REPORT NO. DOT-TSC-RSPA-78-9

ESTIMATION OF TRAFFIC VARIABLES USING POINT PROCESSING TECHNIQUES

R. L. Lopez
P. K. Houpt

Massachusetts Institute of Technology
Electronic Systems Laboratory
Cambridge MA 02139

MAY 1978

FINAL REPORT

DOCUMENT IS AVAILABLE TO THE U.S. PUBLIC THROUGH THE NATIONAL TECHNICAL INFORMATION SERVICE, SPRINGFIELD, VIRGINIA 22161

Prepared for

U.S. DEPARTMENT OF TRANSPORTATION
RESEARCH AND SPECIAL PROGRAMS ADMINISTRATION
Office of Transportation Programs Bureau
Office of Systems Engineering
Washington DC 20590

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT-TSC-RSPA-78-9		

4. Title and Subtitle
ESTIMATION OF TRAFFIC VARIABLES USING POINT PROCESSING TECHNIQUES

5. Report Date
May 1978

6. Performing Organization Code

7. Author(s)
R. L. Lopez and P. K. Houpt

8. Performing Organization Report No.
DOT-TSC-RSPD-78-9

9. Performing Organization Name and Address
Massachusetts Institute of Technology*
Electronic Systems Laboratory
Department of Electrical Eng. and Computer Science
Cambridge MA 02139

10. Work Unit No. (TRAIS)
OS750/R8526

11. Contract or Grant No.
DOT-TSC-849

12. Sponsoring Agency Name and Address
U.S. Department of Transportation
Research and Special Programs Administration
Office of Transportation Programs Bureau
Office of Systems Engineering, Washington DC 20590

13. Type of Report and Period Covered
Final Report
April 1976 - April 1977

14. Sponsoring Agency Code

15. Supplementary Notes
*Under contract to: U.S. Department of Transportation
Transportation Systems Center
Kendall Square
Cambridge MA 02142

16. Abstract

An alternative approach to estimating aggregate traffic variables on freeways--spatial mean velocity and density--is presented.

Vehicle arrival times at a given location on a roadway, typically a presence detector, are regarded as a point or counting Poisson process whose rate is a function of the state of the traffic at every instant of time. Moreover, the traffic state is modeled as a finite-state Markov chain. A sequential point process filter, optimum in the mean-squared error sense, is deisgned to estimate the state from observations of the vehicle arrival-time sequence. Different possibilities for incorporating potential additional information such as speed and headway are explored.

Parameter values for the underlying Markov chain are obtained via a maximum likelihood estimator.

Qualitative behavior of the proposed algorithms is studied with simulated traffic flow data from both macroscopic and microscopic models.

17. Key Words
Estimation, Traffic, Variables, Point Process, Poisson Process, Markov Chain, Filter, Speed, Headway, Maximum Likelihood

18. Distribution Statement

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	80	

Form DOT F 1700.7 (8-72)

Preface

This report is based on the S.M. thesis by Raphael Lopez-Lopez, "Estimation of Traffic Variables using Point Process Techniques," submitted to the M.I.T. Department of Electrical Engineering and Computer Science in February, 1977.

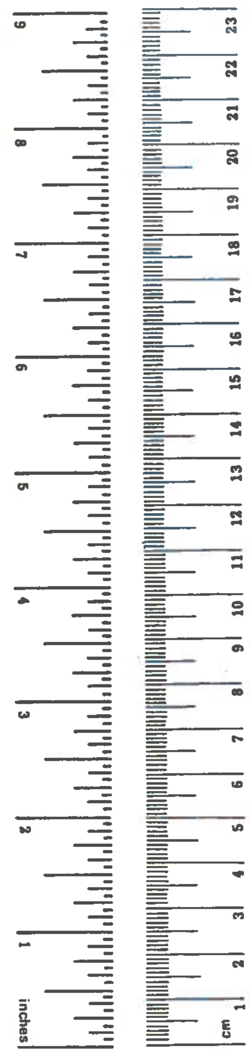
The research was conducted at the M.I.T. Electronic Systems Laboratory with support provided by the U.S. Department of Transportation under contract DOT-TSC-849. The authors gratefully acknowledge the criticisms, remarks, and support provided by Dr. Diarmuid O'Mathuna of the Transportation Systems Center.

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

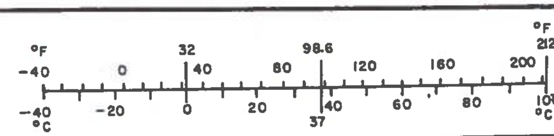
Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



CONTENTS

CONTENTS (Con t)

FIGURES

Figure

FIGURES (Con t)

Figure

Tables

1. INTRODUCTION

A great deal of research on roadway traffic control and surveillance, of which estimation is a very important aspect, is being conducted at present.

To control the state of the traffic, or to be able to detect the occurrence of special situations such as accidents -- which may produce changes in capacity and in other roadway parameters -- it is necessary to know the values of the relevant traffic variables at any instant of time on a given section of the roadway.

1.1 EXISTENT TRAFFIC MODELS

From an analogy to fluid hydrodynamics, a macroscopic model of traffic dynamic behavior has been proposed by Isaksen and Payne [7]. This model consists of partial differential equations, and is, therefore, infinite-dimensional. To be able to apply the highly developed tools of modern control theory (e.g., linear quadratic Gaussian feedback control), it is necessary to simplify the model to obtain a finite-dimensional description of the phenomenon. This is accomplished by dividing the roadway into sections and defining appropriate ensemble averages of the traffic variables for each section (see figure 1.1). With such an approximation, a finite-dimensional model consisting of nonlinear ordinary differential equations on the aggregate variables can be derived, as summarized in [1,10].

The interpretation of these aggregate variables is shown in figure 1.2.

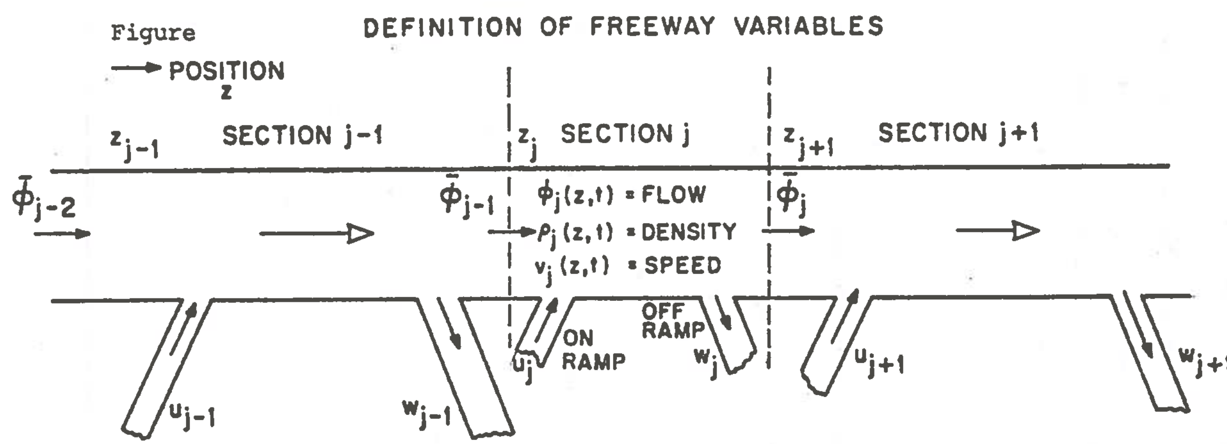
Summarizing, the variables of interest for this model are:

$$\begin{aligned} \text{Flow} \quad & \phi \quad \text{[vehicles/time]}, \\ \text{Density} \quad & \rho \quad \text{[vehicles/distance], and} \\ \text{Velocity} \quad & v \quad \text{[distance/time]}. \end{aligned} \tag{1.1}$$

Flow, velocity, and density are related by the basic equation [1]:

$$\phi = v\rho \tag{1.2}$$

and for steady-state conditions, an empirical relationship between aggregate velocity and density, known as fundamental diagram of traffic, has been observed [8]. See figure 1.3.



CONTINUUM VARIABLES:

$\phi_j(z,t)$ = veh flow (veh/hr) at pos z on section j

$\rho_j(z,t)$ = veh density (veh/mi) at pos z on section j

$v_j(z,t)$ = veh speed (mi/hr) at pos z on section j

$u_j(z,t)$ = entering (on ramp) flow (veh/hr) on link j

$w_j(z,t)$ = exiting (off ramp) flow (veh/hr)

SECTION "MEAN" VARIABLES: $t \in [t, t + \Delta T]$; $\Delta z_j = z_{j+1} - z_j$

$$\bar{\phi}_j(z) \triangleq \frac{1}{\Delta T} \int_t^{t + \Delta T} \phi_j(z,\tau)\, d\tau$$

$$\bar{\rho}_j(t) \triangleq \frac{1}{\Delta z_j} \int_{z_j}^{z_{j+1}} \rho_j(z,t)\, dz$$

$$\bar{v}_j(t) \triangleq \frac{1}{\Delta z_j} \int_{z_j}^{z_{j+1}} v_j(z,t)\, dz$$

FIGURE 1.1 DEFINITION OF TRAFFIC VARIABLES

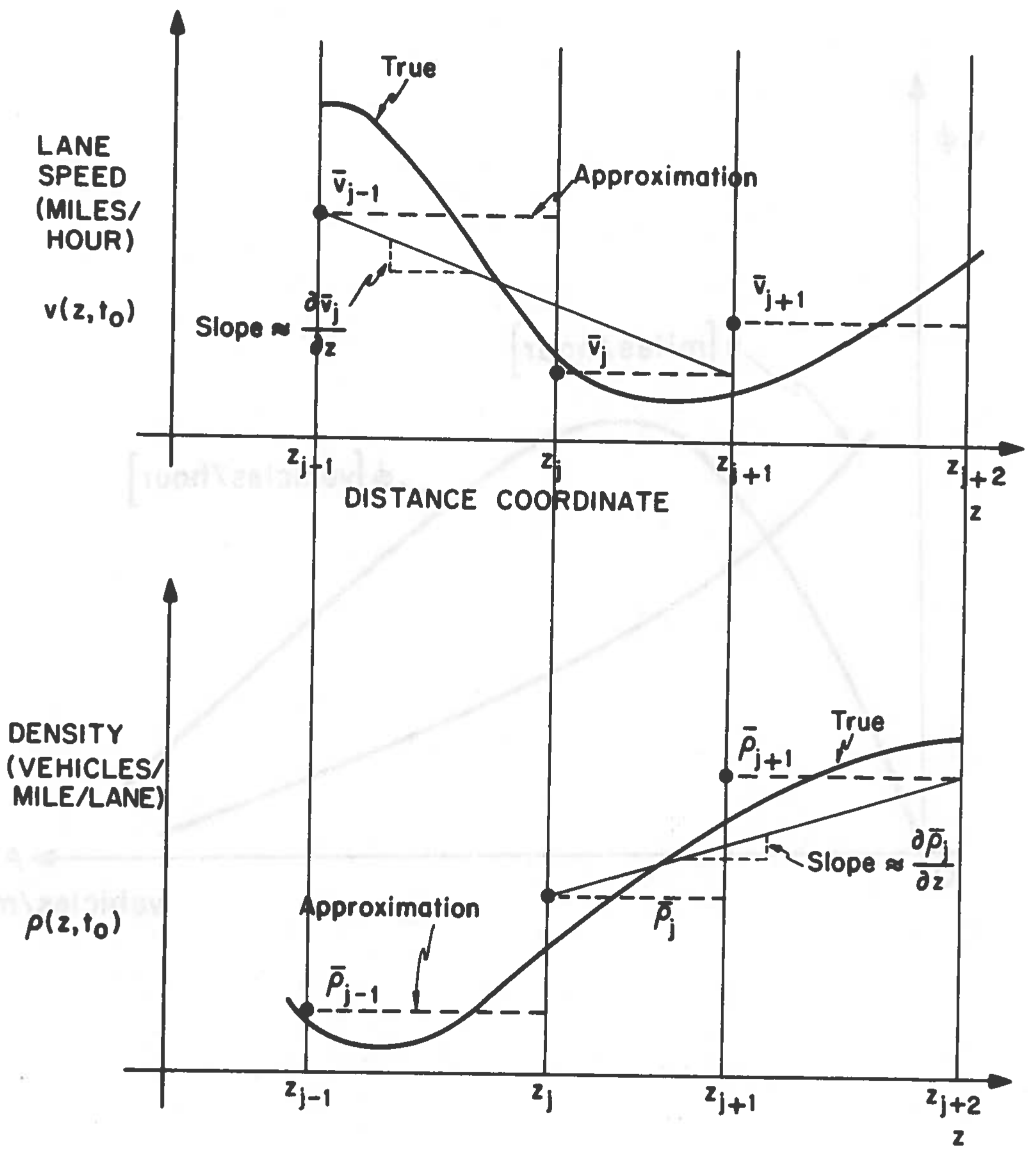


FIGURE 1.2 INTERPRETATION OF SPATIAL-MEAN SECTION VARIABLES IN DYNAMIC MODEL

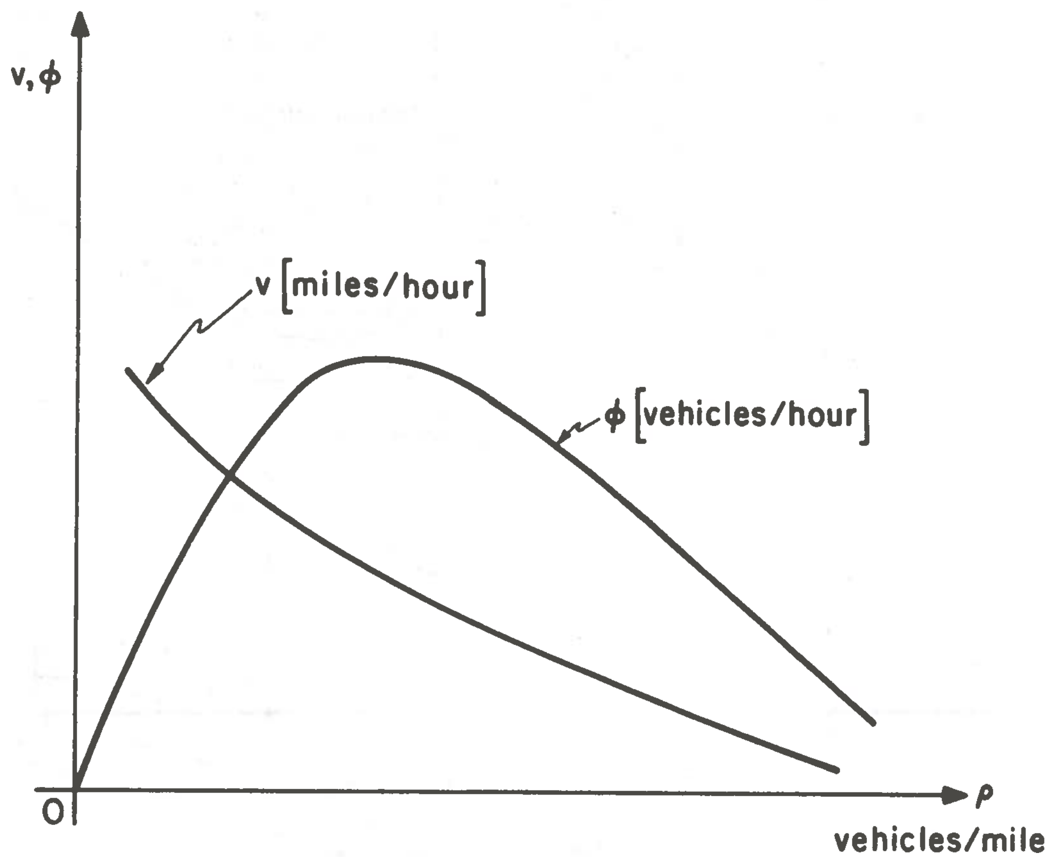


FIGURE 1.3 FUNDAMENTAL DIAGRAM OF TRAFFIC

Since the ultimate goal is to control the traffic flow, our main interest is to obtain estimates of the aggregate variables, and this work is directed toward that end.

