey data: A generalized least squares estimator. *Transportation Research* Vol. 18B No. 4/5. pp. 289-299.
- Cascetta, Ennio and Nguyen, Sang. (1988) A unified framework for estimating or updating origin / destination matrices from traffic counts. *Transportation Research* Vol. 22B No. 6. pp. 437-455.
- Chan, Yupo; Regan, Edward J.; Pan, Wen-Min. (1986) Inferring an origin-destination matrix directly from network flow sampling. *Transportation Planning and Technology* Vol. 11. pp. 27-46.
- Chang, Gang-Len, and Wu, Jifeng. (1993) Recursive estimation of time-varying O-D flows from traffic counts for freeway corridors. Presented at the Transportation Research Board's 70th Annual Meeting, January 10-14, 1993.
- Chatterjee, Arun and Raja, Meenbakshi. (1989) Synthetic models for through trips in small urban areas. *Journal of Transportation Engineering*, Vol. 115, No. 5. pp. 537-555.
- Cremer, M. (1991) Origin-destination matrix: Dynamic estimation. *Concise Encyclopedia of Traffic and Transportation Systems*. Edited by Markos Papageorgiou. Pergamon Press. pp. 310-315.
- Cremer, M. and Keller, H. (1987) A new class of dynamic methods for the identification of origin-destination flows. *Transportation Research* Vol. 21B No. 2. pp. 117-132.
- Cremer, M. and Keller, H. (1984) A systems dynamics approach to the estimation of entry and exit O-D flows. *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*. VNU Science Press. pp. 431-450.
- Cremer, M. and Keller, H. (1981) Dynamic identification of flows from traffic counts at complex intersections. *Proceedings of the Eighth International Symposium on Transportation and Traffic Theory*. University of Toronto Press. pp. 121-142.
- Draper, Norman and Smith, Harry (1981) *Applied Regression Analysis 2nd Edition*. John Wiley and Sones, Inc.

Davis, Gary A. and Nihan, Nancy L. (1991) Stochastic process approach to the estimation of origin-destination parameters from a time series of traffic counts. *Transportation Research Record* 1328. pp. 36-42.

- Dowling, R.G. and May, Adolf D. (1985) Comparison of small-area O-D estimation techniques. *Transportation Research Record* 1045. pp. 9-15.
- Federal Highway Administration (FHWA) (1984) *TRANSYT-7F User's Manual*. University of Florida.
- Feller, W. (1966) An Introduction to Probability Theory and Its Applications Volume 1, 2nd Edition. Wiley Publications. pp. 50-51.
- Fisk, C.S. (1989) Trip matrix estimation from link traffic counts: The congested network case. *Transportation Research* 23B. pp. 331-356.
- Fisk, C.S. (1988) On combining maximum entropy trip matrix estimation with user optimal assignment. *Transportation Research* 22B. pp. 69-76.
- Fisk, C.S. (1984) Game theory and transportation systems modelling. *Transportation Research* 18B. pp. 301-313.
- Fisk, C.S. and Boyce, D.E. (1983) A note on trip matrix estimation from link traffic count data. *Transportation Research* Vol. 17B, No. 3. pp. 245-250.
- Gazis, D.C.- editor (1974) *Traffic Science*. John-Wiley and Sons.
- Geva, I., Hauer, E., and Landau, U. (1983) Maximum-likelihood and bayesian methods for the estimation of origin-destination flows. *Transportation Research Record* 944. pp. 101-105.
- Glen, G.M., and Hague, P. (1975) Cross cordon traffic flow matrices. *Traffic Engineering* and Control Vol. 16 #7/8.
- Gur, Y.J. (1983) Estimating trip tables from traffic counts: Comparative evaluation of available techniques. *Transportation Research Record* 944. pp. 113-117.
- Hall, Van Vliet and Willumsen (1980) Saturn: A simulation assignment model for the evaluation of traffic management schemes. *Traffic Engineering and Control*. Vol. 21 #4.
- Han, A.F., and Sullivan, E.C. (1983) Trip table synthesis for CBD networks: Evaluation of the LinkOD model. *Transportation Research Record* 944. pp. 106-112.
- Hamerslag, R. and Immers, B.H. (1988) Estimation of trip matrices: Shortcomings and possibilities for improvement. *Transportation Research Record* 1203. pp. 27-39.
- Hauer, E., and Shin, B.T. (1981a) O-D matrices from traffic counts. *Traffic Engineering* and Control. Vol. 22 #3.
- Hauer, E., Pagitsas, E., and Shin, B.T. (1981b) Estimation of turning flows from automatic counts. *Transportation Research Record* 795. pp. 1-7.
- Hauer, E. (1978) Correction of license plate surveys for spurious matches. *Transportation Research* 13A. pp. 71-78.

