ITS INSTITUTE
Intelligent Transportation Systems

Real-Time Traffic Prediction for Advanced Traffic Management Systems: Phase I

Final Report

Prepared by:

Gary A. Davis
Yorgos J. Stephanedes
Jeong-Gyu Kang

Department of Civil Engineering
University of Minnesota

CTS 95-05

Technical Report Documentation Page

1. Report No. CTS 95-05	2.	3. Recipients Accession No.
4. Title and Subtitle Real-Time Traffic Prediction for Advanced Traffic Management Systems: Phase I		5. Report Date October 1995 6.
7. Author(s) Gary A. Davis, Yorgos J. Stephanedes, Jeong-Gyu Kang		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Civil Engineering University of Minnesota 500 Pillsbury Drive SE Minneapolis, MN 55455		10. Project/Task/Work Unit No. 11. Contract (C) or Grant (G) No. CTS Project #1992002
12. Sponsoring Organization Name and Address Intelligent Transportation Systems Institute 200 Transportation and Safety Building 511 Washington Ave. SE Minneapolis, MN 55455		13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code
15. Supplementary Notes http://www.its.umn.edu/Publications/ResearchReports/		

16. Abstract (Limit: 250 words)

It has been recommended that Advanced Traffic Management Systems (ATMS) must work in real-time, must respond to and predict changes in traffic conditions, and must included areawide detection surveillance. To support such ATMS, this project developed a tractable, stochastic model of freeway traffic flow and travel demand which satisfies three primary objectives. First, the model should generate real-time estimates of traffic state variables from loop detector data, which can in turn be used as time-varying initial conditions for more comprehensive simulation models, such as KRONOS or FREESIM. Second, the model should generate its own predictions of mainline and off-ramp traffic volumes, as well as calculate the expected error associated with these predictions, thus supporting the use of both deterministic and stochastic optimization for determining traffic management actions. Third, the model should be capable of full on-line implementation, in that it should be capable of estimating required parameters from traffic detector data.

The basic model was developed by combining ideas from the theory of Markov population processes with a new for the relationship between traffic flow and density, producing a stochastic version of a simple-continuum model. Kalman filtering was then applied to the basic model to develop algorithms for (1) estimating from loop detector counts the traffic density in freeway sections broken down by destination off-ramp, (2) predicting main-line and off-ramp traffic volumes from given on-ramp volumes and, (3) computing adaptive estimates of the freeway's origin destination matrix from loop detector counts. Monte Carlo simulation tests were used to evaluate three different methods for off-line estimation of model parameters, as well as to assess the accuracy of the density estimates and volume predictions. The results indicated that the estimation and prediction model tends to be robust with respect to the parameter estimation scheme, and that the model generates a reasonable characterization of estimation and prediction uncertainty. Limited tests with field data tended to confirm the simulation results, and to emphasize the importance of real-time estimation of freeway origin-destination matrices in generating accurate predictions.

17. Document Analysis/Descriptors Advanced traffic management systems, Traffic models, Traffic flow, Markov processes, Freeways, Traffic forecasting, deterministic prediction, Stochastic processes		18. Availability Statement No restrictions. Document available from: National Technical Information Services, Alexandria, Virginia 22312	
19. Security Class (this report) Unclassified	20. Security Class (this page) Unclassified	21. No. of Pages 121	22. Price

Real-Time Traffic Prediction for Advanced Traffic Management Systems: Phase I

Final Report

Prepared by:

Gary A. Davis
Yorgos J. Stephanedes
Jeong-Gyu Kang

Department of Civil Engineering
University of Minnesota

October 1995

Published by:

Intelligent Transportation Systems Institute
University of Minnesota
200 Transportation and Safety Building
511 Washington Ave. SE
Minneapolis, Minnesota 55455

TABLE OF CONTENTS

EXECUTIVE SUMMARY

It has been recommended that Advanced Traffic Management Systems (ATMS) must work in real-time, must respond to and predict changes in traffic conditions, and must included areawide detection and surveillance. To support such ATMS, this project developed a tractable, stochastic model of freeway traffic flow and travel demand which satisfies three primary objectives. First, the model should generate real-time estimates of traffic state variables from loop detector data, which can in turn be used as time-varying initial conditions for more comprehensive simulation models, such as KRONOS or FREESIM. Second, the model should generate its own predictions of mainline and off-ramp traffic volumes, as well as calculate the expected error associated with these predictions, thus supporting the use of both deterministic and stochastic optimization for determining traffic management actions. Third, the model should be capable of full on-line implementation, in that it should be capable of estimating required parameters from traffic detector data.

The basic model was developed by combining ideas from the theory of Markov population processes with a new form for the relationship between traffic flow and density, producing a stochastic version of a simple-continuum model. Kalman filtering was then applied to the basic model to develop algorithms for (1) estimating from loop detector counts the traffic density in freeway sections broken down by destination off-ramp, (2) predicting main-line and off-ramp traffic volumes from given on-ramp volumes and, (3) computing adaptive estimates of the freeway's origin-destination matrix from loop detector counts. Monte Carlo simulation tests were used to evaluate

three different methods for off-line estimation of model parameters, as well as to assess the accuracy of the density estimates and volume predictions. The results indicated that the estimation and prediction model tends to be robust with respect to the parameter estimation scheme, and that the model generates a reasonable characterization of estimation and prediction uncertainty. Limited tests with field data tended to confirm the simulation results, and to emphasize the importance of real-time estimation of freeway origin-destination matrices in generating accurate predictions.

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Problem Statement

Many urban areas in the United States and throughout the world are experiencing increasingly serious congestion pressures on their street and highway networks, while at the same time, environmental, social and fiscal constraints combine to limit the amount of new roadway capacity which can be constructed to relieve these pressures. To help solve this problem, the United States and other developed countries are turning to the deployment of Intelligent Transportation Systems (ITS) as a means of making more effective use of existing infrastructure investments. Many ITS programs include an Advanced Traffic Management System (ATMS), in which an improved ability to monitor traffic conditions is combined with an array of traffic management actions so as to more nearly optimize the use of existing roadway capacity. Three important requirements of an effective ATMS are that (1) its actions should be based on a consideration of their network-wide implications, (2) that they should be based on actual traffic conditions, and (3) that the ATMS should attempt to anticipate future congestions problems and eliminate them, instead of simply responding after a problem has arisen. The most promising approach to developing an ATMS which satisfies these conditions is the hierarchical traffic control structure illustrated in Figure 1.1, in which a prediction of travel demand over a time horizon of 10-30 minutes is used as input to a network-wide optimization model. The optimal actions generated by the optimization model are then used as objectives or constraints by the ATMS components responsible

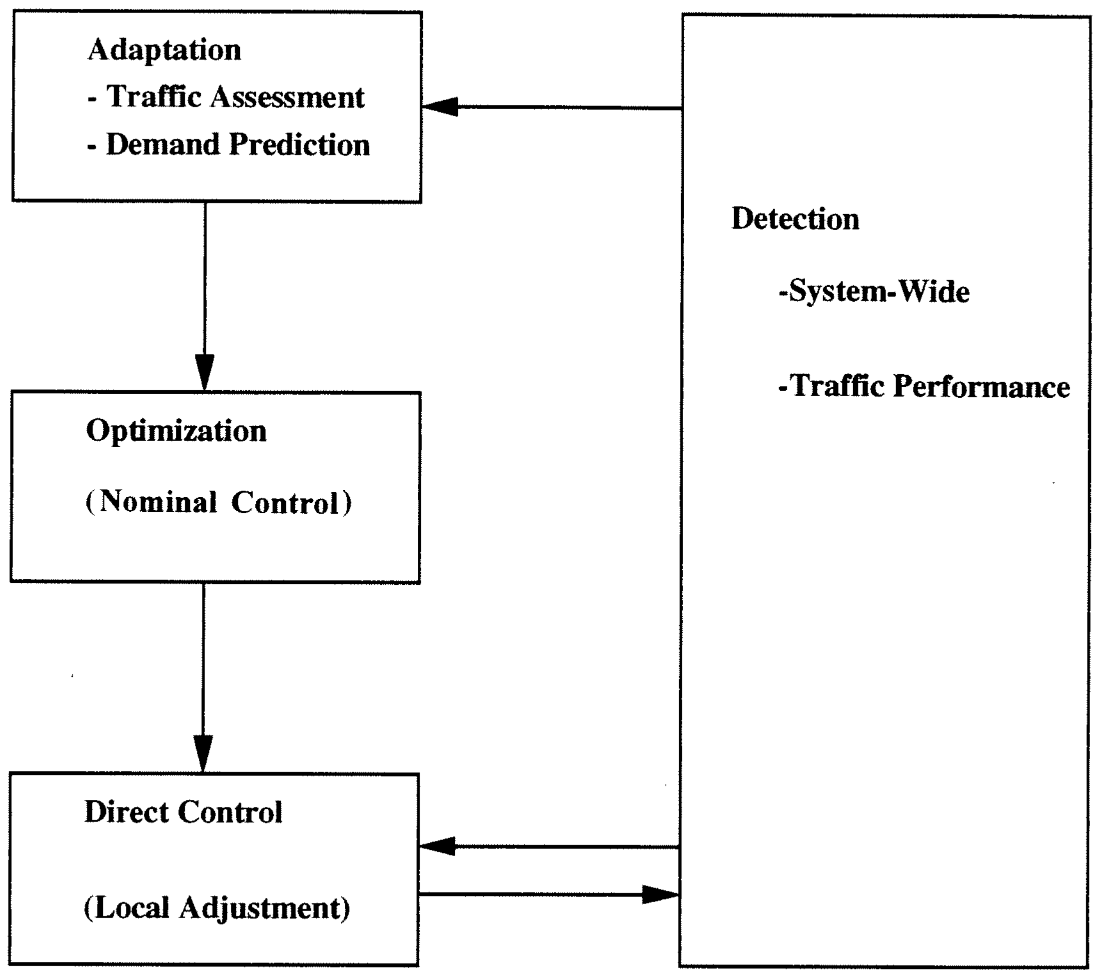


Figure 1.1. Hierarchical Control System for Freeways

for direct, local control of roadways. Thus a fully successful ATMS will require not only optimization and control algorithms, but a method for generating, in real-time, predictions both of travel demand and of the traffic conditions which are likely to result from proposed control actions. However, effective predictions require successful resolution of three important issues.

1.2 Issues in Traffic Prediction

1.2.1 Deterministic Versus Stochastic Predictions

The first issue that arises in traffic prediction is whether these predictions should be deterministic, in which it is assumed that the predicted values will occur with certainty, of if these predictions should be stochastic, in which it is admitted that their is some uncertainty concerning which values will occur. This uncertainty is usually quantified by describing a probability distribution over a range of possible outcomes. In reality, the future is never known with certainty, but deterministic predictions are easier to generate, and when deterministic predictions are used as input to an optimization or control algorithm, they produce "certainty equivalent" control. In certain special cases it can be shown that certainty equivalent controls are optimal even under conditions of uncertainty, but in most traffic management applications it will generate overly optimistic expectations of what traffic management will achieve. This results from the fact that most measures of traffic system performance, such as delay or total travel time, tend to be convex functions of travel demand, in that as travel demand increases these measures increase, and the rate of increase in these measures also increases as the demand increases. As an example, Figure 1.2 graphs average stopped delay at an approach to a signalized intersection as a function of arrival rate, using Webster's two-term formula. It can be seen that as demand (arrival rate) increases, stopped delay gets worse, and it gets worse at an increasing rate.

Jensen's inequality from probability theory implies that when delay, for example, is a convex function of demand, and when the demand prediction is subject to uncertainty, then the expected delay will always be greater than or equal to the delay computed using the deterministic prediction of demand. The magnitude of this discrepancy is determined by the shape of the delay function, and the degree of uncertainty in the demand prediction. Thus when the target measure of traffic system performance is a markedly convex function of demand, and when predicted demand is uncertain, actual improvement from certainty equivalent control will be less than the predicted improvement. This in turn suggests that if stochastic predictions are available, a genuine improvement in system performance could be obtained by switching from a certainty equivalent optimization algorithm to one explicitly considering this uncertainty. If only deterministic predictions are available, this option is not available.

1.2.2 Off-line Versus Real-Time Prediction

Transportation planners have engaged in a form of traffic prediction for several decades now, in which one first estimates or predicts an origin-destination (O-D) matrix, which gives the travel demand between combinations of origin and destination

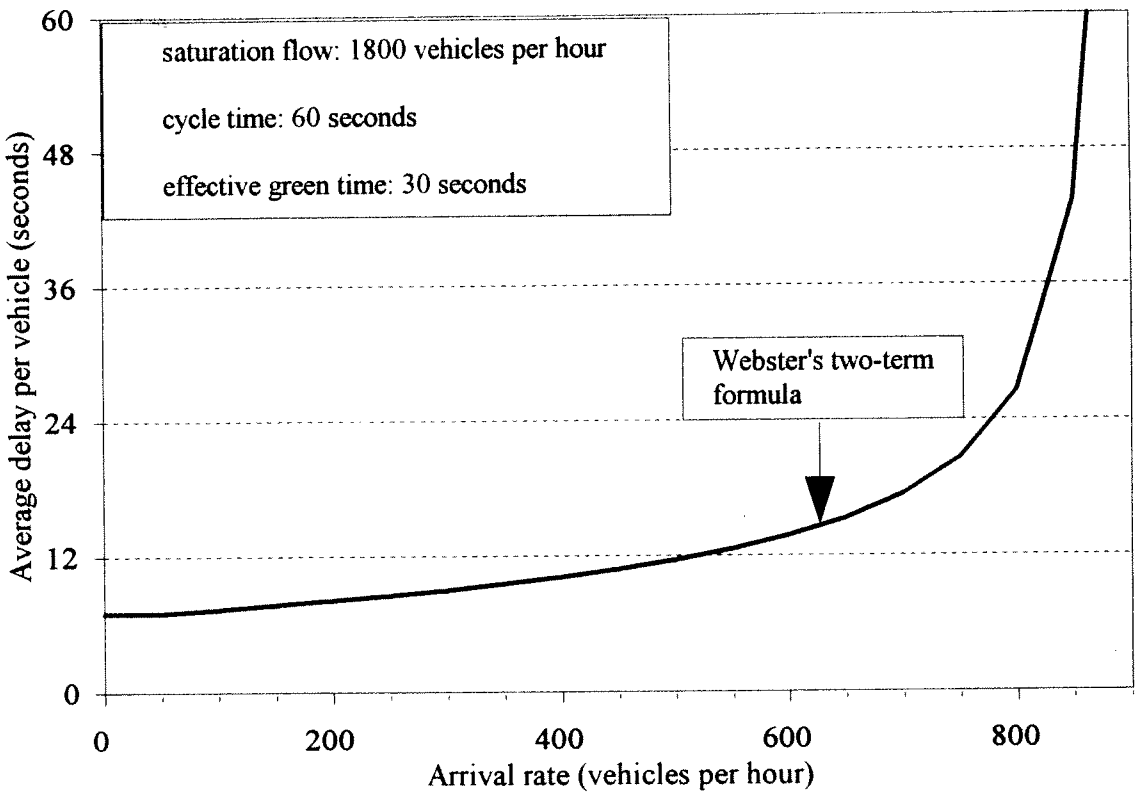


Figure 1.2. Average Delay at an Approach to a Signalized Intersection

points on a transportation network. Traffic is then routed onto the network by assigning the O-D volumes to modes of travel, routes and times-of-day, and the resulting traffic volumes on various road segments or transit routes are then computed by simple summation. Off-line traffic simulation models tend to have a similar structure, in that the demand for travel (arrival rates and turning movement proportions) is treated as fixed, known inputs. Given this demand, traffic is entered onto a network and then propagated according to principles of traffic flow. An important variable for describing traffic flow is the density of traffic on the network links, and in macroscopic models density is computed directly via a conservation equation, while in microscopic models density enters indirectly via a consideration of vehicle spacings.

The prediction problems faced by an ATMS differ from those solved by planning or off-line simulation models in that an ATMS problem is likely to involve a much shorter time horizon, and the prediction must incorporate current traffic conditions. For example, the arrival demand at an approach to a signalized intersection over the next 10 minutes will consist not only of traffic which enters the system from points of origin during the next 10 minutes, but also of traffic which has already entered the system and is enroute to a destination. Thus even if one has a reliable simulation model at hand, its use as a component of an ATMS will require that it should be provided with a sequence of estimated initial conditions reflecting the current traffic conditions. At a minimum, these initial conditions must include the current traffic densities on each link of the network, while microscopic and higher order macroscopic models will require even more information. In addition, an enroute driver may elect to continue following a previously chosen route, but he or she may also chose to switch to some other route, and the outcome of the switching decision will depend on the driver's destination. Thus when generating real-time, demand-responsive predictions each link on the network in effect become a new point of origin, and it is necessary to know not only the original O-D matrix and the traffic densities on the network links, but a current breakdown of the densities by destination. Thus a necessary component for real-time traffic prediction is an on-line filter which generates estimates of initial conditions from available traffic measurements.

1.2.3 Optimal Versus Suboptimal Estimation

To summarize, it has been argued that optimal management of traffic systems is likely to require stochastic rather than deterministic optimization, necessitating stochastic rather than deterministic prediction models, and that real-time prediction will require that the prediction model be continually re-initialized with current estimates of the state of the traffic system. In principle, both optimal state-estimation and optimal stochastic prediction could be accomplished using a stochastic state-space model of traffic flow, coupled with a measurement equation describing how observables such as traffic counts or lane occupancies are related to the system state. Except for very simple traffic networks however, this will result in the need to solve a (very intractable) nonlinear, stochastic filtering problem, suggesting that the use of suboptimal approximations to the optimal filter and/or predictor will be required. In addition, any prediction or filtering model is likely to require estimates of parameters characterizing the demand for travel and the traffic flow properties of the network, and if these estimates can also be generated in real-time, then changes due to alterations in the social or physical environment of the system could be accounted for. Again however, optimal real-time parameter estimation tends to be computationally intensive, and this suggests a practical reliance on suboptimal approaches. Whether or not a suboptimal approach is good enough for practical purposes is an empirical question.

1.3 Objectives of This Project and Organization of This Report

When initially proposed, the principal objective of this project was to develop a real-time prediction model of traffic flow for sections of freeway, by coupling a predictor of on-ramp arrival rates with a macroscopic traffic flow model. Subsequent developments have indicated however that considerable effort is being expended nation-wide to develop traffic simulation models for use in ATMS. In particular, for freeways, work at the University of Minnesota has produced not only the very detailed KRONOS simulator, but advances in the development of high-order macroscopic models. In contrast, real-time estimation of the state variables and parameters which such models would require for real-time implementation has attracted much less attention, and it was felt that this project could make a more important contribution, both in regard to the national ITS effort and as a component of the ITS Institute's effort, by tackling these problems head on. This led to the following set of objectives for the project:

1. To develop a tractable, plausible stochastic state-space model of freeway

traffic flow;

2. To use the above model to develop a tractable filtering algorithm for generating real-time estimates of the destination-specific traffic densities of freeway segments, using only loop detector data;

3. To evaluate a number of procedures for estimating the model's parameters;

4. To evaluate the accuracy of the model in predicting off-ramp and mainline traffic volumes.

Thus the primary function of the model developed here is real-time traffic state and parameter estimation. However, if desired the model can also generate predictions, although not in the detail of KRONOS.

This report is organized as follows. Chapter 2 reviews the state of the art in traffic flow modeling and parameter estimation, while Chapter 3 describes the Markov compartment model which provides the basis for this model development. Chapter 4 describes the development of a tractable approximation to the original Markov model, and how the approximate model is coupled with the Kalman filter to estimate destination-specific densities. Chapter 5 describes an evaluation of several parameter estimation schemes, while Chapter 6 describes tests of the filter and predictor using both simulated and actual traffic data. Finally Chapter 7 lists conclusions and recommendations for further work.

2. REVIEW OF TRAFFIC STATE ESTIMATION AND PARAMETER IDENTIFICATION METHODS

2.1 Introduction

The past two decades have seen important advances in the development and application of computerized tools, that have resulting in improved freeway control strategies. This review of traffic simulation models starts by noting that accurate descriptions of traffic flow dynamics are needed not only for better understanding of traffic behavior, but also for analyzing flow conditions dynamically and devising more effective control strategies. This chapter lays the groundwork for succeeding chapters by outlining the traffic flow model used in this study and providing a review of methods for traffic state estimation and parameter estimation. Section 2.2 describes a macroscopic traffic flow model, and also reviews current thinking on traffic state estimation using detector data. Since the knowledge of the number of vehicles bound for specific destinations on each link of the freeway network is a necessary input to traffic prediction and to the successful development of route guidance (Davis, 1992; Papageorgiou and Messmer, 1991), attention is given to this issue. An additional review is given of the application of state estimation methodologies in transportation. Section 2.3 reviews the estimation of traffic flow and Origin-Destination (O-D) matrices with an emphasis on the statistical properties of the estimators. Also, it is argued that link count data contains more information than is used by current O-D estimation methods, and that this information can be used to improve the practice of O-D estimation.

2.2 Traffic Flow Models and State Estimation

2.2.1 Traffic Flow Models

Existing freeway simulation models fall into two general categories: (a) microscopic and (b) macroscopic. In the transportation field, the microscopic models are used to describe the behavior of individual vehicles in a network, or to analyze the behavior of the driver-vehicle system in a stream of interacting vehicles. The microscopic simulation models consist of a discrete simulation in which each vehicle's path in the network is described by car-following equations (e.g. FRESIM) or by travel time equations (e.g. INTEGRATION). The main applications of microscopic models have been to geometric design, safety analysis, and off-line testing and evaluation of traffic control strategies. Yet in spite of the conceptual appeal of microscopic models, which can track each vehicle's destination, they have not been widely used in developing adaptive traffic operations or control strategies. This is partly because of computational difficulties with networks of realistic size, and partly because of some limitations in their formulation; their assumptions are difficult to validate because human behavior in real traffic is difficult to observe and measure.

The macroscopic continuum (hydrodynamic) models describe traffic dynamics both in time and space via macroscopic traffic variables, such as flow, density, lane occupancy and mean speed. Three relationships are required to describe traffic flow dynamics in a macroscopic way. The first relationship is inherent in the definitions of traffic volume, speed, and density:

$$Q = k \cdot u \tag{2.1}$$

where

$Q = Q(x,t)$ = traffic volume (veh/time) at location x and time t,

$k = k(x,t)$ = density (veh/mile/lane) at location x and t, and

$u = u(x,t)$ = space-mean speed (mile/time) at location x and t.

A second relationship, the conservation equation, has the following general form:

$$\partial k/\partial t + \partial q/\partial x = g(x,t) \tag{2.2}$$

where $g(x,t)$ is a traffic generation rate.

The two fundamental equations (2.1) and (2.2) must be incorporated in all macroscopic continuum models. The continuum models are then further separated into the simple continuum models and the high-order continuum models, based on a third relationship. A simple continuum model uses a relationship between the mean speed and the traffic density under equilibrium condition:

$$u = u_e(k). \tag{2.3}$$

A high-order continuum model consider acceleration and inertia effects by replacing equation (2.3) with a momentum equation. More information on macroscopic models can be found elsewhere (Gerlough and Huber, 1975; May, 1987;1990). Because of their relative simplicity in computation, macroscopic continuum models are more often used in developing traffic control strategies (Payne et al., 1987; Papageorgiou and Messmer, 1991; Kühne, 1991; Stephanedes and Chang, 1993).

