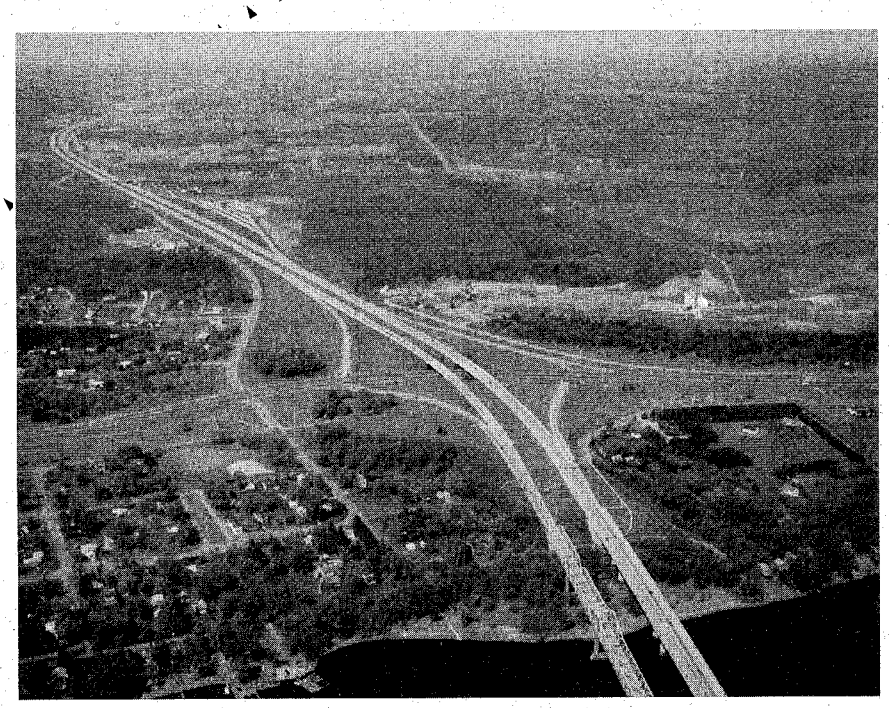
MN DEPT OF TRANSPORTATION
3 0314 00023 5449

Report Number 94-18

MINNESOTA DEPARTMENT OF TRANSPORTATION

research



Dynamic Estimation of Origin-Destination Patterns in Freeways

CTS
HE
336
.T7
D38
1994

UNIVERSITY OF MINNESOTA
CENTER FOR
TRANSPORTATION
STUDIES

Report Documentation Page

1. Report No. MN/RC - 94/18	2.	3. Recipient's Accession No.
4. Title and Subtitle Dynamic Estimation of Origin-Destination Patterns in Freeways		5. Report Date May 1994 6.
7. Author(s) Gary A. Davis, Ph.D.		8. Performing Organization Report No.
9. Performing Organization Name and Address Civil & Mineral Engineering Department University of Minnesota 500 Pillsbury Dr. SE Minneapolis, Mn 55455		10. Project/Task/Work Unit No. 11. Contract(C) or Grant(G) No. (C) Mn/DOT 68092 TOC 63
12. Sponsoring Organization Name and Address Minnesota Department of Transportation Office of Research Administration 200 Ford Building-Mail Stop 330 117 University Avenue St. Paul, Mn. 55155		13. Type of Report and Period Covered Final Report 1991-1992 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract (Limit: 200 words)

Any proposed traffic management action is essentially a forecast that the action will result in certain traffic conditions, but uncertainty concerning the amount and distribution of traffic demand will introduce random error between what is expected and what actually occurs. This report treats the problem of forecasting whether or not a given set of freeway on-ramp volumes are likely to cause over-capacity demand at some point in the freeway mainline. The main source of uncertainty in these forecasts concerns the freeway's origin-destination matrix, and four different methods for estimating this matrix from loop detector data are evaluated using Monte Carlo simulation. Only the method which explicitly modeled freeway traffic flow produced reasonably unbiased and efficient estimates, and it was concluded that successful estimation must be coupled with a good model of freeway traffic flow.

17. Document Analysis/Descriptors Traffic Flow Freeways Origin-Destination Matrix Markov Processes Ramp Metering	18. Availability Statement No restrictions. This document is available through the National Technical Information Services, Springfield, Va. 22161

19. Security Class (this report) Unclassified	20. Security Class (this page) Unclassified	21. No. of Pages 46	22. Price

DYNAMIC ESTIMATION OF ORIGIN-DESTINATION PATTERNS IN FREEWAYS

FINAL REPORT

Prepared by

Gary A. Davis, Ph.D.

Department of Civil and Mineral Engineering
University of Minnesota

May, 1994

Submitted to

MINNESOTA DEPARTMENT OF TRANSPORTATION
OFFICE OF RESEARCH ADMINISTRATION
117 UNIVERSITY AVENUE 2nd FLOOR
ST. PAUL, MINNESOTA 55155

Disclaimer: This report represents the results of research conducted by the author and does not necessarily reflect the view or policy of the Minnesota Department of Transportation. This report does not contain a standard or specified technique.

ACKNOWLEDGEMENTS

The author would like to thank Mr. Ping Yu and Ms. Yihong Gao, graduate students at the University of Minnesota, for programming and conducting some of the tests reported here. He would also like to thank Dr. Eil Kwon, of the University of Minnesota, and Mr. Jim Aswegan, of the Minnesota Department of Transportation, for help in obtaining freeway data. This research was supported by the Minnesota Department of Transportation, but all facts, conclusions and opinions expressed here are strictly the responsibility of the author.

TABLE OF CONTENTS

	Page
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: UNCERTAINTY EFFECTS IN FREEWAY FORECASTING AND MANAGEMENT	9
CHAPTER 3: ESTIMATION OF FREEWAY OD PATTERNS	17
CHAPTER 4: A MARKOVIAN FREEWAY TRAFFIC MODEL	25
CHAPTER 5: MONTE CARLO EVALUATION OF PARAMETER ESTIMATORS	29
CHAPTER 6: CONCLUSIONS	35
REFERENCES	39
APPENDIX: LIST OF SYMBOLS	A1

SUMMARY

Any proposed traffic management action is essentially a forecast that the action will cause certain traffic conditions to result, and uncertainty concerning traffic demand and/or the values of model parameters will introduce random error between what is expected and what actually occurs. This report considers, in some detail, the problem of forecasting whether or not a given set of freeway on-ramp volumes are likely to cause over-capacity demand at some point on the freeway mainline. A simple model for the distribution of freeway traffic is presented, which leads to a straightforward interpretation of the impact of forecast uncertainty as a reduction in practical mainline capacity. The problem of efficient estimation of the proportions of vehicles entering at an on-ramp which are destined for each downstream off-ramp is then taken up. Such proportions provide the basic representation of the distribution of travel demand in freeway models, as well as being interesting in their own right. For instance, if the proportion of traffic entering I-35W south of the Minnesota River and headed for downtown Minneapolis is small, then transit options which terminate in the downtown are not likely to attract a large number of riders from this area. By embedding the demand model in a Markov process model of freeway traffic flow, it was possible to compare four different estimation approaches via Monte Carlo simulation methods. The four estimators were:

(1) an ordinary least-squares (OLS) approach similar to that used to estimate turning movement proportions at intersections,

(2) a version of the expectation-maximization (EM) algorithm for computing maximum-likelihood estimates,

(3) the SYNODM algorithm originally used in the FREQ simulation models,

(4) a nonlinear least-squares (NLS) approach which explicitly modelled traffic flow conditions.

Of these four, only the NLS approach appeared relatively free of bias and/or statistical inefficiency, indicating that successful estimation of these splitting proportions must be coupled with a good model of freeway traffic flow. The Markov process model developed for this project integrates travel demand and traffic flow in a natural way, and provides a general framework for developing the unified travel demand/traffic flow models needed for the Advanced Traffic Management Systems component of IVHS.

INTRODUCTION

This project investigated methods for estimating freeway origin-destination proportions from the traffic count data typically available from freeway surveillance and control systems. It was funded jointly by the Center for Transportation Studies at the University of Minnesota and the Minnesota Department of Transportation. Although originally aimed at developing recursive estimators of these proportions, it was found that the variants of ordinary least-squares which work reasonably for single intersections performed poorly on freeway data. A stochastic model of freeway traffic demand and flow was developed and used to conduct simulation studies aimed at identifying potentially useful estimation approaches. The conclusion was that neglect of the variations in freeway traffic flow was a major cause of the failure of the simple least-squares approaches, and that successful estimation requires a good model of freeway traffic flow.

The report is divided into six chapters. Chapter 1 places this work within the Advanced Traffic Management component of Intelligent Vehicle-Highway Systems. Chapter 2 reviews the relationship between uncertainty in origin-destination proportion estimates and the quality of freeway traffic control, while Chapter 3 presents the four different estimation approaches investigated in this project. Chapter 4 describes the stochastic traffic simulation model developed for this project, and Chapter 5 describes the results of a Monte Carlo simulation study aimed at evaluating the performance of the four candidate estimators. Chapter 6 then presents a summary of the results and the report's conclusions.