Hellinga, B. and Van Aerde, M. (1994a). *A Speed-Delay Study of Highway 401 in Toronto, Ontario*. Compendium of Technical Papers Volume 2 from the CITE '94 Conference held in Windsor, Ontario from June 5-8. pp. 405-420.

- Hellinga, B. and Van Aerde, M. (1994b). *A Simulation Study of the Highway 401 FTMS*. Final report submitted to the Ontario Ministry of Transportation.
- Hellinga, B., and Van Aerde, M. (1994c) An Overview of a Simulation Study of the Highway 401 Freeway Traffic Management System. Canadian Journal of Civil Engineering Vol (21) pp. 439 454.
- Hendrickson, C. and McNeil, S. (1984a) Estimation of origin-destination matrices with constrained regression. *Transportation Research Record* 976. pp. 25-32.
- Hendrickson, C. and McNeil, S. (1984b) Matrix entry estimation errors. *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*. VNU Science Press. pp. 413-430.
- Högberg, P. (1976) Estimation of parameters in models for traffic prediction: A non-linear regression approach. *Transportation Research* Vol. 10. pp. 263-265.
- Hornbeck, R. W. (1975) *Numerical Methods*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632. pp. 100-106.
- Hutchinson, B.G. (1974) Principles of Urban Transport System Planning. McGraw-Hill.
- Irving, J.M., Oakley, C.F., and Ramsey, J.B.H. (1986) The updating of an O-D matrix: a maximum likelihood approach. *Traffic Engineering and Control* Vol. 27, No. 9. pp. 442-446.
- Janson, Bruce N. (1993) Most likely origin destination link uses from equilibrium assignment. *Transportation Research B* Vol. 27B, No. 5. pp. 333-350.
- Janson, Bruce N., and Southworth, Frank. (1992) Estimating departure times from traffic counts using dynamic assignment. *Transportation Research* Vol. 26B, No. 1. pp. 3-16.
- Jeffreys, M. and Norman, M. (1977) On finding realsitic turning flows at road junctions. *Traffic Engineering and Control* Vol. 18. pp. 19-25.
- Jörnsten, K., and Wallace, S.W. (1993) Overcoming the (apparent) problem of inconsistency in origin destination matrix estimations. *Transportation Science* Vol. 27, No. 4. pp. 374-380.
- Kanafani, Adib. (1983) Transportation Demand Analysis. McGraw-Hill Book Company.
- Keller, Hartmut and Ploss, Gerhard. (1987) Real-time identification of O-D network flows from counts for urban traffic control. *Transportation and Traffic Theory*. Editors N.H. Gartner and N.H.M. Wilson. Elsevier Science Publishing Co., Inc. pp. 266-284.
- Kessaci, A., Farges, J.L., and Henry, J.J. (1989) On line estimation of turning movements and saturation flows in PRODYN. *Proceedings of Control, Computers, and Communications in Transportation*. Conference held in Paris, France on Spet. 16-21, 1989.

Korpal, P.R. March 1992. Incident management: the key to successful traffic management in Toronto. *Institute of Transportation Engineers Journal*. pp 58-61.