Both simple and high-order macroscopic traffic flow models employ a steady-state flow-density-speed relationship. Earlier models assumed a single-regime over the complete range of flow conditions, including free flow and congested situations (Greenshields, 1936; Greenberg, 1959; Underwood, 1962; Drew, 1965; Drake et. al, 1967; Munjal and Pipes, 1971). It was shown (Drake et. al., 1967) that most of the single-regime models had deficiencies over some portion of the density range and underestimated speed and flow in the density range from 20 to 60 veh/mile/lane. This is particularly disappointing because they are not able to track faithfully the measured field data near capacity conditions as is often needed. Later models attempted to improve single-regime models by considering two separate regimes, such as a free-flow regime and a congested-flow regime, and attempted to generalize these by introducing additional parameters that could be used to distinguish between roadway environments (Eddie, 1961; Cedar and May, 1976; Easa and May, 1980). However, both single-regime or two-regime flow-density relationships show deficiencies in describing freeway traffic over the full density range (May, 1990), especially in rapidly changing traffic situations.

The application of the existing macroscopic models to ATMS strategies in their current form raises several issues. The first issue is the lack of integration of traveler decision-making and traffic flow (Davis, 1992). Despite some ability to describe traffic dynamics on a freeway, most existing macroscopic models suffer from the problem that no clear relationship exists between travel demand and dynamic traffic volume counts. Since most macroscopic models do not differentiate their component flows by destination they usually specify either a fixed proportion of turns (e.g. FREFLO) or a fixed amount of exit flow (e.g. KRONOS) at an off-ramp in a freeway, or at a junction where there is a route choice. However, in reality both quantities depend on when vehicles bound for the various destinations reach the junction in question (Papageorgiou and Messmer, 1991; Davis, 1992; Vaughan and Hurdle, 1992; Daganzo, 1994).

The second issue concerns the ability of the macroscopic traffic flow models to describe traffic dynamics on a freeway. The simple continuum model is easier to implement than are high-order continuum models because of its simplicity in formulation and numerical computation. However, the simple continuum model has been criticized as not faithfully describing non-equilibrium traffic dynamics because it does not consider acceleration and inertia effects. High-order models, on the other hand, are in principle more realistic as they include the effects of inertia and acceleration of the traffic mass but high-order models have not as yet proved superior to the simple continuum model, at least in medium-to congested flow conditions (Michalopoulos et. al., 1984), and suffer from conceptual inadequacies (Daganzo, 1994). Further, their numerical demands are greater than those of the simple continuum model.

A questionable hypothesis in the simple continuum model approach is whether or not an empirically determined flow-density relationship is valid under time dependent conditions. One possible improvement is to develop a more realistic flow-density relationship, which can describe changing traffic conditions. This goal may be achieved by developing a two-dimensional flow-density relationship based on the assumption that traffic flow, density or speed is a function of not only of the current location's traffic condition but also of downstream traffic conditions. This approach has been explored by Szeto and Gazis (1972), who investigated the problem of coupling between freeway sections. It was assumed that the flow from a section 1 into the immediate downstream section 2 depended on both the density in section 1 and the density in section 2 via a two-dimensional flow-density relationship.

$$q_1(t) = f[k_1(t), k_2(t)]$$

where q and k denote flow and density respectively. The idea of incorporating a downstream condition is partially incorporated in many of traffic flow models; as an acceleration term in car-following model, as an anticipation term in high-order macroscopic traffic flow model, or via special numerical treatments in some simple continuum models. The approach of Szeto and Gazis has been refined by Davis and Kang (1993, 1994) who developed a 2-dimensional flow-density relationship, and by Daganzo (1994) in his cell-transmission model.

In summary, it was argued that major limitations in the application of existing macroscopic models to ATMS strategies are: (a) the lack of an ability to integrate traveler decision making into the traffic flow model, and (b) the lack of a traffic flow model which can realistically describe traffic dynamics and yet is simple to compute or estimate. In Chapter Three these issues will be resolved by introducing an integrated traffic flow/demand model based on Markov processes (Davis, 1992).

2.2.2 Traffic State Estimation

As mentioned, successful implementation of ATMS requires accurate real-time information on various traffic variables. Unfortunately, almost all traffic sensors provide limited traffic data, such as the traffic volumes and lane occupancies provided by magnetic loop detectors, which is aggregated across a network's O-D specific subflows. This fact requires that some traffic variables necessary for developing ATMS strategies, such as section density, speed, or destination-specific density, must be estimated, rather than measured directly. Since the 1970s several researchers have applied the Kalman filter to areas of traffic state estimation. The motivation of the research was to estimate from sensor data the traffic information needed for control system operation. The state dynamics equation and observation equation of a continuous discrete state space model can be represented as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x},\boldsymbol{\beta},t) + \mathbf{w}(t) \tag{2.4}$$

$$\mathbf{y}(t_k) = \mathbf{H}\mathbf{x}(t_k) + \mathbf{v}(t_k) \tag{2.5}$$

where $\mathbf{x}(t_k)$ is the state vector, $\boldsymbol{\beta}$ is a vector of model parameters, $\mathbf{y}(t_k)$ is an observation vector, $\mathbf{w}(t)$ is the state noise process, with covariance matrix $\mathbf{Q}(t)$, and $\mathbf{v}(t_k)$ is the measurement noise process, with covariance matrix $\mathbf{R}(t_k)$.

Gazis and Knapp (1971), Szeto and Gazis (1972), Kurkjian et al. (1980), Akahane and Koshi (1978), Okutani (1987), Smulders (1987), and Payne (1987) all investigated estimating a traffic state variable, say densities and/or mean speeds, from volume, occupancy, and speed measurements. A pioneering density estimation procedure was proposed by Gazis and Knapp (1971). In this paper, the authors introduced a procedure for estimating density by first calculating travel time. The method was complicated because it required the solution of a two-point boundary value problem. Knapp (1973) avoided the solution of a two-point boundary value problem at the expense of storing a great deal of speed and flow data from each detector. This data requirement problem was avoided by Ghosh and Knap (1978) by using an extended Kalman filter based on a linear traffic flow model and a nonlinear measurement. Szeto and Gazis (1972) have also proposed another method that does not require the solution of a two-point boundary value problem. They obtained a pseudo measurement of the density by computing a density value as a function of speed measurements, and then combined this pseudo measurement with the state equation via an extended Kalman Filter. Their method required only the calibration of a static speed-density relationship, which is calibrated as a time invariant state variable through the Kalman filter. Such an assumption may be appropriate in the Lincoln tunnel under homogeneous conditions, but it cannot be expected to hold on freeways, where lane-changing, accidents, and other sources of fluctuation must be accounted for.

It should be pointed out that most of the existing models rely on the following exact conservation-of-vehicles equation as a state equation

$$x(t+\Delta) = x(t) + [q_{in}(t) - q_{out}(t)]/(L \cdot N) + w(t)$$

Here, $x(t)$ represents the density in a specific section at time step t, and $q_{in}(t)$, $q_{out}(t)$ represent the input and output flows between time step t and time step t+1, which are assumed to be given by measurements. The terms L and N represent the length of the section and the number of lanes, respectively. A noise term w(t) is used to model the difference between the actual and measured change in density. Since this conservation equation is valid when $L/\Delta > u_f$ (the state equation has a Markovian property) the measurement interval Δ should be shorter then L/u_f. This enforces a measurement interval of five seconds or less, which is very short for traditional loop detectors, for the filters in Szeto and Gazis (1972) and Ghosh and Knapp (1978). Given that short measurements are available however, this method is still not able to track destination-specific densities because the observed sensor flow is given as a sum across the destinations rather than as destination-specific subflows. Although several macroscopic simulation models which model destination-specific subflows, are available (Mahmassani and Jayakrishnan, 1990; Papageorgiou and Messmer, 1991; Davis, 1993), few applications have been reported for the destination-specific density estimation.

In summary, despite the good theoretical properties of the Kalman-filtering approach in traffic state estimation, several problems hinder a straightforward application of the Kalman filter. The first problem is the lack of a good traffic flow model. Most of the modern estimators obtain $q_{in}(t)$, or $q_{out}(t)$ as state dependent quantities, and questions arise concerning how accurately the traffic flow is replicated by the model. In addition, for a freeway of realistic size, the relations between the entry and exit flows will depend on congestion levels, and thus require an accurate nonlinear traffic flow model to relate link counts to state values.

The filtering problem cannot be seen as apart from a second problem, estimation of parameters that are inherent in the filtering model. To date, all methods for the estimation of traffic flow and/or O-D parameters have employed prediction error minimization (PEM) methods, where one first specifies a model for predicting the network's output (volume or lane occupancy) using the input counts and a trial set of parameters. One then selects as the estimates those parameter values which minimize some measure of the difference between the predicted outputs and the actual ones. A prediction model is thus essential for the estimation of parameters. Our first requirement then is a prediction model that can handle realistic networks and allows for a variety of possible detector configurations.

The third problem is that a lack of the information about the noise covariance matrices $\mathbf{Q}(t)$ and $\mathbf{R}(t_k)$, on which the performance of the Kalman filter depends. Good approximations of these matrices should improve the estimates. If the link volumes are actually stochastic outcomes, as is commonly assumed, then not only the mean of the link flows but also their covariance matrix will depend on the state vector. This in turn means that the error terms are not independent of the state vectors, and that equations (2.4) and (2.5) are misspecified. This suggests that there is more information about the state vector in link count data than is used by past implementations of the Kalman filter in traffic state estimation. However, this approach is hindered by the difficulties in deriving the covariance of the link counts. In Chapter Three this difficulty will be sidestepped by applying the method of large population approximation (Lehoczky, 1980) to the stochastic process describing the link volumes.

2.3 Parameter Estimation for Traffic Flow Models

2.3.1 Approaches to Parameter Estimation

In this section we review approaches to the problem of estimating model parameters from a large set of observations. Before any of the traffic flow models can be used to replicate the behavior of actual freeways they must be calibrated by being provided with estimates of several parameters. Generally these parameters can be divided into two groups. The first group consists of traffic flow parameters, containing quantities such as capacities, free-flow speeds, and jam densities, which describe the evolution of the macroscopic variables mean speed and density. The other group, containing arrival rates at on-ramps and the proportions of on-ramp traffic destined for off-ramps, describes the demand for freeway use.

In the past the required inputs have been estimated by off-line methods, although for ATMS it will often be necessary to have a model of the traffic system available on-line as the system is operating. Such problems can be solved using recursive identification methods, which means that measured input-output data are processed sequentially as they become available, and estimates of the parameters adjusted appropriately. An important practical advantage to using recursive methods is that they result in considerably faster algorithms and are able to track time-varying parameters. Since the O-D proportions of a freeway may be time-varying, the effectiveness of off-line estimation for use in a real-time ATMS is questionable.

There are two disadvantages to recursive identification in contrast to off-line identification. First, the decision of what model structure to use has to be made *a priori*, before starting the recursive identification procedure. In the off-line situation different types of models can be tried out. The second disadvantage is that, with a few exceptions, recursive methods do not give as good of estimates as off-line methods. For long data records, the difference is not usually significant. In general, though, the recursive algorithms are less-efficient than their off-line counterparts (Ljung and Söderström, 1983) because the recursive constraint means that the data is not maximally utilized. Therefore, the statistical properties of the estimators should be evaluated.

From a mathematical standpoint, all traffic flow and/or demand estimators can be thought of as solutions of optimization problems, where a measure of fit is first presented which varies as a function of the values of the parameters. Those parameter values which optimize this measure are identified as "good" estimates, and the optimizing values are usually characterized as the solution of a set of equations and/or inequality constraints. Thus, the problem of computing these estimates is reduced to a problem of solving a system of equations. This next section reviews various methods which have been proposed for estimating freeway traffic flow and O-D parameters.

2.3.2 Traffic Flow Parameter Estimation

Over the past 60 years a number of traffic flow models have been proposed and considerable research has been directed to the estimation of traffic flow model parameters. The analysis of flow-speed-density relationships based on field measurements can be a very difficult task. Unique demand-capacity relationships over time of day and over length of the roadway must be present. Even then the complete range of flow, speed, and density values will probably not have been recorded. Parameter values for flow, speed, and density curves are often difficult to estimate and can greatly vary across sites. Many other factors affect flow-speed-density relationships, such as design speed, access control, speed limits and geometrics. The most common off-line method for the parameter estimation in a freeway traffic flow model is the nonlinear least squares output error method (Cremer and Papageorgiou, 1981; Cremer and May, 1985; Davis, 1993), which minimizes the sum of the squared errors between the output of a parameterized predictor and observations of the output.

Since Szeto and Gazis (1972) applied the Kalman Filter (KF) method for the joint parameter and state estimation for freeway system, several researchers have applied recursive estimation methods to the problem of identifying freeway traffic flow parameters. Most of these are Kalman filter approaches which can be considered as adaptive implementations of optimal filters for state estimation. The state vector is augmented by addition of unknown, but constant parameters, leading to a nonlinear filtering problem due to the occurrence of products between parameters and states. The Extended Kalman Filter (EKF) can in principle be applied to estimate the composite state comprising the original state and parameters. Grewal and Payne (1976) formulated traffic flow parameter estimation as a two-step estimation problem for their macroscopic freeway traffic flow model. First, two traffic flow parameters, which were not identifiable by the filter from the given data, are estimated by least-squares. Next, their remaining two parameters were estimated by using the extended Kalman filter.

Although these approaches showed reasonable estimation results on either simulated data or real data, the problem of supplying a systematic state and measurement noise was largely ignored. It is to be noted that the KF algorithm depends on knowledge of the state and measurement noise covariances. It has been observed from simulations and practical applications that the EKF parameter estimator, may give biased estimates, and may sometimes diverge. The convergence behavior has been investigated by Ljung (1987), who has demonstrated that the convergence difficulties arise from a combination of factors. These include incorrect specification of the noise covariances and the dependence of the Kalman gain on the parameter estimates.

2.3.3 Freeway Origin-Destination Matrices Estimation

Vehicle movement desires are generally summarized in origin-destination (O-D) tables. For a freeway system, origins correspond to on-ramps while destinations are off-ramps. Dynamically updated O-D tables are required for various strategies aimed at optimal usage of existing freeway capacity, such as ramp metering, route guidance and incident management. Historically, these O-D proportions have been estimated by manual counting, but the ongoing deployment of real-time, adaptive traffic control strategies indicates that automatic estimation of these proportions from traffic detector data is becoming increasingly important. As in all statistical estimation exercises, the quality of O-D estimates will depend on the type and quality of the data available. Traffic counts at on-ramps provide reliable estimates of arrival rates, but obtaining O-D matrices directly is extremely difficult and costly, so that indirect O-D estimation of the

O-D matrices from time-series traffic counts has received increasing attention.

As noted in Davis (1993), O-D estimation methods can be classified as either over- or under-determined, depending on whether or not the traffic data at hand is sufficient to produce a unique estimate of the O-D elements. For under-determined approaches, there will be an infinite number of O-D estimates that are consistent with the count data, and one of these is selected by first specifying a prior estimate of the O-D matrix and then selecting as the new estimate that which is consistent with the count data and "closest" to the prior estimate. This problem has been an active area of research for at least 15 years, with good review of earlier work being given by Nguyen (1984) and Spiess (1987). This under-determined approach has been incorporated in several transportation software packages, such as FREQ, The Highway Emulator, FRESIM and INTEGRATION. Unfortunately, these approaches are static in nature, and tend to be biased because of reliance on prior O-D information. The limitations of under-determined approaches were described by Cremer and Keller (1987).

The basic idea of over-determined approach is that traffic flow through a freeway system is treated as a dynamic process in which the sequences of short period exit flow counts depend by causal relationships upon the time-variable sequences of entrance flows. In that way enough information can be obtained from the counts at the entrances and the exits to obtain unique and bias-free O-D flow estimates without further *a priori* information. To date, all methods for the estimation of freeway O-D matrices have employed prediction error minimization (PEM) methods, where one first specifies a model for predicting the freeway's exit counts using the on-ramp input counts and a trial set of O-D parameters. One then selects as the estimated O-D proportions those values which minimize some measure of the difference between the predicted exit counts and the actual ones.

Cremer and Keller (1983) first applied an Ordinary Least Squares (OLS) approach to the problem of estimating turning movement proportions at single intersection, and since then a number of variants on 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 a single intersection, 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 O-D parameters, using only time-series data of the arrival and departure counts.

However, there exist two obvious differences between freeway traffic and intersection traffic. 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. Second, platoon dispersion effects will cause the traffic exiting at an off-ramp to comprise of on-ramp traffic from different time intervals. Davis (1993) reported that the performance of the OLS-based approach incorporating a linear traffic flow model results in poor estimates. This is mainly due to the unrealistic simplicity of the underlying linear traffic flow model. One may improve the results by incorporating a macroscopic freeway traffic flow model.

Unfortunately, most macroscopic models do not differentiate their component flows by destination, and therefore lack explicit O-D routing. Thus, when the flow reaches an off-ramp area in the freeway or junction where there is a route choice, the existing macroscopic models usually specify either a fixed proportion of turns or a fixed amount of exit flow. In other words, they treat on-ramp and off-ramp volumes as boundary conditions for the simulation rather than as outputs, so that the modeling of the destination-specific subflows is not possible. Since the majority of the recursive and Kalman filter based methods for estimating time-varying O-D parameters that have been presented in the literature (Okutani, 1987; Davis, 1987; Chang and Wu, 1994; Zijpp and Hammerslag, 1994; Akiva, 1994) incorporate the linear traffic flow model, these considerations should apply to them all. Thus in order to link on-ramp volumes to off-ramp volumes more accurately it may be necessary to embed the linear traffic distribution model within a nonlinear macroscopic traffic flow model. This approach should provide both on accurate description of traffic propagation on the links and explicit connections of origins to destinations.

A final issue concerns the selection of estimators. Many estimators are based on the Kalman filter because it allows the estimation of time varying parameters. Nihan and Davis (1987) reported a superiority of the Kalman filter based method over the other methods on freeway data. Zijpp and Hammerslag (1994) tested several methods including least squares, constrained optimization, and the Kalman filter based methods on both simulated and observed data for a freeway. While their state model is limited to the linear model, the Kalman filter based method led to the best results.

3. DEVELOPMENT AND TESTING OF A STOCHASTIC FREEWAY TRAFFIC FLOW MODEL

3.1 Introduction

A dynamic traffic flow model plays an important part in traffic prediction and control. Application of methods from stochastic systems theory for parameter estimation, filtering and control first requires the formulation of a state-space model describing the evolution of the system of interest. In a manner analogous to the deterministic state space model used in earlier research (Michalopoulos et al., 1984; 1991, Cremer and May, 1985), an appropriate stochastic state space model for traffic system can be based on a Markovian network model (Davis, 1991).

In section 3.2, a stochastic version of the simple continuum model will be formulated as a Markov population process. In section 3.3, a new transition rate function will be developed. In section 3.4, the stochastic model will be evaluated.

3.2 Markov Compartment Model of Freeway Traffic Flow (MARCOM)

A compartmental system is defined (Jacquez, 1985) as "a system which is made up of a finite number of macroscopic subsystems, called compartments, each of which is well mixed, and the compartments interact by exchanging materials. There may be inputs from the environment into one or more of the compartments, and there may be outputs from one or more of the compartments into the environment." The development and use of compartment models began in the 1940s, with the use of tracer experiments in research into the physiology of animals and humans. Later Conlisk (1976) introduced "interactive Markov Chains" were future transitions depend on the current distribution of the population in the compartments. This is also termed a partial independence model (Lehoczky, 1980). Although Markov compartment models have been applied to a problems in a sociology, chemistry, and biochemistry, practical applications in transportation have been limited.

Karmeshu and Pathria (1981) developed a Markov compartment model for highway traffic and provided a "large population" analysis. Using the method of diffusion approximations, a tractable approximation to their master equation was derived. This equation led to the derivation of ordinary differential equations describing the evolution of the means and variances of the compartment populations. Davis (1992) recently proposed an integrated traffic demand/flow model for general networks using Markov compartment models. The essence of this idea is to treat each section of roadway as a Markovian compartment, where the material is composed of vehicles and the stochastic nature of material transfer is caused by the random movements of vehicles according to a discrete time Markov chain. Conditional on the current compartment population sizes, each particle makes its exit independently of every other one, so that if x_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 x_k and p_k.

Now imagine that a section of freeway has been divided into subsections (segments), such that on-ramps join the freeway only at the upstream boundaries of segments, off-ramps diverge from the freeway only at the downstream boundaries of segments, and mainline detectors are located at the downstream boundaries of segments. In addition, the number of lanes, grade, and other geometric characteristics are constant within the segment. Assume that the freeway has m origins, indexed by i=1,..,m, s destinations indexed by j=1,..,s, and n sections, indexed by k=1,..,n. By convention, origin 1 is taken to be the upstream mainline boundary of the original freeway segment, while destination n is taken to be the downstream mainline boundary. Next, define the following variables:

$x_{oi}(t)$ = total remaining vehicles at origin i at time t,

$x_{dj}(t)$ = total vehicles which have exited the segment at destination j by time t,

$x_{kj}(t)$ = vehicles in segment k bound for destination j, at time t,

$y_l(t)$ = total vehicles counted at counter l up to time t.