CHAPTER 1

INTRODUCTION

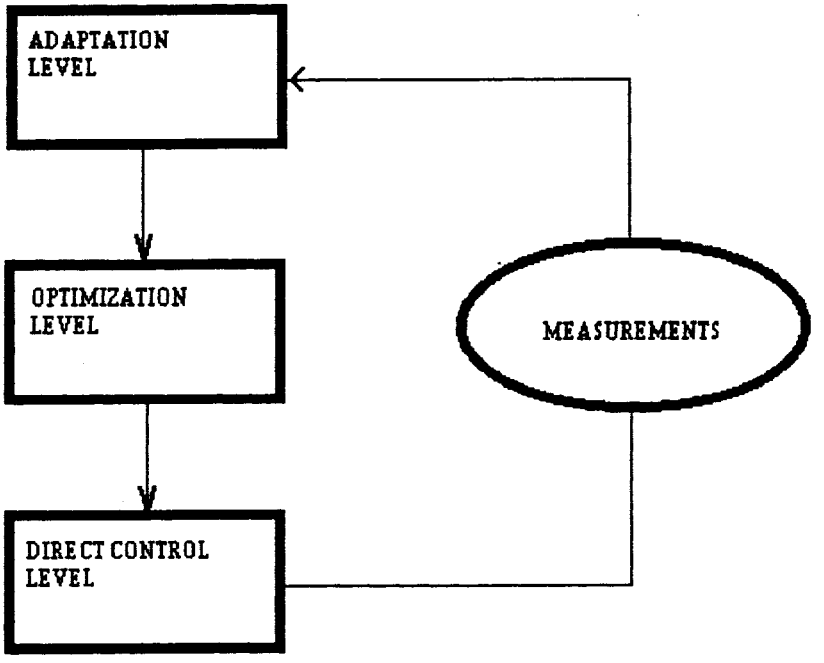
One of the major transportation challenges of the next decade will be alleviation of the growing congestion experienced by many, if not most, of the nation's urban roadways. The possibility of providing major increases in the physical capacity of roadways is sharply constrained by both social and economic conditions, leading to a search for more innovative solutions. In the Advanced Traffic Management Systems (ATMS) component of the Intelligent Vehicle and Highway Systems (IVHS) effort, the objective is to combine an improved ability to measure traffic conditions with an improved understanding of traffic behavior and a broader range of management actions to provide congestion relief without requiring capacity increases. ATMS can be productively viewed as the application of advanced systems control methods to the real-time management of traffic networks, and this leads immediately to issues concerning the optimality of any proposed traffic management scheme. If a scheme is suboptimal, there is always the possibility that by shifting to a more nearly optimal one, congestion relief can be achieved within the limits of the existing infrastructure. Optimal schemes, on the other hand, define the boundary at which further improvements require infrastructure change.

Current traffic-responsive management techniques, such as those deployed for freeway ramp metering and the SCAT and SCOOT intersection control models, can be characterized as heuristic, myopic and demand-insensitive, since they employ frequent updates of management actions, based primarily on local measurements, with no emphasis on influencing route choice or other expressions of travel demand. Such approaches have demonstrated an ability to respond

effectively to short-term fluctuations in traffic conditions, but potential exists for improving them by integrating with actions taken over longer space-time horizons and by incorporating vehicle routing considerations. Figure 1 displays a three-level hierarchical traffic management model which has been proposed as a framework for ATMS. The lowest, or "Direct" level corresponds to the myopic methods described previously. Here actions (such as resetting traffic signals) are updated frequently (e.g. ever 20 seconds) with the update decision being based primarily on local, real-time measurements of traffic conditions. The middle, or "Optimization" level consists of a parameterized model relating the joint impact of traffic demand and traffic management actions to traffic conditions, together with an algorithm for finding that set of actions which, given the current demand, optimizes some measure of system performance. The solution is recomputed at somewhat longer time intervals (e.g. every 5-15 minutes), and then passed to the Direct level to provide nominal objectives which the Direct level seeks. The highest, or "Adaptation" level in turn provides inputs for the Optimization level. These consist of the traffic demand forecasts and/or model parameter estimates required by the Optimization level's model, and it is at this highest level that issues concerning traveller route selection and demand management are considered. The expectation of ATMS then is that by integrating longer-range, system-wide considerations with the more direct, myopic actions, noticeable gains in congestion relief can be achieved.

Clearly, the practical benefits of the three-level model will depend critically on the quality of the nominal solutions passed by the Optimization level to the Direct level, and this in turn will depend on the realism embodied in the Optimization level's traffic demand/flow model. However, almost all existing traffic models can be characterized as deterministic/certainty

FIGURE 1: A Three-Level Hierarchical Model for Advanced Traffic Management



equivalent, that is, that the models treat traffic variables as deterministic quantities, and that any model parameters are treated as being known with certainty. This approach has the advantage that the more computationally tractable methods of deterministic optimization can then be used to find nominal solutions, but it lacks realism in two ways. First, traffic demand is ultimately determined by the particular decisions of individual drivers and, in the absence of detailed information on each individual, these decisions will not be perfectly predictable. Second, the values used for traffic model parameters will be determined by statistical estimation, and hence known only with degrees of uncertainty characteristic of the estimation methods. Since any traffic management action is, in essence, a forecast that certain outcomes should result as consequences of that actions, the presence of demand and/or parameter uncertainty will introduce random error between what is expected and what actually occurs. If the variance of this error is large enough, this in turn could lead to a perception that the traffic management action is ineffective, even though that action might be optimal on the average. In such cases either the uncertainty must be explicitly considered in the optimization model, or steps must be taken to reduce it. The former approach would require replacing the deterministic optimization usually considered for the Optimization level with stochastic optimization methods, while the second approach requires either more detailed information concerning individual driver actions, or the use of more efficient parameter estimators.

This study will not address these issues in their most general level, but rather will consider, in some detail, the simpler problem of predicting whether or not a given set of freeway on-ramp volumes is likely to cause the demand on any freeway mainline section to exceed that section's capacity. Freeway surveillance and control systems are probably the most widespread

existing examples of ATMS, and the justification for these systems stems from the fact that when traffic demand exceeds capacity, queuing of vehicles occurs, and the traffic upstream from the bottleneck suffers slow and erratic speeds, leading to increased delay, increased carbon monoxide production and increased fuel consumption. By metering the flow of vehicles from nearby on-ramps, the surveillance system essentially shifts the queue off of the freeway mainline and onto the on-ramps where, by involving a smaller number of vehicles for shorter periods of time, the total negative effect of the queuing is lessened. Avoiding over-capacity demands is arguably the most common nominal objective of existing freeway management systems, and is also explicitly considered in optimization routines such as FREQ and FREFLO. This problem of explicitly relating the distribution of travel demand to traffic conditions takes on even more importance when one seeks to describe driver re-routing in response to traffic information.

This line of argument suggests that there is an intimate relationship between accuracy of the parameter estimates used to drive a management model and the resulting traffic conditions. It is useful then to discuss some the different ways that statisticians use to assess the quality of an estimator, and how these bear on the problem of freeway control. An estimator is said to be unbiased if, on the average, the estimator equals the true value of the parameter being estimated. Thus an unbiased estimator is not guaranteed to produce the true value for any particular data set, but if estimates are computed for a large number of different data sets, the average of these estimates should equal the true value. A somewhat different, but related concept is consistency. An estimator is said to be consistent if, as the amount of data in a given data set becomes arbitrarily large, the estimate converges to the true value. Thus, if an estimator is consistent, we can get as close as we like to the true value of the parameter by simply collecting more data.

Among the class of consistent estimators is a subclass of asymptotically Normal estimators, which have the property that, for large data sets, the distribution of differences between the estimates around the true value can be approximated by a Normal probability distribution. Asymptotically Normal estimators are also said to be asymptotically unbiased, in that if any bias is present, it tends to zero as the amount of data increases. Finally, under reasonably general conditions it is often possible to compute a lower bound for the variance of an asymptotically Normal estimator. An estimator whose variance actually achieves this lower bound is said to be efficient, and when comparing two candidate estimators, the one with the lower variance is said to be more efficient than the one with the larger variance. This can be interpreted as meaning that, for any given data set, the estimate computed using a consistent, efficient estimator is more likely to be closer to the true parameter value than an estimate computed using an inconsistent, biased, or inefficient estimator. Under reasonably general conditions, it can be shown that estimates computed using the principle of maximum likelihood (ML) tend to be consistent, asymptotically Normal and asymptotically efficient. For the problem of traffic management, these properties can be interpreted as meaning that, by using a consistent estimator and enough data, we can essentially eliminate uncertainty due to parameter estimation error and produce a certainty equivalent control problem. If restricted to a finite data set, by using an asymptotically Normal estimator we can quantify the effect of parameter uncertainty, and by using an efficient estimator we can minimize the effect of this uncertainty.