It is clear, in view of equation (1.2), that the state of the roadway traffic is determined by knowledge of any two of the three variables (1.1).

1.2 DETECTORS

Monitoring roadway traffic is in most cases accomplished by means of loop detectors placed along the road. A loop detector (schematically shown in figure 1.4) is basically a tuned circuit which is detuned each time a car passes over it, thus producing a signal registering the event (figure 1.5a). This signal is passed through a threshold detector and sampled at a uniform rate (figure 1.5b).

The fraction of time that a sensor is activated over a period of time is called occupancy. Thus, for a T seconds interval:

$$OCC = \frac{\text{number of "on" samples in T}}{\text{total number of samples in T}} . \qquad (1.3)$$

This variable, usually expressed in percent, gives an approximate measure of the density in the roadway.

Although more sophisticated monitoring methods such as radar, ultrasound, television cameras, and more complex roadway detectors have been suggested, it is likely that loop detectors will prevail due to their simplicity and ease of installation and maintenance. Therefore, we will focus our attention on data obtained through this medium.

1.3 PREVIOUS ESTIMATION SCHEMES

Traditional estimation schemes [2,3,4,5,12,13] make use of initially noisy measurements or rough estimates of variables such as velocity and flow, which are then passed through some sort of filter to obtain more accurate estimated values of the state. These noisy measurements are computed from presence (loop) detectors, and some strong assumptions have to be made to obtain estimates. For instance, a standard length of vehicles and a constant loop detector size are hypothesized [12,13] to obtain a rough estimate of the speed of the vehicle passing over the sensor. This estimate is:

$$v = \frac{d + \ell}{\tau}$$

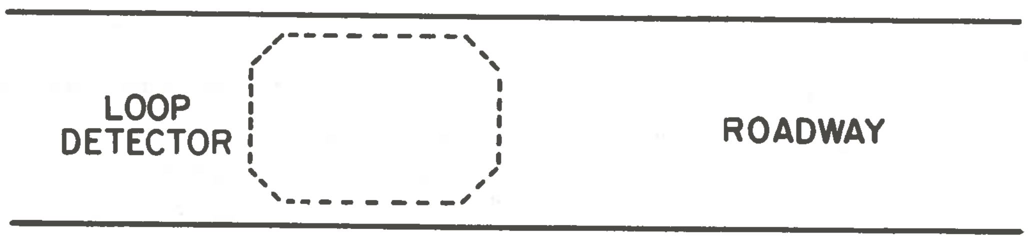


FIGURE 1.4 SCHEMATIC REPRESENTATION OF LOOP DETECTOR

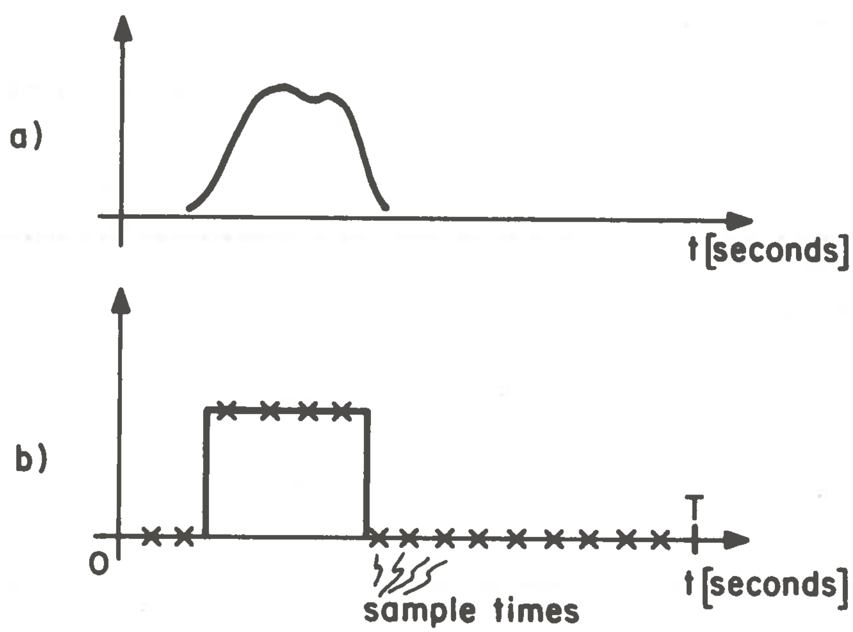


FIGURE 1.5 SIGNAL FROM DETECTOR

a) Analog signal

b) Signal after threshold detection

where

d = vehicle length (standard),

ℓ = loop detector length, and

τ = time over detector,

The quality d + 1 is known as effective length (figure 1.6).

Note that the estimate (1.4) is biased because d is not the actual vehicle length but an assumed constant for all vehicles. Furthermore, the time over detector, τ, is not known exactly due to the sampling intervals because of restricted data handling capabilities.

Kalman filtering techniques [4,13] have been proposed for estimating the traffic variables from the initial noisy measurements mentioned above. This approach requires making approximations at the modeling stage (deriving a finite-dimensional traffic flow model), at the filtering stage (linearizing the traffic model at each step, in what is known as extended Kalman filtering), and at the measuring stage (data for these filters are not directly available from detectors, and some approximations, as in (1.4), have to be used).

Also, the extended Kalman filter is a suboptimal estimator.

1.4 NEW FILTERING APPROACH

As noted above, velocity and density are not directly measured in real-life systems. All that is available is the vehicle-counting process.

Therefore, it seems a logical step to try to design some kind of filter whose input is the point process itself (arrival times), and whose outputs are accurate estimates of the variables of interest.

The procedure is as follows:

Traffic flow is modeled as a doubly stochastic Poisson process (to be defined in the next section), and a recursive filter is used to estimate the state of the traffic from the arrivals of vehicles at a sensor. This filter is very similar in structure to the Kalman filter as we shall see below. Furthermore, the point process filter that will be used here is optimum in the mean-square sense.

Thus, this approach requires making no approximations for the filtering or measuring stages. Approximation is done only at the modeling stage (probabilistic model for arrival process).

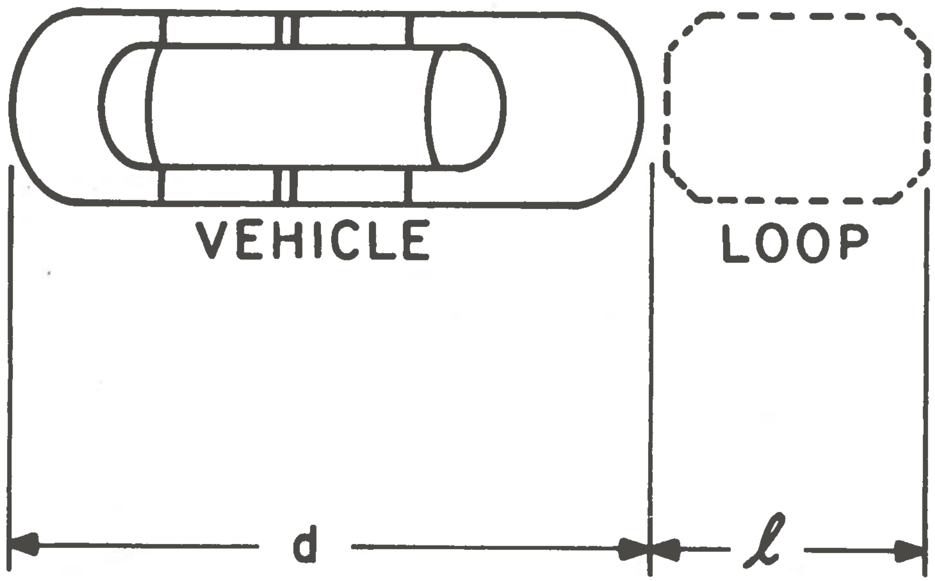


FIGURE 1.6 EFFECTIVE LENGTH

of Vehicle and Detector

Summarizing, the objective of this work will be to use a point process approach (as outlined in [6,14,16,18], for instance) in designing a filter to determine velocity and density from counting of vehicles.

1.5 APPLICATIONS TO SIMULATION AND IMPROVED DYNAMIC MODELS

Modeling the traffic flow as a point process is useful for simulation purposes also. Typically, for simulation of microscopic models, it is assumed that the arrivals of vehicles behave as a homogeneous Poisson process. A more realistic model might be what is termed as doubly stochastic Poisson process [17,18]; that is, a Poisson process with intensity parameter which is itself a random process (in particular, a finite state Markov chain for the purposes of this work); i.e., the Poisson intensity parameter is varying in accordance to the traffic conditions, as will be studied here.

2. MATHEMATICAL MODELS

In this section, we introduce and give a precise definition of the point process models that will be considered here, together with the recursive filter algorithms for the estimation of the traffic state.

2.1 DOUBLY STOCHASTIC POISSON PROCESSES

In modeling a traffic system, one is faced with a number of uncertainties that result from the particular behavior of drivers on the road, their desired speeds and destinations, their moods, the weather conditions, the time of the day, etc. Also it is clear that traffic on weekdays is different from traffic on holidays or weekends.

Therefore, if one is to describe the traffic behavior (the arrival times of vehicles at a given point) by a Poisson process, it is clear that the intensity of this process (in cars per second for instance) should be also varying in a random fashion. This is in essence the nature of a doubly stochastic Poisson process: the intensity of the process is randomized by allowing it to be influenced by an exterior stochastic process -- in our case the traffic state. We give now a formal definition

2.1.1 Definition

A doubly stochastic Poisson process ia a Poisson process $\{N_t, t \geq t_0\}$ with intensity $\{\lambda_t(x_t), t \geq t_0\}$, where $\{x_t, t \geq t_0\}$ is itself a random process.

x_t is known as the information process, and it corresponds to the state of the traffic $(v(t), \rho(t))$ at time t.

N_t is the vehicle arrival process or point process (which is directly measured by the sensors).

λ_t, the intensity of the point process, is a function of the state of the traffic; i.e. $\lambda_t(x_t) = \lambda(v(t), \rho(t))$.

It is worth noting that the above terminology is not completely standard. We are following Snyder's [18] closely. From this definition, N_t is a conditionally Poisson process given the state; i.e., if $x_t = j$ for $t_1 \leq t < t_2$, the process is Poisson with intensity $\lambda_t(j)$:

$$\Pr\{k \text{ arrivals in } [t_1, t_2) \mid x_t = j,\ t_1 \leq t < t_2\}$$

$$= \frac{[m(t)]^k}{k!} \exp[-m(t)]$$

where $$m(t) = \int_{t_1}^{t_2} \lambda_t(j)\,dt .$$

Thus, modeling the vehicle arrival process as doubly stochastic Poisson requires making the usual assumptions for a Poisson process; namely:

i) Independence of events (arrival of vehicles), and

ii) Arbitrarily small probability of two or more arrivals occurring simultaneously.

Although these are good approximations for light-traffic conditions, with vehicles moving freely and almost no interactions, it is clear that high-density traffic is fairly correlated, and arrivals do not take place so randomly. The model we propose here will have to be tested against such conditions to determine its range of validity.

2.1.2 Nature of Information Process

The name given to x_t is due to the fact that this variable is, in most instances, the quantity one is interested in because it conveys desired information as the traffic state does in our application. At this stage, we are faced with the problem of hypothesizing a convenient structure for x_t, a structure which makes sense physically, and at the same time, enables us to obtain mathematically tractable filtering expressions. With this in mind, we chose the following.

Hypothesis: The stochastic process x_t forms a finite-state Markov chain. The rationale for this model is

a) There is a very strong relation between dynamic systems and Markov processes.

b) With the Markovian assumption, the resursive point process filters take an explicit form which is readily implementable in a computer.

It is clear that this involves discretizing the state space of velocity density into a finite number of regions, and associating to each of these regions, an intensity parameter that will drive the vehicle arrival process as well as transition rates which will characterize a Markov chain (see figure 2.1).

Once the way in which the space is to be discretized has been determined, we will have to choose one out of the many possible discrete-state models, specially regarding the choice of the number of Markov chain states necessary to obtain a good filter performance.

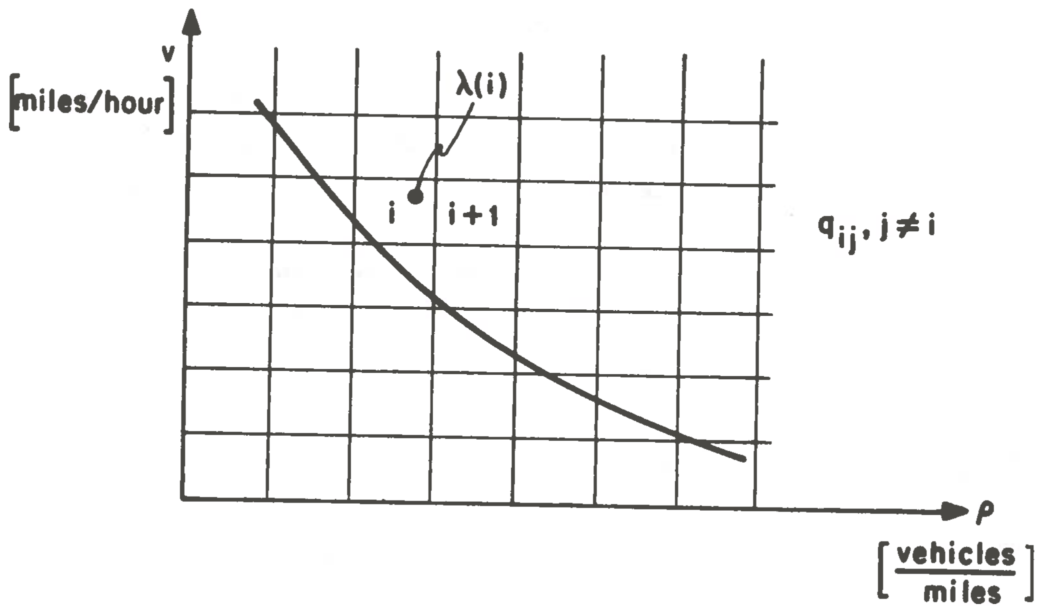


FIGURE 2.1 DISCRETIZATION OF STATE-SPACE

2.1.3 Problem Statement

As mentioned above, we want to know the state of the traffic given the arrival time point process N_t. In more precise terms, the problem is: Find the best estimate (in the mean-square sense) of the intensity parameter γ_t given all the past information on the process N_t, i.e. find $\hat{\lambda}_t$, such that

$$E\{(\lambda_t-\hat{\lambda}_t)^2\} \leq E\{(\lambda_t-\lambda_t^*)^2\} \tag{2.1}$$

for all other estimates λ_t^*; where $E\{\cdot\}$ denotes expectation.