- Lam, William H.K., and Lo, H.P. (1991) Estimation of origin-destination matrix from traffic counts: A comparision of entropy maximizing and information minimzing models. *Transportation Planning and Technology* Vol. 16. pp. 85-104.
- Landau, U., Hauer, E., and Geva, I. (1982) Estimation of cross-cordon origin-destination flows from cordon studies. *Tranportation Research Record* 891. pp. 5-10.
- Lapin, L.L. (1983) *Probability and Statistics for Modern Engineering*. Brooks/Cole Publishing Company.
- LeBlanc, L.J. and Farhangian, K. (1982) Selection of a trip table which reproduces observed link flows. *Transportation Research* Vol. 16B, No. 2. pp. 83-88.
- Maher, M.J. (1987) Bias in the estimation of O-D flows from link counts. *Traffic Engineering and Control* Vol. 28, No. 12. pp. 624-627.
- Maher, M.J. (1983) Inferences on trip matrices from observations on link volumes: A bayesian statistical approach. *Transportation Research* Vol. 17B, No. 6. pp. 437-447.
- Martin, P.T., and Bell, M.C. (1992) Network programming to derive turning movements from link flows. *Transportation Research Record 1365*. pp. 147-154.
- McNeil, S. and Hendrickson, C. (1985a) A regression formulation of the matrix estimation problem. *Transportation Science* Vol. 19, No. 3. Operations Research Society of America. pp. 278-292.
- McNeil, S. and Hendrickson, C. (1985b) A note on alternative matrix entry estimation techniques. *Transportation Research* Vol. 19B, No. 6. pp. 509-519.
- Mekky, A. (1979) On estimating turning flows at road junctions. *Traffic Engineering and Control* Vol. 20. pp. 486-487.
- Mountain, L.J. and Westwall. (1983) The accuracy of estimation of turning flows from automatic counts. *Traffic Engineering and Control* Vol. 24, No. 1. pp. 3-7.
- Nguyen, Sang. (1984) Estimating origin-destination matrices from observed flows. *Transportation Planning Models*. Edited by M. Florian. Elsevier Science Publishers B.V. (North-Holland). pp. 363-380.
- Nihan, Nancy L. and Davis, Gary A. (1989) Application of prediction-error minimization and maximum likelihood to estimate intersection O-D matrices from traffic counts. *Transportation Science* Vol. 23, No. 2. Operations Research Society of America. pp. 77-90.
- Nihan, Nancy L. and Davis, Gary, A. (1987) Recursive estimation of origin-destination matrices from input/output counts. *Transportation Research* Vol. 21B, No. 2. pp. 149-163.
- Nihan, Nancy L. (1982) Procedure for estimating freeway trip tables. *Transportation Research Record* 895. pp. 1-5.

Norman, M., Hoffmann, N., and Harding, F. (1979) Non-iterative methods for generating a realistic turning flow matrix for a junction. *Traffic Engineering and Control* Vol. 20. pp. 587-589.

- Rakha, H., Van Aerde, M., Case, E.R., and Ugge, A. (1989) Evaluating the benefits and interactions of route guidance and traffic control strategies using simulation. VNIS'89 IEEE Conference Proceedings, pp. 296-303.
- Robertson, D.I. (1969) *TRANSYT: A Traffic Network Study Tool*, Road Research Laboratory Report No. 253, Crowthrone, U.K.
- Robillard, P. (1975) Estimating the O-D matrix from observed link volumes. *Transportation Research* Vol. 9. pp. 123-128.
- Sheffi, Yosef and Barnhart, Cynthia. (1987) XNET: Extended traffic assignment model. *Journal of Transportation Engineering*, Vol. 113, No. 4. pp. 450-463.
- Spiess, H. (1987) A maximum likelihood model for estimating origin-destination matrices. *Transportation Research* Vol. 21B, No. 5. pp. 395-412.
- Stokes, R.W. and Morris, D.E. (1986) Use and effectiveness of synthetic origin-destination data in a macroscopic freeway simulation model. *ITE Journal*, April 1986. pp. 43-47.
- Stokes, R.W. and Morris, D.E. (1984) Application of an algorithm for estimating freeway trip tables. *Transportation Research Record* 976. pp. 21-25.
- Turnquiest, Mark and Gur, Yehuda. (1979) Estimation of trip tables from observed link flows. *Transportation Research Record* 730. pp. 1-6.
- Van Aerde, M. (1994) *INTEGRATION: A model for simulating integrated traffic networks. User's guide for model version 1.5c.* Transportation Systems Research Group,
 Department of Civil Engineering, Queen's University, Kingston, Canada.
- Van Aerde, M., Hellinga, B., Lei, Y., and Rakha, H. (1993) *Vehicle Probes as Real-Time ATMS Sources of Dynamic O-D and Travel Time Data*. Presented at the Advanced Traffic Management Systems Conference, St. Petersburg, Florida. October 3-8.
- Van Aerde, M., MacKinnon, G., and Hellinga, B. (1991a) *The Generation of Synthetic O-D Demands from Real-Time Vehicle Probe Data: Potential and Limitations.* Proceedings of the VNIS-91 Conference, Dearborn, Michigan.
- Van Aerde, M., Rilett, L., and Krage, M. (1991b) *INTEGRATION Model Simulation of the Traffic Responsive Route Guidance Functions of the TRAVTEK Experiment*. Proceedings of the VNIS-91 Conference, Dearborn, Michigan.
- Van Aerde, M. and Yagar, S. (1988) *Dynamic Integrated Freeway/Traffic Signal Networks: A Routing-Based Modelling Approach*. Transportation Research A, Vol.(22A), Number 6, pp. 445-453.
- Van Aerde, M. 1985. Modeling of traffic flows, assignment and queuing in integrated freeway/traffic signal networks. Ph.D. Thesis, Department of Civil Engineering, University of Waterloo, Waterloo, Canada.