We will assume that the total number of vehicles in the system is fixed, so that $N=\Sigma_i x_{oi}(t)+\Sigma_k\Sigma_j x_{kj}(t)+\Sigma_j x_{dj}(t)$ is constant at all times t. Let

$$\mathbf{x}(t) = (x_{o1}(t),..,x_{om}(t),x_{11}(t),x_{12}(t),..,x_{ns}(t),x_{d1}(t),..,x_{ds}(t))^T$$

be a column vector containing the various compartment populations, and

$\mathbf{y}(t)=(y_1(t), \ldots, y_p(t))^T$

be a column vector containing the count totals. Letting $\mathbf{e}_g$ denote a column vector with all elements equal to zero except for position g, and letting g, h index arbitrary elements of the vector $\mathbf{x}$, it will be assumed that over a very short time interval of length Δ, transitions of the form

$$\begin{bmatrix} \boldsymbol{x}(t+\Delta) \\ \boldsymbol{y}(t+\Delta) \end{bmatrix} - \begin{bmatrix} \boldsymbol{x}(t) \\ \boldsymbol{y}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_h - \boldsymbol{e}_g \\ \boldsymbol{H}\boldsymbol{e}_g \end{bmatrix}$$

occur with probability $x_g q_{g,h}(\mathbf{x}(t))\Delta + o(\Delta)$. It will further be assumed that transitions with

$$[\mathbf{x}(t+\Delta)^T, \mathbf{y}(t+\Delta)^T]^T - [\mathbf{x}(t)^T, \mathbf{y}(t)^T]^T = \mathbf{0}$$

occur with probability $1-\Sigma_{h\neq g} x_g q_{g,h}(\mathbf{x}(t))\Delta + o(\Delta)$, and all other transitions have a probability which is o(Δ). Note that a $\mathbf{x}(t+\Delta)-\mathbf{x}(t) = \mathbf{e}_h - \mathbf{e}_g$ corresponds to the transition of a vehicle from compartment g to compartment h. By defining **H** such that

$\mathbf{H}_{lg}$ = 1 if counter l registers departure from g

0 otherwise,

$\mathbf{y}(t+\Delta)-\mathbf{y}(t) = \mathbf{H}\mathbf{e}_g$ corresponds to an increment in the counter registering departures from compartment g. The vehicle movements follow a closed, continuous time Markov

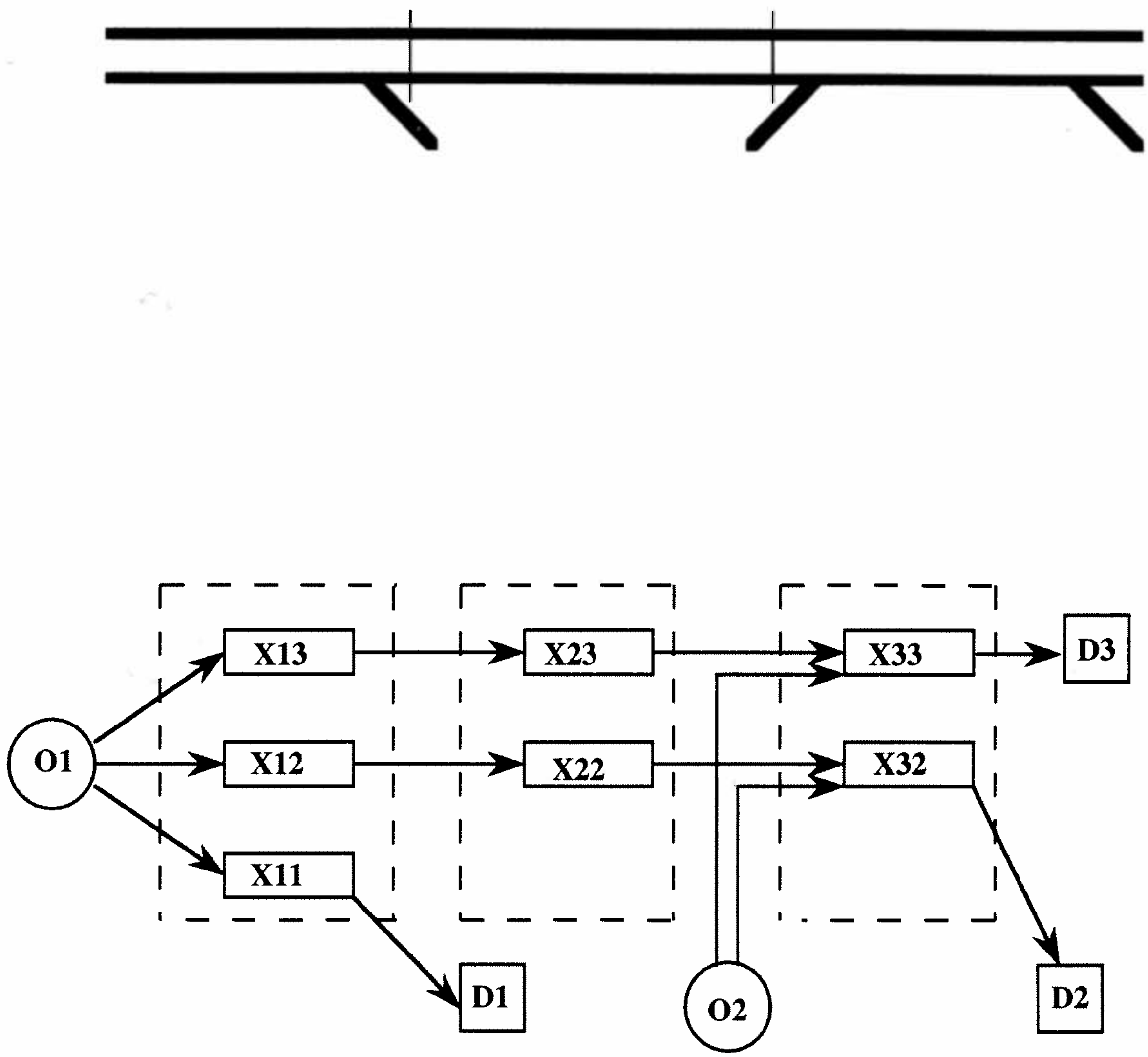


Figure 3.1. Markov Compartment Model of Freeway Traffic Flow

compartment model, (or equivalently, a nonlinear birth and death process), with the state vector augmented to include vehicle counts. In order to facilitate the interpretation of this Markov compartment model of freeway traffic flow, a freeway system with 2-origin and 3-destination is depicted in Figure 3.1. First, traffic arrives at origins and on-ramp traffic is assigned to have destination j with O-D splitting probability b_{ij}. This is equivalent to a birth process. Finally, each destination-specific flow $x_{kj}(t)$ propagates to next segment or exits to off-ramps according to the transition functions $q_{g,h}(\mathbf{x}(t))$. Meanwhile, counter crossings are cumulated in the counter $\mathbf{y}_l(t)$.

A continuous-time discrete-state Markov process can also be defined in terms of particles having exponentially distributed residency times within the compartments. At the end of its residency time a particle is transferred to one of the other compartments with probabilities according to a Markov-chain transition matrix. Therefore, the continuous time Markov compartment model (MARCOM) for freeway traffic flow just described can be expressed as a simple recursive process, well suited for computer simulation (Table 3.1).

Table 3.1. MARCOM Freeway Traffic Simulation Model

Step 0: Given O-D splitting probabilities b_{ij} and destination specific variables $x_{kj}(0)$, let $t=0$, $i=1,...,m$, $k=1,...,n$, $j=1,...,s$.

Step 1: Generate the next arrival time at origin i destined for j, Δ_{ij}, as an exponential outcome with parameter $\lambda_{oi} \cdot b_{ij}$, where λ_{oi} = arrival rate from on-ramp i.

Step 2: Calculate the mainline transition rates $x_{kj}q_{k,h}(\mathbf{x}(t))$, where $q_{k,h}(\mathbf{x}(t))$=mainline transition intensity which will be derived in section 3.3.

Step 3: Generate the next transition time at each compartment k destined for j, Δ_{kj}, as an exponential outcome with parameter $x_{kj}q_{k,h}(\mathbf{x}(t))$.

Step 4: Pick a minimum next arrival time Δ_{min} among (Δ_{ij}, Δ_{kj}).

Step 5: Let $t = t + \Delta_{min}$, update state variable $x_{kj}(t)$ and counter $y_{lg}(t)$:

$x_{kj}(t+\Delta_{min}) = x_{kj}(t) + 1$, if it is a birth compartment;

$x_{kj}(t+\Delta_{min}) = x_{kj}(t) - 1$, if it is a death compartment; and

$y_{lg}(t+\Delta_{min}) = y_{lg}(t) + 1$, if detector l register departures from g.

Step 6: Go to *Step 1*.

3.3 Development of a New Transition Rate Function for Freeway Traffic Flow

The MARkov COMpartment model (MARCOM) for freeway traffic flow not only models the random departure and distribution of traffic but also the propagation of traffic from on-ramps to off-ramps, in a way that preserves the random allocation of arriving vehicles to off-ramps. To implement MARCOM, transition intensity functions $q_{g,h}(\mathbf{x}(t))$ are necessary that reflects traffic condition in a natural way. For the transitions from the origin sources to mainline sections, it is reasonable to use transition intensities of the form $q_{oi}b_{ij}$, where

q_{oi} = constant arrival intensity from on-ramp i,

b_{ij} = probability a vehicle is destined for off-ramp j, given it arrives at on-ramp i.

If the origin populations $x_{oi}(t)$ are large enough so that the total arrivals during the time period of interest is a small proportion of the original total, we can take the quantity $x_{oi}(t)q_{oi}$ as being a constant λ_i, giving Poisson arrival rates at the freeway origins (Whittle, 1986).

To develop functions giving the transition rates within the mainline sections, assume that at time t, the vehicles in section k have speeds assigned as independent, identically distributed random outcomes from a common speed distribution. It is further assumed that distances from the downstream boundary of section k are assigned as independent, identically distributed outcomes from a uniform random variable with probability density $1/L_k$, where L_k is the length of section k. Let s_{kl} denote the location of vehicle l in section k, u_{kl} denote the speed of vehicle l in k, and let $f_k(u)$ denote the probability density function for the speeds. Now clearly, vehicle l will exit section k only if

$$s_{kl} < u_{kl} \cdot \Delta \tag{3.1}$$

so that

$$p_k = Prob[vehicle\ l\ exits\ section\ k] = \int_0^{L_k} \int_0^{\Delta u_{kl}} \frac{f_k(u)}{L_K} duds \tag{3.2}$$

it follows that

$$p_k = \bar{u}_k \cdot \Delta / L_k \tag{3.3}$$

where $\bar{u}_k$ denotes the space mean speed in section k. The formulation can then be closed by requiring the space mean speeds $\bar{u}_k$ to depend directly on $\mathbf{x}(t)$ via a form of the equilibrium speed-density relations of traffic flow theory, giving a version of the simple continuum model. When $\bar{u}_k \leq L_k/\Delta$ the Markovian property is preserved and the flow rate is expressed as the product of density and space-mean speed. As formulated though, this model will show a tendency to "lock up" when the densities in a section rise above the critical density, because the exit probabilities go to zero as $\bar{u}_k$ goes to zero (Ross, 1988). Although Markov traffic models can be extended to produce analogues of high-order continuum models (Cremer and May, 1985; Davis, 1993), a simpler solution is to employ the device originally due to Szeto and Gazis (1972), and allow the flow across the boundary of two sections to depend on both the upstream and downstream densities. It is assumed that the exiting probability from section k is the product of a passing probability for section k, given by equation (3.3), and the probability an exiting vehicle is not blocked by one in the downstream segment.

$$[\text{Exiting prob.}]_k = [\text{passing prob.}]_k \cdot [\text{non-blocking prob.}]_{k+1}$$

In this study, the Bell-shaped speed-density function, parameterized by free-flow speed and critical density, is used in the passing probability function, and a function parameterized by jam density and a value r is used as a non-blocking probability. The continuous two-dimensional exiting probability function then takes the form

$$\begin{aligned} p(d_k,d_{k+1}) &= u_f e^{-\frac{1}{2}\left(\frac{d_k}{d_c}\right)^2}\left[1-\left(\frac{d_{k+1}}{d_{jam}}\right)^r\right]\frac{\Delta}{L_k}, \quad d_k \le d_c \\ &= \frac{Q_0}{d_c}\left[1-\left(\frac{d_{k+1}}{d_{jam}}\right)^r\right]\frac{\Delta}{L_k}, \quad d_k > d_c \end{aligned} \tag{3.4}$$

where u_f is a free-flow speed, d_c is a critical density, Q_0 is a capacity flow, d_{jam} is a jam density, and $d_k=x_k/(N_kL_k)$ is the traffic density in section k.

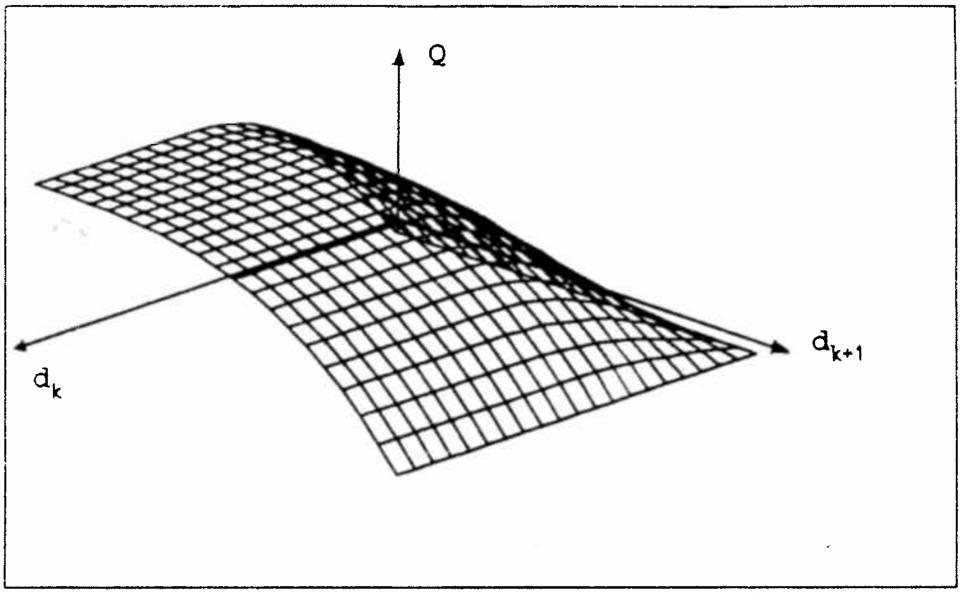
From the relationship of $p_k = \bar{u}_k \cdot \Delta / L_k = (Q_k \cdot \Delta)/(d_k \cdot L_k)$, the form of the two-dimensional transition rate is then

$$
\begin{aligned}
Q(d_k,d_{k+1}) &= d_k u_f e^{-\frac{1}{2}\left(\frac{d_k}{d_c}\right)^2}\left[1-\left[\frac{d_{k+1}}{d_{jam}}\right]^r\right], \quad d_k \le d_c \\
&= Q_0\left[1-\left[\frac{d_{k+1}}{d_{jam}}\right]^r\right], \quad d_k > d_c
\end{aligned}
\tag{3.5}
$$

Table 3.2 shows the flow condition across section boundaries as a combination of upstream and downstream traffic conditions. For a constant downstream density d_{k+1}, equation (3.5) gives an increasing cross-boundary flow as the upstream density d_k increases, up to the point where d_k equals the critical density. The cross boundary flow then remains constant, thus modeling the upstream section as (approximately) an oversaturated finite-server queue. As the downstream density d_{k+1} approaches the jam density d_{jam}, the cross boundary flow goes to zero, with the sensitivity of this effect being governed by the exponent r. It should be noted that equation (3.5) gives an equilibrium flow-density relationship when the upstream density d_k is equal to the downstream density d_{k+1}. Figure 3.2 displays a plot of equation (3.5) as calibrated for an actual segment of freeway.

Table 3.2. Flow Conditions Across Section Boundaries

		Upstream Density	
		Low	High
Downstream Density	Low	Homogeneous	Shockwave
	High	Queuing	Homogeneous



Q	:	flow (vph)
d_k	:	density at section k
d_{k+1}	:	density at section k+1

Figure 3.2. Two-Dimensional Flow vs. Density Relationship

3.4 Testing of Markov Compartment Model of Freeway Traffic Flow (MARCOM)

As stated earlier, one of the ultimate objectives of this research is to estimate destination-specific traffic densities on freeways. The overall solution strategy is to first describe a Markovian traffic model, then approximate the Markovian model with a time-varying linear stochastic model, and finally apply the theory of Kalman filtering in order to estimate the destination-specific densities. Three questions then arise concerning this approach: (1) how reasonable is the underlying Markovian traffic model, (2) how accurate is the approximation and (3) how well does the resulting Kalman filter perform? The first question becomes important since destination-specific densities are almost impossible to observe in practice, so the accuracy of the Kalman Filter will have to be assessed using simulated data. Although the basic idea behind equation (3.5) is not new, the traffic flow model which results is still somewhat novel, and it was first desired to see if a model based on equation (3.5) could produce reasonable behavior at bottleneck. To this end, a computer program implementing the algorithm in Table 3.1, called MARCOM, was written and used to generate flows for the given freeway section. MARCOM can be used to generate a series of simulated on-ramp volumes, distributing these on-ramp volumes to off-ramps using the above birth-death model and then propagate these volumes down the freeway and out the off-ramps in a manner consistent with the embedded traffic flow model. In this section, a preliminary testing of the Markov traffic flow model was performed using both hypothetical data and actual freeway data. First, in order to examine whether the proposed models behave reasonably for a wide range of hypothetical flow situations, a qualitative evaluation of the model was done. Next, a comparison with field data was performed.

3.3.1 Qualitative Testing of MARCOM

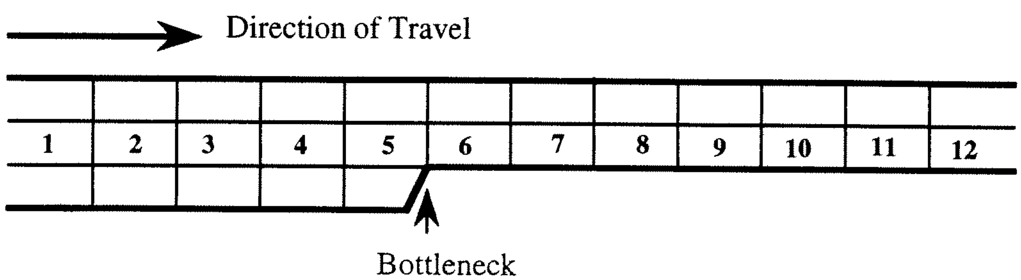
Behavior of MARCOM at a Lane Drop Bottleneck

Figure 3.3 shows a hypothetical 5.6 km (3.5 mile) freeway section where the number of lanes is reduced from three to two behind the fifth of twelve subsections (the lengths of the subsections were uniformly chosen to be 457.2 m (1,500 feet)). A 60 minute simulation was started with demand of 3,000 vph and then increased to 4,800 vph, which exceeds the capacity of the two-lane section by approximately 20 percent, and finally decreased to 1,200 vph. One can analyze the behavior of traffic near this bottleneck by plotting the volume and density trajectories of the test section at 5 minute intervals. As illustrated in Figures 3.4 and 3.5, MARCOM provides a reasonable description of queue buildup and dissipation at the bottleneck section in that:

(1) Congestion starts above the bottleneck and moves in the upstream direction, while the density within the bottleneck remains around the critical density.

(2) The volumes in the bottleneck are limited to the capacity during congestion building and dissipation. Further, vehicles do not pass upstream subsections at a rate greater than the capacity of the bottleneck during congestion.