When originally proposed, this project intended to develop ordinary least-squares (OLS) estimators of freeway origin-destination (OD) patterns and then, using well-known formulas to quantify the uncertainty characterizing these estimates, to produce a chance-constrained version

of the linear program commonly used to find optimal time-of-day ramp metering rates. These chance-constrained metering rates were then to be compared to FREQ-generated rates using the FREQ10 freeway simulator. As it turned out however, OLS tended to produce very poor estimates of freeway OD parameters, a surprising contrast to the performance of OLS on similar problems involving single intersections. Thus the bulk of the project was spent attempting to identify the causes of OLS's failure, and then finding a remedy for these. This led to the conclusion that the linear traffic model implicit in the OLS approach was the cause of this poor performance, and that estimation of freeway OD patterns cannot be usefully separated from the problem of fitting freeway traffic flow models. Chapter 2 presents a derivation of the chance constraints needed to guarantee that a given set of metering rates does not produce over-capacity demand, and provides a useful interpretation of the effect of demand and parameter uncertainty as reductions in the practical capacity which can be assumed for a freeway section. Chapter 3 treats the problem of estimating freeway OD patterns in some detail, and presents the four estimation approaches tested in this project. Chapter 4 describes the stochastic, macroscopic freeway simulation model developed during this project, and Chapter 5 describes the Monte Carlo simulation study conducted to evaluate the OD estimation approaches. Finally, Chapter 6 presents the project's conclusions.

CHAPTER 2

UNCERTAINTY EFFECTS IN FREEWAY FORECASTING AND MANAGEMENT

In this Chapter, we will derive some basic relationships concerning the effect of uncertainty on the quality of freeway management. The ideas will first be illustrated with a simple, hypothetical example consisting of a stretch of one-way road forking into a Y-intersection. Let q denote the total traffic volume entering this fork, let b denote the proportion of this traffic taking the right-hand branch, and let c denote the capacity of the right-hand branch. Assume that the traffic manager has no effect on b, but can control q, and seeks to maximize q, subject to the traffic on the right-hand branch not exceeding the capacity c. Let x denote the traffic volume taking the right hand branch, and if the distribution of traffic between the two branches is deterministic, so that $x=bq$, the solution to this problem is simply $q=c/b$. Now suppose that the actual number of vehicles taking the right hand branch is random, which can be modelled by treating x as a binomial random variable with parameters q and b. The expected value of x is still bq, but by exploiting the approximate symmetry of the binomial distribution for large values of q, it follows that the solution $q=c/b$ produces a situation where the probability that x exceeds c is approximately equal to 0.5. That is, roughly half of the time the demand x will exceed the capacity c, even though on the average demand equals capacity. For any value of q greater than c there will be a nonzero probability that x also exceeds c, so the best that can be done under random demand is to require that the proposed value of q satisfy a reliability constraint of the form

$$\text{Prob}[x > c] \leq \alpha. \tag{1}$$

That is, the solution should not produce capacity exceedences more than 100α % of the time. Generally, reliability constraints require working with the full probability distribution of x but, if q is large enough, x can be approximated by a Normal random variable, and the reliability constraint then reduces to the algebraic inequality

$$bq \leq c - z_\alpha[qb(1-b)]^{1/2} \tag{2}$$

where z_α is the standard Normal deviate corresponding to a right tail probability of α, and qb(1-b) is the variance of the random demand x. From a practical standpoint, equation (2) shows that the effect of randomly varying demand is to reduce the target capacity by a quantity equal to the right-hand side of (2), which depends both on the desired reliability level and on the variability of x.

So far, it has been assumed that the splitting probability b is known with certainty, but in practice this parameter must be estimated from some data set, and so will only be known with a degree of uncertainty characteristic of the estimation procedure. To model this effect, assume that b has been estimated by an unbiased estimator $\hat{b}$, that $\hat{b}$ is, at least approximately, Normally distributed with a mean of b and a variance of σ^2, and that $\hat{b}$ is uncorrelated with x. This would be the case, for instance, with maximum likelihood or method of moments estimators computed from reasonably large sets of historical data, and where the underlying data generating process satisfies certain plausible mixing conditions (Ljung, 1978; Ljung and Caines, 1979). Then $\hat{x}=\hat{b}q$ is an unbiased forecast of x with forecast error variance $qb(1-b)+q^2\sigma^2$, and the condition

$$\text{Prob}[x > c] = \text{Prob}[x-\hat{x} > c-\hat{x}] \leq \alpha \tag{3}$$

can be approximated by the constraint

$$\hat{b}q \leq c- z_\alpha[q\hat{b}(1-\hat{b})+q^2\sigma^2]^{1/2}. \tag{4}$$

Thus uncertainty in the actual value of the parameter b, coupled with a reliability constraint, produces a further reduction of the right-hand branch's target capacity. Ignoring this parameter uncertainty would produce a tendency to select values for q which violate the capacity constraints more often than expected, possibly leading to a sense that the control is not adequate. On the other hand, using an inefficient estimator of b would produce large values for the estimator variance σ^2, and hence lead to selection of unnecessarily conservative values for q. This suggest that, at least in principle, improved traffic control could result from an explicit recognition of traffic demand uncertainty, coupled with the use of efficient statistical estimation procedures.

Now to extend these ideas to freeway control problems, first consider a linear model for distributing the vehicles departing from on-ramps to the off-ramps, which is appropriate for traffic networks where only one route connects each origin-destination pair and where the travel times between origin and destination can be ignored. Deterministic versions of the model have been employed in deriving freeway ramp-control strategies (e.g. Chen et al. 1974; Papageorgiou, 1980) and in developing intersection turning proportion estimators (e.g. Cremer and Keller, 1983), while the stochastic version employed here has been presented and discussed in Nihan and Davis (1989). To specify the model, first assume that the control period of interest has been divided into S time intervals (say, of 5 minutes each), and let $t=1,..,S$ index these intervals. Assume further that the segment of freeway to be controlled has a total of m on-ramps, indexed by $i=1,..,m$ and n off-ramps, indexed by $j=1,..,n$. By convention, the upstream boundary of the segment will be the first "on-ramp" ($i=1$) while the downstream boundary will be the last "off-ramp" ($j=n$). Finally, assume that the segment of freeway to be controlled has been divided into p sections such that on-ramps enter the freeway only at the upstream boundaries of sections

while off-ramps diverge only at the downstream boundaries of sections. Figure 2 depicts a length of freeway divided into 10 sections. Let $k=1,..,p$ index the sections, and define the following variables

$q_i(t)$ = traffic departing from on-ramp i during time interval t,

$\mathbf{q}(t)$ = a m-dimensional vector whose elements are the $q_i(t)$,

$y_j(t)$ = traffic arriving at off-ramp j during time interval t,

$x_k(t)$ = traffic demand for freeway section k during time interval t,

c_k = capacity of section k,

$x_{ij}(t)$ = traffic arriving at on-ramp i and destined for off-ramp j, during interval t,

b_{ij} = probability a vehicle which departs from on-ramp i is destined for off-ramp j,

$\mathbf{B}$ = an mxn dimensional matrix whose elements are the b_{ij}.