As it is well known, this estimate is the conditional expectation (Snyder [18]) given the process N_t:

$$\hat{\lambda}_t = E\{\lambda_t | N_\sigma, \quad t_0 \leq \sigma < t\} \tag{2.2}$$

where t_0 is the intial time.

Note that this is a filtering problem. Its solution will be given by a recursive filter which updates the estimate acting on an innovations process each time a new arrival takes place.

2.1.4 Filter Equations

The Markov chain x_t is defined by a matrix $Q(t) = [q_{ij}(t)]$ known as the infinitesimal generator of x_t (see for instance Karlin [9]), where:

$$q_{ij}(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathcal{P}r\{x_{t+\Delta t} = j | x_t = i\}, \; i \neq j \tag{2.3}$$

$$q_{ii}(t) = -\sum_{j \neq i} q_{ij}(t) \; . \tag{2.4}$$

Let $\hat{\pi}_i(t)$ denote the probability of being at state i, at time t, given all the past information on the process N_t:

$$\hat{\pi}_i(t) = \mathcal{P}_r\{x_t = i \mid N_\sigma, t_0 \leq \sigma < t\} \; . \tag{2.5}$$

It can be shown [14,18] that the opitmal recursive filter has the following form:

i) For interarrival times

$$\nu_i(t) = \lambda_t(i) - \hat{\lambda}_t \qquad \text{(innovations process)}, \tag{2.6}$$

$$\frac{d}{dt} \hat{\pi}_i(t) = \sum_j q_{ij}(t)\hat{\pi}_j(t) - \hat{\pi}_i(t) \; _i(t) \tag{2.7}$$

$$\hat{\pi}_i(t_0) = \pi_i(t_0) \qquad \text{(a priori probability)} \tag{2.8}$$

ii) Update equation (σ = arrival time)

$$\hat{\pi}_i(\sigma^+) = \hat{\pi}_i(\sigma^-) \frac{\lambda_{\sigma-}(i)}{\hat{\lambda}_{\sigma-}} . \tag{2.9}$$

iii) Minimum mean-square estimate of the intensity parameter

$$\hat{\lambda}_t = \sum_j \hat{\pi}_j(t)\lambda_t(j) \tag{2.10}$$

Note that (2.7) is nonlinear. Nevertheless it can be solved via a set of linear equations [14]:

$$\frac{d}{dt} p(t) = R^T(t)p(t), \quad \sigma_1 < t < \sigma_2 \tag{2.11}$$

$$p(\sigma_1) = \hat{\pi}(\sigma_1^+)$$

where σ_1, σ_2 are two consecutive arrival times, and the matrix $R(t) = [r_{ij}(t)]$ is given by:

$$r_{ij}(t) = q_{ij}(t), \quad i \neq j \tag{2.12}$$

$$r_{ii}(t) = q_{ii}(t) - \lambda_t(i) \quad . \tag{2.13}$$

Finally, one can compute the conditional probability of being at state i as:

$$\hat{\pi}_i(t) = \frac{p_i(t)}{\sum_j p_j(t)} \quad . \tag{2.14}$$

The implementation of this filter is shown schematically in figure 2.2.

2.1.5 Preliminary Example

Before attempting to work with real data, it was considered useful getting a qualitative picture of the point process filter features by studying a simple two-dimensional case.

For this elementary case, λ_t is a two-state, continuous-time, Markov chain which for sufficiently small Δt we could represent as in figure 2.3, where the values of the arcs are the probabilities of making a transition in Δt seconds.

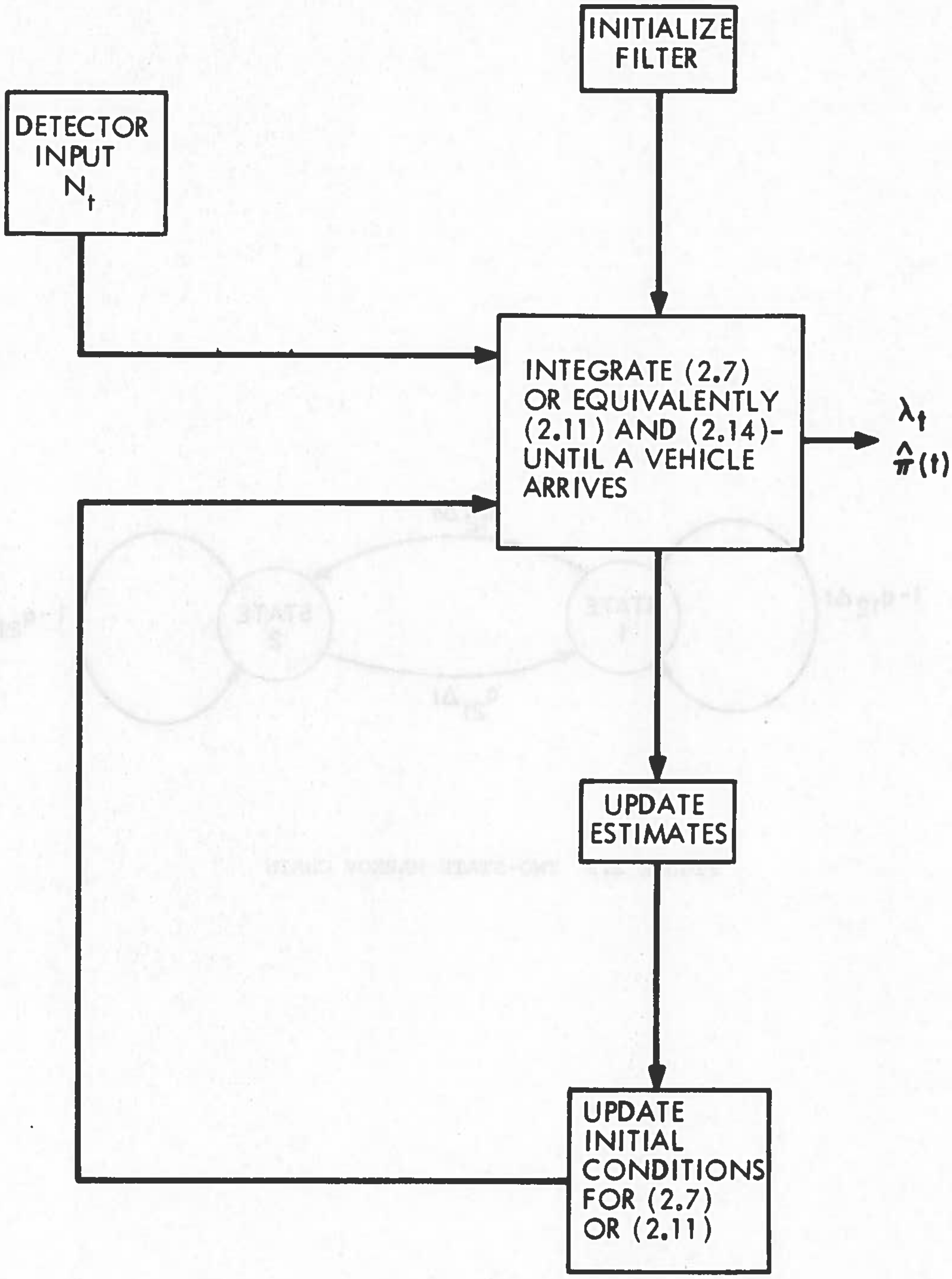


FIGURE 2.2 FILTER IMPLEMENTATION

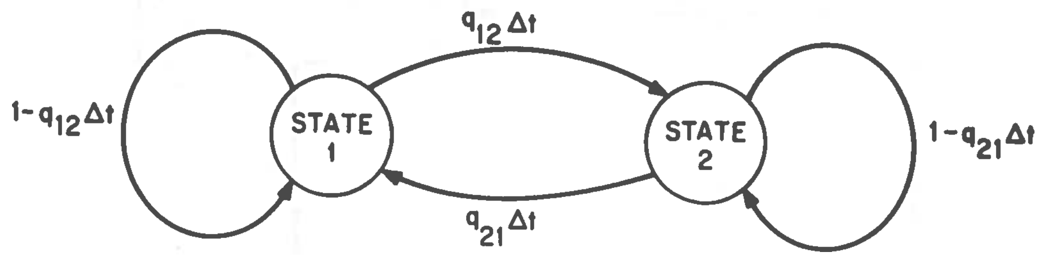


FIGURE 2.3 TWO-STATE MARKOV CHAIN

This chain is assumed to be stationary and defined by the transition rates matrix (infinitesimal generator)

$$Q = \begin{bmatrix} -q_{12} & q_{12} \\ q_{21} & -q_{21} \end{bmatrix} . \tag{2.15}$$

The arrival process N_t is a Poisson process given the state (i.e., conditionally Poisson). Therefore, for simulation purposes, we used the fact that:

$$\mathcal{P}r\ \{k \text{ counts in } [t_1,t_2)\,|\,x_\sigma=i, t_1\le\sigma<t_2\} = \frac{[\lambda(i)\ \ (t_2-t_1)]^k}{k!}\, e^{-\lambda(i)(t_2-t_1)} \tag{2.16}$$

In this case, the filter equations reduce to:

$$\left.\begin{aligned} \dot{\hat{\pi}}_1(t) &= [-q_{12}\hat{\pi}_1(t) + q_{21}\hat{\pi}_2(t)] - \hat{\pi}_1(t)[\lambda(1)-\hat{\lambda}_t] \\ \hat{\pi}_1(0) &= \pi_1(0) \\ \dot{\hat{\pi}}_2(t) &= [q_{12}\hat{\pi}_1(t)-q_{21}\hat{\pi}_2(t)] - \hat{\pi}_2(t)[\lambda(2)-\hat{\lambda}_t] \\ \hat{\pi}_2(0) &= 1 - \pi_1(0) \end{aligned}\right\} \tag{2.7'}$$

where

$\lambda(1)$ = intensity at state 1,

$\lambda(2)$ = intensity at state 2,

$\pi_1(0)$ = initial a priori probability of being at state 1

and

$$\hat{\lambda}_t = \hat{\pi}_1(t)\lambda(1) + \hat{\pi}_2(t)\lambda(2) . \tag{2.10'}$$

To solve (2.7'), we use the linear system:

$$\begin{bmatrix} \dot{p}_1(t) \\ \dot{p}_2(t) \end{bmatrix} = \begin{bmatrix} -q_{12}-\lambda(1) & q_{21} \\ q_{12} & -q_{21}-\lambda(2) \end{bmatrix} \begin{bmatrix} p_1(t) \\ p_2(t) \end{bmatrix} \tag{2.11'}$$

for $\sigma_1<t<\sigma_2$ (σ_1, σ_2 are two consecutive arrival times).

$$\left.\begin{aligned} &\begin{bmatrix} p_1(\sigma_1^+) \\ p_2(\sigma_1^+) \end{bmatrix} = \begin{bmatrix} \hat{\pi}_1(\sigma_1^+) \\ \hat{\pi}_2(\sigma_1^+) \end{bmatrix} \\ &\text{and} \\ &\hat{\pi}_1(t) = \frac{p_1(t)}{p_1(t) + p_2(t)} \\ &\hat{\pi}_2(t) = 1 - \hat{\pi}_1(t) \ . \end{aligned}\right\} \qquad (2.14')$$

Results

The filter was simulated for various combinations of values of Q and intensities $\lambda(1)$ and $\lambda(2)$ as well as for different initial probabilities $\pi_1(0)$.

The following facts were observed after the preliminary simulations:

i) When the intensities of the two states are very similar, that is, when $\lambda(2)-\lambda(1)$ is small, the filter was unable to track the true intensity adequately. This means that when performing the state-space discretization, we will have to take into account that utilizing a very fine grid may not prove very useful.

ii) The effect of the initial (a priori) probability distribution disappears after relatively short time. Thus, it will not be necessary to know exactly the state of the traffic at the beginning of the filtering process.

iii) When the chain is jumping from one state to another very fast; that is, when q_{12} and q_{21} are large as compared to $\lambda(i)$, although the filter response is slow, it is able to do a reasonable job in estimating the true value of the intensity. On the other hand, for actual traffic situations, this case is not likely to occur: traffic dynamics are relatively slow, and fast changes back and forth between light and moderate traffic conditions, for instance, do not normally take place.

Two sample outputs, both for large $\lambda(2)-\lambda(1)$, are shown. Figure 2.4 depicts the case of fast changing states, and figure 2.5 that of slowly changing states.