(a) Geometrics



(b) Demand Pattern

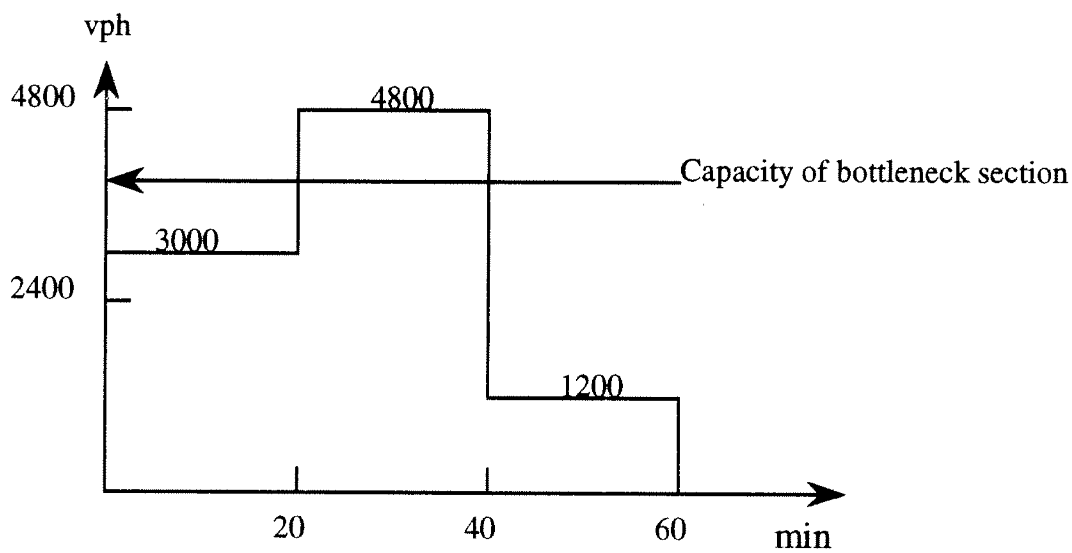


Figure 3.3. Geometrics and Demand Pattern of Freeway Section with a Bottleneck

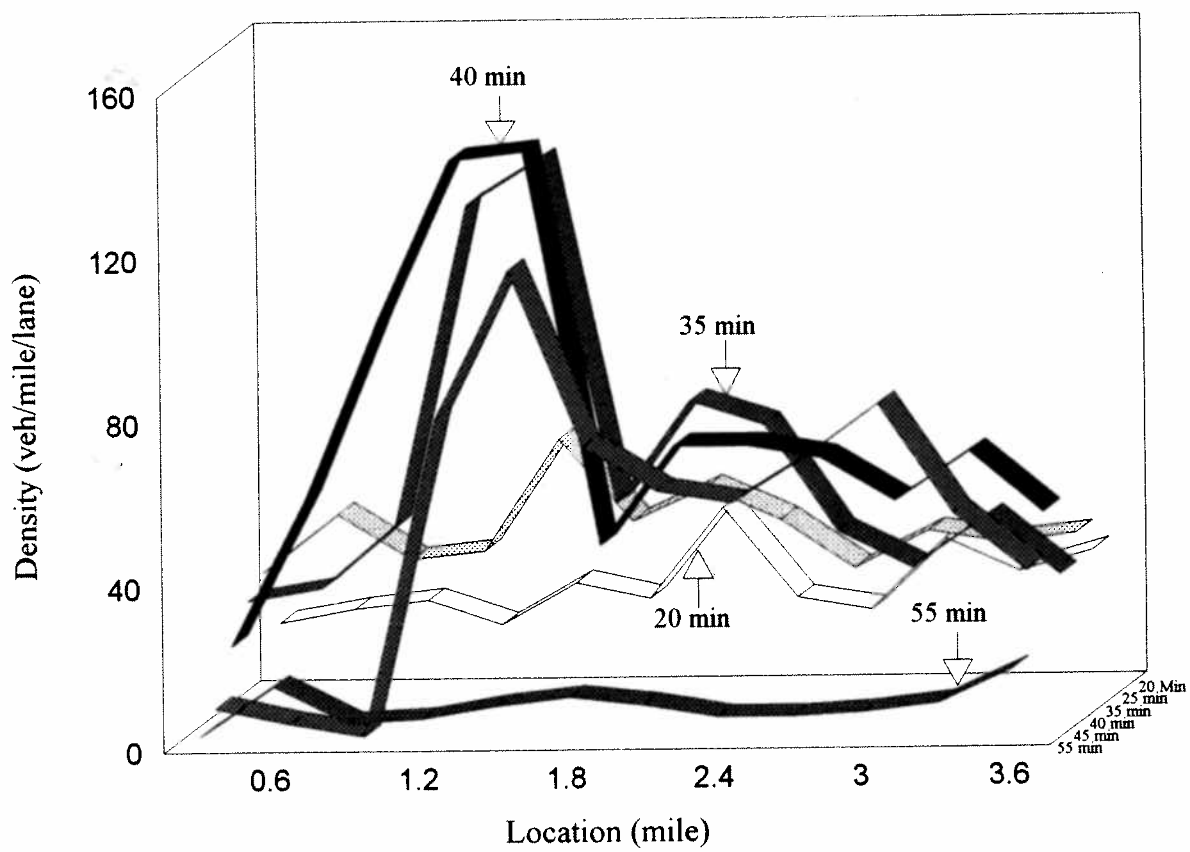


Figure 3.4. Density Trajectories at a Bottleneck Section

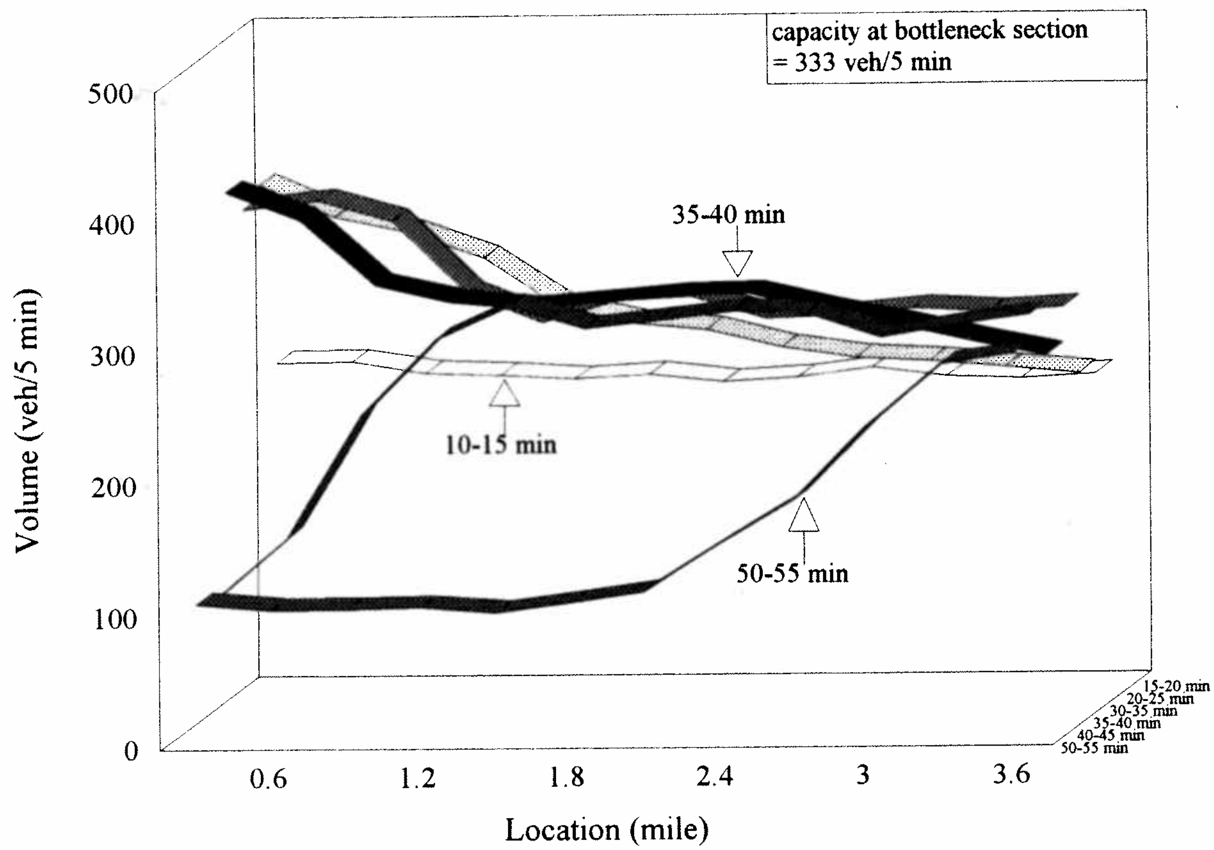


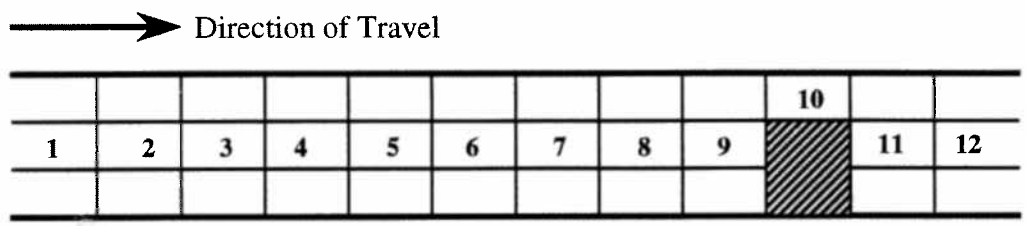
Table 3.5. Volume Trajectories at a Bottleneck Section

Behavior of MARCOM at an Incident Section

To test the ability of MARCOM on more complex traffic flow behavior, another test was performed for a hypothetical geometry. Figure 3.6 shows the geometrics and demand-capacity variation of a 7.24 km (4.5 mile) freeway section with part of the 10th subsection is closed because of an incident (the lengths of the subsections were uniformly chosen as 609.6 m (2,000 feet)). When the reduced capacity at the incident area is less than the through demand, congestion may develop and propagate upstream. A 150 minute simulation with constant demand of 4,200 vph was done with traffic flow parameters of u_f=104.6 kph (65.0 mph), d_c=37.8 veh/km/lane (60.8 veh/mile/lane), d_{jam}=99.4 veh/km/lane (160 veh/mile/lane). In order to create an incident at the 10th subsection, two lanes were closed during the interval 30-60 minutes and one lane was closed during the interval 60-120 minutes. Figure 3.7 shows the simulated density trajectories for selected subsections during the entire simulation period, at one minute intervals. Figure 3.8 further describes the evolution of the flow-density relationship at subsection eight, which is upstream of the incident subsection. Again, MARCOM reasonably describes the expected behavior of traffic near the incident area in that:

(1) During time 30-60 minutes, congestion starts above subsection 10 and propagates in the upstream (subsections 8 and 4) direction, while the density values within incident area (subsection 10) remain uniform at the critical density. As congestion moves into the upstream subsections from the incident area, the density values in these subsections rapidly increase until they reach the right-hand side value of the basic flow-density

(a) Geometry



Length=12*2,000 ft =24,000 ft (4.54 mile)

Capacity=6,150 vph

(b) Demand and Capacity of Incident Section

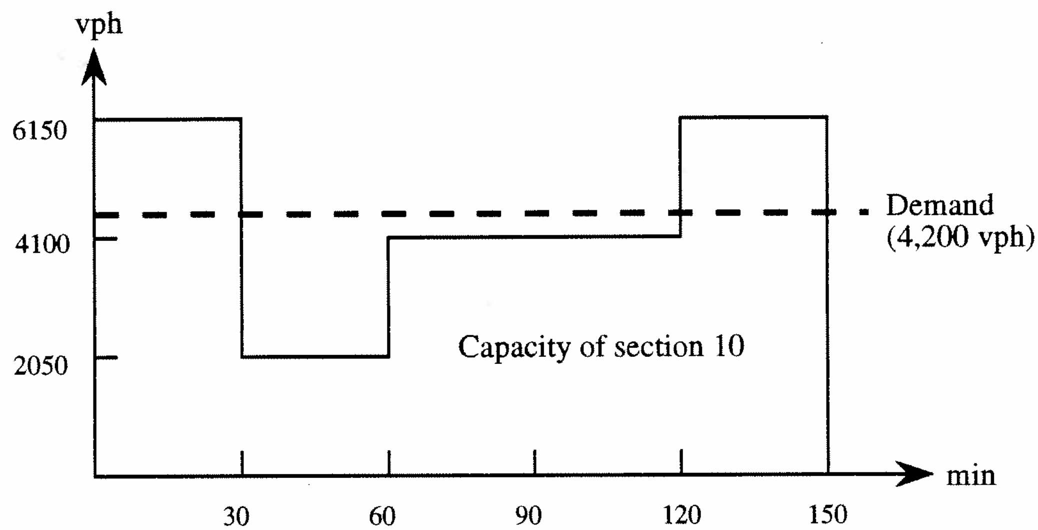


Figure 3.6. Geometrics and Demand Pattern of Freeway Section with an Incident

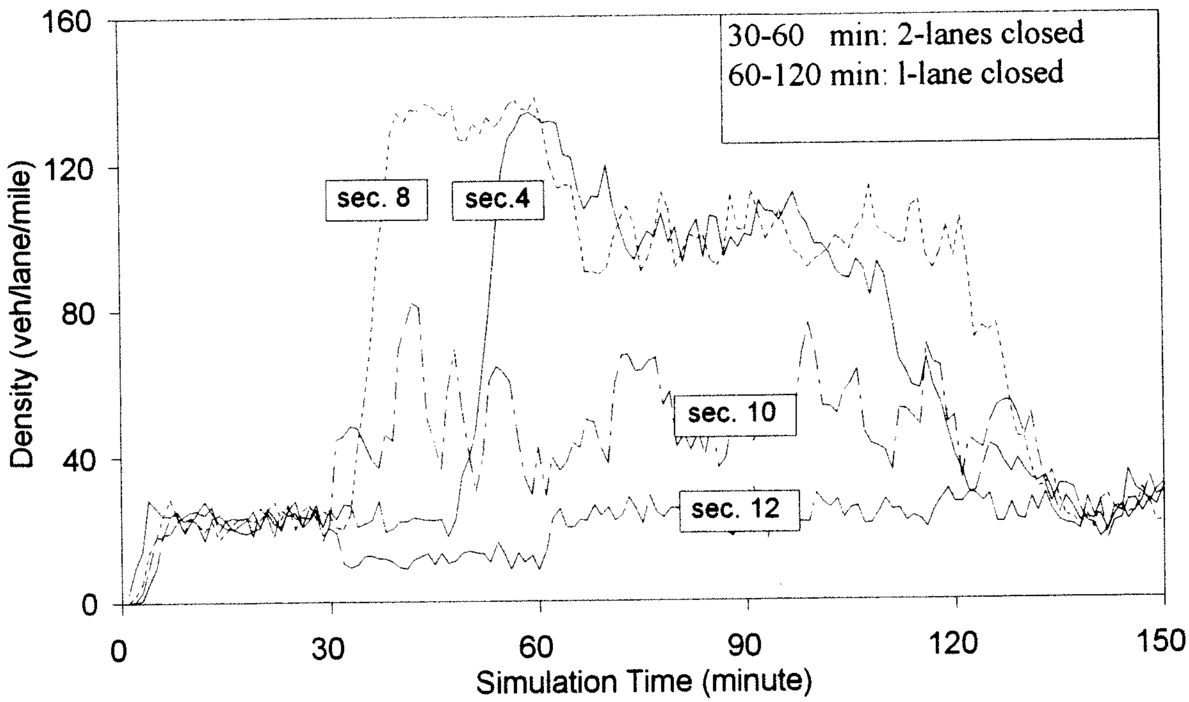


Figure 3.7. Density Distribution During Incident Simulation

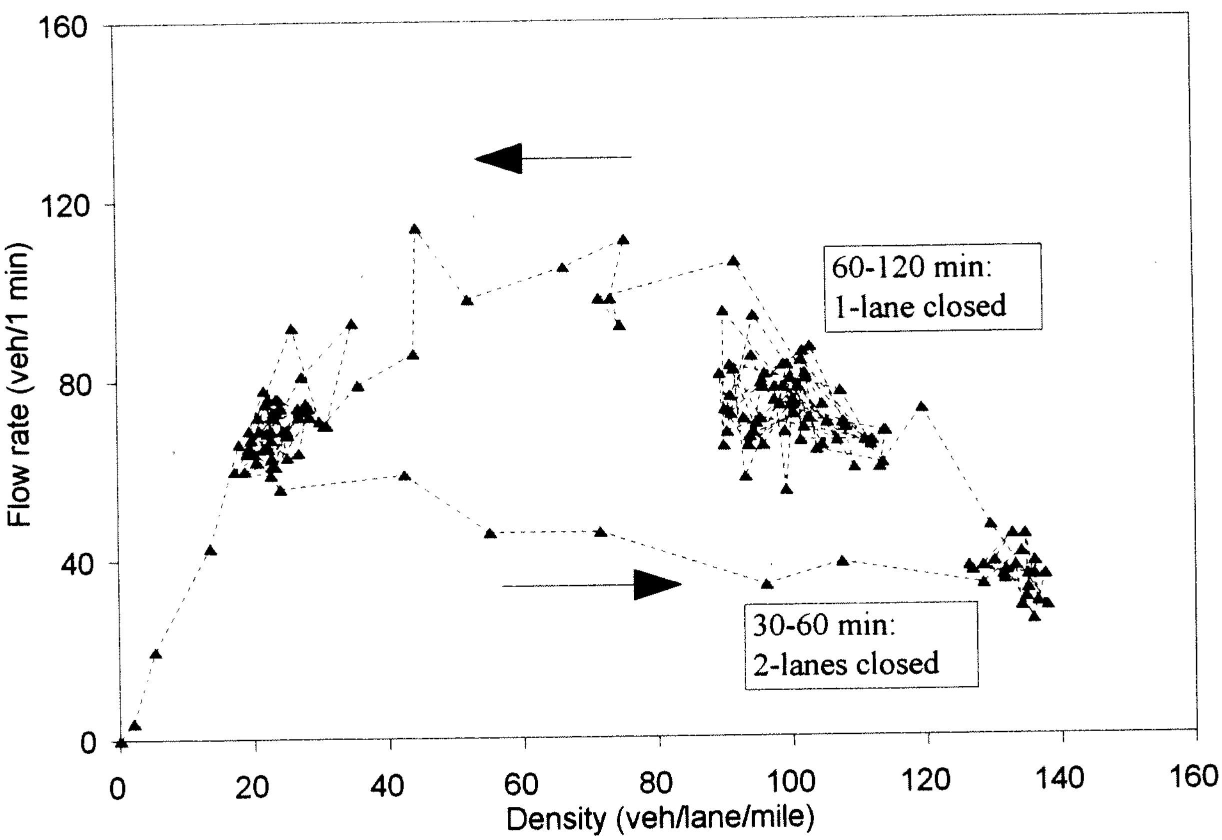


Figure 3.8 Flow vs. Density Relationship During Incident Simulation

curve, where flow corresponds to the capacity of incident area. Further, the density values remain fairly uniform. However, one can see that subsection 12, which is downstream of incident area, exhibits no congestion during the incident period. When the incident was removed, the congested queue dissipated reasonably.

(2) Flow-density plots at subsection eight remained in the uncongested region until it was affected by the incident downstream (Figure 3.8). Once this subsection was affected by the incident, the flow gradually reduced to the capacity of subsection 10 while density values increased sharply. This congestion building process was illustrated by the lower flow-density trajectories in Figure 3.8. Once the flow reached the capacity of the incident area it stayed there until the capacity of downstream recovered. The congestion dissipating process after the removal of incident is illustrated by upper flow-density trajectories in Figure 3.8.

(3) It should be noted that an apparent discontinuity around the critical density (37.8 veh/lane/km (60.8 veh/lane/mile)) was due to rapid queue propagation upstream of the incident section. This corresponds to field observation (Banks, 1991; Hall et. al., 1993).

While the idea of the proposed 2-dimensional flow-density relationship adopted in MARCOM is quite simple, it allows modeling complex traffic patterns such as mainline congestion propagation due to incidents and/or geometric bottlenencks. It should be pointed out that such an effective description of complex traffic patterns is possible without any special modeling or boundary treatment. Further, the ability to describe O-D routing in MARCOM could improve the description of traffic in the ramp areas where a significant amount of lane-changing and/or spillback occurs. Since efficient modeling of those problems actually need destination information, traffic dynamics, or route choice, the basic idea in MARCOM could be a valuable tool in improving freeway traffic simulation tools.

3.3.2 Field Testing of MARCOM

Field implementation of MARCOM was accomplished by using data collected from Interstate highway 35W in Minneapolis. The following section describes the process of calibration and running of the MARCOM simulation program.

Example Network and Data

Figure 3.9 shows a 4.0 km (2.5 mile) long, seven-origin, four-destination segment of northbound Interstate highway I-35W. This freeway section is separated into 10 subsections. Five-minute cumulative volume and lane occupancy measurements during a three-hour morning peak period (6:00-9:00 a.m.) for mainline stations, on-ramp and off-ramp stations are obtained from the Minnesota Department of Transportation (MNDOT).

To run the stochastic simulation model MARCOM, it is necessary to know the on-ramp arrival rates λ_i, the origin-destination splitting probabilities b_{ij}, and the parameters governing the flow-density relation (3.5). The arrival rates can simply be estimated as those values that reproduce the corresponding five-minute arrival counts, allowing the arrival rates to vary for each five-minute interval. To determine the parameters for equation (3.5), the lane occupancy measurements were converted to

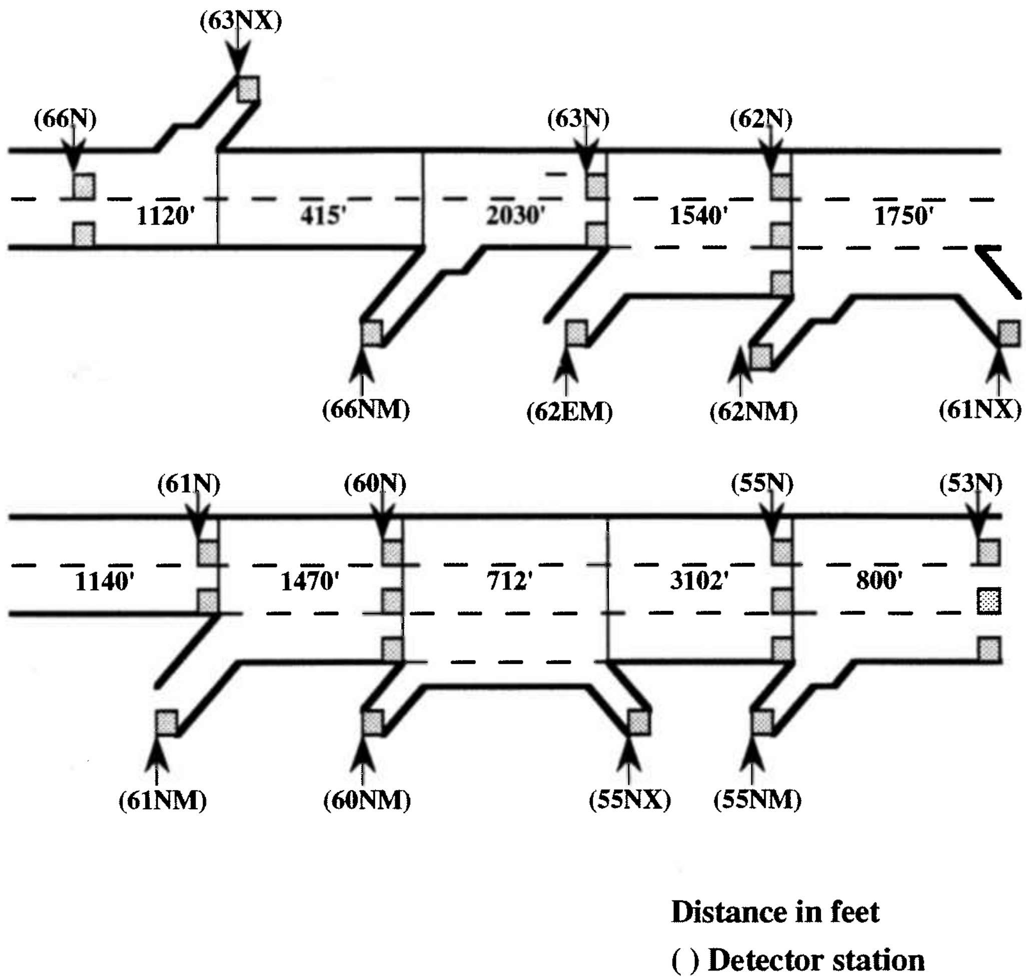


Figure 3.9. Geometrics of 7-Origin, 4-Destination Section (I-35W, Northbound)

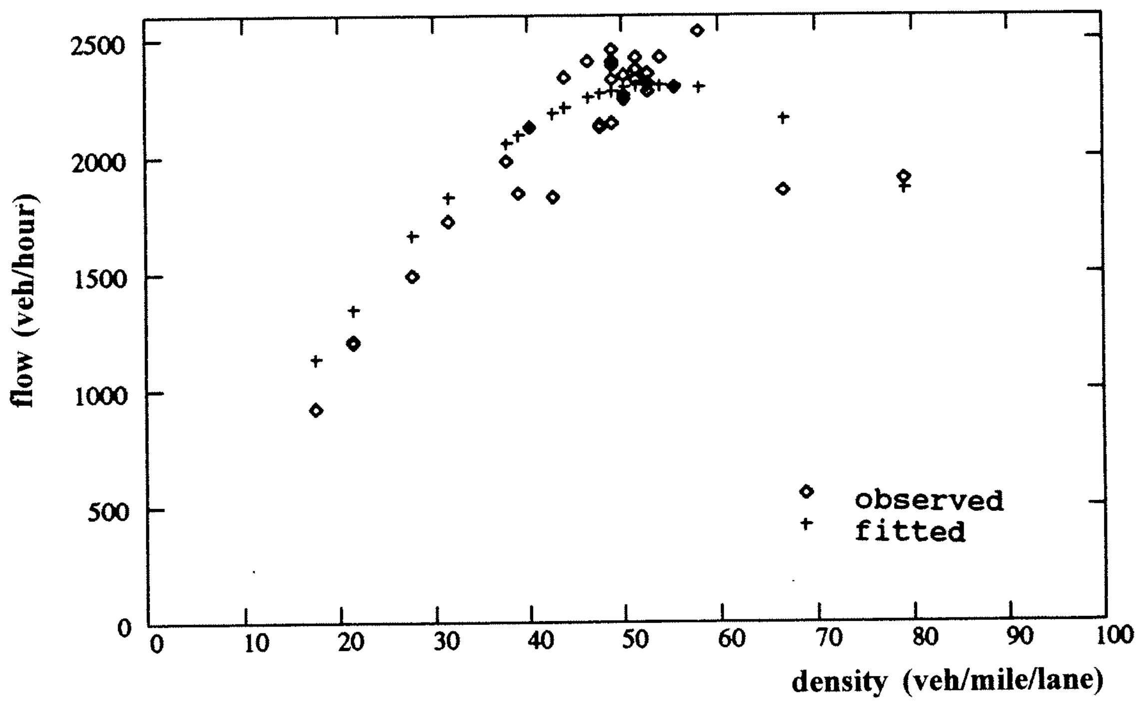


Figure 3.10. Fitted and Observed Steady State Flow vs. Density

approximate density values, and then the parameters u_f, d_c, d_{jam} and r in equation (3.5) were estimated using nonlinear least-squares by setting d_k=d_{k+1}, corresponding to the notion of approximate homogeneous flow. Figure 3.10 shows the observed and fitted flow-density curve obtained for the estimates u_0=107.2 kph (66.6 mph), d_c=40.1 veh/km/lane (64.5 veh/mile/lane), d_{jam}=74.6 veh/km/lane (120 veh/km/lane) and r=3. Finally, the O-D splitting probabilities b_{ij} were estimated by first using the estimated traffic flow parameters to numerically solve the mean value equation of MARCOM for a trial sets of b_{ij} values. For a given set of origin counts this produced estimated destination counts, and those b_{ij} values which minimized the sum of squared errors between forecasted and actual counts were obtained by embedding this routine in a nonlinear optimization program. These estimates were then used as inputs to MARCOM which simulated the Markov compartment process to generate simulated traffic counts for various time intervals, as well as destination-specific section populations, $x_{kj}(t)$.