In order to be legitimate distribution probabilities, the b_{ij} must satisfy

$$0 \le b_{ij} \le 1.0, \qquad i=1,..,m,\ j=1,..,n; \tag{5a}$$

$$\Sigma_j\, b_{ij} = 1.0, \qquad i=1,..,m. \tag{5b}$$

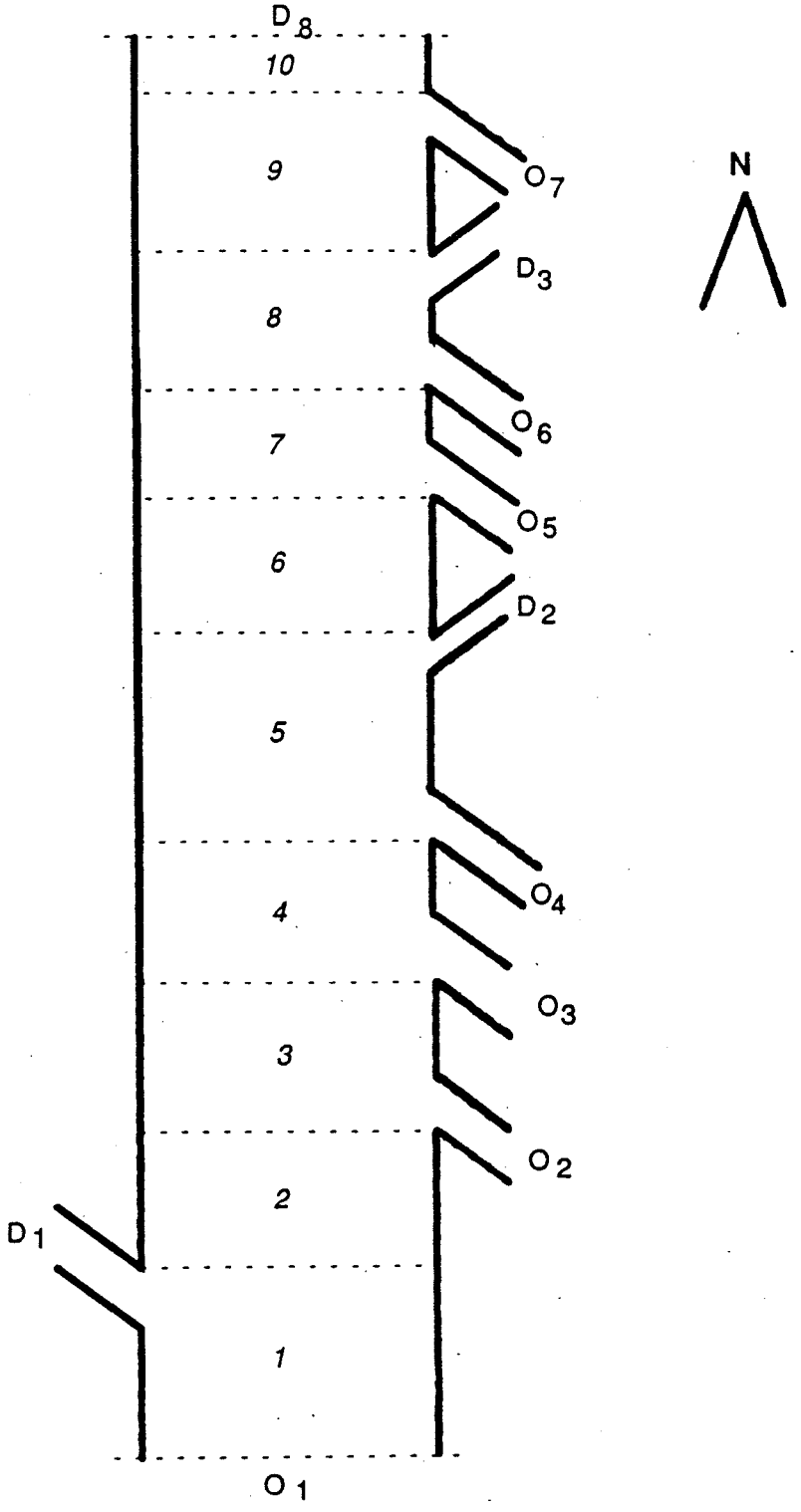
Individual vehicles are assumed to select their destinations independently of one another, so that for each on-ramp i, the $x_{ij}(t)$ are generated as the outcome of independent multinomial random vectors, with number of "trials" equal to $q_i(t)$ and choice probability distribution $(b_{ij}, j=1,..,n)$. The off-ramp volumes are then given by

$$y_j(t) = \Sigma_i x_{ij}(t) \tag{6}$$

and the freeway section demands are given by

$$x_k(t) = \Sigma_i\, \Sigma_j\, \delta_{ijk}\, x_{ij}(t), \tag{7}$$

FIGURE 2: Schematic of a 10-Section Segment of Freeway.



where

δ_{ijk} = 1, is section k lies between on-ramp i and off-ramp j,

0, otherwise.

It is then straightforward to show that the expected values and variances of the $y_j(t)$ and $x_k(t)$ are given by

$$E[y_j(t) \mid \mathbf{q}(t), \mathbf{B}] = \Sigma_i\, b_{ij}\, q_i(t) \tag{8a}$$

$$Var[y_j(t) \mid \mathbf{q}(t), \mathbf{B}] = \Sigma_i\, b_{ij}(1-b_{ij})q_i(t) \tag{8b}$$

$$E[x_k(t) \mid \mathbf{q}(t), \mathbf{B}] = \Sigma_i\, \beta_{ik}\, q_i(t) \tag{8c}$$

$$Var[x_k(t) \mid \mathbf{q}(t), \mathbf{B}] = \Sigma_i\, \beta_{ik}(1-\beta_{ik})q_i(t), \tag{8d}$$

where $\beta_{ik}=\Sigma_j \delta_{ijk} b_{ij}$ gives the probability that a vehicle departing from on-ramp i uses section k. Appealing to the Central Limit Theorem, the quantities $y_j(t)$ and $x_k(t)$ can be treated as approximately Normal random variables, with means and variances given in (8a-8d). As with the simple example presented earlier, if the on-ramp volumes are selected so that $\Sigma_i\, \beta_{ik} q_i(t)=c_k$, the symmetry of the Normal distribution implies that

$$Prob[x_k(t) > c_k] = 0.5. \tag{9}$$

so that the actual demand for section k will exceed its capacity about 50% of the time, even though on the average the demand equals capacity. Reliability constraints now take the form

$$\Sigma_i\, \beta_{ik}\, q_i(t) \le c_k - z_\alpha\, [\Sigma_i\, \beta_{ik}(1-\beta_{ik})q_i(t)]^{1/2}, \tag{10}$$

and again the practical effect of the random distribution of vehicles to off-ramps is to reduce the capacity which metering algorithms can assume for the freeway sections. The magnitude of these capacity reductions increases both as the desired probability of avoiding capacity exceedence

increases and as the variability of the section demands increases.

To generalize the problem of uncertainty concerning the values of the parameters b_{ij} (and hence β_{ik}), let $\hat{\beta}_{ik}$ denote an estimator of β_{ik}, and let $\mathbf{Q}_k$ denote the covariance matrix of the estimates $\hat{\boldsymbol{\beta}}_k=(\hat{\beta}_{1k},..,\hat{\beta}_{mk})^T$. It will be assumed that $\hat{\boldsymbol{\beta}}_k$ is a Normal random vector with expected value equals to the true $\boldsymbol{\beta}_k$ and covariance matrix $\mathbf{Q}_k$, and that $\hat{\boldsymbol{\beta}}_k$ is uncorrelated with the current x_k. These assumptions can be relaxed, but the resulting technicalities tend to obscure the drift of the argument. Let $\hat{x}_k(t)=\Sigma_i\ \hat{\beta}_{ik}\ q_i(t)$ be a forecast of the traffic demand for section k given a set of on-ramp volumes, and it is straightforward to verify that the forecast is unbiased

$$E[\hat{x}_k(t)]=\Sigma_i\ E[\hat{\beta}_{ik}]\ q_i(t)\ =\ E[x_k(t)] \tag{11}$$

and that the variance of the forecast error is given by

$$\sigma_k^{\ 2}\ =\ E[x_k(t)-\hat{x}_k(t)]^2\ =\ \Sigma_i\beta_{ik}(1-\beta_{ik})q_i(t)\ +\ \mathbf{q}^T(t)\mathbf{Q}_k\mathbf{q}(t) \tag{12}$$

The right-hand side of (12) contains two terms, the first being the forecast variance due to the random distribution of vehicles to off-ramps (demand uncertainty) while the second gives the variance due to uncertainty in the parameter estimates (parameter uncertainty). Note that if the parameters are known perfectly (so that the elements of $\mathbf{Q}_k$ are zero) the forecast variance reduces to the variance in $x_k(t)$, given in (8d). Appealing to the approximate Normality of the forecast error $x_k(t)-\hat{x}_k(t)$, and replacing the β_{ik} in (3) with their estimates, the reliability constraint for guaranteeing that the probability of capacity exceedence is no greater than α now becomes

$$\Sigma_i\ \hat{\beta}_{ik}\ q_i(t)\ \le\ c_k\ -\ z_\alpha\ [\Sigma_i\ \hat{\beta}_{ik}(1-\hat{\beta}_{ik})q_i(t)+\mathbf{q}^T(t)\mathbf{Q}_k\mathbf{q}(t)]^{1/2} \tag{13}$$

and it can be seen that both demand and parameter uncertainties cause reductions in the section's target capacity.

To summarize, if one of the goals of a freeway ramp-metering system is to prevent the

traffic demand on a given freeway section from exceeding that sections's capacity, and if the ability to predict a section's demand is subject to uncertainty, then in order to guarantee that a given set of metering rates does not cause capacity exceedences more than a specified percentage of the time, it is necessary to treat each section as having a target capacity that is lower than its nominal value. The actual capacity reductions depend in a nonlinear manner on the on-ramp volumes and also on the estimated on-ramp to off-ramp OD pattern, or splitting probabilities, but roughly speaking, the greater the uncertainty in the demand forecasts, the greater the capacity reductions needed to achieve a given level of reliability. At least for unbiased parameter estimators, the magnitudes of these capacity reductions also depend on the statistical efficiency of the procedure used to estimate the splitting probabilities. To put this another way, by shifting from a less efficient estimation procedure to a more efficient one, and hence reducing the quantity $\mathbf{q}^T(t)\mathbf{Q}_k\mathbf{q}(t)$ appearing in (13), one might "gain" some capacity. Where feasible, such gains are likely to be very cost effective when compared to physical expansion of the roadway.

CHAPTER 3

ESTIMATION OF FREEWAY OD PATTERNS

This section reviews the various methods which have been proposed for estimating freeway OD matrices. Attention is first given to "synthetic OD" methods, which can be viewed as special cases of maximum-entropy fitting of trip tables, and then second, to methods based on linear least-squares estimation. The third approach treated is maximum-likelihood (ML) estimation based on the EM algorithm, and the fourth approach is nonlinear least-squares based on macroscopic traffic flow models.