2.2 MARKED POINT PROCESSES

It is evident that even in the case of monitoring roadway traffic with presence detectors only, more information than the simple arrival time of a vehicle is avilable. For instance, one is also measuring headway, or time

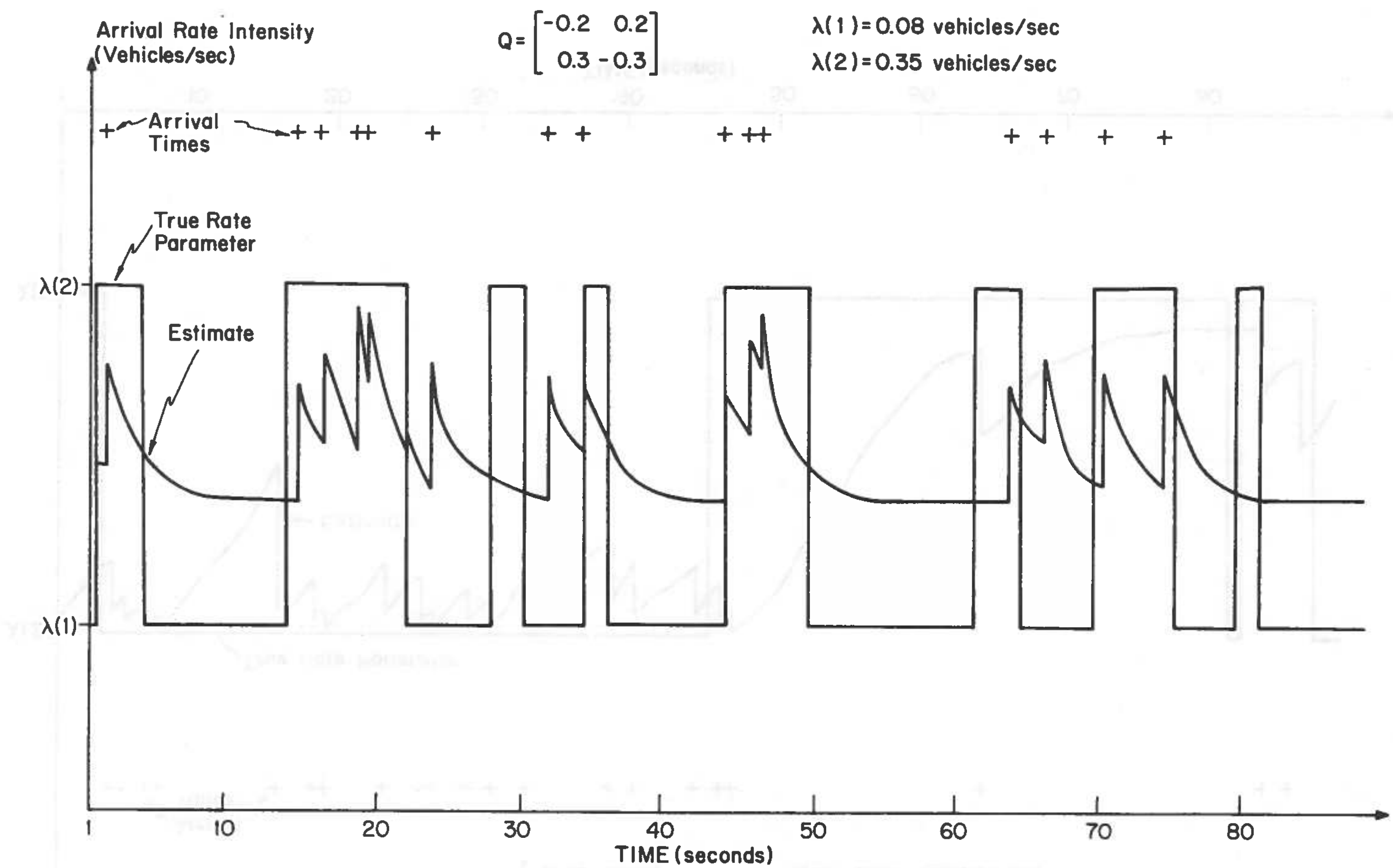


FIGURE 2.4 TWO-STATE FILTER SIMULATION

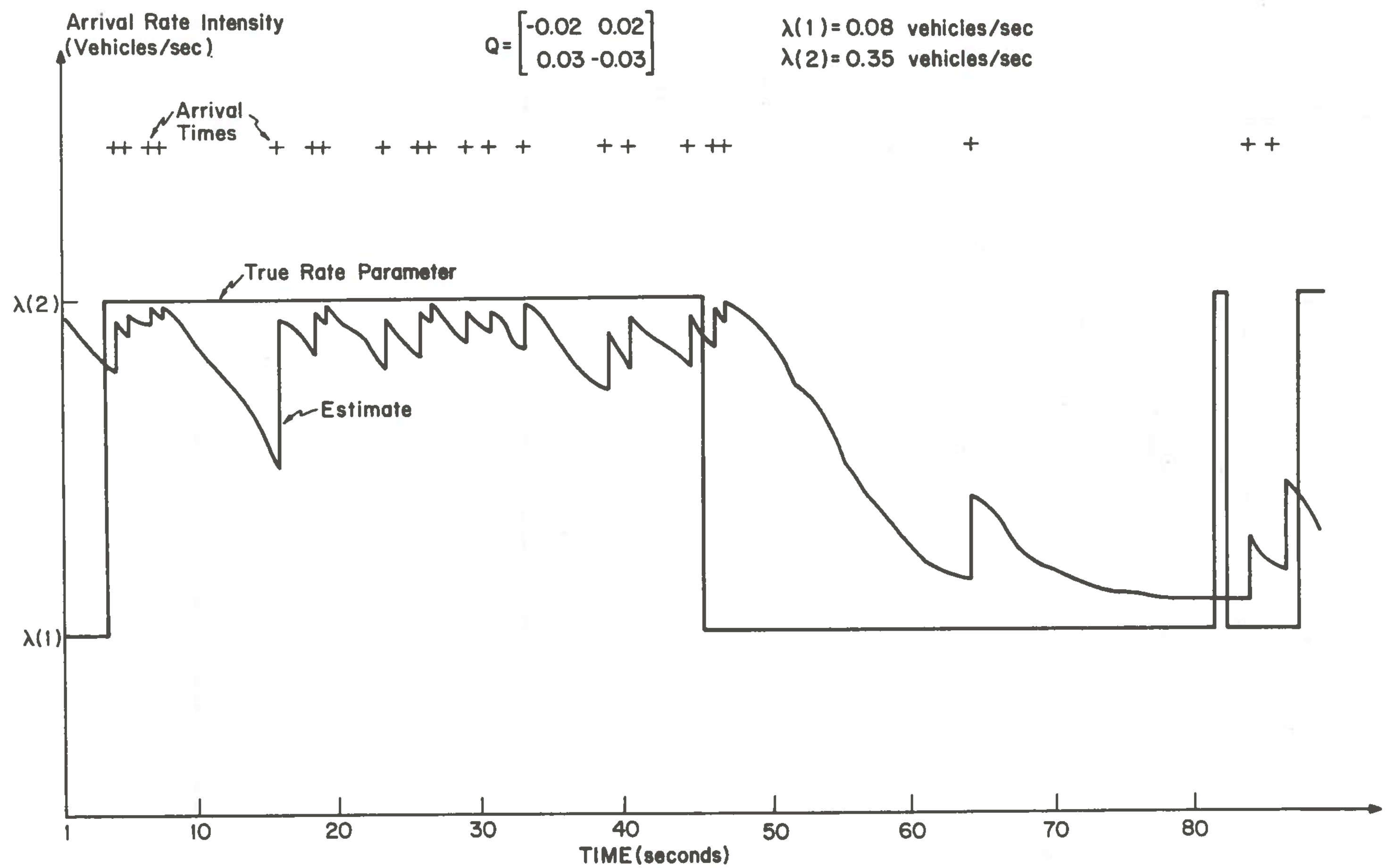


FIGURE 2.5 TWO-STATE FILTER SIMULATION (CONCLUDED)

separation between two adjacent vehicles, which under homogeneous traffic conditions is inversely proportional to the density. Velocity can also be measured using a detector pair.

This "extra information" can be incorporated into the model described in the previous section by associating with each point of the Poisson process an auxiliary variable called a mark, which could also be influenced by the information process x_t. This results in the so-called Marked point processes [18], which can be also thought of as a generalization of a compound Poisson process, in that the former do not require any assumptions of statistical independence between the process producing the marks and the point process itself.

For the purposes of this work, we will assume the marks to be contained in a denumerable set of R^n:

$$\{u_1, u_2, \ldots\} , \quad u_i \varepsilon R^n \tag{2.17}$$

where u_i is one of the possible marks. For instance, we can associate a mark $u_i \varepsilon R^2$,

$$u_i = \begin{bmatrix} v_i \\ h_i \end{bmatrix} \tag{2.18}$$

with every arrival time, v_i and h_i being discretized versions of the velocity and headway, respectively, of the particular vehicle crossing the detector at a given time. The discretization is necessary to have a denumerable mark space.

The intensity of the process now depends not only on the state of the Markov chain x_t, but on the particular mark u_i also. This will be denoted as $\lambda_{t,i}(j)$; i.e. the intensity for mark u_i when $x_t = j$.

The quantity $\lambda_{t,i}(j)$ can be given the following interpretation: Given that the information process x_t is at state j, the probability of there being a point in $[t, t+\Delta t)$ with mark u_i is approximately $\lambda_{t,i}(j)\Delta t$, for Δt small.

Following the same notation as in 2.14, the filtering equations are:

i) For interarrival times

$$\frac{d}{dt}\hat{\pi}_j(t) = \sum_j q_{kj}(t)\hat{\pi}_k(t) - \hat{\pi}_j(t)\sum_i (\lambda_{t,i}(j) - \hat{\lambda}_{t,i}) \tag{2.19}$$

$$\hat{\pi}_j(t_0) = \pi_j(t_0) \quad . \tag{2.20}$$

ii) Update equation (σ = arrival time with mark u_i)

$$\pi_j(\sigma^+) = \hat{\pi}_j(\sigma^-)\frac{\lambda_{\sigma^-,i}(j)}{\hat{\lambda}_{\sigma-,i}} \quad . \tag{2.21}$$

iii) Minimum mean-square estimate of the intensity

$$\hat{\lambda}_{t,i} = \sum_j \hat{\pi}_j(t)\lambda_{t,i}(j) \quad . \tag{2.22}$$

The implementation of this filter is completely analogous to that of the previous section.

2.3 DISCRETE-TIME POINT PROCESSES

A further approximation in our models can be done by discretizing the time, which enables us to make use of the discrete-time point process approach as developed by Segall [16]. Here, the intensity parameter forms a finite-state, discrete-time Markov chain, and the recursive filter is given by a set of nonlinear difference equations acting over an innovations process.

2.3.1 Discrete Time Versus Continuous Time

Working in discrete time has both advantages and disadvantages. On one hand, the discrete-time filtering scheme, involving only difference equations, is much more easily implementable in a computer, which may be desirable for on-line control and surveillance purposes as compared with its continuous-time counterpart that requires integration of a system of differential equations. Furthermore, this system can be of very large order as a result of the discretization of the state space. (For instance, the grid resulting by dividing the range of possible values of the density and velocity in, say, 20 and 15 parts, respectively -- which is not a very fine grid -- already translates into a 300-state Markov chain.)

On the other hand, the discrete-time approach involves the assumption of having no more than one arrival time in every time interval Δt, where the discrete times are $n\Delta t$, $n = 0,1,2 \ldots$. This assumption implies that the time step Δt will have to be small enough as to prevent the possibility of having two or more arrivals in one time interval.

2.3.2 Choice of Discrete Time Justification

Working in discrete time was decided because of the following reasons.

i) The ease of implementation of the filtering algorithms, as discussed above,

ii) Even though, theoretically, the probability of having more than one arrival in an arbitrarily small nonzero time interval is greater than zero (for Poisson processes), in the real worlds, because of the vehicles physical dimensions, it is always possible to choose a time interval Δt, such that no more than one vehicle can cross a detector in that time. (In this sense, the traffic behavior can never be truly Poisson.)

iii) As explained in the first section, data from loop detectors are available at discrete times only due to the sampling process.

iv) Use of a digital computer for this work and eventually for traffic control and surveillance, will make time discretization unavoidable.

2.4 DISCRETE-TIME FILTERS

In this section, we present the discrete-time equivalent of section 2.1.4 as developed by Segall in [15] and [16]. The same notation as before will be used, with the variable t now taking only integer values $t = 0,1,2, \ldots$.

2.4.1 Mathematical Model

A discrete-time point process $\{N_t, t = 0,1,2, \ldots\}$ is simply a binary sequence (ones and zeroes) describing the occurrence of some type of event -- in our case, the arrival of a vehicle. Here the occurence of the event is indicated by $N_t = 1$ while non-occurence is indicated by $N_t = 0$, $\{N_t = 1\}$ shows that an event occurs at time t and $\{N_t = 0\}$, that there is no occurrence. If we define

$$\mathcal{B}_{t-1} = \sigma\{n(0), n(1), \ldots, n(t-1), x(0), x(1), \ldots, x(t)\} \tag{2.23}$$

where $\sigma(\alpha)$ denotes the α-algebra generated by α, the discrete-time equivalent of the intensity process λ_t is

$$\lambda_t(x_t) = Pr\{N_t = 1|\mathcal{B}_{t-1}\} = 1 - Pr\{N_t = 0|\mathcal{B}_{t-1}\} \tag{2.24}$$

x_t now forming a discrete-time, finite-state Markov chain, defined by stochastic matrix $Q(t) = [q_{ij}(t)]$, where

$$q_{ij}(t) = Pr\{x_{t+1} = j|x_t = i\} \quad . \tag{2.25}$$

(Note that $\sum_j q_{ij}(t) = 1$.)

It follows from (2.24) that

$$E\{N_t|\mathcal{B}_{t-1}\} = \lambda_t(x_t) \quad . \tag{2.26}$$

2.4.2 Estimation

The estimation problem is precisely the same one stated before (section 2.3.1), and its solution is given by the following recursive algorithm (Segall [16]):

i) Innovations process:

$$\nu(t|t-1) = N_t - \hat{\lambda}_t \tag{2.27}$$

and

$$\text{ii)} \quad \underline{\hat{\pi}}(t+1|t) = Q^T(t)\hat{\pi}(t|t-1) + \underline{K}(t, \hat{\pi}(t|t-1))\ \nu(t|t-1) \tag{2.28}$$

where

$$K(t, \hat{\pi}(t|t-1)) = \frac{S^T(t)\underline{\hat{\pi}}(t\ t-1) - Q^T(t)\underline{\hat{\pi}}(t|t-1)\underline{\hat{\pi}}^T(t\ t-1)\underline{\lambda}_t}{\hat{\lambda}_t(1 - \hat{\lambda}_t)}$$

and $$S(t) = [s_{ij}(t)] = [q_{ij}(t)\lambda_t(i)] \quad . \tag{2.29}$$

iii) Minimum mean squared error estimate of λ_t

$$\hat{\lambda}_t = \underline{\hat{\pi}}^T(t|t-1)\underline{\lambda}_t \ , \tag{2.30}$$

(We are using vector notation. For instance, if N is the number of states,

$$\underline{\lambda}_t^T = [\lambda_t(t), \lambda_t(2), \ldots, \lambda_t(N)] \tag{2.31}$$

where the upper index T denotes transpose.)

Note. Equation (2.29) requires an additional independence assumption. The definition of $S(t) = [s_{ij}(t)]$ is given by Segall [16] as

$$s_{ij}(t) = Pr\{x_{t+1} = j, N_t = 1 | x_t = i\} \tag{2.32}$$

since, in general, the events $\{x_{t+1} = j\}$ and $\{N_t = 1\}$ could be conditionally dependent, given that the system was at state $x_t = i$. On the other hand, x_{t+1} and N_t are independent conditioned on x_t, equation (2.29) follows immediately from (2.25) and (2.26). This independence assumption is implicit in Snyder's development of doubly stochastic Poisson process filters [17], and for the traffic problem we are interested in it makes good sense also: given the state of the traffic at the present time, the events of having a car arrival at a detector and the state taking a particular value Δt seconds later can well be thought of as being independent events.

3. PARAMETER ESTIMATION: STOCHASTIC MATRIX Q

Once a mathematical structure has been hypothesized, the second stage in the process of modeling a physical phenomenon is estimating values for the parameters appearing in the structure, which in our case are the stochastic matrix Q (equation 2.25) and the intensities $\lambda(j)$ - or $\lambda_i(j)$ if marks are used.

The case of assigning values to the intensities will be treated in the next section. Here, we concentrate our attention to the problem of estimating Q.

In the sequel, we assume that N, the number of states of the Markov chain, is a given fixed number. Obtaining an optimum value for N in a rigorous, theoretical fashion is practically impossible at this stage of the research work. For this reason, we will proceed in a heuristic way with a "good" value for N found by trying a set of different values, and choosing the one for which the filter performs best, in a statistical sense.

3.1 Q FROM DYNAMIC MODEL OF TRAFFIC

One possibility for the estimation of Q is trying to construct a Markov chain which is equivalent to a given dynamic model of the traffic. This problem also includes finding a suitable definition of two systems being equivalent.