The resulting comparisons of 5-minute volumes generally showed good agreement between simulated and actual data. As indicated by the error measures in Table 3.3, MARCOM provided a reasonable reproduction of traffic flows.

Table 3.3. Mean Error of the Simulated Volume Results (6:00-9:00 A.M.)

Detector Station	63N	62N	61N	55N	53N
MAPD[a]	2.2	2.0	2.0	2.1	2.2
MAE[b]	13	21	18	27	28

Note:
a: Mean absolute percentage difference between observed and simulated volumes
$= [\sum_{t-1,N} 100 \cdot | y(t) - \hat{y}(t) | / y(t)] / N$, where N = 36

b: Mean absolute error between observed and simulated volumes (veh/5 min)
$= \sum_{t-1,N} | y(t) - \hat{y}(t) | / N$, where N = 36

4. LARGE POPULATION APPROXIMATION AND FORMULATION OF KALMAN FILTER FOR DESTINATION-SPECIFIC DENSITY ESTIMATION

4.1 Introduction

As stated earlier a chief goal of this research is to produce estimates of the unobserved destination-specific segment populations, $x_{kj}(t)$, using vehicle counts $\mathbf{y}(t)$ which have been aggregated across the origin-destination subflows. This can be done by the application of stochastic filtering theory, given the formulation of a state space model describing the evolution of the system of interest. A state space model consists of two components, the first being a state dynamic equation describing how the system's state variables change over time, the second being an observation equation describing how observable quantities are related to the state variables. In a manner analogous to the deterministic models used in earlier research (Michalopoulos et al., 1984;1991; Cremer and May, 1985), a stochastic freeway traffic flow model was developed in Chapter Three. This model, which is based on a Markovian network model, is a stochastic state space model. The problem of estimating the unobserved destination-specific populations $x_{kj}(t)$ using the counts at time t can now be solved using filtering theory. However, the resulting filtering problem will be nonlinear and often intractably large for realistic networks. Therefore, the original nonlinear stochastic traffic flow model will be approximated by the sum of a nonlinear deterministic process and a linear, time-varying Gaussian stochastic process, via the method of Large and a linear, time-varying Gaussian stochastic process, via the method of Large Population Approximation (LPA). Using this approximation, a Kalman filter which tracks a freeway section's density broken down by destination, using traffic counts, will be derived in section 4.3.

4.2 Large Population Approximation of a Stochastic Traffic Flow Model

Approximating one Markov process with another has a rather long history including Conlisk's (1976) introduction of "interactive Markov chains" in discrete time and his study on such processes using deterministic approximations. An "interactive Markov Chains" is a population process in which the individual transition probabilities depend on the population's distribution over the various states. For this class of the models, Lehoczky (1980) has shown that if the transition probabilities are extensive, the interactive Markov chain converge weakly to a diffusion process, with specified drift and scale terms, as the population size becomes infinite, at least over a finite time horizon. Lehoczky in turn draws from work by Kurtz (1978) on approximating populations process with diffusions. Ingenbleek and Lefevre (1985) have used approximations similar to the ones developed below to study a discrete time epidemic model. This approach was also used by Karmeshu and Pathria (1981) on a proposed Markov compartment model for highway traffic to provide an asymptotic analysis using a diffusion approximation. The assumption of a large population seems reasonable for traffic systems, and as mentioned earlier, MARCOM is a particular form of an distribution of the vehicles over the various sections. Therefore, given reasonable conditions on the functions $q_{g,h}(\mathbf{x}(t))$, Lehoczky's (1980) argument can be adapted to this case in order to show that as N, the total vehicles in the system, becomes large, the random vectors $[\mathbf{x}(t)^T, \mathbf{y}(t)^T]^T$ can be approximated by the sum of a nonlinear deterministic process and a linear, time-varying Gaussian stochastic process

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \approx \begin{bmatrix} \bar{x}(t) \\ \bar{y}(t) \end{bmatrix} + z(t) \tag{4.1}$$

where $[\bar{x}(t)^T, \bar{y}(t)^T]^T$ are the deterministic, mean values resulting from the deterministic approximation (Lehoczky, 1980) and satisfy the ordinary differential equation

$$\begin{aligned} \frac{d\bar{x}_g(t)}{dt} &= \sum_h \bar{x}_h(t) q_{h,g}(\bar{x}(t)) = f_{1,g}(\bar{x}(t)) \\ \frac{d\bar{y}_l(t)}{dt} &= \sum_g H_{lg} \bar{x}_g(t) \sum_{u \neq g} q_{gu}(\bar{x}(t)) = f_{2,l}(\bar{x}(t)) \end{aligned} \tag{4.2}$$

$\mathbf{z}(t)$ is a zero mean Gaussian random vector with covariance matrix $\mathbf{P}(t)$. It follows then that $[(x(t)-\bar{x}(t))^T, (y(t)-\bar{y}(t))^T]^T$ will have a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{P}(t)$, and $\mathbf{P}(t)$ is the unique symmetric nonnegative definite solution of the matrix Ricatti equation

$$\frac{d\boldsymbol{P}(t)}{dt} = \boldsymbol{F}(\bar{\boldsymbol{x}}(\boldsymbol{t}))\boldsymbol{P}(t) + \boldsymbol{P}(t)\boldsymbol{F}(\bar{\boldsymbol{x}}(\boldsymbol{t}))^{T} + \boldsymbol{G}(\bar{\boldsymbol{x}}(\boldsymbol{t})) \tag{4.3}$$

Here $\boldsymbol{F}(\bar{\boldsymbol{x}}(t))$ denotes the Jacobian matrix of the right-hand side of equation (4.2) with respect to $\bar{\boldsymbol{x}}(t)$, and is given by

$$\boldsymbol{F}(\bar{\boldsymbol{x}}(t)) = \begin{bmatrix} \dfrac{\partial \boldsymbol{f}_1}{\partial \bar{\boldsymbol{x}}(t)} & \boldsymbol{0} \\ \dfrac{\partial \boldsymbol{f}_2}{\partial \bar{\boldsymbol{x}}(t)} & \boldsymbol{0} \end{bmatrix}_{\bar{x}(t)} \tag{4.4}$$

where $\boldsymbol{f}_1$ and $\boldsymbol{f}_2$ are vectors containing the functions defined in (4.2). $\boldsymbol{G}(\bar{\boldsymbol{x}}(t))$ is a covariance matrix of a noise process with

$$\boldsymbol{G}(\bar{\boldsymbol{x}}(t)) = \begin{bmatrix} \boldsymbol{G}_1(\bar{\boldsymbol{x}}(t)) & \boldsymbol{G}_2(\bar{\boldsymbol{x}}(t)) \\ \boldsymbol{G}_2^T(\bar{\boldsymbol{x}}(t)) & \boldsymbol{G}_4(\bar{\boldsymbol{x}}(t)) \end{bmatrix} \tag{4.5}$$

and $G_k(\bar{x}(t))$ has the elements of

$$
\begin{aligned}
G_{1,ii}(\bar{x}(t)) &= \sum_{j \neq i} q_{ij}(\bar{x}(t)) - x_i(t) q_{ij}(\bar{x}(t)) \ for\ i=j, \\
G_{1,ij}(\bar{x}(t)) &= -[x_i(t) q_{ij}(\bar{x}(t)) + x_j(t) q_{ji}(\bar{x}(t))] \ for\ i \neq j, \\
G_{2,ij}(\bar{x}(t)) &= [x_j q_{ij}(\bar{x}(t))] \boldsymbol{H}^T, \\
G_{4,ij}(\bar{x}(t)) &= \boldsymbol{H}[diag(x_i \textstyle\sum_{i \neq j} q_{ij}(\bar{x}(t))] \boldsymbol{H}^T
\end{aligned}
$$

The transition intensities, $q_{ij}(\bar{x}(t))$ were defined in Chapter 3.

In summary, it has been argued that the Markov process $[\boldsymbol{x}(t)^T, \boldsymbol{y}(t)^T]^T$ for any t can be approximated by the sum of a nonlinear deterministic process (drift term) and a linear, time varying Gaussian stochastic process (scale term)

$$
\begin{bmatrix} \boldsymbol{x}(t) \\ \boldsymbol{y}(t) \end{bmatrix} \approx N\left(\begin{bmatrix} \bar{\boldsymbol{x}}(t) \\ \bar{\boldsymbol{y}}(t) \end{bmatrix}, \boldsymbol{P}(\bar{\boldsymbol{x}}(t)) \right) \tag{4.6}
$$

The probability distributions of $[\boldsymbol{x}(t)^T, \boldsymbol{y}(t)^T]^T$ are therefore Gaussian with mean and variance given by equations (4.2) and (4.3). In passing, we note that the distributional results will allow for the construction of a likelihood function from interactive Markov process data, so these results will be important for both state and parameter estimation.

4.3 Formulation of Kalman Filter for Destination-Specific Density Estimation

As before, imagine that a section of freeway has been divided into subsections, such that on-ramps join the freeway only at the upstream boundaries of the section, off-ramps diverge from the freeway only at the downstream boundaries of section, and mainline detectors are located at the downstream boundaries of sections. The task is estimation of the number of vehicles in a freeway subsection, broken down by the destination, $x_{kj}(t)$. Since the only observations are the counts of arrival vehicles $\mathbf{q}(t)$ and passing vehicles $\mathbf{y}(t)$, one needs to estimate the destination-specific density information from these observations. Therefore, the problem can be formally stated as: Given a set of traffic flow measurements $\{\mathbf{q}(t)$ and $\mathbf{y}(t)\}$, for t=1,...,N characterized by the measurements model, find the linear minimum (error) variance estimate of the destination-specific traffic densities $\mathbf{x}(t)$ characterized by the state-space model. That is, find the best estimate $\mathbf{x}(t)$ given the measurement data up to time t, $\mathbf{Y}(t) := \{\mathbf{q}(1),\ldots,\mathbf{q}(t),\mathbf{y}(1),\ldots,\mathbf{y}(t)\}$.

The estimation procedure is based on Kalman filter that first requires the formulation of a state-space model describing the evolution of the system of interest. Coupled with an observation equation of the form

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t), \tag{4.7}$$

the system equation in Table 4.1 define a time-varying state-space model with normally distributed state noise. The conditional means and conditional variances are available by solving the mean value equation (4.2) and the Ricatti equation (4.3) for a given time. Since the approximating random process is a Gaussian process, complete estimation requires a knowledge of only these first and second moments. Therefore optimal estimation is possible and no nonlinear filter is superior to the best linear filter in the mean-square error sense, as long as the approximation is valid. The procedure for implementing the Kalman filter for given values of the model parameters can be viewed as a predictor-corrector algorithm, as in standard numerical integration.

Given an initial estimates $\hat{x}(0 \mid 0) = \bar{x}(0)$, $\hat{y}(0) = \mathbf{0}$, $P(0 \mid 0) = P(0)$, we are now in the prediction phase of the algorithm. The differential equations (4.2) and (4.3) are then solved numerically to give predicted compartment totals, $\bar{x}(t)$ and cumulative counts, $\bar{y}(t)$, along with variances and covariances, $\mathbf{P}(t)$, for any future time t. Once a measurement becomes available at time t_k, we then calculate the Kalman gain $\mathbf{K}(t_k)$ and the innovation covariance $\mathbf{HP}(t_k|t_{k-1})\mathbf{H}^T$. We then determine the innovation $e(t_k) = y(t_k) - \bar{y}(t_k)$, and a measurement update of compartment totals and their covariance terms is performed. Here we update the state using the new information in the most recent measurement. The old, or predicted state $\hat{x}(t_k \mid t_{k-1})$, is used to form the filtered state estimate $\hat{x}(t_k \mid t_k)$. The innovation is weighted by the gain $\mathbf{K}(t_k)$ to correct the old state estimate $\hat{x}(t_k \mid t_{k-1})$ and the associated error covariance $\mathbf{P}(t_k|t_{k-1})$. The differential equations (4.2) and (4.3) can then be restarted with $\bar{x}(0) = \hat{x}(t_k \mid t_k)$, $\mathbf{P}(0)=\mathbf{P}(t_k|t_k)$ and $\bar{y}(0) = \mathbf{0}$, and the recursion continued until the next count becomes available. The time and measurement update equations are collected for reference in Table 4.1.

Table 4.1. Kalman Filter for Discrete-Measurements of Continuous-Time Systems

System Model:

$$\begin{bmatrix} d(\boldsymbol{x}(t)-\bar{\boldsymbol{x}}(t)) \\ d(y(t)-\bar{y}(t)) \end{bmatrix} = \boldsymbol{F}(\bar{\boldsymbol{x}}(t)) \begin{bmatrix} (\boldsymbol{x}(t)-\bar{\boldsymbol{x}}(t))dt \\ (y(t)-\bar{y}(t))dt \end{bmatrix} + \boldsymbol{G}(\bar{\boldsymbol{x}}(t))^{\frac{1}{2}} d\boldsymbol{B}(t)$$

Measurement Model:

$$z(t_k) = \boldsymbol{H} \begin{bmatrix} \boldsymbol{x}(t_k) \\ y(t_k) \end{bmatrix} + \boldsymbol{\nu}(t_k) = [\boldsymbol{0},\boldsymbol{I}] \begin{bmatrix} \boldsymbol{x}(t_k) \\ y(t_k) \end{bmatrix} + \boldsymbol{\nu}(t_k), \quad \boldsymbol{\nu}(t_k) \sim N(\boldsymbol{0},\boldsymbol{R}(t_k))$$

Initial conditions:

$$\hat{\boldsymbol{x}}(0 \mid 0) = \bar{\boldsymbol{x}}(0),\ \hat{y}(0 \mid 0) = \boldsymbol{0},\ \boldsymbol{P}(0 \mid 0) = \boldsymbol{P}(0)$$

Time update:

-State Estimate Propagation

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} f_1(\bar{\boldsymbol{x}}(t)) \\ f_2(\bar{\boldsymbol{x}}(t)) \end{bmatrix}$$

where

$$\frac{d\bar{x}_g(t)}{dt} = f_{1,g}(\bar{\boldsymbol{x}}(t)) = \sum_h \bar{x}_h(t) q_{h,g}(\bar{\boldsymbol{x}}(t))$$

$$\frac{d\bar{y}_l(t)}{dt} = f_{2,l}(\bar{\boldsymbol{x}}(t)) = \sum_g H_{lg} \bar{x}_g(t) \sum_{u \neq g} q_{g,u}(\bar{\boldsymbol{x}}(t))$$

-Error Covariance Propagation

$$\frac{d\boldsymbol{P}(t)}{dt} = \boldsymbol{F}(\bar{\boldsymbol{x}}(t))\boldsymbol{P}(t) + \boldsymbol{P}(t)\boldsymbol{F}(\bar{\boldsymbol{x}}(t))^T + \boldsymbol{G}(\bar{\boldsymbol{x}}(t))$$

Measurement Update at Time t_k:

-Gain Matrix

$$\boldsymbol{K}(t_k) = \boldsymbol{P}(t_k \mid t_{k-1})\boldsymbol{H}^T[\boldsymbol{H}\boldsymbol{P}(t_k \mid t_{k-1})\boldsymbol{H}^T + \boldsymbol{R}(t_k)]^{-1}$$

-State Estimate Update

$$\hat{\boldsymbol{x}}(t_k \mid t_k) = \hat{\boldsymbol{x}}(t_k \mid t_{k-1}) + \boldsymbol{K}(t_k)\ [y(t_k) - \bar{y}(t_k)]$$

-Error Covariance Update

$$\boldsymbol{P}(t_k \mid t_k) = [\boldsymbol{I} - \boldsymbol{K}(t_k)\boldsymbol{H}]\boldsymbol{P}(t_k \mid t_{k-1})$$

-Error Covariance Propagation

$$\frac{d\boldsymbol{P}(t)}{dt} = \boldsymbol{F}(\bar{\boldsymbol{x}}(t))\boldsymbol{P}(t) + \boldsymbol{P}(t)\boldsymbol{F}(\bar{\boldsymbol{x}}(t))^T + \boldsymbol{G}(\bar{\boldsymbol{x}}(t))$$

<u>Measurement Update at Time t_k:</u>

-Gain Matrix

$$\boldsymbol{K}(t_k) = \boldsymbol{P}(t_k \mid t_{k-1})\boldsymbol{H}^T[\boldsymbol{H}\boldsymbol{P}(t_k \mid t_{k-1})\boldsymbol{H}^T + \boldsymbol{R}(t_k)]^{-1}$$

-State Estimate Update

$$\hat{\boldsymbol{x}}(t_k \mid t_k) = \hat{\boldsymbol{x}}(t_k \mid t_{k-1}) + \boldsymbol{K}(t_k)\ [y(t_k) - \bar{y}(t_k)]$$

-Error Covariance Update

$$\boldsymbol{P}(t_k \mid t_k) = [\boldsymbol{I} - \boldsymbol{K}(t_k)\boldsymbol{H}]\boldsymbol{P}(t_k \mid t_{k-1})$$

5. DEVELOPMENT AND TESTING OF PARAMETER ESTIMATORS

5.1 Introduction

Chapters Three and Four dealt with estimation of destination-specific traffic densities using sensor data, by first describing a Markovian traffic model, then approximating the Markovian model with a linear stochastic model, and applying the Kalman filter to the linear model. Although the form of the proposed filtering model is known, it contains unknown parameters which need to be estimated. This chapter deals with the problem of estimating (or identifying) parameters for the filtering model using sensor data. The model's parameters may be divided into two groups:

(a) traffic flow parameters (i.e., free flow speed, capacity, critical density, jam density and exponent r) governing the transition functions (3.5),

(b) demand parameters (i.e., arrival rates and O-D proportions).

Here we are not developing new identification techniques but applying existing techniques to a particular class of problems. In section 5.2, the off-line identification of the model's parameters is formulated as a prediction-error minimization problem. In section 5.3, two off-line nonlinear least-squares (NLS) estimators and a quasi-maximum likelihood (QML) estimator for identifying traffic flow and O-D parameters are proposed. Section 5.4 describes a Monte Carlo Evaluation of two different approaches to estimate freeway O-D proportions b_{ij} and the traffic flow parameters, the objective being to decide which of the methods, under practically useful conditions, tend to be unbiased and to assess their relative statistical efficiency. In section 5.5, the identification of model parameters is considered from a recursive (on-line) standpoint.

5.2 Approaches to Parameter Estimation

In this section we formulate the identification of the model's parameters as an off-line least squares problem. We wish to give an appreciation for off-line schemes for two reasons. First, these schemes can be employed to give initial parameter estimates for use in an on-line scheme. Second, many on-line schemes in current use can be thought of as sequential implementations of off-line algorithms. For example, the recursive least squares approach used by Nihan and Davis (1989) is an exact counterpart of the linear ordinary least-squares algorithm, the Extended Kalman Filter approach described by Chang and Wu (1993) can be viewed as a recursive implementation of a nonlinear weighted least-squares approach, while the Kalman Filter method tested by Ashok and Ben-Akiva (1993) is a recursive implementation of a linear, multi-lag least squares approach. In particular, there is a natural connection between the efficiency of an off-line estimator and the convergence rate of its recursive counterpart, in that the standard error of estimate for the off-line estimator obtained with a sample of size N is a lower bound for the standard error of the recursive estimator after N iterations.

The idea underlying this approach is as followed. Traffic on a freeway is now considered as a causal process on which certain stimuli act as inputs while any that the freeway has m origins, indexed by i=1,2,...,m, s destinations indexed by j=1,2,...,s, and n sections, indexed by k=1,...,n. Let us introduce

a state vector $\mathbf{x} = (x_{11}, x_{12},..., x_{1s},..., x_{k1}, x_{k2},..., x_{ns})^T$,

an input vector at origins $\mathbf{q} = (q_1, q_2,..., q_m)^T$,

an output vector at off-ramps and mainline $\mathbf{y} = (y_1, y_2, ..., y_s, y_{s+1},..., y_{s+p})^T$,

for the whole freeway section,

and vector of unknown parameters $\boldsymbol{\beta} = [\mathbf{w}^T, \mathbf{b}^T]^T = [u_f, d_c, d_{jam}, b_{11},..., b_{ms}]^T$.

Furthermore, let **q**(t), **y**(t), t=1,...,N be the time sequences of measured input and output data collected from traffic sensors. The most common approach for the identification of a system without *a priori* information is the least sqHxres prediction error method, which minimizes the discrepancy between the model and the observation with respect to a quadratic output error function, and the parameter identification problem can now be formulated as the following least squares output problem:

Given the time sequence of measured input and output counts **q**(1),.....,**q**(N), **y**(1),.....,**y**(N), a trial sets of parameter values $\boldsymbol{\beta}$, and the traffic flow model for predicting the output counts, and initial state of traffic state **x**(0). Find the set of parameters $\boldsymbol{\beta}$ which minimizes following least squares criterion

$$S(\beta) = \sum_{t=1,N} [y(t)-\hat{y}(t,\beta)]^T[y(t)-\hat{y}(t,\beta)] \tag{5.1}$$

where $\hat{y}(t,\beta)$ denotes the vector of predicted outputs. Attempting to minimize S with respect to β leads to a nonlinear least squares (NLS) problem, since our traffic flow model for predicting the output counts are nonlinear in both the state variables and the parameters. This problem can be solved using any of a number of a standard routines as long as the problem is well-defined, in the sense that at least a locally unique minimizing value of β exists.

5.3 Development of Off-line Version of Parameter Estimators

In the sequel, two NLS based estimators, which are called as, respectively, Joint NLS and Two-step NLS, and a quasi maximum likelihood estimator are developed to identify traffic flow and O-D parameters.

5.3.1 Joint NLS Estimator

Given the arrival rates **q**(t) and a trial set of β values including traffic flow and O-D parameters, the mean value equation (4.2) is solved numerically. For a given set of origin counts this produces forecasted off-ramp and mainline counts $\hat{y}(t,\beta)$ that are aggregated over the measurement intervals. A final set of parameter vector β which minimizes the sum of the squared errors criterion (5.1) can be computed iteratively by a nonlinear optimization routine.

5.3.2 Two-step NLS Estimator

Two-step NLS estimation of traffic flow and O-D parameters is formulated by separately identifying the two groups of parameters. Let $d_k(t)$ and $d_{k+1}(t)$ denote the average density of the upstream and the downstream segment during time interval t, respectively, and $m_k(t)$ denotes the mainline counts at the end of segment n during time interval k. A set of traffic flow parameters $\mathbf{w} = (u_f, d_c, d_{jam})^T$ are first identified from mainline volume/density measurements, and then an O-D parameter vector **b** is identified from on-/off-ramp counts. Note that this method assumes the availability of density measurements as well as volume counts.