As in all statistical estimation exercises, the quality of OD estimates will depend on the type and quality of the data available. Ideally, one would take a representative sample of the drivers at each on-ramp and then identify the off-ramp to which he or she is headed, so that good estimates of the splitting probabilities could be obtained as simple proportions. In practice though, collecting such "complete" data requires time-consuming and error-prone activities such as origin-destination surveys or license plate studies, and since the splitting probabilities can change according to time-of-day and season, such methods are not useful if timely, up to date estimates are required. On the other hand, traffic volume counts at on-ramps, off-ramps and mainline locations are relatively easy to obtain, leading to the possibility of obtaining splitting parameter estimates from traffic count data. From a mathematical standpoint, all OD estimators can be thought of as solutions to optimization problems, where a quantitative criterion is first presented which varies as a function of the values of the OD parameters. Those parameter values

which optimize (either maximize or minimize, depending on the context) this criterion are identified as "good" estimates, and the optimizing values are characterized as the solutions of a set of equations and/or inequality constraints. Thus the problem of actually computing these estimates is reduced to the computationally more tractable problem of solving a system of equations.

If one has available on ramp counts $q_i(t)$, and off-ramp counts $y_j(t)$, then clearly the $x_{ij}(t)$ must satisfy the conservation equations

$$\Sigma_j \, x_{ij}(t) = q_i(t) \tag{15a}$$

$$\Sigma_i \, x_{ij}(t) = y_j(t), \tag{15b}$$

or, treating the b_{ij} as constants,

$$\Sigma_i \, b_{ij} = 1.0 \tag{16a}$$

$$\Sigma_j \, b_{ij} \, q_i(t) = y_j(t). \tag{16b}$$

Thus the $x_{ij}(t)$, or the b_{ij}, could in principle be determined from traffic counts as the solutions to a system of linear equations. The solvability of the above systems however depend on the relation between the number of equations available and the number of unknown parameters b_{ij}. From a practical standpoint, the cases of interest are where there are either too few or too many equations to permit a unique solution for the b_{ij}, since the case of exactly the right amount can be treated as a special case of either of these others. The first case usually occurs when only one set of traffic counts is available (data is available for only one time interval), so that the system of equations is underdetermined, and an infinite number of solutions exists. In this case, the problem is usually posed in terms of solving the first set of equations (15a-15b). The second case usually occurs when repeated counts have been made, allowing a set of equations to be set up

for each time-period for which counts are available. In this case, the problem is usually posed in terms of solving the second set of equations (16a-16b), with the b_{ij} being treated as unknown constants.

The problem of estimating OD parameters from an underdetermining set of traffic counts has in fact been an active area of research for at least 15 years, with a good review of earlier work being given by Nguyen (1984), and a more recent interpretation of the problem can found in Spiess (1987). Here, although one cannot estimate the OD parameters from scratch, if one has an existing OD estimate, one can use the counts to update this prior estimate. The estimation problem can then be formulated mathematically as one of finding a new OD estimate that reproduces the available count data, and is "closest" in some sense, to the prior estimate. For freeway OD problems, the distance measure most often employed has been the entropy criterion

$$S = -\Sigma_i \Sigma_j \, x_{ij} \, [\log(x_{ij}/X_{ij}) - 1], \qquad (16)$$

where X_{ij} denotes the prior estimate, and the goal is to find the values for x_{ij} which maximize (16) while reproducing the ramp counts. This approach appears to have first been used in the "synthetic OD" (SYNODM) method used in the FREQ freeway simulation programs, which simply starts with a noninformative prior estimate of $X_{ij} = 1$ or 0, for feasible and unfeasible OD flows respectively, and then uses iterative proportional fitting (as in fitting a gravity model) to adjust these until the ramp counts are reproduced (Jovanis, et al. 1978). This maximum entropy approach has been generalized by Willumsen and his associates (Van Zuylen and Willumsen, 1980) to produce the ME2 model, which in turn has been incorporated in several transportation software packages, such as The Highway Emulator, FRESIM, and INTEGRATION.

Unfortunately, the maximum entropy approach has two weaknesses which limit its

usefulness for ATMS. First, because of its tendency to produce estimates which remain close to the prior estimate, the maximum entropy approach (and underdetermined approaches in general) will not generally correct for poor prior estimates, and if the prior estimates are biased, the resulting maximum entropy estimates will tend to retain that bias. In particular, if two on-ramps lead to the same off-ramps (as would be the case where two on-ramps are not separated by an intervening off-ramp), and if noninformative prior estimates are used, the maximum entropy estimates of the splitting probabilities for the two on-ramps will be identical, independent of the actual parameter values. This appears to be the cause of the rather disappointing performance of SYNODM reported by Stokes and Morris (1983). Second, because it was developed primarily as a deterministic, rather than as a statistical, procedure it is difficult to assess the consistency and efficiency of this approach.

The maximum entropy and related OD estimators were developed with the assumption that only a limited amount of count data, such as ADT estimates, are available for OD estimation. The increased prevalence of automatic surveillance and control systems, both on freeways and on urban arterials however means that time-series data of traffic counts are likely to be available, so that methods originally developed for fitting dynamic systems models might be employed. With repeated observations, the equations (15a-15b) become overdetermined, so that instead of seeking an exact solution, one seeks a solution which minimizes the error between the right and left hand sides. The simplest measure of error is the ordinary sum of squares

$$\Sigma_t(y_j(t)-\Sigma_i b_{ij} q_i(t))^2,\ j=1,..,m, \tag{17}$$

and it is well known the estimate of $\mathbf{b}_i$ which minimizes (17) is simply

$$\hat{\mathbf{b}}_i = [\Sigma_t \mathbf{q}(t)\mathbf{q}(t)^T]^{-1} (\Sigma_t \mathbf{q}(t) y_j(t)),\ i=1,..,m, \tag{18}$$

the ordinary least-squares (OLS) estimates. Cremer and Keller (1983) first applied this approach to the problem of estimating turning movement proportions at single intersections, and since then a number of variants on the least-squares approach have appeared in the literature (Cremer and Keller, 1987; Nihan and Davis, 1987, 1989; Bell, 1991). In particular, Nihan and Davis (1989) show that for single intersections, the OLS estimator is both unbiased and consistent, and that recursive versions of OLS are consistent and asymptotically unbiased. Generally, it appears that if the variability of travel times between origin and destination can be ignored, the OLS-based methods can give useful estimates of the b_{ij} parameters, using only time-series data of the arrival and departure counts. A Monte Carlo comparison of OLS-based methods and the Expectation-Maximization (EM) method (Nihan and Davis, 1989) for computing maximum likelihood estimates indicated that while the EM method showed considerably less sampling variability (i.e. greater efficiency) than did OLS-based methods, the EM method also showed both a noticeable bias in its estimates and a tendency to consume large amounts of computer time.

For freeways however, the performance of the OLS-based approach appear to be inferior to its performance on single intersections. For example, Table 1 displays results obtained during using OLS to estimate the splitting probabilities for a four-origin, two-destination segment of Interstate Highway I-35W, using actual traffic count data. Since, for this segment, origin 4 was downstream from destination 1, the parameter b_{14} was constrained to equal zero, but all other values were left unconstrained. Whereas OLS estimates for single intersection data come close to satisfying the constraints (5a-5b), the freeway estimates in Table 1 show considerable deviation. Although it would be possible to constrain the OLS procedure to guarantee satisfaction of (5a-5b), it is clear that if an unconstrained parameter estimator produces nonsensical estimates,

this is evidence of a flaw in the underlying traffic model used to generate these estimates. The

TABLE 1

OLS Estimates for Typical Freeway Data

		Off-ramp	
		1	2
	1	0.0	0.79
On-ramp	2	-.10	2.22
	3	0.34	2.43
	4	0.0	1.35

poor performance of OLS could be due to bias, inefficiency, or both, and there is no guarantee that constraining procedures would do more than disguise these underlying problems. Thus it was felt that this study should attempt to correct these problems, not mask them.