Assuming that the traffic flow is reasonably well described by a system of ordinary differential equations such as Payne-Isaksen's previously mentioned, the problem could be stated in more precise terms as follows: Given a dynamic system consisting of a set of ordinary differential equations with continuous, finite-dimensional state space, find a discrete-time, finite-state Markov chain whose posterior statistics evolve in the same way. This problem can be solved in the following manner:

Let the dynamic system be given by

$$x(t) = f(t,x,u,\eta) \tag{3.1}$$

where $x(t) \in R^n$ is the state,

$u(t) \in R^m$ is a deterministic input, and

$\eta(t) \in R^p$ is a random variable of known statistics, and assume that all the possible trajectories of (3.1) lie within a closed set $A \subset R^n$. Assume also that we are given a partition of the set A into N disjoint regions $(A_1, A_2, \ldots, A_N)$ whose union equals A $(\bigcup_{i=1}^{N} A_i = A)$. Then, given that the state at time t is x_1; i.e., $x(t_1) = x_1$, one can compute the probability of the state being in the j^{th} region $(j = 1,2, \ldots, N)$ at time $t_1+\Delta t$ as

$$Pr\{x(t_1+\Delta t)\varepsilon A_j | x(t_1)=x_1\} = \int_{A_j} p(z|x(t_1)=x_1)dz \tag{3.2}$$

where $z = x(t_1+\Delta t)$, and $p(z|x(t) = x_1)$ is the probability density function of $x(t_1+\Delta t)$ conditioned on the event $\{x(t_1)=x_1\}$, t_1 and Δt being fixed quantities.

We are interested in the probability (see eq. 2.25),

$$Pr\{x(t_1+\Delta t)\varepsilon A_j \; x(t_1)\varepsilon A_i\} = q_{ij}(t_1) \; . \tag{3.3}$$

If we make the further assumption that $x(t_1)\varepsilon A$ means that $x(t_1)$ can be anywhere in region i with equal probability, a direct application of Bayes' rule leads to the expression

$$q_{ij}(t_1) = \frac{1}{K_i} \int_{A_i} [\int_{A_j} p(z|x(t_1)=x_1)dz]dx_1 \tag{3.4}$$

$$i, j=1,2, \ldots, N$$

where K_i is a proportionality constant equal to the volume of region A_i; i.e.,

$$K_i = \int_{A_i} dz \tag{3.5}$$

for $z\varepsilon R^n$.

Since the probability density function of x depends in general on the deterministic input u(t), it is clear that the elements of the stochastic matrix Q(t) will also be functions of u(t),

$$q_{ij} = q_{ij}(t,u(t)) \tag{3.6}$$

and even if the system (3.1) is time-invariant, the resulting Markov chain will be nonhomogeneous,

$$q_{ij} = q_{ij}(u(t)) \; . \tag{3.7}$$

Remarks

1) The above procedure requires computing the posterior statistics of x(t), which can be very difficult -- if not impossible -- for systems other than linear with additive white Gaussian noise.

2) Integration of equation (3.4) will have to be done numerically in most of the instances.

3) A simplifying approximation can be made as follows: Identify each region A_i by a single point x_i (e.g., its center of mass), that is

$$x(t) \varepsilon A_i \quad x(t) = x_i \ .$$

Then, one has to compute the evolution of the posterior statistics for N points only, and (3.4) reduces to

$$q_{ij}(t) \approx \int_{A_j} p(z|x(t)=x_i)dz \ .$$

Equation (3.4) is valid for any value of N, but for (3.4') the larger the value of N -- i.e., the smaller the regions A_i -- the better the approximation.

4) Assuming that a model of the form (3.1) is available for the traffic flow and that we can solve (3.4'), we still have the problem of measuring the input variable u(t), which includes the input and output ramps flows as well as the flows in the "first" and "last" links of the roadway. This could be accomplished with the use of loop detectors.

3.2 MAXIMUM LIKELIHOOD ESTIMATION OF Q

An alternative way of obtaining a value for Q is from direct observation of the process of interest. In this section, we derive firstly an estimator for Q and discuss afterward the necessary assumptions for its application to the traffic problem.

3.2.1 Maximum Likelihood Estimator

Suppose we want to estimate the stochastic matrix Q (eq. 2.25) of a homogeneous, discrete-time, finite-state Markov chain from observation of a particular realization of the process. Let N be the number of states.

Thus, we observe a sequence,

$$x(1) = K_1,\ x(2) = K_2,\ \ldots,\ x(M) = K_M, \qquad M>>N$$

where K_i is one of the N states. As it is well known (see for instance [20]), the maximum likelihood estimate of q_{ij}, which we will denote by $\hat{q}_{ij}$, is the value that maximizes the joint probability density of the observations:

$$p(\underline{x};Q) = Pr\{x(1)=K_1,\ x(2)=K_2,\ \ldots,\ x(M)=K_M\} \qquad (3.8)$$

Using Bayes' rule and the Markovian property:

$$p(\underline{x};Q) = Pr\{x(M)=K_M | x(M-1)=K_{M-1}\} \cdot Pr\{x(M-1)=K_{M-1} | x(M-2)=K_{M-2}\}$$

$$\cdots Pr\{x(2)=K_2 | x(1)=K_1\}\ Pr\{x(1)=K_1\} \qquad (3.9)$$

which, by the definition of q_{ij}, becomes

$$p(\underline{x};Q) = q_{K_{M-1},\ K_m}\ q_{K_{M-2},\ K_{M-1}} \cdots q_{K_1,\ K_2}\ P(1) \qquad (3.10)$$

with the initial probability

$$P(1) \triangleq Pr\{x(1) = K_1\}\ . \qquad (3.11)$$

Maximizing $p(\underline{x};Q)$ is equivalent to maximizing its logarithm. Thus, we form the log-likelihood function

$$\ln p(\underline{x};Q) = \sum_i \sum_j N_{ij}\ \ln q_{ij} + \ln q_{ij} + \ln P(1) \qquad (3.12)$$

where N_{ij} is defined as the number of observed transitions from state i to state j.

Recall that

$$q_{ii} = 1 - \sum_{j\neq i} q_{ij}\ . \qquad (3.13)$$

(All sums run from 1 to N.) Then, (3.12) can be rewritten as

$$\ln p(\underline{x};Q) = \sum_i \sum_{j\neq i} N_{ij}\ \ln q_{ij} + \sum_i N_{ii}\ \ln(1 - \sum_{k\neq i} q_{ik}) + \ln P(1)\ . \qquad (3.14)$$

Now, $\frac{d}{dq_{ij}} \ln p(\underline{x};Q) = 0$ implies

$$\hat{q}_{ij} = \frac{N_{ij}}{N_{ii}} (1 - \sum_{k\neq i} \hat{q}_{ik}) \qquad (3.15)$$

$$i = 1,2,\ \ldots,\ N$$

$$j,k \neq i$$

which specifies N systems of N-1 equations each (one for each i).

After some manipulation, (3.15) gives

$$\hat{q}_{ij} = \frac{N_{ij}}{T_i}, \qquad i \neq j \tag{3.16}$$

$$q_{ii} = 1 - \sum_{k \neq i} q_{ik}$$

with $T_i = \sum_k N_{ik}$ (3.17)

being the total observed time spent at state i.

3.2.2 $\hat{Q}$ for Traffic Problem

Equation (3.16) defines the maximum likelihood estimate of Q from an observed sequence. To apply it to the traffic problem a number of assumptions are required:

1) The Markov chain was assumed to be homogeneous, which implies that the input flow to the roadway has to be constant. Clearly, this is not true for real-world systems, but we can make use of the fact that traffic dynamics are relatively slow and assume Q to be constant over a fixed period of time (e.g., 15 minutes, 1 hour, etc.); that is, a piecewise homogeneous Markov chain. Although the algorithm (3.16) is pretty well behaved, as will be discussed in Section 5, the fact that Q is assumed to be constant over short periods of time only may translate into not having enough data for a good estimation of Q. This problem can be solved in principle by noting that traffic patterns develop as a function of the day and the time of the day, particularly in roadways connecting suburban and industrial centers (see for instance [19]). Thus, Q can be estimated for a given time interval by taking data for that time over a large number of days with similar characteristics (e.g., weekdays, holidays, etc.).

2) Perfect observations of the state were assumed in (3.16). Thus, a good calibration of the model may need observations of the traffic flow via aerial photography for instance, but if a very long string of data are desirable, this method could be expensive. A heuristic alternative may be starting with a rough estimate of Q obtained from a not-too-large number of data and using the outputs of the filter (estimates of the state) as new observations when operating on-line. The validity of this approach and

its effects on the filter are a subject for further research.

3.3 SUMMARY

We have described two possible approaches for the estimation of Q: approximating a dynamic system by a Markov chain, and maximum likelihood estimation. For this work, we chose the latter because

i) Of its simplicity as compared with the mathematical complexities of the approach in 3.1, and

ii) The available dynamic models for traffic are also approximations whose validity has yet to be established, and therefore using them as precise descriptions of the real world may not prove very useful.

4. PARAMETER ESTIMATION: INTENSITIES λ

In the models described in section 2 there is an underlying intensity process driving the actual point process or arrival-time sequence. This intensity is a function of the state of the traffic -- the Markov chain x_t -- and possibly of a set of marks (see section 2.2).

In this section, we discuss how values were assigned to these intensities as well as the relevant assumptions involved in the process.

Note that although all the work will be done in discrete time, the discussion to follow and the terminology will correspond to the continuous-time case. Two reasons in so doing are, first, that things are more easily understood in continuous time within our traffic control problem framework, and second, that we want to avoid introducing new terminology.

4.1 DOUBLY STOCHASTIC POISSON PROCESS

As seen in section 2, for each state j (j = 1,2, ... N) of the Markov chain x_t, there is a corresponding value $\lambda(j)$ of the intensity process.

From the definition 2.1.1, a doubly stochastic Poisson process is conditionally Poisson given the state. The mean value of this process for the homogeneous case we are interested in (i.e., $\lambda(j)$ constant over time) is

$$\lambda(j)\Delta t \qquad \text{[vehicles]} \tag{4.1}$$

and thus, $\lambda(j)$ has units of vehicles/time, say, vehicles/hour.

In brief, $\lambda(j)$ corresponds to the value of the flow at state j; i.e., (eq. 1.2),

$$\lambda(j) = \lambda(x_t(j)) = \rho(j) \cdot v(j) = \phi(j) \quad . \tag{4.2}$$

Having a partition of the state space ρ-v, equation (4.2) implies that single values are assigned to both ρ and v -- and hence to ϕ -- whenever the point (ρ,v) lies within region j.

For this work, the middle point of the region was chosen, i.e., if region j is such that

$$\rho' \leq \rho < \rho''$$

$$v' \leq v < v''$$

then

$$\lambda(j) = \frac{1}{4} (v' + v'') \; (\rho' + p'') \qquad \text{vehicles/hour} \quad . \qquad (4.3)$$

Thus, this first filter produces basically an estimate of the aggregated flow in the roadway, although some additional information is also available: the conditional probabilities of being at each state given the past sequence of arrival times (eq. 2.5). It is expected on the average the probability of being at the true state to be larger than that of being at any other state. Hence, we can in principle obtain an estimate of the state by looking at these probabilities.

On the other hand, in view of figure 1.3, it is clear that two different states can have the same intensity parameter associated with them and, because of that, given the true value of λ, the filter will not be able to distinguish between two possible states (1) and (2).

A tentative solution could be to obtain, separately, information on the velocity -- which need not be very accurate -- that will help us decide between two possible states. Referring to figure 4.1, if somehow we obtain an estimate $\hat{v}_1$ of the velocity, we can choose one out of the two possible densities $\hat{\rho}_1$ and $\hat{\rho}_2$, thus determining an estimate of the state of the traffic (in this case state (1): $\hat{\rho}_1$ and the corresponding velocity $\hat{\lambda}_1/\hat{\rho}_1$).

To focus on resolution of this ambiguity resulting from single parameter estimation, we study now the marked point process filters, which enable us to incorporate information other than arrival times.

4.2 MARKED POINT PROCESSES

Recall that here the underlying intensity process is a function not only of the state j but also of the mark u_i, which we denote by $\lambda_i(j)$, and that marks are assumed to belong to a denumerable set on R^p. Moreover, we will work with a finite mark space:

$$\{u_1, u_2, \ldots, u_M\} \; , \qquad u_i \in R^p \qquad (4.4)$$

where M is the total number of possible marks.

It is convenient to define an intensity parameter associated with each state regardless of the marks as:

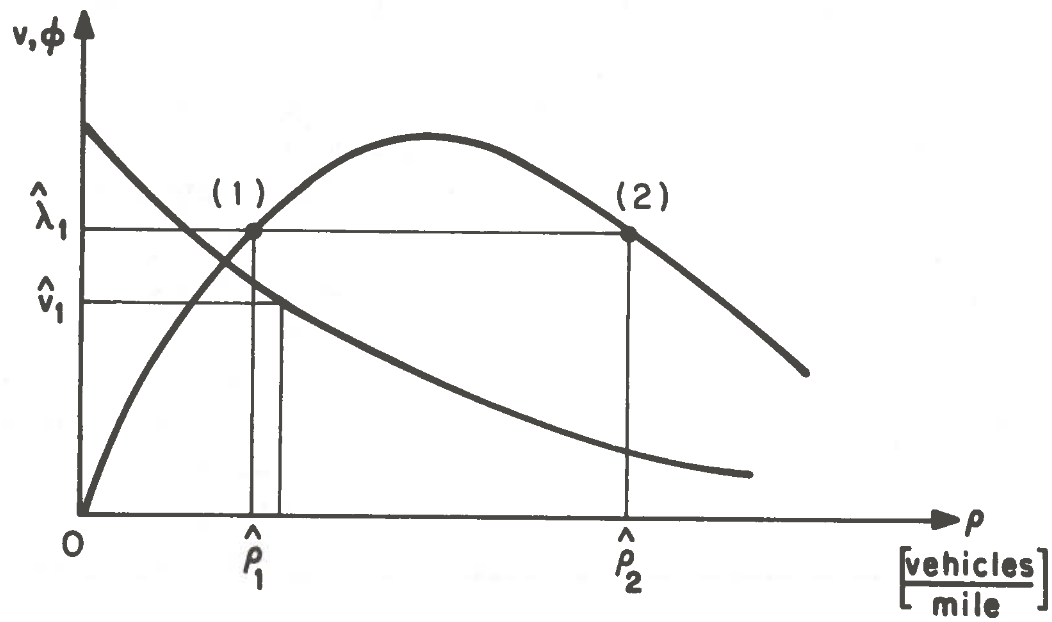


FIGURE 4.1 NONUNIQUENESS OF PROBLEM

$$\Lambda(j) = \sum_{i=1}^{M} \lambda_i(j) \quad . \tag{4.5}$$

From section 2 (see also [18]), the probability of having one arrival with any mark is approximately $\Lambda(j)\Delta t$ for small Δt, when the system is at state j. Furthermore, the arrival process disregarding the marks is conditionally Poisson with intensity $\Lambda(j)$ given the state $x_t = j$. Thus, the discussion of the last section is still valid if we substitute $\lambda(j)$ by $\Lambda(j)$, and equations (4.1) to (4.3) apply for $\Lambda(j)$ also. In particular,

$$\Lambda(j) = \frac{1}{4}\,(v' + v'')\,(\rho' + \rho'') \quad . \tag{4.3'}$$

Having determined $\Lambda(j)$ by equation (4.3'), our problem is now to specify the type and number of marks such that (4.5) holds.