Van Vliet, D. and Willumsen, L.G. (1981) Validation of the ME2 model for estimating trip matrices from traffic counts. *Proceedings of the 8th International Symposium of Transportation and Traiffic Theory*. University of Toronto Press. pp. 641-655.

- Van Zuylen, H.J. and Branston, D.M. (1982) Consistent link flow estimation from counts. *Transportation Research* Vol. 16B, No. 6. pp. 473-476.
- Van Zuylen, H.J. (1981) Some improvements in the estimation of an O-D matrix from traffic counts. *Proceedings of the 8th International Symposium of Transportation and Traiffic Theory*. University of Toronto Press. pp. 656-671.
- Van Zuylen, H.J. and Willumsen, L.G. (1980) The most likely trip matrix estimated from traffic counts. *Transportation Research* Vol. 14B. pp. 281-293.
- Van Zuylen, H.J. (1979) The estimation of turning flows on a junction. *Traffic Engineering and Control*, November 1979. pp. 539-541.
- Watling, David P. and Maher, Micheal J. (1992) A statistical procedure for estimating a mean origin-destination matrix from a partial registration plate survey. *Transportation Research* Vol. 26B, No. 3. pp. 171-193.
- Watling, D.P. (1990) Recent developments in techniqes in analysing partial registration plate data. *Precedings of the Planning and Transport Research and Computation Conference*. 18th Summer Annual Meeting in Transportation Planning Methods held at the University of Sussex, Brighton in Sept. 1990. PTRC Education and Research Services Ltd.
- Williams, Garth (1984) Linear Algebra with Applications. Allyn and Bacon, Inc.
- Willumsen, L.G. (1992) Origin-destination matrix: Static estimation. *Concise Encyclopedia of Traffic and Transportation Systems*. Edited by Markos Papageorgiou. Pergamon Press. pp. 315-322.
- Willumsen, L.G. (1984) Estimating time-dependent trip matrices from traffic counts. *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory.* VNU Science Press. pp. 397-411.
- Wilson, A.G. (1970) Entropy in Urban and Regional Modelling. Pion Limited, London, England.
- Wilson, A.G. and Kirkby, M.J. (1975) *Mathematics for Geographers and Planners*. Clarendon Press, Oxford, England pp. 242-243.
- Yagar, Sam. (1988) Generating partial origin-destination tables for streamlined application of corridor models. *Transportation Research Record* 1194. pp. 135-138.
- Yang, H., Sasaki, T., and Iida, Y. (1992) Estimation of origin-destination matrices from link traffic counts on congested networks. *Transportation Research* Vol. 26B, No. 6. pp. 417-434.
- Yang, H., Iida, Y, and Sasaki, T. (1991) An analysis of the reliability of an origindestination trip matrix estimated from traffic counts. *Transportation Research* Vol. 25B, No. 5. pp. 351-363.