Recall that generally, least-squares methods are less efficient than maximum-likelihood (ML) methods, so that if the problems shown by OLS using real freeway data are due to inefficiency, replacing the OLS estimators with ML estimators could solve the problem. If it were possible to observe the on-ramp to off-ramp volumes $x_{ij}(t)$, then the maximum likelihood estimators for the probabilities b_{ij} would be

$$\hat{b}_{ij} = \frac{\sum_t x_{ij}(t)}{\sum_t q_i(t)} \tag{18}$$

and these estimates tend to be very efficient, (Nihan and Davis, 1989). In practice however, usually only the on-ramp volumes $q_i(t)$, the off-ramp volumes $y_j(t)$, and possibly some section volumes $x_k(t)$ will be available from the surveillance and control system. Since these counts are sums of multinomial outcomes, computing the likelihood functions for this aggregated data would require computing the convolution of several multinomial distributions, a task which is computationally prohibitive for all but very simply cases (e.g. Kalbfleish, et al. 1983). This situation can be viewed however as an instance of an incomplete data model, and Nihan and Davis (1989) have presented a procedure based on the EM algorithm (Dempster, Laird and Rubin, 1977) for computing the ML estimates. For this algorithm, one begins with a trial estimate of the splitting probabilities **B**, and then computes approximations of the conditional expectations of the $x_{ij}(t)$

$$\hat{x}_{ij}(t) = E[x_{ij}(t) \mid \mathbf{B}, \mathbf{y}(t)]. \tag{19}$$

For large $\mathbf{q}(t)$, the $x_{ij}(t)$ are approximately Normally distributed, so these conditional expectations can be approximated by

$$E[\mathbf{x} \mid \mathbf{y}] = E[\mathbf{x}] + C(\mathbf{x},\mathbf{y})C^{-1}(\mathbf{y},\mathbf{y})\,(\mathbf{y}-E[\mathbf{y}]), \tag{20}$$

where

$C(\mathbf{x},\mathbf{y})$ denotes the matrix of cross-covariances between the random vectors **x** and **y**,

$C(\mathbf{y},\mathbf{y})$ denotes the covariance matrix of the random vector **y**.

It can be shown (Nihan and Davis, 1989) that the elements of these matrices are given by

$$c(x_{ij}(t),y_k(t)) = \begin{cases} -b_{ij}\, b_{ik}\, q_i(t), & j \neq k, \\ b_{ij}(1-b_{ij})q_i(t), & j = k, \end{cases} \tag{21}$$

$$c(y_j(t), y_k(t)) = \begin{cases} -\Sigma_i b_{ij} b_{ik} q_i(t), & j \neq k, \\ \Sigma_i b_{ij}(1-b_{ij}) q_i(t), & j = k. \end{cases} \tag{22}$$

The estimates of **B** are then recomputed using (18) but with the new $\hat{x}_{ij}(t)$ substituted for $x_{ij}(t)$, and these two steps are iterated until a convergence criterion is satisfied.

The three estimation approaches discussed so far, SYNODM, OLS and the EM algorithm, are all based on the linear traffic model discussed in Chapter 2. Although appropriate for single intersections, the failure of OLS on freeway data suggest that the linear model may not be adequate for freeway traffic. Now two obvious difference between freeway traffic and intersection traffic are that first, the travel times between freeway origins and destinations can vary both as functions of the distance separating these points, and also as functions of the intervening traffic conditions and, second, platoon dispersion effects will cause the traffic exiting at an off-ramp during a time interval to be a mixture of traffic entering during different prior intervals. Thus to more accurately link on-ramp volumes to off-ramps volumes it may be necessary to embed the linear traffic distribution model within a macroscopic traffic flow model. This is the topic of the next section. However, existing traffic flow models such as KRONOS and FREFLO treat on-ramp and off-ramp volumes as boundary conditions to the simulation rather than as outputs, so that the modeling of destination-specific subflows is not possible. A combined traffic demand and traffic flow model, developed during this project, which does model such destination-specific subflows is described next.

CHAPTER 4

A MARKOVIAN FREEWAY TRAFFIC MODEL

In Nihan and Davis (1989), it was possible to obtain some idea concerning the relative efficiency of several estimators of intersection splitting probabilities using Monte Carlo simulation. For a freeway section however, it may be necessary to not only model the random departure and distribution of traffic but also the propagation of traffic from on-ramp to off-ramp, in a manner that both preserves the random allocation of arriving vehicles to off-ramps and is also consistent with current knowledge on traffic flow. It turns out that a unified traffic demand/flow model can be constructed using a class of stochastic process models called Markov population models (Karemusha and Pathria, 1981; Davis, 1992). The essence of this idea is to treat each section of freeway as a Markovian compartment, which can be thought of as a container from which particles can make random exits. Conditional on the current compartment population sizes, each particle makes its exit independently of every other one, so that if N_k denotes the current population in compartment k, P_k denotes the exit probability and y_k denotes the number of exiting particles over some short time interval, y_k is a binomial random variable with parameters N_k and P_k. To transfer this idea to traffic flow phenomena, imagine dividing the control period into T short time intervals of length Δ (e.g. typically 1 second $\leq \Delta \leq$ 10 seconds) and let $\tau = 1,..,T$ index these intervals. Let $N_k(\tau)$ denote the number of vehicles in section k at the beginning of interval τ, $y_k(\tau)$ denote the number of vehicles which exit from k during interval τ. What is now needed is a method of determining the exit probabilities P_k so that they

reflect traffic conditions in a natural way. To do this, let L_k denote the length of section k, and let the vehicles in section k at the beginning of interval τ be indexed by $l=1,..,N_k(\tau)$. To each of these vehicles assign a speed u_{kl}, and a location s_{kl} which denotes the distance from vehicle l to the downstream boundary of section k. Assume that the speeds u_{kl} are independent, identically distributed random variables with density function $f_k(u)$ while the locations s_{kl} are independent, identically distributed random variables with density function $g_k(s)$, where the form of these density functions may depend on current traffic conditions. Now clearly, vehicle l will exit section k only if

$$s_{kl} < u_{kl}\,\Delta \tag{23}$$

so that

$$P_k = Prob[vehicle\ l\ exits\ section\ k] = \int_0^{L_k}\int_0^{\Delta u_{kl}} f_k(u)g_k(s)\,du\,ds \tag{24}$$

Generally, the above double integral will be difficult to evaluate, but if the vehicle locations are uniformly distributed, i.e. if $g_k(s)=1/L_k$, it follows that

$$P_k = \frac{\Delta \bar{u}_k}{L_k} \tag{25}$$

where $\bar{u}_k$ is the space-mean speed of the vehicles in section k. By letting M_k denote the number of lanes in section k and $r_k(\tau)=N_k(\tau)/L_kM_k$ denote the traffic density in section k, $\bar{u}_k$ can be computed using a speed-density function, or $\bar{u}_k$ can be included as a dynamic variable which changes via some form of a momentum equation. For this study, the version of Payne's (Payne,

1979) momentum equation described in Cremer and May (1985) was used to determine the exit probabilities, producing the STOMAC (STOchastic MACroscopic) simulation model described below.

STOMAC Freeway Simulation Model

State Variables:

$N_{kj}(\tau)$ = # of vehicles in section k destined for off-ramp j;

$\bar{u}_k(\tau)$ = space-mean speed of vehicles in section k

Dynamics:

1. $P_k(\tau) = \Delta\bar{u}_k(\tau)/L_k$

2. $y_{kj}(\tau) = \text{binomial}\ (N_{kj}(\tau),\ P_k(\tau))$

3. $q_{ij}(\tau) = \text{Poisson}\ (\ b_{ij}q_i(\tau))$

4. $N_{kj}(\tau+1) = N_{kj}(\tau) + y_{k-1,j}(\tau) - y_{kj}(\tau) + \Sigma_i w_{ik} q_{ij}(\tau)$

5. $r_k(\tau) = (\Sigma_j\ N_{kj}(\tau))/M_k L_k$

6. $\bar{u}_k(\tau+1) = \bar{u}_k(\tau) + \Delta\bar{u}_k(\tau)(\bar{u}_{k-1}(\tau)-\bar{u}_k(\tau))/L_k$

$+ (\Delta/T)(\bar{u}_e(r_k(\tau))-\bar{u}_k(\tau)) - (\nu\Delta)(d_k r_{k+1}(\tau)-r_k(\tau))/(L_k T(r_k(\tau)+\kappa M_k)$

Special Terms:

T,κ,ν = momentum equation parameters,

$d_k = M_k/M_{k+1}$

$\bar{u}_e(r)$ = equilibrium speed/density function,

w_{ik} = 1,if on-ramp i joins section k; 0, otherwise.

STOMAC can be used to generate a series of simulated on-ramp volumes, distribute these on-ramp volumes to off-ramps using the above multinomial model and then propagate these volumes down the freeway and out the off-ramps in a manner consistent with the embedded traffic flow model. This in turn makes it possible to gain some idea of the statistical properties of splitting parameter estimators via Monte Carlo simulation.