Before doing this, some notation is necessary: Let $V = \{v_1, v_2, \ldots, v_r\}$ and $P = \{\rho_1, \rho_2, \ldots, \rho_{\delta}\}$ be the sets of middle points of the discretized regions in the velocity and density spaces, respectively. Clearly, $r \cdot \delta = N$, the total number of states (see figure 4.2).

Thus, equation (4.3') says that Λ is a function

$$\Lambda : V{\times}P \rightarrow R \tag{4.6}$$

and we want to determine i functions

$$\lambda_i : V{\times}P \rightarrow R$$

that is, the $M \cdot r \cdot \delta$ scalars:

$$\lambda_i(v_k, \rho_\ell) \quad , \qquad \begin{array}{l} i = 1,2,\ \ldots,\ M \\ k = 1,2,\ \ldots,\ r \\ \ell = 1,2,\ \ldots,\ \delta \quad . \end{array} \tag{4.7}$$

We explore two possibilities.

4.2.1 <u>Velocity Mark</u>

First, we associate with each arrival time the speed of the vehicle at that instant of time. Thus, p=1 in (4.4).

This speed has to be discretized and that we accomplish with the same discretization used for the state space of velocity-density. (For instance, in the case depicted in figure 4.2 there would be three possible marks (velocities) v_1, v_2, and v_3.)

At this stage, we have to assume homogeneous conditions for the traffic state to use the velocity information. This means that in a given section of the roadway at a particular time, all the vehicles are assumed to be traveling at the same speed, which is therefore equal to the aggregate speed v.

With this assumption, we can define λ_i of equation (4.7) as follows:

$$\lambda_i(v_k, \rho_\ell) = \begin{cases} 0 & \text{for } i \neq k \\ \Lambda(v_k, \rho_\ell) & \text{for } i = k \end{cases} \tag{4.8}$$

$$i = 1,2, \ldots, r$$
$$k = 1,2, \ldots, r$$
$$\ell = 1,2, \ldots, s$$

with $\Lambda(v_k, \rho_\ell)$ given by (4.3').

This is so because the spatial homogeneity assumption implies that if the system is at state (v_k, ρ_ℓ), the only mark that can possibly occur is v_k. Observe that (4.8) satisfies equation (4.5).

<u>NOTE</u>: For further work in this area the homogeneity assumption, which is indeed very unrealistic, may be relaxed as follows.

Suppose we had perfect observations of the process for a long enough time interval. Let $n_A(j)$ be the total number of arrivals that occurred when the system was at state j, each one with a particular velocity mark v_i. Then, one can compute the relative frequencies of occurrence of the marks conditioned on the state, and let

$$\lambda_i(j) = \frac{n_i(j)}{n_A(j)} \Lambda(j) \tag{4.9}$$

where $n_i(j)$ is the observed number of arrivals with mark i when the system was at state j, and $\Lambda(j)$ is given by (4.3'). It is clear that (4.9) satisfies (4.5).

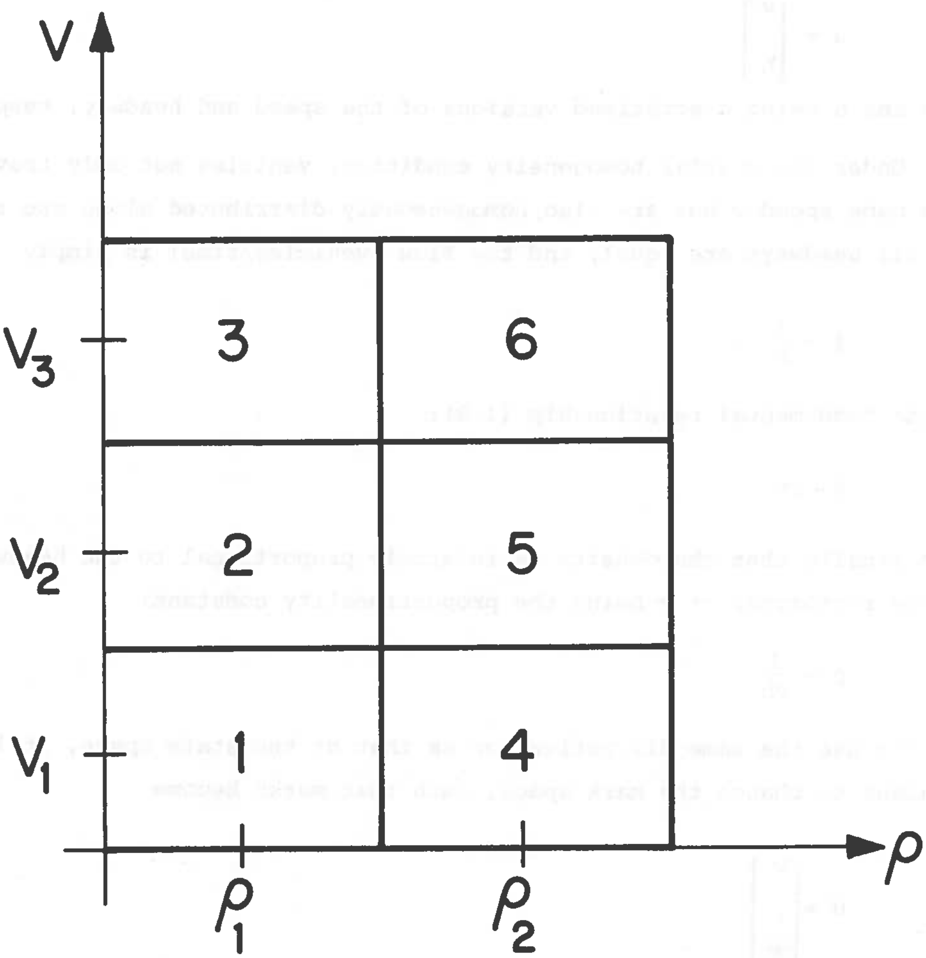


FIGURE 4.2 DISCRETIZATION OF STATE-SPACE

$$V = \{v_1, v_2, v_3\}$$

$$\rho = \{\rho_1, \rho_2\}$$

4.2.2 Velocity Plus Headway Mark

A second possibility is associating with each arrival time not only the speed of the vehicle but also the headway (h); that is, the time interval separating it from the vehicle that crossed the detector immediately before. Thus, marks are in this case

$$u = \begin{bmatrix} v \\ h \end{bmatrix} \tag{4.10}$$

with v and h being discretized versions of the speed and headway, respectively.

Under the spatial homogeneity condition, vehicles not only travel at the same speed v but are also homogeneously distributed along the roadway. Thus, all headways are equal, and the flow [vehicles/time] is simply

$$\phi = \frac{1}{h} \quad . \tag{4.11}$$

From the fundamental relationship (1.2):

$$\phi = \rho v$$

we see finally that the density is inversely proportional to the headway, with the reciprocal of v being the proportionality constant:

$$\rho = \frac{1}{vh} \quad . \tag{4.12}$$

To use the same discretization as that of the state space, it is more convenient to change the mark space, such that marks become

$$u = \begin{bmatrix} v \\ \frac{1}{vh} \end{bmatrix} \tag{4.13}$$

Then, the number of possible marks equals that of the states; i.e., M=N. Furthermore, under the spatial homogeneity condition, mark

$$u = \begin{bmatrix} v_k \\ \rho_\ell \end{bmatrix}$$

can only occur if the system is at stage (v_k, ρ_ℓ); or, alternatively, mark j can only occur at stage j. Thus, equation (4.7) becomes

$$\lambda_i(j) = \begin{cases} 0 & \text{for } i \neq j \\ \Lambda(j) & \text{for } i=j \end{cases} \qquad (4.14)$$

$$i, j=1,2, \ldots, N$$

with $\Lambda(j)$ given by (4.3').

4.3 SUMMARY

In this section, we have shown ways of determining values for the intensity process under a spatial homogeneity condition, for the various presented models: Doubly stochastic Poisson process (eq. 4.3); Marked point process with velocity mark (eq. 4.8), and Marked point process with velocity and headway marks (eq. 4.14). Also a heuristic approach for the relaxation of the homogeneity hypothesis was outlined (eq. 4.9).

5. EXAMPLE. SIXTEEN-STATE MARKOV CHAIN

In this section, we construct a Poisson process driven by a 16-state Markov chain and test the mathematical algorithms discussed so far.

It is clear that the validity of a proposed model is solely determined by its actual performance when applied to real data. Nevertheless, we considered this example to be a necessary step before dealing with the traffic problem because:

i) It is desirable to evaluate the performance of the filters when applied to point processes that are "actually" Poisson. One expects the performance of the filters when used with traffic data to deteriorate as a result of the simplifying assumptions and approximations. Thus, an example can provide an idea of the goodness of the algorithms.

ii) Similarly, we want to evaluate the filters when the underlying information process is a true Markov chain.

iii) It is also necessary to test the maximum likelihood estimation algorithm for Q when there actually exists an underlying Q matrix against which the estimate can be compared. This is useful in determining the number of observations required for a good estimate, as well as the asymptotic behavior of the algorithm.

iv) Finally, we want to compare the different approaches (doubly stochastic, marked point processes with velocity marks, etc.), and determine which one performs the best.

5.1 MODEL

To work with a reasonably large Markov chain, we arbitrarily chose a 16-state birth-and-death process (see for instance Karlin [9]). In this kind of process, the state can only jump to the two adjacent states; i.e.,

$$\begin{aligned} q_{ij} &\neq 0 \quad \text{for } j=i-1, \\ &\qquad\qquad j=i+1 \\ q_{ij} &= 0 \quad \text{otherwise, } i\neq j \\ &\qquad\qquad i, j=1,2, \ldots, N \;. \end{aligned} \tag{5.1}$$

5.1.1 Generation of Markov Chain

All the simulations were done in continuous time; only the final point process (arrival times) was discretized.

A realization of a continuous-time Markov chain is performed as follows:

First, given the initial a priori probabilities, an initial state is chosen using a random number generator: If z is a random variable uniformly distributed between zero and one, we take a sample of z and determine the initial state i as shown in figure 5.1. Now, given that the Markov chain is at state i, we have to find the time T at which it elaves the state, as well as the new stage j. As shown by Karlin [9], the time T is exponentially distributed with parameter

$$a = \sum_{k \neq i} q_{ik} \tag{5.2}$$

Thus, by sampling an exponential distribution, we obtain the time T. Alternatively, we can sample the uniform random variable z, and let

$$T = -\frac{1}{a} \ln z \tag{5.3}$$

which is exponentially distributed. Finally, given that the chain leaves state i at time T, it goes to state j with probability (see [9])

$$p_j = \frac{q_{ij}}{\sum_{k \neq i} q_{ik}} \qquad j=1,2, \ldots, N \quad j \neq i \quad . \tag{5.4}$$

Thus, the new state j is determined by means of a sample of z and the probabilities p_j, as was done for the initial state in figure 5.1.

5.1.2 Generation of Arrival Process

Once a realization of the Markov chain has been obtained, an arrival-time sequence is generated as follows:

Assume j to be the state at time zero. Given that the chain is at state j, the arrival sequence is a Poisson process with intensity λ(j). Let T be the time spent at state j before jumping to another state. Thus, we sample the uniform random variable z, and compute

$$t_i = -\frac{1}{\lambda(j)} \ln z \quad . \tag{5.5}$$

If $t_1 < T$, we have an arrival and genrate t_2 similarly. If $t_1 + t_2 < T$ we have a second arrival, etc. When, for some n

FIGURE 5.1 SELECTION OF INITIAL STATE

π_j = A priori probabilities

$j = 1,2,...,N$

z = Random variable uniformly distributed between 0 and 1

$$t_1 + t_2 + \ldots + t_n > T$$

we have n-1 arrivals in [0,T); disregard t_n and start again at T by sampling an exponential distribution with parameter $\lambda(k)$, if k is the next state of the Markov-chain realization.

5.2 MAXIMUM LIKELIHOOD ESTIMATION OF Q

Q was estimated from a particular realization of the Markov chain using the algorithm developed in section 3.5, equation (3.16). Note that the proof was given in discrete time. For a treatment of the continuous-time case, see [18].

A number of experiments were performed for various matrices Q and different realizations, with the following results:

1) In all the instances, the algorithm behaved asymptotically unbiased, with the error approaching zero as the observation time was increased. Since looking at the error for every single element of Q gets complicated when the order of Q is large, a single number for the difference between Q and its estimate $\hat{Q}$ was defined as

$$\text{Error} = \sum_{i,j} \left| \frac{\hat{q}_{ij} - q_{ij}}{q_{ij}} \right| \quad . \tag{5.6}$$

Figure 5.2 shows the error for two different realizations of the same chain; i.e., same matrix Q, as a function of the observation time. In general, different realizations of the same process yielded about the same estimate for a sufficiently large observation. (To give an idea of the orders of magnitude, all q_{ij}, $i \neq j$, were less than unity [sec^{-1}], and the observation time was also in seconds.)