(A) Traffic flow parameter estimation:

Given a sequence of density measurements, $\{d_k(t), d_{k+1}(t), t=1,...,N\}$, at mainline segments, and given the time sequences of mainline counts for location k, $\{m_k(t), t=1,...,N\}$ as outputs, one finds the traffic flow parameter vector **w** by minimizing following sum-of-squares function with respect to **w**

$$S_k(w) = \sum_{t=1,N} [m_k(t) - \hat{m}_k(t,w)]^2 \qquad (5.2)$$

where the predicted output $\hat{m}_k(t,w)$ is obtained from two-dimensional flow-density relationship

$$m_k(t,\mathbf{w}) = \begin{cases} d_k u_f e^{-\frac{1}{2}(\frac{d_k}{d_c})^2} \left[1 - \left(\frac{d_{k+1}}{d_{jam}}\right)^r\right], & d_k \le d_c \\ \frac{d_c u_f}{\sqrt{e}} \left[1 - \left(\frac{d_{k+1}}{d_{jam}}\right)^r\right], & d_k > d_c \end{cases} \tag{5.3}$$

where u_f is a free-flow speed, d_c is a critical density, and d_{jam} is a jam density. Given density measurements and a trial set of **w** values, equation (5.3) is evaluated. A final set of traffic flow parameter values, which minimizes the sum of squared errors between forecasted and actual mainline counts is found by embedding this routine in a nonlinear optimization routine. It should be mentioned that the quality of the traffic model parameter estimates is influenced by both measurement intervals and detector locations.

(B) O-D parameter estimation

To estimate the O-D parameters, it is assumed that estimates of the traffic flow parameters $\hat{\mathbf{w}}$ are known from the first step and fixed. Given an initial set of O-D estimates **b** and the time sequences of measurement {**q**(t), t=1, ..., N} at on-ramps as inputs, and given the time sequences of counts {**y**(t), t=1, ..., N} at off-ramps as outputs, one finds the O-D parameter vector **b** by minimizing the following sum-of-squares function with respect to **b**

$$S_2(\boldsymbol{b}) = \sum_{t=1,N} [y(t)-\hat{y}(t,\hat{w},\boldsymbol{b})]^T[y(t)-\hat{y}(t,\hat{w},\boldsymbol{b})] \qquad (5.4)$$

where the predicted output counts $\hat{y}(t,\hat{w},\boldsymbol{b})$ are obtained by solving the mean value equation (4.2) numerically, given traffic flow parameters $\hat{w}$ and a trial set of b_{ij} values.

5.3.3 Quasi-Maximum Likelihood (QML) Estimation Algorithm

To apply the principle of maximum likelihood to an estimation problem requires that the probability distribution of the data, as a function of the parameters, be given. The maximum likelihood method can be seen as a special case of the prediction error criterion (Ljung, 1987), where one desire to minimize the negative logarithm of the probability density of the observation sequence, with respect to unknown parameters. More formally, let **Y** be a random variable whose probability density $p(\mathbf{y},\beta)$ depends on an unknown parameter vector β. To estimate β from an observation **y** choose the values of β that maximize the likelihood function $L(\mathbf{y},\beta)=p(\mathbf{y},\beta)$.

The first step in applying the maximum likelihood method is to determine the likelihood function. Let $\{\mathbf{y}(0),\mathbf{y}(1),\ldots,\mathbf{y}(N)\}$ sequence of observation vectors. Since the innovations (the differences between predicted and actual observations) for the large population approximation are Gaussian, the approximate conditional likelihood (given $\mathbf{y}(0)$) is then

$$L(y_n;\beta)=\prod_{t=1,n}\frac{1}{\sqrt{(2\Pi)^n\ |\ S(t,\beta)\ |\ )}}\exp[-\frac{1}{2}(y(t)-\hat{y}(t,\beta))^T S(t,\beta)^{-1}(y(t)-\hat{y}(t,\beta))] \quad (5.5)$$

where $\hat{y}(t,\beta)$ denotes the conditional mean of **y**(t) given **y**(t-1) and **S**(t,β) is the corresponding conditional covariance matrix, and these are computed by the Kalman filter. Rather than maximize this function directly it will be more convenient to minimize following negative of log-likelihood function with respect to β

$$LL(\beta) = \sum_t [\log\ |\ S(t,\beta)\ |\ + (y(t)-\hat{y}(t,\beta))^T S(t,\beta)^{-1}(y(t)-\hat{y}(t,\beta))] \quad (5.6)$$

Since the likelihood (5.5) is an approximation of likelihood for MARCOM, this produces a quasi-maximum likelihood (QML) estimator (White, 1982).

The state dynamics equation in Table 4.1 together with the observation equation

$$\mathbf{y}(t_k) = \mathbf{H}\mathbf{x}(t_k) + \mathbf{v}(t_k)$$

comprise a linear state-space model, and at least in principle the methods of linear systems identification (Ljung and Söderström, 1983) might be employed to develop parameter estimators. It turns out however that when only partial (**H** is not an identity matrix) and error prone counts are available, the Kalman filter must be employed to compute predicted observations $\hat{y}(t,\beta)$, and the corresponding conditional covariance matrix $\mathbf{S}(t,\boldsymbol{\beta})$.

The procedure for evaluating the likelihood function for a particular value of the parameter vector $\boldsymbol{\beta}$ is as follows. The initial state vector and covariance matrix are first calculated (or assumed given) for $\boldsymbol{\beta}$. The prediction of conditional means $\hat{y}(t \mid t-1)$ and conditional covariance $\mathbf{S}(t,\boldsymbol{\beta})$ are obtained via the Kalman filter in Table (4.1). The value of the likelihood function is then obtained by evaluating the equation (5.6). The maximum likelihood estimates of the traffic flow and O-D parameters can thus be computed by minimizing the function (5.6) with respect to $\boldsymbol{\beta}$.

5.4 Evaluation of Off-line Parameter Estimators via Monte Carlo Tests

One can distinguish two basic motivations for parameter estimation in stochastic systems as description-oriented or applications-oriented (Goodwin and Sin, 1984). In the description-oriented approach the main emphasis is placed on obtaining an understanding of the process, whereas in the applications-oriented approach the main emphasis is on achieving certain objectives (such as estimation or control of the underlying system). This distinction may influence the choice of algorithms and the choice of the performance criteria.

The first part of this section will be focused on the description-oriented approach, where the attention will be given to the properties of the estimated parameters. In evaluating the proposed estimators via Monte Carlo simulation, focus

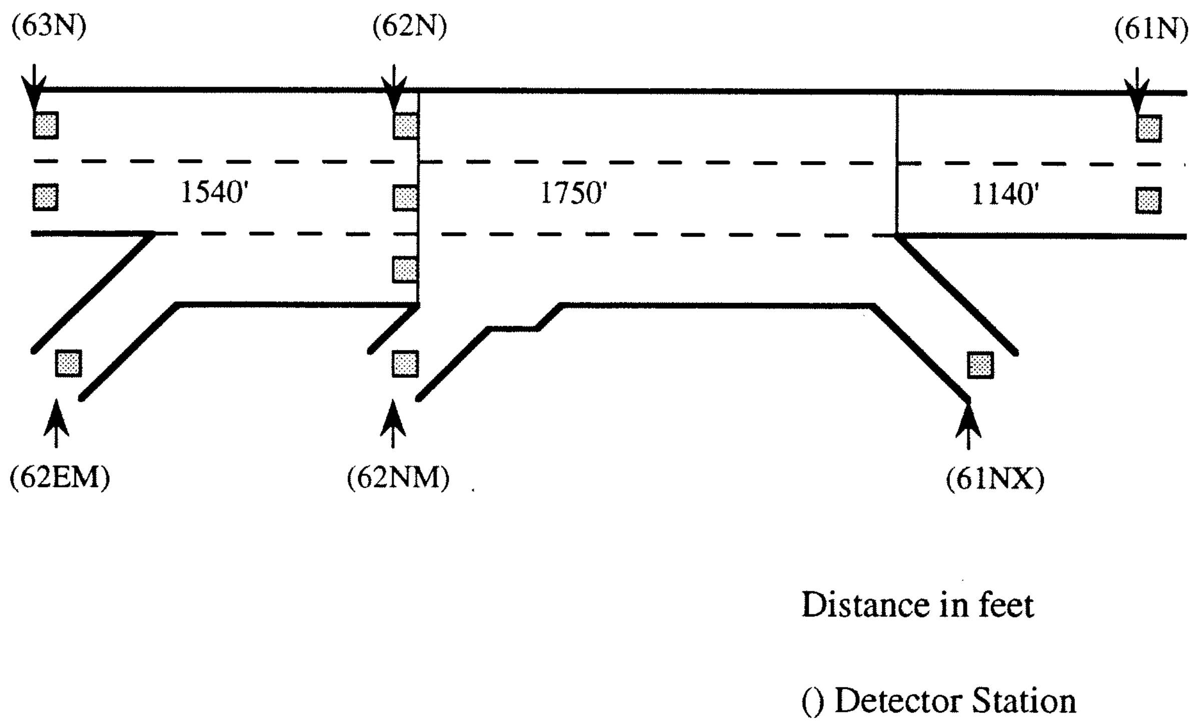


Figure 5.1. Geometrics of 3-Origin, 2-Destination Section (I-35W, Minneapolis)

will be given to the properties of the proposed candidate estimators, i.e., 2-step NLS, Joint-NLS, and QML. Next, our focus will be given to the applications-oriented approach, which considers the degree to which the ultimate objective (destination-specific density estimation) is achieved by the given estimators. This will be done using a model sensitivity analysis.

5.4.1 Monte Carlo Example

A relatively simple freeway section of I-35W northbound (Figure 5.1) was selected as an example for the Monte Carlo study. Since this section is composed of 3-origin and 2-destination, there are six O-D pairs to be estimated. The O-D proportions are represented by:

$$[b_{11},b_{12},b_{21},b_{22},b_{31},b_{32}]$$

Data Generation

MARCOM was run to generate 50 data sets for the example freeway system under the presumed O-D proportions, b_{ij}. Each data set consists of a simulated three-hour sequence of five-minute periods. MARCOM took the actual on-ramp counts and the presumed **b** vector as inputs, and then computed simulated counts. With the actual on-ramp demands for a morning peak period (6:00 am - 9:00 am), MARCOM generated traffic data in the low to moderately congested regions, which was similar to the actual situation of the morning peak period at the example section. The simulated counts were aggregated to produce 5-minute volume counts for mainline segments and off-ramps, as well as mainline segment densities.

From each simulated data set it is possible, using a given estimator, to compute estimates of the traffic flow parameters and the O-D proportion vector **b**, which were used to generate this data set. FORTRAN programs NLS2STEP, NLSJOINT, and QMLJOINT were written and implemented on a SUN Sparcstation 1+, which evaluated the minimization criterion in equations (5.1), (5.4) and (5.6), respectively. The traffic flow parameter estimates in 2-step NLS were then computed using the NLS routines E04FDF and E04YCF, while the others were computed using the Quasi-Newton search routine E04JBF (NAG, 1987). The set of estimates from all 50 data sets form a pseudorandom sample of each estimator, from which one can compute sample means and standard deviations. The efficiencies of each estimator can then be estimated via the sample standard deviations. Tendencies of the estimators to show bias can be evaluated by comparing the sample means to the "true" values used in generating the data.

Comparison and Evaluation of Results

Table 5.1 displays the means, standard deviations, and t-statistics computed across the 50 simulated data sets for the candidate estimators. Tendencies of the estimators to show bias can be evaluated by comparing the sample means to "true" values used in generating the data. The efficiencies of different estimators can be compared via the sample standard deviations. The [ ] frame a t-test statistic for the null hypothesis that the sample average equals the true value. The critical value for a sample size of 50 at 0.05 level is 2.01. If the absolute value of the computed t-value is greater than 2.01 the estimator appears biased at the 0.05 level, and the asterisk, *, in Table 5.1 denotes rejection of the null hypothesis. Table 5.1 also reports the root-mean-squared (RMS) distance between the estimates and their true values, which is computed via

$$RMS_{\beta} = \sqrt{(\beta - \bar{\beta})^2 + \sigma_{\beta}^2}$$

The RMS distance combines the estimation error due to inefficiency with that due to bias to provide a single measure of these two tendencies. Initially, all the estimators failed to give consistent estimates (large variance) of the jam density (d_{jam}) using the given data. It is difficult to obtain a good estimate of d_{jam} without traffic data over the range of all possible densities, $0 \le d \le d_{jam}$. In a separate test, the estimability of the jam density was improved using a wider range of traffic data that were generated by an incident simulation, indicating that all the given parameters are identifiable if one has a full range of density measurements. Obtaining a full range of densities, however, on a ramp-controlled freeway is not likely except during non-recurrent congestion such as that caused by incidents. This estimability problem might be partially relieved by the fact that jam density is a certain fixed value determined physically, and so is highly transferable. This estimability problem of the jam density was also reported (Van Aerde, 1995). Therefore, once jam density has been estimated from some data set, it can be taken as fixed and considered as constant thereafter. A discussion of the results in Table 5.1 follows.

(a) All estimators produce, on the average, biased estimates of traffic flow parameters. Again this bias might be reduced using a wider range of traffic data. With respect to the RMS criterion, the QML algorithm is superior to the NLS-based algorithms, mainly because of higher efficiency.

(b) The Two-step NLS estimator, on the average, was clearly the worst performer. The quality of the O-D estimates by the NLS estimator (NLS^b column in Table 5.1) was significantly improved by using true traffic flow parameters, indicating that reducing the ignorance concerning the traffic flow parameters is one way of improving the O-D estimates by the Two-step NLS estimators.

(c) Generally, the QML based estimator produces unbiased O-D estimates with lower standard deviations than did the NLS-based estimators. Like other estimators, the QML estimator produced the highest variability in the O-D parameter b_{31}, apparently due to the low O-D flows. The flow from origin three to destination one, q_{31}, is 1.2 vehicles per 5 minute, which is very low.

Table 5.1. Comparison and Evaluation of Off-line Parameter Estimators

Parameter	True Value	2-step NLS	Joint NLS	NLS[b]	QML
u_f	65.0	63.98[a] (5.91) [1.21] {6.00}	63.71 (3.97) [2.28]* {4.18}	Known 65.0	62.12 (1.88) [10.81]* {3.44}
k_{max}	58.34	60.85 (7.04) [-2.52]* {7.47}	60.82 (4.71) [-3.72]* {5.32}	Known 58.34	60.00 (2.27) [-5.16]* {2.81}
b_{11} q_{11}	0.16 (23.2)	0.178 (0.044) [-2.84]* {0.048}	0.156 (0.053) [0.556] {0.053}	0.142 (0.031) [3.890]* {0.036}	0.164 (0.032) [-0.82] {0.032}
b_{21} q_{21}	0.19 (31.5)	0.163 (0.054) [3.54]* {0.060}	0.188 (0.073) [0.168] {0.074}	0.206 (0.040) [-2.77]* {0.043}	0.180 (0.037) [-0.88] {0.039}
b_{31} q_{31}	0.10 (1.2)	0.216 (0.206) [-3.97]* {0.237}	0.180 (0.256) [-2.20]* {0.268}	0.098 (0.161) [0.06] {0.161}	0.150 (0.180) [-1.96] {0.187}

Note:
b_{ij}: O-D proportions
q_{ij}: Average O-D flow (veh/5 min) = $q_i(k) \cdot b_{ij}(k)$, where $q_i(t)$ is arrival demand at on-ramps
a : averages of estimated traffic flow parameters, sample size 50
( ): standard deviations of parameters
[ ]: t-test statistic, the cut-off value t_{49} is 2.01 (α=0.05)
{ }: root-mean-squared distance : $RMS_\beta = \sqrt{(\beta - \bar{\beta})^2 + \sigma_\beta^2}$
b : 2-step (Joint) NLS estimates by using true traffic flow parameters

(d) In this descriptive approach, it was obvious that the QML estimator had the best statistical properties, which is consistent with theoretical expectations. However, the price for these desirable statistical properties was a near 40-fold increase in CPU time. The average CPU time consumptions of each estimator on a SUN Sparcstation 1+ are reported in Table 5.2. Although the trade-off between statistical and numerical efficiency appears unforgiving, in the applications oriented approach what happens with the parameter estimates will be of secondary importance as long as they achieve the ultimate objective, i.e., estimating destination-specific density. Next, our focus will be given to this applications oriented approach. This issue will be pursued via a model sensitivity analysis.

Table 5.2. Average CPU Time of Each Estimator

	2-step NLS	Joint NLS	QML
CPU Time in seconds	64.34	191.83	8324.75

5.4.2 Sensitivity Analysis

Here the natural question concerns how the traffic flow parameters estimates affect the quality of estimates of destination-specific density. In order to investigate the performance differences of the filtering models for different parameter estimates the following sensitivity analyses were performed. First, traffic flow and O-D parameters were estimated by the three candidate parameter estimators using one day's simulation data. Next, a filtering model using each of the sets of estimated parameters was run. Finally, the discrepancy between the filtering model's results and simulation results was quantified by the "Mean squared error" index. Table 5.3 displays the results from three estimators along with those using true parameters.

These investigations show that the value of performance criterion differed generally by less than 10 % when the nominal (true) parameter set was replaced by an estimated parameter set. It is seen that the filtering model is not very sensitive to small parameter changes from the different estimators. This demonstrates a robustness of the filtering model to parameter estimation.

Table 5.3. Sensitivity of the Filtering Model for Different Parameter Sets

	True value	2-step NLS	Joint NLS	QML
X_{11}	12.7[a]	14.2	13.2	13.9
X_{12}	12.0[a]	13.2	12.8	13.7
X_{21}	4.8[a]	5.7	4.6	4.7
X_{22}	26.7[a]	25.6	25.3	25.5
X_{32}	29.3[a]	28.7	29.4	28.8
m_1	116.2[b]	118.8	158.9	127.3
y_1	152.7[b]	142.8	209.6	141.9
y_2	58.4[b]	55.3	60.6	56.4

Note:
x_{kj}: estimated vehicles in section k bound for destination j, at the end of every 5 minute
m_k: predicted 5-minute vehicle counts at mainline segment k
y_j: predicted 5-minute vehicle counts at destination j
a : Mean squared estimation error = $\sum_{t-1,N}(x_{kj}(t)-\hat{x}_{kj}(t))/N$, where N = 36
b : Mean squared prediction error = $\sum_{t-1,N}(y(t)-\hat{y}(t))/N$, where N = 36.

5.5 Development and Testing of Recursive Version of O-D Parameter Estimator

5.5.1 Formulation of the adaptive filter

Basically, a recursive estimator generates a sequence of O-D estimates **b**(t) by updating the estimate from the previous interval, **b**(t-1), using only data collected during interval t. An advantage possessed by recursive, but not by off-line algorithms, is that recursive algorithms can be modified so as to track parameters which show occasional stepwise changes across time. That is, if O-D parameters are piecewise constant, and the interval between step change is long compared to the convergence rate of the algorithm, these changes can be tracked. Since the O-D pattern in urban freeways during a morning peak period is likely to be changing in time, recursive estimation of time-varying freeway O-D matrices could improve the operational efficiency necessary for real-time applications. For example, the ability of the proposed filtering model in predicting volumes at off-ramps would be improved.

One common implementation of recursive parameter identification is phrased as a state estimation problem by introducing constant parameters as augmented state variables (Grewal and Payne, 1976; Chang and Wu, 1993). This leads to a nonlinear filtering problem. The extended Kalman filter can then be applied to estimate the composite state comprising the original state vector and the parameters. The subject of parameter-adaptive filtering, estimating unknown parameters as well as state variables simultaneously, can be viewed as an application of nonlinear estimation theory.

To formulate the adaptive state estimator, the proposed system needs to be structured into a standard state-space model. First, it is assumed that the set of unknown O-D parameters evolve as a random walk process as shown in equation (5.7):

$$\dot{\boldsymbol{b}}(t) = \boldsymbol{w}_3(t) \tag{5.7}$$

Combining original system equation in Table 4.1 with equation (5.7), we have the augmented state equations

$$\begin{bmatrix} \frac{d\tilde{\boldsymbol{x}}(t)}{dt} \\ \frac{d\tilde{\boldsymbol{y}}(t)}{dt} \\ \frac{d\boldsymbol{b}(t)}{dt} \end{bmatrix} = \begin{bmatrix} \boldsymbol{F}(\bar{\boldsymbol{x}}(t),\boldsymbol{b}(t))\tilde{\boldsymbol{x}} \\ \boldsymbol{F}(\bar{\boldsymbol{x}}(t),\boldsymbol{b}(t))\tilde{\boldsymbol{y}} \\ \boldsymbol{0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{w}_1(t) \\ \boldsymbol{w}_2(t) \\ \boldsymbol{w}_3(t) \end{bmatrix} \tag{5.8}$$

where $\tilde{\boldsymbol{x}}(t) = \boldsymbol{x}(t) - \bar{\boldsymbol{x}}(t)$ and the measurement equations

$$z(t_k) = \boldsymbol{H}_b \begin{bmatrix} \bar{\boldsymbol{x}}(t_k) \\ \bar{\boldsymbol{y}}(t_k) \\ \boldsymbol{b}(t_k) \end{bmatrix} + \nu(t_k) = [\boldsymbol{0},\boldsymbol{I},\boldsymbol{0}] \begin{bmatrix} \bar{\boldsymbol{x}}(t_k) \\ \bar{\boldsymbol{y}}(t_k) \\ \boldsymbol{b}(t_k) \end{bmatrix} + \nu(t_k), \quad \nu(t_k) \sim N(\boldsymbol{0},\boldsymbol{R}(t_k)) \tag{5.9}$$

The error covariance matrix **P**(t) evolves according to

$$\frac{dP(t)}{dt} = F(\bar{x}(t),b(t))P(t)+P(t)F(\bar{x}(t),b(t))^T + \begin{bmatrix} G_1(\bar{x}(t)) & G_2(\bar{x}(t)) & 0 \\ G_2^T(\bar{x}(t)) & G_3(\bar{x}(t)) & 0 \\ 0 & 0 & Q \end{bmatrix} \quad (5.10)$$

Here $F(\bar{x}(t),b(t))$ denotes the Jacobian matrix of the right-hand side of equation (4.2) with respect to $[\bar{x}(t),\bar{b}(t)]$, and $G(\bar{x}(t))$ is a covariance matrix of the diffusion process given in equation (4.5). The matrix **Q** will have diagonal elements and the strength of the noise should correspond roughly to the possible range of O-D parameter variation. When actual counts become available at some time t_k, the extended Kalman filter algorithm is used to give a measurement update.

5.5.2 Monte Carlo Test of the Recursive O-D Parameter Estimator

The same freeway section used in the Monte Carlo tests evaluating off-line estimators (Figure 5.1) was selected for the Monte Carlo study of the recursive estimator.

Data Generation

In order to generate plausible time-varying O-D proportions the proposed recursive parameter adaptive filtering model was run using a real data of morning peak period from 6:00 am to 9:00 am. This generated a sequence of 36 five-minute O-D proportions, $b_{ij}(k)$, which will be referred to "true time-varying O-D proportions". Next, MARCOM took the actual five-minute on-ramp counts and the true time-varying O-D proportions as input, and generate 50 data sets, each consisting of a simulated three-hour sequence of five-minute on-ramps, off-ramps and mainline counts, for the example freeway system.

From each simulated data set the proposed parameter-adaptive estimator was used to estimate time-varying O-D proportions and destination-specific densities using a FORTRAN program TVODEST, implemented on a SUN Sparcstation 1+. From the estimates of all 50 data sets one can compute sample means and standard deviations. Figures 5.2-5.4 show the true time-varying O-D proportions along with the 95 percent confidence produced by the estimator. In each case the estimation range tracks the true (simulated) O-D proportions reasonably well, with the O-D proportions that have higher O-D subflow, $q_i(k) \cdot b_{ij}(k)$, being estimated somewhat better. This indicates that the parameter-adaptive filtering model is performing properly in tracking time-varying O-D proportions.