CHAPTER 5

MONTE CARLO EVALUATION OF PARAMETER ESTIMATORS

Figure 2 on page 13, Chapter 2, depicts a 7-origin, 4-destination segment of northbound Interstate highway I-35W, which was selected as a subject for the simulation study. This segment is about 2.5 miles (4.0 km.) long, and the somewhat complicated origin-destination pattern results from State Highway 62 and I-35W sharing about 1.5 miles (2.4 km.) of right-of-way. Five-minute automatic traffic counts for each on-ramp, off-ramp and several mainline locations were obtained from the Minnesota Dept. of Transportation (MNDOT) for a typical morning peak period running form 6:00 AM to 9:00 AM. In order to run the simulation program STOMAC, it is necessary to know both the test matrix of splitting probabilities, **B**, and also the parameters entering in STOMAC's momentum equation. Plausible values for the momentum equation parameters were obtained from Cremer and May's (1985) report, while the "true" values for the matrix **B** were estimated from actual count data using nonlinear least-squares (NLS). This was done by writing a FORTRAN subroutine which took the actual on-ramp counts and a trial estimate of the **B** matrix as inputs, and then computed forecasted off-ramp counts by performing the STOMAC recursion with a 5-second Δ and with the Poisson and binomial random numbers used in STOMAC being replaced by their expected values. The five-second forecasted off-ramp counts were then aggregated to produce forecasted 5-minute counts, $\hat{y}_j(t)$, which in turn permitted computation of the sum-of-squares function

$$\Sigma_t \Sigma_j \, (y_j(t)-\hat{y}_j(t))^2. \qquad (26)$$

The values for **B** which minimized (26) were then computed iteratively by the Quasi-Newton optimization routine E04JAF, contained in the NAG Workstation Library (NAG, 1986). The results were then rescaled slightly to produce values satisfying constraints (5a-5b), and are displayed under the 'True' column in Table 2. It should not be argued that these are necessarily good estimates of this freeway section's actual **B** matrix, due to the arbitrary selection of values for the momentum equation parameters, since in practice one would estimate both the momentum parameters and **B** jointly. For Monte Carlo evaluation of statistical properties however, what is important is knowing exactly what parameter values were used to generate the simulated data, and so these values were deemed acceptable as inputs for the Monte Carlo simulation. Using STOMAC, 50 simulated series of 5-minute on-ramp and off-ramp counts were then generated, using the actual on-ramp counts for STOMAC's Poisson means. The Poisson and binomial random number generators were also available in the NAG Workstation Library, with all computer programs being written in FORTRAN and all computations being performed on a SUN Sparcstation 1+ computer. This estimation scenario corresponds to a situation where the b_{ij} may be changing slowly over a period of months, but where the estimates from a recent preceding day are considered representative of the current day.

From each simulated data set, it is possible, using a given estimator, to compute an estimate of the **B** matrix used to generate this data set, and the set of estimates from all 50 data sets form a pseudo-random sample of this estimator, from which one can compute sample means and standard deviations. The efficiencies of different estimators can then be compared via the sample standard deviations, while tendencies of the estimators to show bias can be evaluated by comparing the sample mean to the 'True' values used in generating the data. In this study, four

TABLE 2
Performance of Freeway OD Parameter Estimators

		OLS			NLS		
	True	Mean	S.Dev	RMS	Mean	S.Dev	RMS
b11	.056	.055	.003	.003	.055	.004	.004
b12	.134	.133	.041	.041	.136	.022	.022
b13	.019	.025	.056	.056	.020	.030	.030
b14	.790	.671	.123	.171	.798	.042	.043
b22	.291	.388	.321	.335	.303	.055	.056
b23	.290	.250	.346	.348	.295	.080	.080
b24	.419	.993	.902	1.07	.409	.062	.063
b32	.223	.225	.055	.055	.234	.034	.036
b33	.148	.140	.047	.048	.145	.028	.028
b34	.629	.618	.146	.146	.619	.045	.046
b42	.194	.203	.129	.129	.195	.038	.038
b43	.240	.208	.138	.142	.232	.055	.056
b44	.566	.646	.354	.363	.574	.059	.060
b53	.263	.248	.077	.078	.260	.047	.047
b54	.737	.825	.191	.210	.743	.054	.054
b63	.284	.297	.376	.376	.281	.029	.029
b64	.716	.841	1.07	1.08	.715	.058	.058
			EM			**SYNODM**	
b11	.056	.056	.002	.002	.056	.002	.002
b12	.134	.176	.012	.044	.177	.002	.043
b13	.019	.137	.022	.120	.118	.002	.099
b14	.790	.630	.031	.163	.699	.003	.141
b22	.291	.193	.041	.106	.184	.002	.107
b23	.290	.167	.064	.139	.122	.002	.168
b24	.419	.639	.074	.232	.694	.003	.276
b32	.223	.182	.018	.045	.184	.002	.039
b33	.148	.134	.043	.045	.122	.002	.026
b34	.629	.685	.042	.070	.694	.003	.065
b42	.194	.201	.071	.071	.184	.002	.010
b43	.240	.220	.092	.094	.122	.002	.118
b44	.566	.579	.099	.100	.694	.003	.128
b53	.263	.105	.031	.161	.154	.002	.109
b54	.737	.895	.031	.161	.846	.002	.109
b63	.284	.123	.084	.182	.154	.002	.130
b64	.716	.877	.084	.182	.846	.002	.130

different parameter estimators were evaluated: the NLS procedure described above (but without rescaling to guarantee satisfaction of (5a-5b)), together with OLS, the EM algorithm and SYNODM. For SYNODM, the five-minute on-ramp and off-ramp counts were totaled across the entire three-hour period to produce one set of counts. SYNODM was included as a benchmark against which to judge the time-series based estimators. Table 2 displays the sample means (b_{ij}) and standard deviations (σ_{ij}) computed for each of these procedures, along with the root-mean-squared (RMS) distance between the estimates and their true values. This latter quantity is defined via

$$RMS_{ij} = \sqrt{(b_{ij}-\bar{b}_{ij})^2+\sigma_{ij}^2} \tag{27}$$

and combines the estimation error due to inefficiency with that due to bias, to provide a single measure of these two tendencies.

Table 2 shows that OLS, the EM algorithm and SYNODM tended to produce estimates which can, on the average, differ from the true values. However, t-tests of the difference between the true values and the OLS averages tended to turn up nonsignificant, due to the OLS estimates having relatively large standard deviations. This suggests that the OLS estimator is in fact unbiased, and that differences between the average estimates and true values are due to sampling variability. This was not the case for the EM algorithm or SYNODM, both of which showed at least an order of magnitude less variability than OLS. Here both SYNODM and the EM algorithm tended to converge on values differing, on average, from the true values. SYNODM's tendency to produce identical estimates for on-ramps leading to the same set of off-

ramps can also be observed. With respect to the RMS criterion, both the EM algorithm and SYNODM appear to be superior to OLS, mainly because of lower sampling variabilities. SYNODM in fact, was almost immune to sampling variability, producing nearly identical estimates for each data set. In contrast to the other three estimators, the NLS procedure appears to produce estimates which are, on the average, close to their true values while showing a degree of efficiency comparable to the EM algorithm. This suggests that reducing ignorance concerning underlying traffic processes is one way of improving demand estimates. The variability shown by the OLS estimates is large enough to make them essentially useless in practice. For instance, appealing to the approximate Normality of the OLS estimates, an approximate 95% confidence interval for b_{14}=.566 would be [-0.62,1.35], which one knows prior to collecting data. This variablility provides a possible explanation of the tendency of unconstrained OLS to produce nonsensical values when used on real freeway data sets. In sum, neither of the linear model time-series procedures, OLS and the EM algorithm, could be said to perform better than SYNODM. When coupled with an exact model of the underlying traffic flow dynamics however, the least-squares criterion did tend to produce estimates with an improved efficiency and little tendency toward bias. This suggests that, at least for freeway traffic, it may not be possible to identify the travel demand model independent of the traffic flow model.

Using the estimates in Table 2, it is possible next to employ equation (13) to obtain a rough idea of magnitude of the capacity reductions which can be attributed to demand and parameter uncertainty. Since equation (13) assumes that the estimation procedure is unbiased, it is applicable only the OLS and NLS methods, and Table 3 displays the capacity reductions that could be obtained using a typical set of 5-minute on-ramp volumes for freeway sections 6 and

10. The quantities in Table 3 were computed assuming no cross-covariance among the parameter estimates, corresponding to the statistically ideal case where the matrices $\mathbf{Q}_k$ in (13) would be diagonal. The reliability level of $\alpha=0.05$ was selected, and Table 3 gives the capacity reductions in units of vehicles/lane/hour (v/l/h). Assuming a nominal capacity of 2000 v/l/h, Table 3 shows that if the splitting probabilities are known exactly, a 95% reliability would require using target capacities of around 1911 v/l/h in section 6 and 1923 v/l/h in section 10. When the splitting probabilities have been estimated by OLS, the target capacities drop to 1468 v/l/h in section 6 and 1567 v/l/h in section 10, while the use of NLS estimates would cause the target capacities to drop to 1807 v/l/h in section 6 and 1862 v/l/h in section 10. To put this another way if, under this data generation scenario, one is using OLS estimates and not metering with target capacities around 500 v/l/h lower than the nominal capacities, one could generate over capacity demands more than 5% of the time, possibly leading to a perception that the metering is less effective than expected. By switching from OLS estimates to NLS estimates one could however, gain practical capacity increases of around 300-350 v/l/h, without sacrificing the 95% reliability level.

TABLE 3

Typical Capacity Reductions

(in vehicles/lane/hour)

Section	Demand Uncertainty	Parameter Uncertainty	
		OLS	NLS
6	89	532	19
10	77	433	139

CHAPTER 6

CONCLUSIONS

To summarize, this report first argued that uncertainty in the ability to forecast the effects of a set of freeway management actions can lead to a lowering of the quality of the resulting traffic flow. Two sources of uncertainty were identified, one source being due to the lack of knowledge concerning the off-ramps selected by individual drivers entering the freeway, the other being due to a lack of precision in the estimates of the proportions of drivers selecting a given off-ramp from a given on-ramp. Using a stochastic freeway simulation model it was possible to provide estimates, for a particular freeway segment, of the uncertainties incurred by four different procedures for estimating these parameters. For the unbiased estimators OLS and NLS, computations of the loss of freeway capacity resulting from these uncertainties indicated, at least for this example, that demand uncertainty led to a loss of around 100 v/l/h, while the losses due to parameter uncertainty were sensitive to the efficiencies of the parameter estimation procedures. OLS estimation caused losses of roughly an additional 400 v/l/h while the more efficient NLS procedure caused additional losses of around 50-100 v/l/h.

The results of this study suggest two main conclusions, one concerning the impact of uncertainties on traffic management, and one concerning appropriate approaches to traffic modelling. First, it appears that both demand and parameter uncertainties can, in practical situations, be large enough to affect to influence the outcome of traffic management actions. Interestingly enough, the capacity reductions attributed to demand uncertainty reported here, of the range of 100 v/l/h for a Minnesota freeway, correspond to the traffic flows of around 1900 v/l/h produced by ramp-metering on a section of Seattle freeway (Nihan and Davis, 1984). Even

if this source of uncertainty is recognized however, the naive use of certainty equivalent traffic management models, where estimated parameter values are used as if they were known with certainty, can still lead to poor performance. The degree of performance degradation will depend on the properties of the parameter estimator, and on the context in which the parameter estimates are used. For instance, if the OD splitting probabilities can be treated as relatively stable over time, one might use data sets from a large number of days together with a consistent parameter estimator to produce a situation in which certainty equivalent methods would be justified. If, on the other hand, one is attempting to deal with time-varying splitting probabilities, then a combination of an efficient parameter estimator together with explicit recognition of the resulting parameter uncertainty is likely to be necessary. This in turn suggests that an ATMS based on a poorly-fitted traffic model is not likely to reach its full potential, and that issues concerning the appropriate use of estimates may deserve a more central place in ATMS research.

The second conclusion concerns the relationship between traffic flow and travel demand models. Historically, traffic flow and travel demand have tended to follow separate developmental paths, despite the fact that both classes of models deal with overlapping aspects of the same general phenomenon. This separation has produced fairly sophisticated traffic flow models, such as KRONOS and FREFLOW, which contain crude representations of traveller OD routing, and fairly sophisticated traffic assignments models which contain crude representations of traffic flow, such as the common BPR delay functions. In freeway modelling, this has led to a tendency to treat problems of estimating travel demand as somehow separate from problems of testing and deploying a traffic flow model, and this separation is implicit in the OLS, EM and SYNODM estimators tested earlier. The results of this study suggest that such a separation is

probably not desirable, and may not even be possible. The only estimation procedure which generated OD estimates free of troubling properties was the NLS estimators, which used an exact model of traffic flow to connect the on-ramp volumes to the off-ramp volumes. In practice however, the traffic flow model will not be known exactly, and must also be identified from traffic data. This leads to problems involving the simultaneous estimation of traffic flow and travel demand parameters, and the Markovian compartment model developed in this study as a simulation tool also provides a natural framework for doing such estimation. This line of investigation is being actively pursued.

For more general traffic networks, with multiple routes connecting any given OD pair, the problem of joint modelling of travel demand and traffic flow is likely to be even more acute, and leads to questions concerning appropriate modelling strategies. The Markovian compartment model developed in this study combines travel demand and traffic flow considerations in a natural way, and in fact can be generalized to handle the problems arising in general networks. Such models offer a plausible framework for the development of ATMS, and the identification of an approach for unified travel demand/traffic flow modelling may prove to be the most important contribution made by this study.

REFERENCES

Bell, M. (1991) "The real time estimation of origin-destination flows in the presence of platoon dispersion." *Transp. Res.* 25B, 115-125.

Chen, C., Cruz, J., and Paquet, J., (1974) "Entrance ramp control for travel rate maximization in expressways." *Transp. Res.* 8, 503-508.

Cremer, M. and Keller, H. (1983) "Dynamic identification of flows from traffic counts at complex intersections." *8th Int. Symp. on Transp. and Traffic Theory*, V. Hurdle et al. eds., Univ. of Toronto Press, Toronto, Canada, 121-142.

Cremer, M. and Keller, H. (1987) "A new class of dynamic methods for the identification of origin-destination flows." *Trans. Res.*, 21B, 117-132.

Cremer, M. and May, A. (1985) *An Extended Model of Freeway Traffic*, UCB-ITS-RR-85-7, Inst. of Transp. Studies, Univ. of Cal., Berkeley, CA.

Davis, G. (1992) "Integrated traffic assignment and flow models via Markovian networks." *Proceedings of the 25th ISATA Conference*, Florence, Italy.

Dempster, A., Laird, N. and Rubin, D., (1977) "Maximum likelihood from incomplete data via the EM algorithm." *J. R. Stat. Soc. Ser. B,* 39, 1-38.

Jovanis, P. Yip, W.-K., and May, A. (1978) *FREQ6PE: A Freeway Priority Entry Control Simulation Model*, UCB-ITS-RR-78-9,Inst. of Transp. Studies, U. Cal. Berkeley.

Kalbfleisch, J., Lawless, J. and Vollmer, W., (1983) "Estimation in Markov models from aggregate data," *Biometrics*, 39, 907-919.

Karemusha, and Pathria, R. (1981) "A stochastic model for highway traffic." *Transp. Res.* 15B,

285-294.

Ljung, L. (1978) "Convergence analysis of parametric identification methods." *IEEE Trans. Automatic Control*, AC-23(5), 770-782.

Ljung, L. and Caines, P., (1979) "Asymptotic Normality of prediction error estimators for approximate system models." *Stochastics*, 3(1), 29-46.

NAG (1986) *NAG Workstation Library*, Numerical Algorithms Group, Ltd., Oxford, UK.

Nguyen, S. (1984) "Estimating origin-destination matrices from observed flows," in M. Florian (ed.) *Transportation Planning Models,* Amsterdam, Elsevier Science, 363-380.

Nihan, N. and Davis, G. (1987) "Recursive estimation of origin-destination matrices from input/output counts." *Trans. Res.* 21B, 149-163.

Nihan, N. and Davis, G. (1989) "Application of prediction-error minimization and maximum likelihood to estimate intersection O-D matrices from traffic counts." *Transp. Sci.* 23(2), 77-90.

Papageorgiou, M. (1980) "A new approach to time-of-day control based on a dynamic freeway traffic model." *Transp. Res.* 14B, 349-360.

Payne, H. (1979) "FREFLO: a macroscopic simulation model of freeway traffic." *Transp. Res. Rec.* 722, 68-77.

Speiss, H. (1987) "A maximum likelihood model for estimating origin-destination matrices," *Transp. Res.* 21B, 395-412.

Stokes, R. and Morris, D. (1983) "Application of an algorithm for estimating freeway trip tables," *Transp. Res. Rec.*, 21-25.

Van Zuylen, H. and Willumsen, L. (1980) "The most likely trip matrix estimated from traffic counts," *Transp. Res.*, 14B, 281-293.

APPENDIX

List of Symbols Used in This Report

b_{ij} = probability a vehicle entering at i is destined for j;

$\hat{b}_{ij}$ = estimate of b_{ij};

$\overline{b}_{ij}$ = mean of estimates of b_{ij};

B = matrix of the b_{ij};

c_k = capacity of section k;

C(.,.) = matrix-valued function giving the covariance matrix between two random vectors;

c(.,.) = scalar-valued function giving the covariance between two random variables;

i = index of on-ramps, i=1,..,m;

j = index of off-ramps, j=1,..,n;

k = index of freeway mainline sections, k=1,..,p;

l = index of individual vehicles;

m = number of on-ramps;

M_k = number of lanes in sections k;

n = number of off-ramps;

N_k = number of vehicles in section k;

N_{kj} = number of vehicles in section k destined for j.

p = number of mainline sections;

P_k = probability of exiting section k;

q_i = traffic volume for on-ramp i;

$\mathbf{Q}_k$ = covariance matrix of $\boldsymbol{\beta}_k$;

r_k = traffic density in section k;

S = number of time intervals;

s_{kl} = location of vehicle l in section k;

T = number of sub-intervals;

t = time index;

u_{kl} = speed of vehicle l in section k;

$\overline{u}_k$ = space-mean speed of vehicles in section k;

x_k = traffic demand for section k;

y_j = traffic volume on off-ramp j;

z_α = standard Normal deviate with tail probability α;

α = critical probability value;

β_{ik} = probability a vehicle entering at i uses section k;

$\hat{\beta}_{ik}$ = estimate of β_{ik};

$\boldsymbol{\beta}_k$ = vector of β_{ik};

Δ = duration of time step;

$\sigma^2, \sigma_k^2, \sigma_{ij}^2$ = variance terms;

τ = time index;