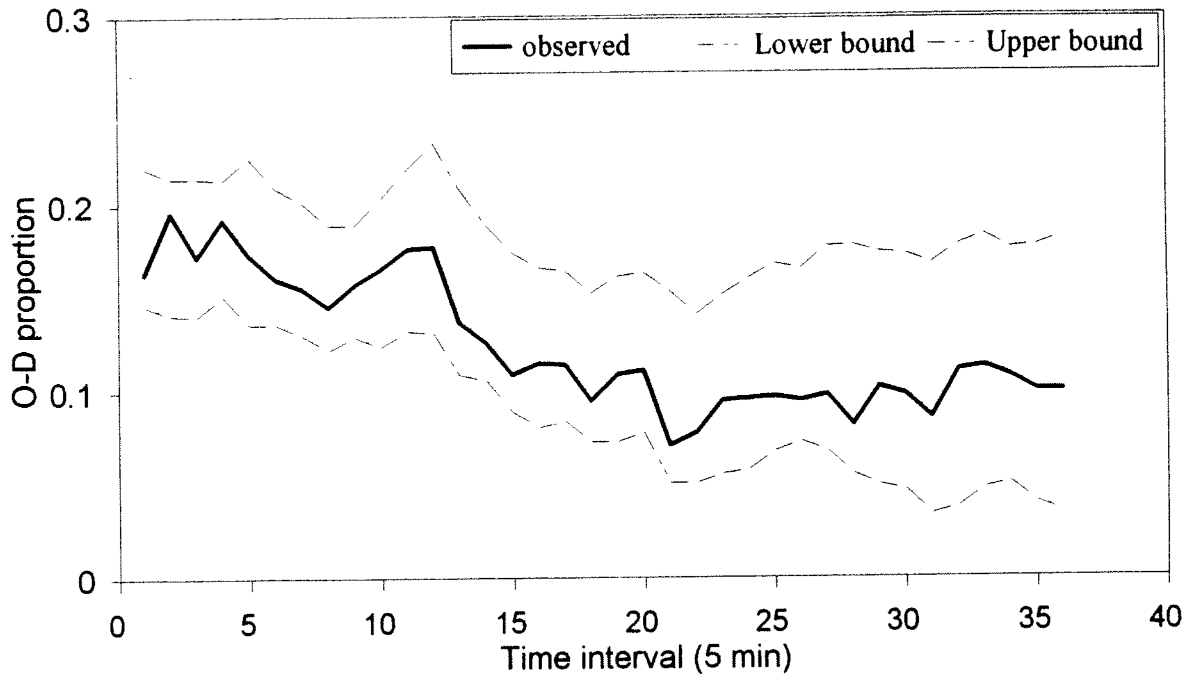


Figure 5.2. Confidence Interval of Time-varying O-D Pair b_{11}

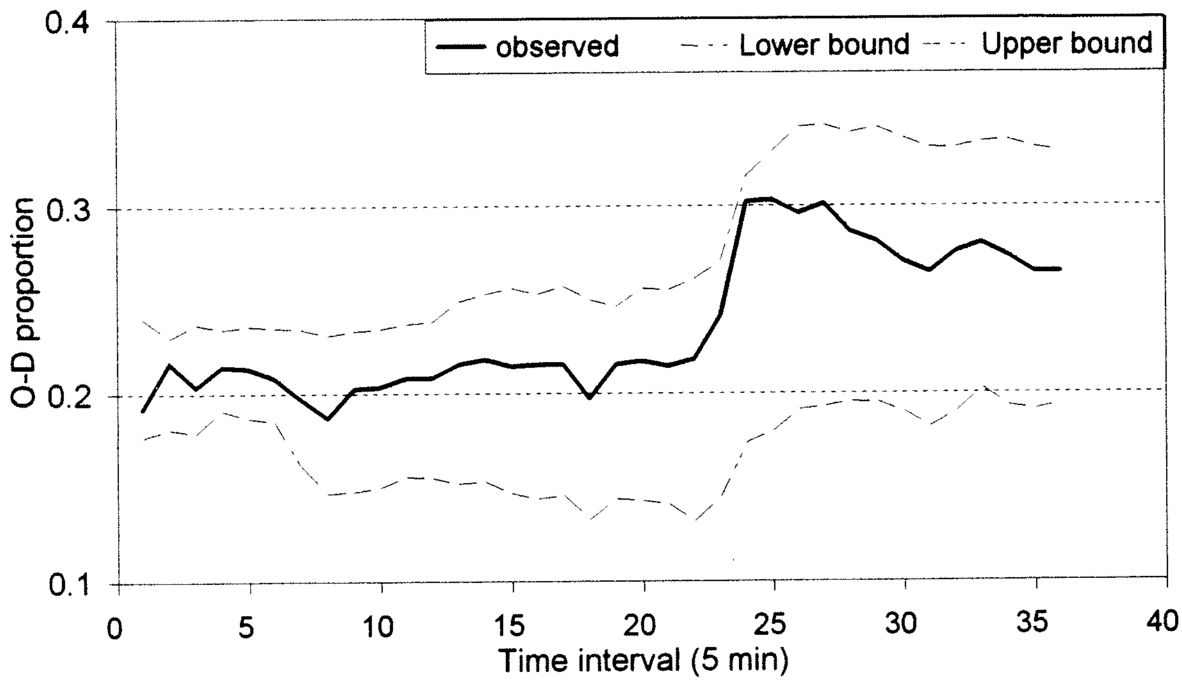


Figure 5.3. Confidence Interval of Time-varying O-D Pair b_{21}

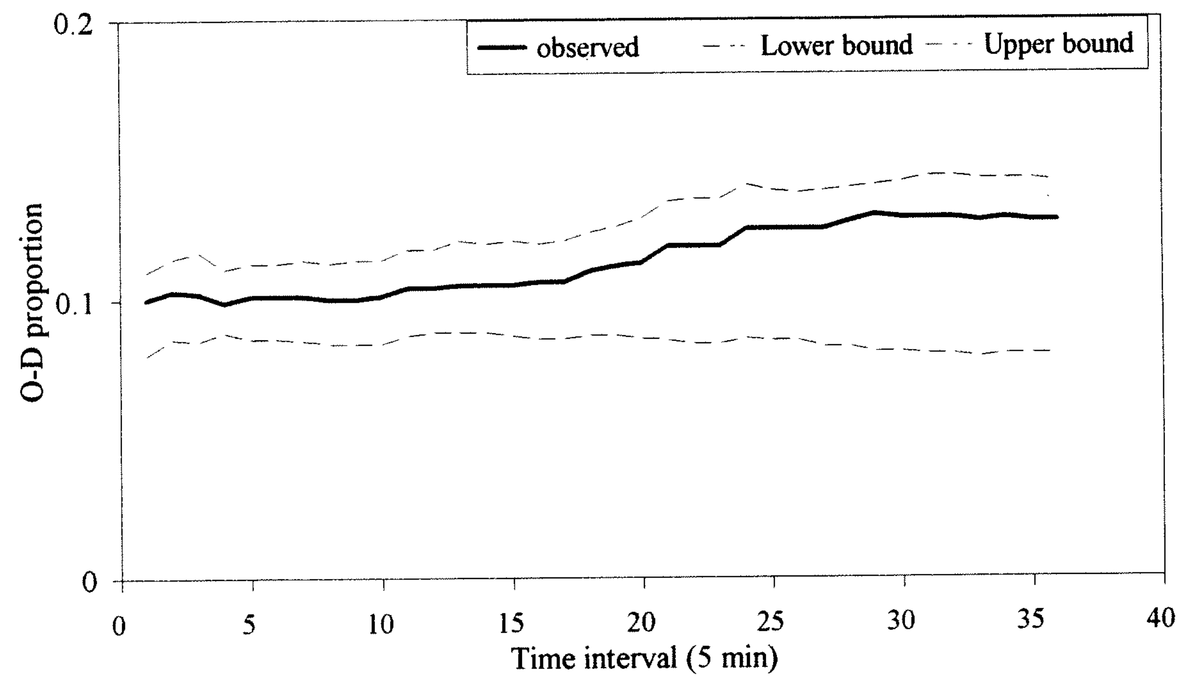


Figure 5.4. Confidence Interval of Time-varying O-D Pair b_{31}

6. PRELIMINARY TESTING AND VALIDATION OF THE FILTERING MODEL

6.1 Introduction

The objective of this chapter is the preliminary testing of the performance of the Kalman filter estimates of destination-specific densities. Since destination-specific densities are almost impossible to observe in practice, the accuracy of the Kalman filter will have to be assessed using simulated data. However, simulation experiments only give limited insight whether a method would work in practice. Therefore the Kalman filter will also be tested by predicting off-ramp volumes using actual freeway data.

6.2 Testing with Simulation Data

6.2.1 Experiment 1: Simulation Test with Short Section

A 1.2 km (0.75 mile) segment of northbound interstate highway I-35W was selected for the preliminary test of the proposed filtering model using simulation (Figure 5.1). The segment contained two on-ramps and one off-ramp, giving a total of three origins and two destinations, and was divided into three sections. Five-minute cumulative volume and lane occupancy measurements during the three hour morning peak period were obtained from the Minnesota Department of Transportation (MNDOT) for stations marking the upstream and downstream segment boundaries, as well as one intermediate station at the boundary between the first and second subsections. Five-minute on-ramp and off-ramp counts were available as well. Since the segment contained two origins and three destinations, and was divided into three sections, the state vectors are comprised of 5 compartment populations and 6 count totals, as follows

(a) 5 destination specific densities: $x_{11}(t)$, $x_{12}(t)$, $x_{21}(t)$, $x_{22}(t)$, $x_{32}(t)$

(b) 3 origin counts: $q_1(t)$, $q_2(t)$, $q_3(t)$

(c) 2 destination counts: $y_1(t)$, $y_2(t)$

(d) 1 mainline count: $m_1(t)$

First, simulated data were obtained by running the stochastic traffic flow model, MARCOM, shown in Table 3.1. The arrival rates for the simulation were simply estimated as those values which reproduced the corresponding actual five-minute arrival counts, while the traffic flow parameters and O-D parameters for the simulation were estimated by the Joint-NLS estimator described in section 5.3. The origin-destination proportions and traffic flow parameters are estimated as b_{11}=0.16, b_{21}=0.19, b_{31}=0.1, u_f=104.6 kph (65.0 mph), d_c=37.8 veh/km/lane (60.8 veh/mile/lane), d_{jam}=99.4 veh/km/lane (160 veh/mile/lane), and r=3. The assumed parameters and on-ramp counts were then used as inputs to MARCOM stochastic simulation model, and a three hour simulation was performed. Instantaneous destination-specific section densities, $x_{kj}(t)$, at the end of every 5-minute interval, and 5-minute cumulative volume counts, $y_i(t)$ and $m_l(t)$, at designated detectors were generated from the simulation output.

Next, the proposed Kalman filter model was run to estimate the destination-specific densities using simulated counts ($\mathbf{q}_i(t)$, $\mathbf{m}_k(t)$, $\mathbf{y}_j(t)$) as measurements. Since the true states (simulated destination-specific densities, $x_{kj}(t)$) are available, one can calculate the estimation error and the Mean Absolute Error (MAE) between simulated and estimated values. One can also calculate the prediction error (innovation) between predicted counts and measured counts. The low values of the MAE for x_{kj} in Table 6.1 indicates good performance of the filtering model in tracking the simulated destination-specific densities although it is based on an approximation of the original stochastic processes.

To ensure that the estimator is operating properly, further tests need to be done. A necessary and sufficient condition for a Kalman filter to be optimal is that the innovation sequence is zero-mean and not serially correlated (Candy, 1986). If we assume that the innovation sequence is ergodic and Gaussian, then we can use the sample mean, $\hat{m}_e$, as test statistic to estimate m_e, the population mean. To test that the mean of the *i*th component of the innovation vector, $e_i(t)$, is equal to zero at the 5 percent significance level (α=0.05), one can use the test statistic

$$\tau_i = 1.96\sqrt{\frac{\hat{R}_e(i)}{N}}$$

where $\hat{R}_e(i)$ is the sample variance (assuming ergodicity) given by

$$\hat{R}_e(i) = \frac{1}{N}\sum_{t=1,N} e_i^2(t)$$

Then the decision rule is

accept the hypothesis $m_e(i) = 0,\ if\ |\hat{m}_e(i)| \leq \tau_i$,

reject the hypothesis $m_e(i) = 0,\ if\ |\hat{m}_e(i)| > \tau_i$,

where $\hat{m}_e(i)$ is the sample mean of the prediction errors for measurement component i. Table 6.1 clearly indicates that the innovation sequence means are not significantly different from zero.

To test for serial correlation in the innovations, one can use the normalized sample covariance test statistic (Candy, 1986). The ith component of normalized covariance is given by

$$\hat{\rho}_e(i,k) = \frac{\hat{R}_e(i,k)}{\hat{R}_e(i)} = \frac{1}{N}\sum_{t=k+1,N} [e_i(t)-\hat{m}_e(i)][e_i(t+k)-\hat{m}_e(i)]/\hat{R}_e(i)$$

and under the null hypothesis, $\rho_e(i,k)=0$, $\hat{\rho}_e(i,k)$ is approximately normally distributed with mean equal to zero and variance equal to 1/N. Thus if $|\hat{\rho}_e(i,k)| > 1.96/\sqrt{N}$, one

Table 6.1. Performance of the Filtering Model and Zero-mean Test (simulated data)

	MAE[a]	$m_e(i)$[b]	$\tau(i)$	H_0
X_{11}	2.0	0.6	1.2	
X_{12}	2.3	-0.6	1.2	
X_{21}	1.6	-0.5	0.7	
X_{22}	3.8	-0.3	1.7	
X_{32}	4.2	1.0	1.8	
m_1	8.3	-1.5	3.5	Accept
y_1	9.2	-2.1	2.5	Accept
y_2	6.1	-3.1	4.0	Accept

Note:
x_{kj}: estimated vehicles in section k bound for destination j, at the end of every 5 minute
m_k: predicted 5-minute vehicle counts at mainline section k
y_j: predicted 5-minute vehicle counts at destination j
a : Mean Absolute Error between simulated and estimated values: $\sum_{t=1,N} | y(t) - \hat{y}(t) | / N$, where N = 36.
b : sample mean of innovation

Table 6.2. Normalized Covariance Coefficients of the Innovations (simulated data)

	1[a]	2	3	4	5	6	7	8	9	10
m_1[b]	0.07	-0.15	-0.02	-0.34	-0.17	-0.19	0.17	0.16	-0.05	0.31
y_1[c]	-0.11	-0.08	0.21	-0.01	-0.04	-0.04	0.10	-0.07	-0.08	0.01
y_2	0.09	-0.08	-0.12	-0.10	0.13	0.00	0.08	-0.15	-0.06	-0.07

Note:
a : covariance lag
b : mainline station
c : off-ramp station
critical value for each cell = $\pm 1.96/\sqrt{N} \approx 0.33$ (N=36)

would reject the hypothesis $\rho_e(i,k)=0$, at the α=0.05 level of significance. Table 6.2 displays the normalized covariance coefficients, which were computed from the innovation sequences, for lags 1 through 10. The sequences of the three detector stations are white, since 0 percent of the sample covariances exceed the confidence limit.

6.2.2 Experiment 2: Simulation Test with Longer Section

To assess the filtering model's accuracy for a longer section of freeway, a second series of simulation tests was done for the 4.0 km (2.5 mile) long, seven-origin, four-destination segment of I-35W depicted in Figure 3.9. Again three hours of simulation data were generated and the Kalman filter model was run using the same procedure as in the first experiment.

Figures 6.1 and 6.2 show the simulated destination-specific traffic densities

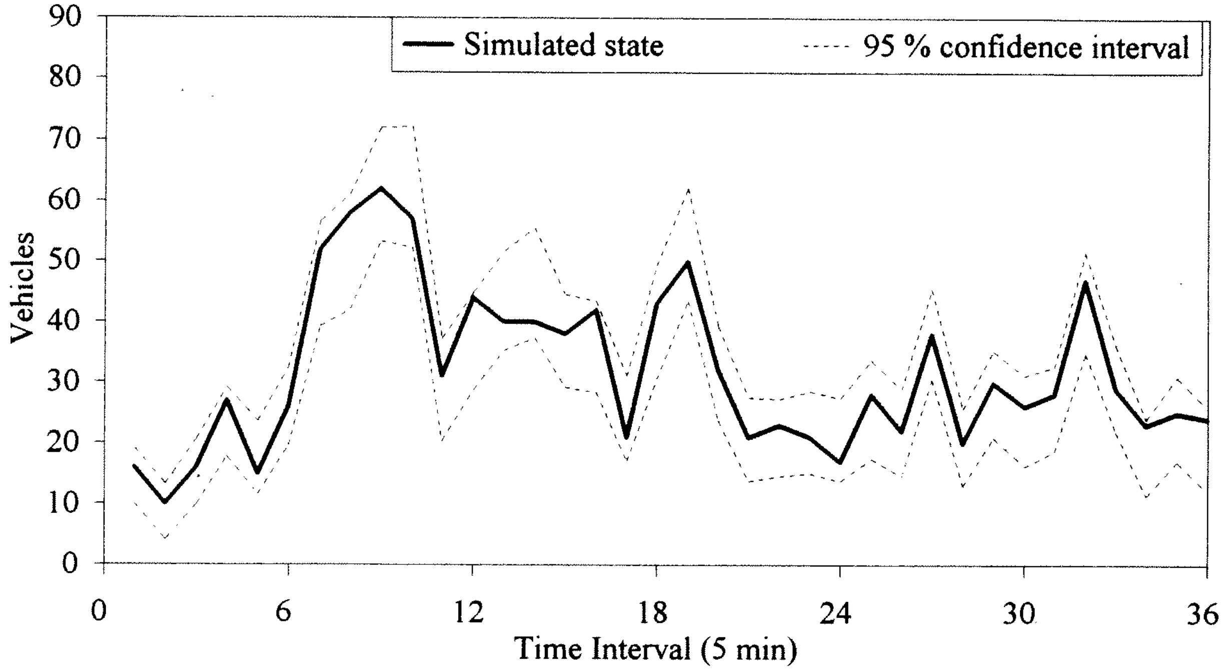


Figure 6.1. Simulated State and Confidence Interval (state variable: x_{54})

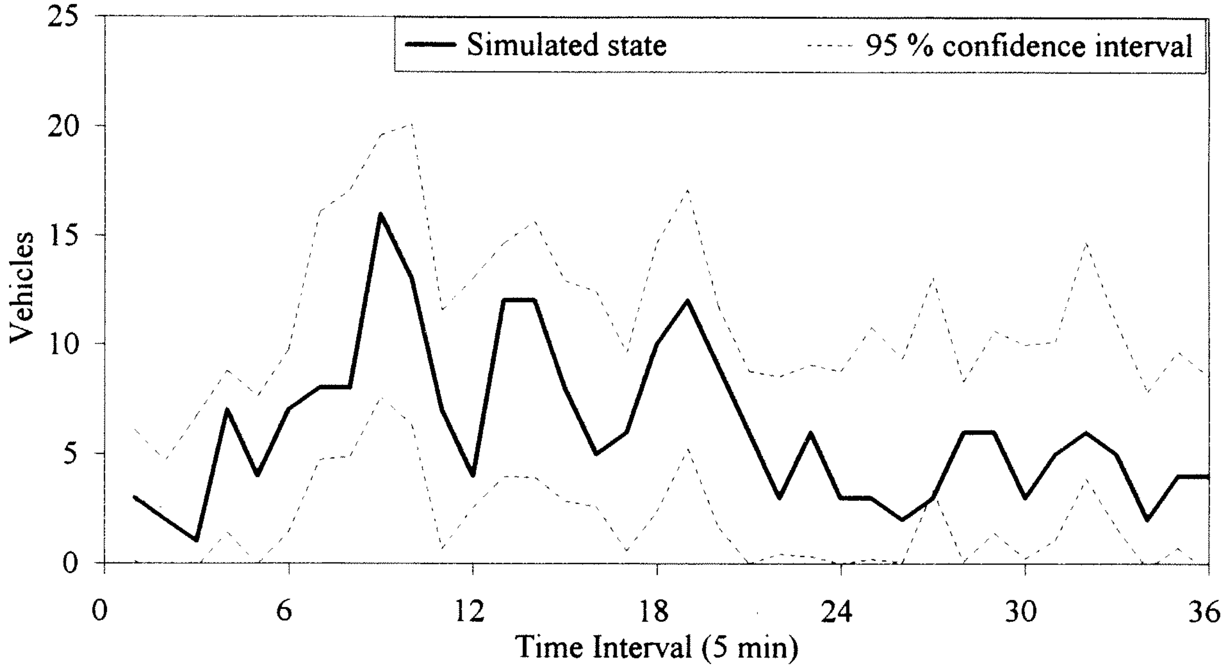


Figure 6.2. Simulated State and Confidence Interval (state variable: x_{52})

along with the approximate 95 percent confidence interval ($\approx 1.96\sqrt{P_i(t_k|t_{k-1})}$ about the estimated sequences) produced by the Kalman filter. If the covariance estimates of the filter are reasonable, then 95 percent of the known observation sequences should lie within the constructed intervals. In each case the estimation range tracks the simulated values reasonably well, with the larger volume flows being tracked somewhat better. This indicates that the filter, which is based on the approximation of the original stochastic process, is performing properly for this longer segment of freeway.

6.3 Preliminary Testing of a Parameter Adaptive Filter with Field Data

The primary focus of the this last set of tests is to evaluate the advantage of the adaptive filterings over the original filtering model. A 1.2 km (0.75 mile) segment of northbound interstate highway I-35W, which was used in experiment 1 (Figure 5.1), was selected as the test site. For convenience of comparison, the original Kalman filter model with fixed O-D proportion is referred as **Model A**, while the parameter-adaptive filtering model incorporating time-varying O-D is named **Model B**. Since both Model A and Model B are very similar, they were implemented with the identical initial parameter settings to improve comparability. The origin-destination proportions for Model A were fixed as b_{11}=0.16, b_{21}=0.19, b_{31}=0.1, while those of Model B were recursively estimated. The traffic flow parameters of u_f=104.6 kph (65.0 mph), d_c=37.8 veh/km/lane (60.8 veh/mile/lane), d_{jam}=99.4 veh/km/lane (160 veh/mile/lane), and r=3 were used for both Model A and Model B. Next, both models were run to generate predicted mainline and off-ramp counts when fed by actual on-ramp counts.

Figure 6.3 shows actual 5 minute mainline counts at the end of segment 3, which is the second destination, along with predicted counts by Model A and Model B. The mainline volumes are tracked reasonably well by both Model A and Model B, while Model B appears to predict the fluctuations slightly better. However, Figure 6.4 shows that Model B is superior to Model A in tracking the off-ramp fluctuations. The test results are further summarized in Table 6.3. The statistical test indicates that the innovation sequences of each station for both models are zero-mean. Table 6.4 displays the normalized covariance coefficients for the innovations. Model A results show that the innovation sequences of mainline stations (m_1 and y_2) were not serially correlated, while that of the off-ramp (y_1) station is. However, the test results of Model B lead to the acceptance of the null hypothesis that every component of the innovations are not serially correlated.

Using the MAE and serial correlation test, it was possible to show important differences in off-ramp volume prediction ability between the original filter (Model A) and the parameter-adaptive filter (Model B).

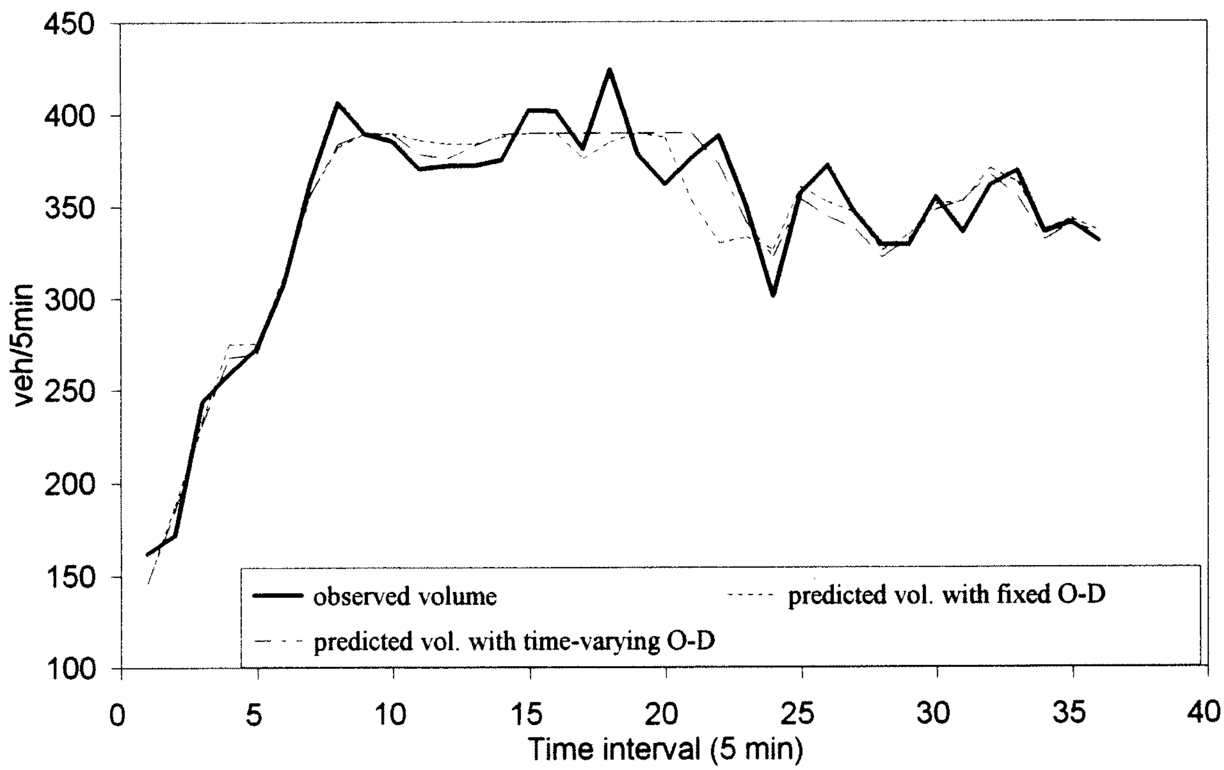


Figure 6.3. Actual Volume vs. Predicted Volume with Fixed and Time-varying O-D (mainline: station 61N)

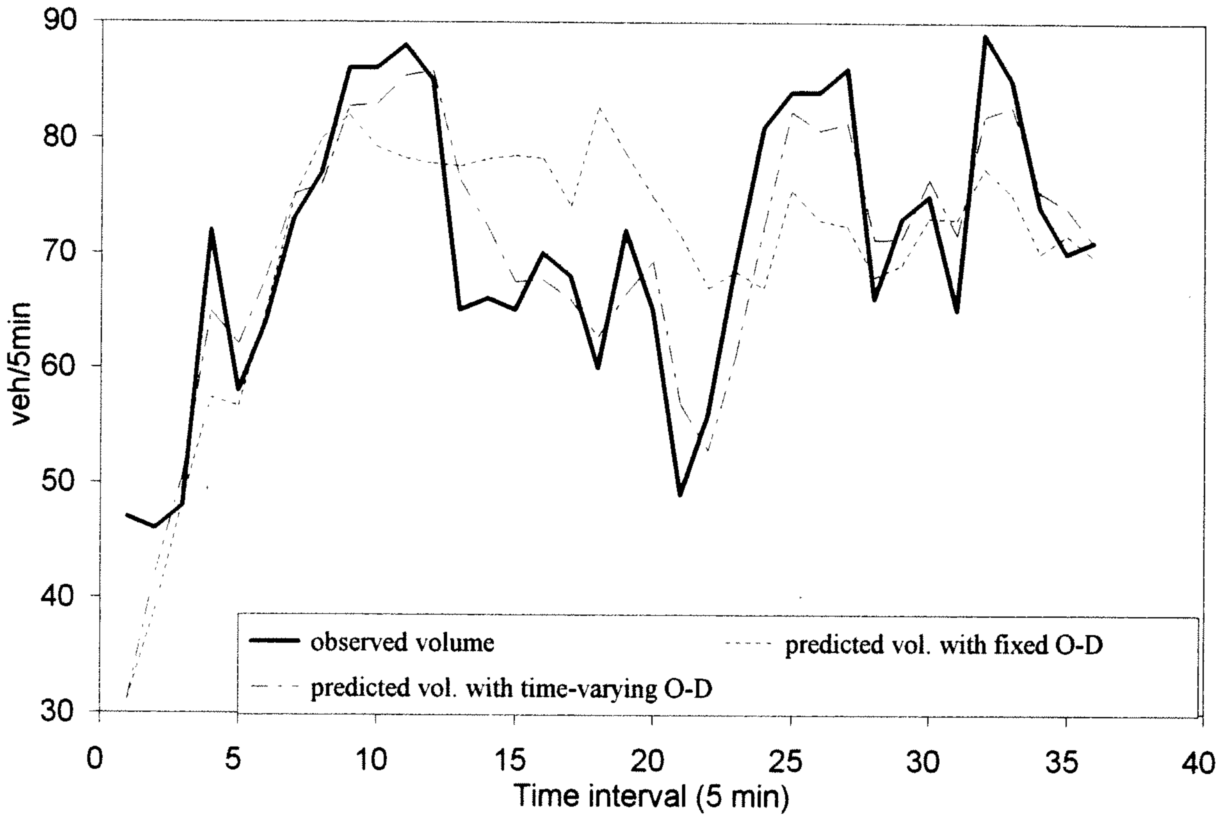


Figure 6.4. Actual Volume vs. Predicted Volume with Fixed and Time-varying O-D (off-ramp: station 61NX)

Table 6.3. Performance of the Filtering Model and Zero-mean Test (field data)

	MAD[c]	$m_e(i)$	$\tau(i)$	H_0
m_1-A[a]	8.6	1.1	3.7	Accept
y_1-A	8.1	0.1	3.2	Accept
y_2-A	12.7	1.6	5.6	Accept
m_1-B[b]	8.6	1.1	3.7	Accept
y_1-B	4.3	0.6	1.8	Accept
y_2-B	10.9	1.2	4.4	Accept

Note:
a : results from **Model-A** (fixed O-D)
b : results from **Model-B** (time-varying O-D)
m_1: mainline detector station at subsection 1
y_1: destination detector station at off-ramp
y_2: destination detector station at subsection 3
c : Mean absolute error $\sum_{k-1,N} | y(k)-\hat{y}(k) | /N$, where N is the number of measured points

Table 6.4. Normalized Covariance Coefficients of the Innovation (field data)

(Table 6.4-A): Model-A (fixed O-D proportions)

	1[a]	2	3	4	5	6	7	8	9	10
m_1[b]	-0.25[c]	0.26	-0.26	0.15	-0.08	0.04	0.03	-0.19	0.14	-0.10
y_1	0.52	0.30	0.27	0.01	0.03	-0.05	0.02	0.00	-0.19	-0.18
y_2	0.14	-0.34	-0.14	0.18	0.07	-0.13	0.11	0.04	-0.17	0.07

(Table 6.4-B): Model-B (time-varying O-D proportions)

	1[a]	2	3	4	5	6	7	8	9	10
m_1[b]	-0.24[c]	0.25	-0.25	0.14	-0.09	0.03	0.04	-0.18	0.14	-0.09
y_1	0.16	-0.12	0.03	-0.24	-0.11	-0.06	0.20	0.20	-0.08	-0.02
y_2	-0.04	-0.27	-0.07	0.03	-0.17	-0.32	0.24	0.16	-0.14	0.22

Note:
a: covariance lag
b: measurement stations
c: critical value for each cell = $\pm 1.96/\sqrt{N}$ $(N=36) \approx 0.33$

7. SUMMARY AND CONCLUSIONS

7.1 Summary

In Chapter 1 it was argued that successful Advanced Traffic Management Systems will need to not only predict the effects of management actions, but also quantify the uncertainty associated with those predictions. Ignoring this uncertainty is likely to produce overly optimistic expectations for an ATMS, while explicitly considering this uncertainty, through the use of stochastic rather than deterministic, optimization methods, could result in a genuine improvement of traffic system performance. Although a considerable effort is now underway nationwide to develop simulation and prediction models for possible use in ATMS, relatively little effort is being devoted to either determining the initial conditions these models will need if they are to be used in real-time, or to quantify these models' prediction uncertainties. Chapters 3 and 4 described a stochastic, macroscopic model of freeway traffic flow constructed using the theory of Markov population processes, and it was argued that this model possesses two important practical features. First, the model supports in a natural way the use of Kalman filtering to estimate in real-time the inputs required by more detailed freeway simulation models, such as KRONOS and FREESIM. Second, the stochastic model can be used as a predictor in its own right and when so used it not only generates point predictions but also computes the variance (uncertainty) associated with those predictions. Thus it can support the use of stochastic optimization methods in the optimization level of the hierarchical traffic control model depicted in Figure 1.1.

Chapters 5 and 6 described a number of tests conducted to evaluate the performance of freeway filtering and prediction models, and to test several methods for estimating the model's parameters using traffic data. Using Monte Carlo simulation it was found that there is a trade-off between the statistical efficiency of a parameter estimation procedure, as measured by the estimator's standard error, and the numerical efficiency, as measured by the CPU time needed to compute set of parameter estimates. In fact, the estimation method with the best statistical properties, quasi-maximum likelihood, was so slow that its use in a real-time context is questionable. However, sensitivity tests showed that the choice of parameter estimator has little effect on the accuracy of the filtering and prediction models. This is especially promising since the two-step nonlinear least-squares estimator can be readily adapted for on-line use. Monte Carlo simulation also indicated that filtered estimates of destination-specific traffic densities can be generated from loop detector counts, so that the model is capable of generating real-time estimates of the initial conditions needed by other simulation and predictions models, and also that it can generate reasonable short term predictions of freeway mainline and off-ramp volumes. Additional tests with field data produced an interesting finding concerning the relation between estimated demand parameters and the accuracy of off-ramp volume predictions. Predictions generated using a constant freeway origin-destination (O-D) matrix showed discernable flaws, and these flaws vanished when the constant O-D matrix was replaced by real-time estimates generated using the extended Kalman filter.

7.2 Conclusions and Future Work

The major conclusions of this research are:

(1) Markov population models, coupled with a large population approximation can be used to construct a tractable, yet accurate stochastic model which integrates freeway traffic flow and travel demand effects in a natural way.

(2) The flow-density relation incorporating both upstream and downstream densities provides a realistic yet tractable implementation of the simple-continuum model of traffic flow.

(3) The application of Kalman filtering theory to the above models produced an algorithm for real-time estimation of both the destination-specific traffic densities in freeway sections as well as the freeway's origin-destination matrix.

(4) The filtering and prediction model appears to be reasonably robust with regard to the method used in estimating traffic flow and origin-destination parameters.

At present, it is recommended that future work on this problem follow two lines of investigation. First, the freeway model described in this report should be developed into a fully real-time, implementable version. This will require developing an on-line version of the two-step nonlinear least-squares parameter estimator, and then testing the model on longer, more realistic segments of freeway. For real-time parameter estimation, only the on-line version of the traffic-flow parameter estimator needs development, as the adaptive O-D estimator described in Chapter 6 provides real-time estimates of the freeway origin-destination matrix. Second, to be useful in an integrated ATMS, the stochastic modelling approach needs to be extended to networks of signalized and nonsignalized intersections. Phase II of this project will address both of these topics.

REFERENCES

Ashok, K. and Ben-Akiva, M.E. (1993) Dynamic origin-destination matrix estimation and prediction for real-time traffic management systems. in C. Daganzo (ed.) *12th International Symposium on Transportation and Traffic Theory, Amsterdam*, Elsevier Science, pp. 465-484.

Banks, J. H. (1991) The two-capacity phenomenon: some theoretical issues. *Transpn. Res. Rec.* **1320**, 234-241.

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

Beck, J. F. and Arnold, K. J. (1977) *Parameter Estimation in Engineering and Science.* New York, John Wiley & Sons.

Candy, J. V. (1986) *Signal Processing: The Model Based Approach.* New York, McGraw-Hill Book Company.

Cascetta, E. and Nguyen, S. (1988) A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transpn. Res.*, **22B**, 437-455.

Chang, G. L. and Wu, J. (1993) Recursive estimation of time-varying O-D Flows from traffic counts in freeway corridors. submitted to *Transpn. Rec.*

Conlisk, J. (1976) Interactive Markov chains, *Journal of Mathematical Sociology 4*, pp. 157-185

Cremer, M. and Keller, H. (1987) A new class of dynamic methods for the identification of origin-destination flows. *Transpn. Res.*, **32B**, 117-132.

Cremer, M. and May, A. (1985) *An Extended Model of Freeway Traffic.* Research Report UCB-ITS-RR-85-7, Inst. of Transp. Studies, University of California, Calif.

Cremer, M. and Papageorgiou, M. (1981) Parameter identification for a traffic flow model. *Automatica*, **17(6)**, 837-843.

Daganzo, C. F. (1994) The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transpn. Res.*, **28B**, 269-287.

Davis, G. (1989) *A Stochastic Dynamic Model of Traffic Generation and its Application to the Maximum Likelihood Estimation of Origin-Destination Parameters.*, Doctoral Dissertation, Dept. of Civil Engineering, Univ. of Washington, Seattle, WA.

Davis, G. (1992) Integrated traffic assignment and flow models vial Markovian networks. *Paper Presented at 33rd ORSA/TIMS Joint National Conference*, Orlando, FL.

Davis, G. (1993) Estimating freeway origin-destination parameters and impact of uncertainty on ramp control. *ASCE J. of Transportation Engineering*. Vol. **119**, No. 4, Jul./Aug., 489-503.

Davis, G. and Kang, J. G. (1994) Estimating destination-specific traffic densities on urban freeways for advanced traffic management. *Transpn. Res. Rec.*, **1457**, 143-148.

Davis, G. and Kang, J. G. (1993) Filtering and prediction on freeways using Markov models. *Paper Presented at Pacific Rim Transtech Conference*, Seattle, WA.

Eisenfield, J. (1985) Remarks on Bellman's structural identifiability. *Math. Biosci., 77*, pp.229-243.

Drake, J., Scofer, J. and May, A. D. (1965) A statistical analysis of speed-density hypotheses: *Proc. of Third International Symposium on Theory of Traffic Flow*, New York.

Edie, L. C. (1961) Car-following and steady-state theory for noncongested traffic. *Oper. Res.*, **9(1)**, 66-76.

Eisenfield, J. (1986) A simple solution to the compartmental structural-identifiability problem. *Math. Biosci., 79*, pp. 209-220.

Gabard, J. F. (1987) Traffic simulation model: microscopic models. in *Systems and Control Encyclopedia* (M. Singh, ed.) Pergamon Press, Oxford.

Gazis, D. C. and Knapp, C. H (1971) On-line estimation of traffic densities from time-series of flow and speed data. *Transportation science*, **5(3)**, 283-301.

Gerlough, D. L. and Huber, M. J. (1975) *Traffic Flow Theory. Special Report 165.* TRB, National Research Council, Washington, D.C.

Gelb, A. (1974) *Applied Optimal Estimation.* Analytic Sciences, Cambridge, MA.

Ghosh, D. and Knapp, C. H. (1978) Estimation of traffic variables using a linear model of traffic flow. *Transpn. Res.*, **12**, 395-402.

Goodwin, G. C. and Sin, K. S. (1984) *Adaptive Filtering Prediction and Control.* Prentice-Hall,INC., Englewood Cliffs.

Greenshields, B. D. (1934) A study of traffic capacity: *Proc. Highway Res. Board*, **14**, 448-477.

Greenberg, H. (1959) An analysis of traffic flow. *Oper. Res.*, **7(1)**, 79-85.

Grewal, M. and Payne, H. (1976) Identification of parameters in a freeway traffic model. *IEEE Trans. on Systems, Man and Cybernetics*, **SMC-6**, 176-185.

Halati, A, Torres, J. J., Cohen, S. L. (1990) Freesim-Freeway Simulation Model. *Paper Presented at TRB Annual Meeting*, Washington, D.C..

Hsu, P. and Banks, J. (1993) Effects of location on congested-regime flow-concentration relationships for freeways. *Transpn. Res. Rec.*, **1398**, 17-23.

Ingenbleek, J. F. and Lefevre, C. (1985) A Discrete time model of interactive diffusion. *J. of Mathematical Sociology*. **11**, 25-42.

IVHS America (1991) *Proc. 1st Ann. Mtg. IVHS America*, Reston, Virginia, 17-20 March.

Jacquez, J. A. (1972). *Compartmental Analysis in Biology and Medicine. Kinetics of Distribution of Tracer-labeled Materials*. Elsevier Pub. Co., New York.

Jacquez, J. A. (1985) *Compartmental Analysis in Biology and Medicine*. 2nd ed., Univ. of Michigan Press.

Jacquez, J. A. and Greif, P. (1984) Numerical parameter identifiability and estimability: integrating identifiability, estimability, and optimal sampling design. *Math. Biosci. 77*, pp. 201-227.

Karmeshu and Pathria, R. K. (1981) A stochastic model for highway traffic. *Transportation Research*. **15B**, pp. 285-294.

Kühne, R. D. (1991) Freeway control using a dynamic traffic flow model and vehicle re-identification techniques. *Paper Presented at TRB Annual Meeting*, Washington, D.C..

Kurkjian, A., Gershwin, S., Houpt, P., Willsky, A. and Chow, E. (1980) Estimation of roadway traffic density on freeways using presence detector data. *Transpn. Science*, **14**, 232-261.

Kurtz, T. G. (1978) Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and Their Applications*. **6**, 223-240.

Leo, C. J. and Pretty, R. L. (1992) Numerical simulation of macroscopic continuum traffic flow models. *Transpn. Res.*, **26B**, 207-220.

Lehoczky, J. (1980) Approximations for interactive Markov chains in discrete and continuous time. *Journal of Mathematical Sociology*, **7**, 139-157.

Ljung, L. (1987) *System Identification: Theory for the User*. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Ljung, L. and Söderström, T. (1983) *The Theory and Practice of Recursive Identification*. Cambridge, MA, MIT Press.

Luenberger, D. G. (1966) Observers for multivariable system. *IEEE Trans. Autom. Control*. **AC-11**, 190-197.

Luenberger, D. G. (1971) An introduction to observers. *IEEE Trans. Autom. Control*. **AC-16**, 596-602.

Mahmassani, H. and Chang G-L (1987) On boundedly-rational user equilibrium in transportation systems. *Transportation Science*, **21**, 89-99.

Mahmassani, H. and Jayakrishnan, R (1990) System performance and user response under real-time information in a congested traffic corridor. *Transpn. Res.*, in print.

May, A. D. (1987) Freeway simulation models revisited. *Transpn. Res. Rec.*, **1132**, 94-99.

May, A. D. (1990) *Traffic Flow Fundamentals*. Prentice-Hall,INC., Englewood Cliffs.

Michalopoulos, P., Kwon, E. and Kang, J. G. (1991) Enhancement and field testing of a dynamic freeway simulation program. *Transpn. Res. Rec.*, **1320**, 203-215.

Nahi, N. E. and Trivedi, A. N. (1973) Recursive estimation of traffic variables: section density and average speed. *Transpn. Science*, **7**, 269-286.

Nihan, N. and Davis, G. (1987) Recursive estimation of origin-destination matrices from input/output counts. *Transpn. 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. *Transpn. Science*, **23**, 77-90.

Nguyen, S. (1984) Estimating origin-destination matrices from observed flows. In M. Florian (Ed.) T*ransportation Planning Models*. Elsevier Publishers, 363-380.

Numerical Algorithms Group (1986) *NAG workstation library, Version I.* Oxford, UK.

Okutani, I. (1987) The Kalman filtering approaches in transportation and traffic problems, *Transportation and Traffic Flow Theory (Gartner, N.H. and Wilson, N.H.M., Editors),* Elsvier Science Publishing Co., Inc.

Papageorgiou, M. (1983) *Application of Automatic Control Concepts to Traffic Flow Modeling and Control.* Berlin, Springer-Verlag.

Papageorgiou, M. and Messmer, A. (1991) Dynamic network traffic assignment and route guidance via feedback regulation. *Transpn. Res. Rec..* **1306**, TRB, National Research Council, Washington, D.C., 49-58.

Payne, H.J. (1979) FREFLO: A macroscopic simulation model of freeway traffic. *Transpn. Res. Rec.*, **722**, 68-77.

Payne, H.J., Brown, D, and Cohen, S.L. (1987) Improved techniques for freeway surveillance. *Transpn. Res. Rec..* **1112**, TRB, National Research Council, Washington, D.C., 52-60.

Rathi, A. K., Lieberman, E. B. and Yedlin, M. (1987) Enhanced FREFLO: Modeling of Congested Environments. *Transpn. Res. Rec..* **1112**, TRB, National Research Council, Washington, D.C., 61-71.

Ross, P. (1988) Traffic dynamics. *Transportation Research*, **22B**, 421-435.

Rissanen, J. and Caines, P. E. (1979) The strong consistency of maximum likelihood estimators for ARMA processes. *Annals of Statistics.* **7**, 297-315.

Seber, G. and Wild, C. (1989) *Nonlinear Regression.* Wiley and Sons, New York.

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

Smulders, S. (1987) Modeling and filtering of freeway traffic flow. *Transportation and Traffic Flow Theory (Gartner, N.H. and Wilson, N.H.M., Editors),* Elsvier Science Publishing Co., Inc.

Stephanedes, Y. and Kwon, E. (1993) Adaptive demand-diversion prediction for integrated control of freeway corridors. *Transpn. Res.*, **1C**, 23-42.

Stephanedes, Y. and Chang, K. (1993) Optimal control of freeway corridors, ASCE *Journal of Transpn. Engineering.* Vol. **119**, No. 4, Jul./Aug., 504-514.