Figure 5.3 shows the error in estimating three different Markov chains. That with the largest error is the chain with the largest number of jumps (q_{ij} relatively large), and the smallest error was produced in estimating the chain with the smallest number of transitions (q_{ij} relatively small). Finally, the error in estimation a fourth-order Markov chain ($q_{ij} \neq 0$, all i,j) is shown in figure 5.4 for comparison. Note that the necessary time to obtain approximately the same accuracy is 10 times as small as in the previous cases, although the number of parameters to be estimated is reduced only in half:

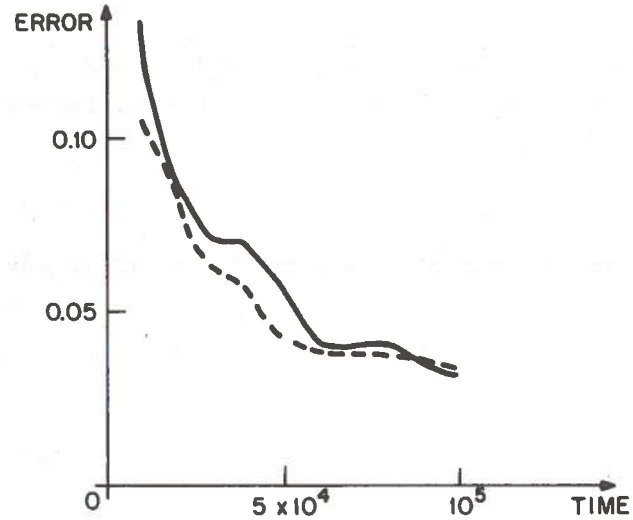


FIGURE 5.2

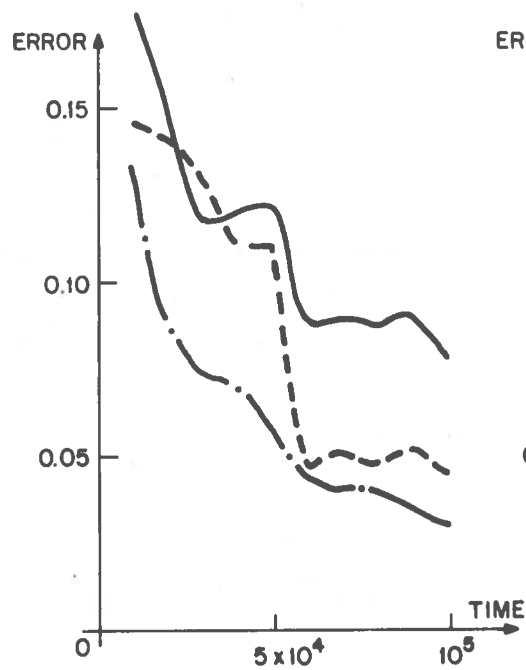


FIGURE 5.3 (CONTINUED)

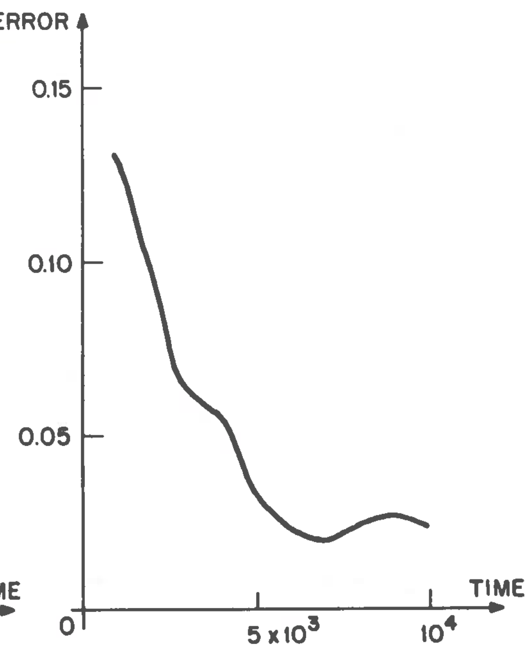


FIGURE 5.4 (CONCLUDED)

ILLUSTRATIONS OF ERRORS IN MAXIMUM LIKELIHOOD ESTIMATION OF Q

12 for the 4-order chain, and 30 for the 16-order chain (the main diagonal elements are not included, as they are linear functions of the off-diagonal terms).

2) Experiments were also conducted to determine if the performance of the algorithm were improved by means of Monte-Carlo trials; i.e., if instead of observing one realization over a period of time T, one takes the average over N different realizations of duration $\frac{T}{N}$ each. It turned out that for all practical purposes, the performance of both approaches was the same, with the former (observed over an extended period of time) doing slightly better. This result can obviously hold true only if one is dealing with an ergodic chain, where the limiting-state probabilities are independent of the initial state. Otherwise, if one had classes of trapping states, the Monte-Carlo method should be used.

Relating these results to the traffic problem, as discussed in section 3.2.2, one could observe the process over a short period of time on several days and take each observation as a Monte-Carlo trial of the same chain, thus estimating Q by averaging over the number of days.

5.3 FILTERS

The filters of section 2 were tested by generating a point process as in section 5.1 using the matrix Q, but giving the estimate $\hat{Q}$ to the filter as if it were the "true" underlying matrix. Three cases were studied.

5.3.1 No Marks

This type is the doubly stochastic Poisson process of section 2, and was generated as in section 5.1, with the final continuous-time point process being discretized so as to obtain a sequence of ones and zeroes at discrete intervals of time. As expected, more than one arrival occurred in a single time interval from time to time, and these cases were counted as having only one arrival. On the average, 1 out of 15 intervals had more than 1 arrival (2 arrivals in most cases), although this statistic is irrelevant for the traffic problem since it is pretty much dependent on the particular values chosen for q_{ij}. As a matter of fact, as we shall see in section 7, the traffic simulations very rarely produced more than one arrival in a one-second time step.

5.3.2 Velocity Marks

Using the same point process as in the previous section, a velocity mark was associated with each arrival by keeping track of the true state at each instant of time. Thus, if an arrival occurred and the true state was (v_k, ρ_ℓ) -- see section 4 for notation -- mark k was associated with that arrival. In other words, the filter was given perfect information about the velocity, and each time an arrival occurred it has to choose only among those states in the same class of equivalence; i.e., for mark k only states

$$(v_k, \rho_\ell) , \qquad \ell = 1,2, \ldots, s$$

were possible.

5.3.3 Velocity and Density Marks

Similarly, a velocity and density mark

$$u = \begin{bmatrix} v_k \\ \\ \rho_\ell \end{bmatrix}$$

was associated with an arrival taking place when the true state was (v_k, ρ_ℓ); that is, perfect information about the state was provided to the filter with each arrival, and thus, the estimates were perfect each time arrivals occurred.

5.3.4 Results

The three filters were applied to the same realization of the point process. As expected, the more the information provided to the filter, the better the results.

Two criteria were selected to evaluate the filter's performance. One was the mean-squared error

$$e_1 = \frac{1}{L} \sum_{i=1}^{L} (\lambda_i - \hat{\lambda}_i)^2 \tag{5.7}$$

where L is the number of discrete-time observations. The other was the percentage of time the filter picked the right state

$$e_2 = \frac{L_r}{L} \% \tag{5.8}$$

where L_r is the number of correctly estimated states. This number was computed by looking at the conditional probabilities given the past of the point process (eq. 2.5) and choosing the state with the largest probability; i.e., choose

$x_t = i$ if $\hat{\pi}_i(t) > \hat{\pi}_j(t)$, for all $j \neq i$.

Results are shown in Table 5.1 and typical realizations are illustrated in figures 5.5 to 5.7.

TABLE 5.1 COMPARISON OF DIFFERENT FILTERING SCHEMES

FILTER TYPE \ PERFORMANCE CRITERIA	e_1	e_2 %
No Marks	0.0214	26.63
Velocity Marks	0.0107	47.73
Velocity and Density Marks	0.0069	66.33

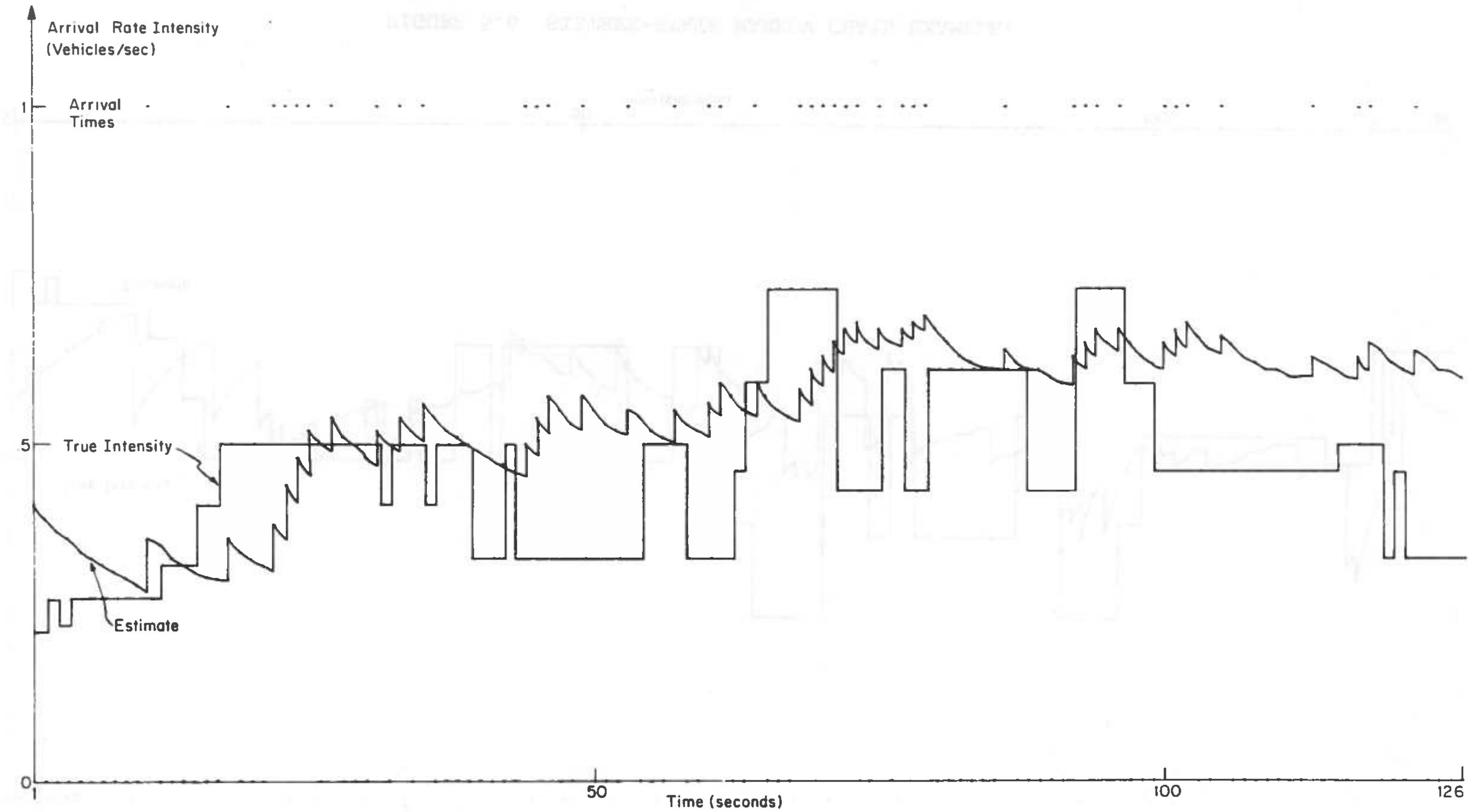


FIGURE 5.5 SIXTEEN-STATE MARKOV CHAIN EXAMPLE:

FILTER WITHOUT MARKS

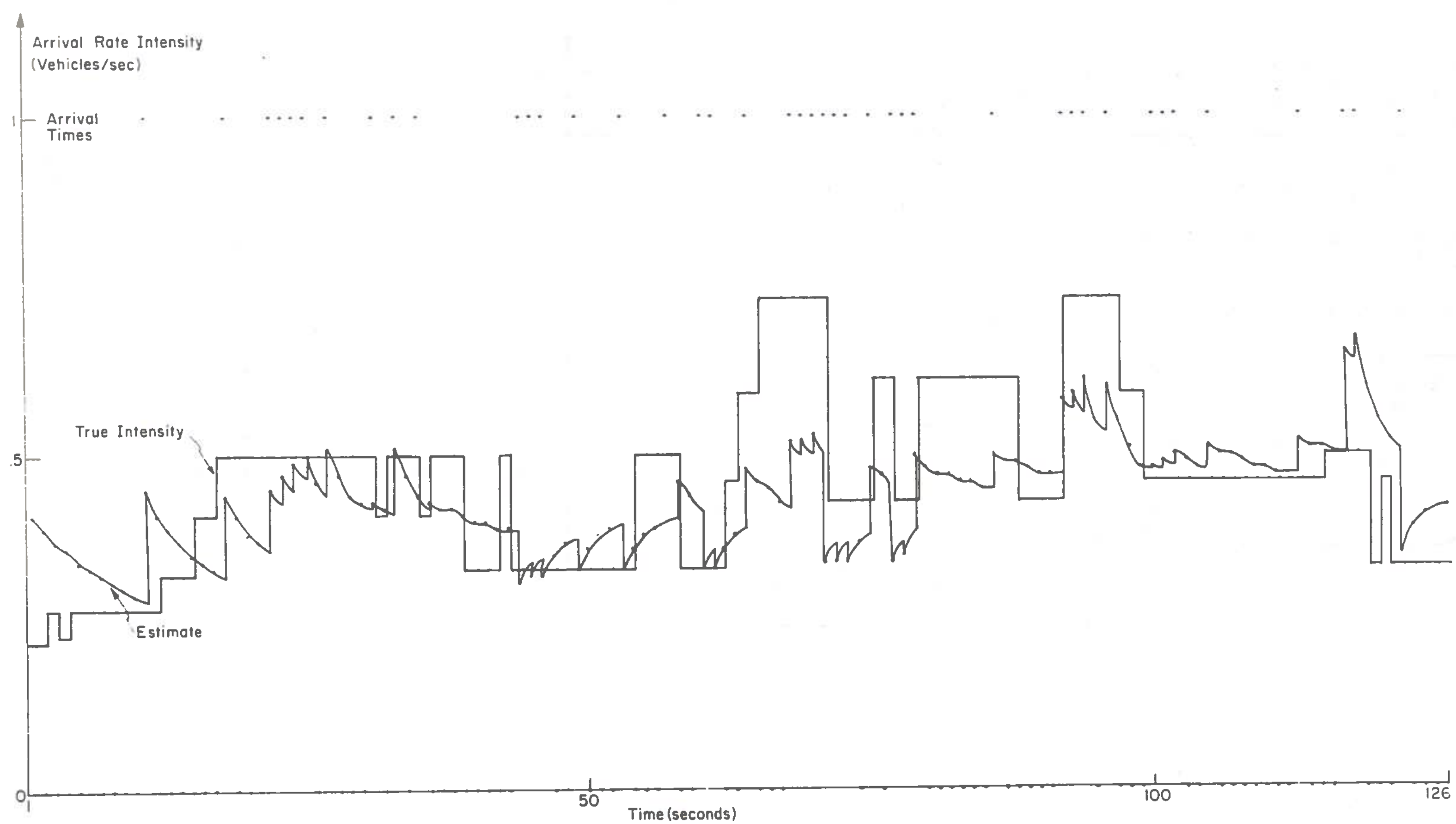


FIGURE 5.6 SIXTEEN-STATE MARKOV CHAIN EXAMPLE:

FILTER WITH VELOCITY MARKS

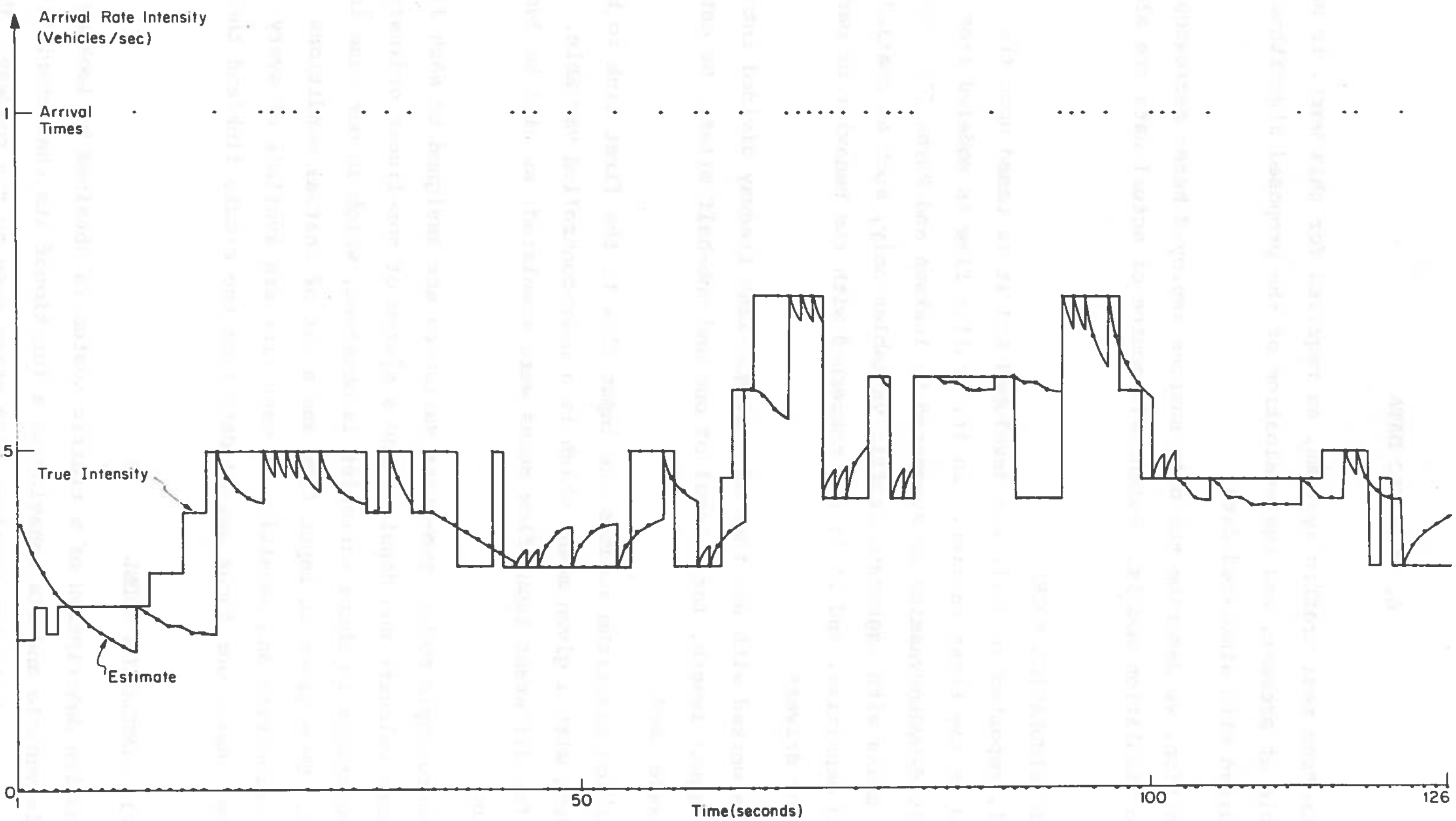


FIGURE 5.7 SIXTEEN-STATE MARKOV CHAIN EXAMPLE:
FILTER WITH VELOCITY AND HEADWAY MARKS

6. TRAFFIC DATA

Street data from real traffic systems, as required for this work, is not locally available at present, and the evaluation of the proposed algorithms had to be performed with simulated data.

In this section, we describe the data sources employed here: macroscopic and microscopic simulation models. Potentials source of actual data are alos discussed.

6.1 MACROSCOPIC SIMULATION MODEL

This model, reported in [13], was developed and it is based upon the ideas discussed in the first section. In it, traffic flow is modeled from analogies to fluid-hydrodynamics as suggested by Isaksen and Payne [7]. Thus, the simulation deals with aggregate traffic variables only, such as spatial mean speeds and densities, and it is not concerned with the behavior of particular vehicles or drivers.

The case we worked with was that of a single-lane freeway divided into three links of equal length, for a total of one and one-half miles. No entrance or exit ramps were used.

The simulation algorithm assumes the input flow to the first link to be a Poisson process with a given mean, which is a user-controlled variable. Several cases for different input flow means were simulated, as will be described in the following

For the macroscopic model, two-state variables are assigned to each link; namely, aggregate velocity and density; and a system of non-linear ordinary differential equations in these variables is obtained, which in our case is of sixth order. Thus given an input flow and a set of initial conditions, the values of velocities and densities on each link are availble at every instant of time. Here, our focus was on data from the middle link and their behavior.

6.2 MICROSCOPIC SIMULATION MODEL

An alternative description of a traffic system is obtained by looking at every single vehicle and its behavior as a functionof its characteristics, the driver's type, and the interactions with other cars on the roadway. Using the ideas of car-following theory, a microscopic model was developed which incorporates features such as different types of vehicles and drivers, passing routines, simulation of lane-blocking accidents, presence detectors, etc. For a complete description of the model, see [11].

The configuration used for this work was a two-lane freeway, one and one-half miles long divided into three links of equal length. This division is made for the purposes of computing aggregate variables -- spatial mean velocity and density. No entrance or exit ramps were used, and a presence detector was located at the downstream end of the middle link, left-hand lane. Note that although we were interested in the arrival process on one lane only, the model requires simulating two lanes because of the car-passing algorithms.

As with the previous case, this model uses a Poisson process input flow.

Given the mean of the input flow and initial conditions, the following information is available from a particular simulation:

i) Point process: Time at which a vehicle crosses the detector is registered, as well as its velocity at that particular moment.

ii) Aggregate variables: At a given instant of time, spatial mean velocity and density for each link can be obtained as follows:

Velocity

$$v(t) = \frac{1}{N(t)} \sum_{i=1}^{N(t)} v_i(t) \qquad \text{miles/hour}, \qquad (6.1)$$

Density

$$\rho(t) = \frac{1}{L} N(t) \qquad \text{cars/mile} \qquad (6.2)$$

where

$N(t)$ = the number of vehicles occupying the link at time t.

$v_i(t)$ = the speed of the $i^{\underline{th}}$ vehicle at time t, and

L = the length of the link

6.3 COMPARISON OF SIMULATION MODELS

Both macrosimulation and microsimulation models were used under different input-flow conditions to generate data for the evaluation of the point process filters. A few comments on these models are in order.

i) Although the models have been shown to describe qualitative behavior of traffic flow under different scenarios, they have not yet been calibrated against real traffic data to determine their validity. Thus, simulations under similar initial conditions and input flow are expected to be only roughly equivalent.

ii) Since the dynamics of the models are very slow, it is necessary to have very long simulation times to be able to study some interesting cases.

iii) Due to its nature, computer simulations of the microscopic model are more expensive than those of the macroscopic case. This difference can become very large particularly if congested (high-density) conditions are required.

6.4 POTENTIAL SOURCES OF REAL DATA

As mentioned in section 3.2, obtaining a good parameter initialization for the filters may require using aerial photography to observe the state of the traffic, or another sophisticated method such as television cameras. Once the filter is working on line, the estimates produced could be used as additional data to update the parameter estimates.

7. RESULTS

We describe in the present section the different cases studied, the parameter estimation, and the filtering results.

7.1 SELECTION OF REPRESENTATIVE CASES

The following instances were studied:

1) Constant input flow

Both macroscopic and microscopic models were allowed to evolve under constant input-flow conditions

$$\phi_{in} = 1500 \text{ vehicles/hour}$$

and initial conditions

$$v_0 = 50 \text{ miles/hour, and}$$

$$\rho_0 = 20 \text{ cars/mile .}$$

The macroscopic model was observed for a simulated period of 1000 seconds, and the microscopic model for 200 seconds.

2) Flow pulse

The conditions in this case were as follows:

$$\phi_{in} = \begin{cases} 1500 \text{ miles/hour, } 0 \leq t < 100 \text{ seconds} \\ 4000 \text{ miles/hour, } t \geq 100 \text{ seconds} \end{cases}$$

$$v_0 = 50 \text{ miles/hour}$$

$$\rho_0 = 20 \text{ miles/hour}$$

with simulation times of 200 seconds in both cases.

Also, a 1000-seconds macrosimulation was performed with a square-wave input flow of 200 seconds, and maximum and minimum values of 4000 and 1500 vehicles-per-hour, respectively.

3) High density

Congested cases of traffic flow are also of interest, and those were simulated with the microscopic model by starting the system at a very high density and low velocity, with constant input flow:

$$\phi_{in} = 1000 \text{ vehicles/hour}$$

$$v_0 = 4 \quad \text{miles/hour}$$

$$\rho_0 = 225 \quad \text{vehicles/mile} \quad .$$

4) <u>Lane-blocking accidents</u>

An alternative way of obtaining relatively large densities is by simulating a stalled vehicle in one of the lanes of the microscopic model:

$$\phi_{in} = 1500 \text{ vehicles/hour}$$

$$v_0 = 50 \quad \text{miles/hour}$$

$$\rho_0 = 20 \quad \text{vehicles/mile} \; ,$$

with a stalled vehicle at time $t = 60$ seconds on the right-hand lane.

In all of the previous cases, a one-second sampling time interval was used, which proved to be sufficiently small in the sense that multiple arrivals in a single interval occurred very infrequently (approximately once every 1000 seconds).

7.2 ESTIMATION OF MARKOV CHAIN

A grid was defined in the velocity-density plane as follows:

$$\rho_{min} = 0 \quad \text{vehicles/hour}$$

$$\rho_{max} = 228 \quad \text{vehicles/hour}$$

$$\Delta\rho = 12 \quad \text{vehicles/hour}$$

$$v_{min} = 0 \quad \text{miles/hour}$$

$$v_{max} = 65 \quad \text{miles/hour, and}$$

$$\Delta v = 5 \quad \text{miles/hour}$$

Note that although this defines a 247-state Markov chain, each one of the cases described in the last section occupies only a relatively small region of the space, and thus, it is not necessary to work with all of the states at once.

By keeping track of the true state of the traffic as defined by the <u>aggregate</u> variables, different Q matrices were estimated for every case using the maximum likelihood algorithm. For comparison, figure 7.1 shows how states were occupied under three different situations.

As mentioned before, simulations under equivalent conditions result in only roughly similar Markov chains for the microscopic and macroscopic models (see figures 7.2a and 7.2b for two examples).

7.3 ESTIMATION OF TRAFFIC VARIABLES

Having estimated the Markov chains describing the evolution of the traffic state, the filtering schemes of section 2 were tested. This was done using Q estimated for both models an dthe point processes (arrival time of vehicles) produced by the microscopic simulation model. Thus, for every matrix $\hat{Q}$, the filters were applied to several different point processes, and the traffic state was estimated.

The implementation of the unmarked case (doubly stochastic Poisson process) was straightforward. The cases of velocity marks and velocity plus headway marks were treated basically in the same way described in section 5 (the 16-state example) since the velocity of a vehicle crossing the detector was known, and headway is only the time different between two adjacent vehicles. However, some problems arose.

For one thing, velocity of individual vehicles has a wider range of variation that that of the aggregate speed of the roadway, which was used to estimate Q. (Similar behavior occurs with the spatial mean density and the "density" obtained as 1/vh -- see equation (4.12) --, the latter spreading over a wider range.)

Thus, it may happen that at a given time, a car arrives at the detector with velocity of, for instance, 36 miles/hour and headway of 2 seconds -- relative to the car that crossed before it -- which translates into a density

$$p = \frac{1}{vh} = 50 \text{ vehicles/mile} \quad .$$

However, the discretized state corresponding to the point (50, 36) does not necessarily have a nonzero probability of being occupied, in the Markov chain defined by $\hat{Q}$ (see figure 7.2). For this reason, it would make no sense to give to the filter the information that the car arrived with such a mark.

To handle these cases, the filter was allowed to saturate; i.e., before giving the mark information to th e filter, a check was made to determine if

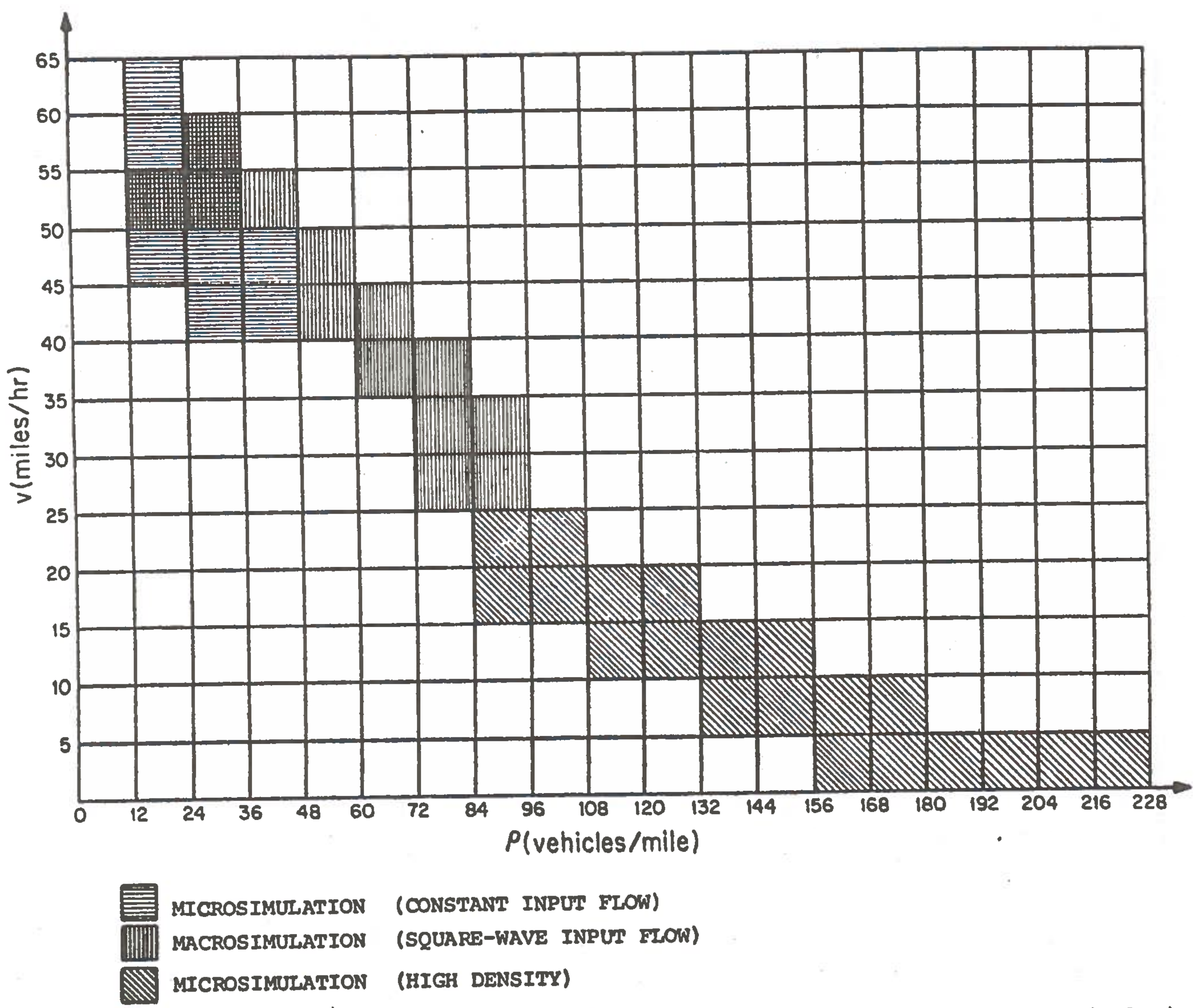


FIGURE 7.1 State-Space Trajectories Resulting From Simulation Run

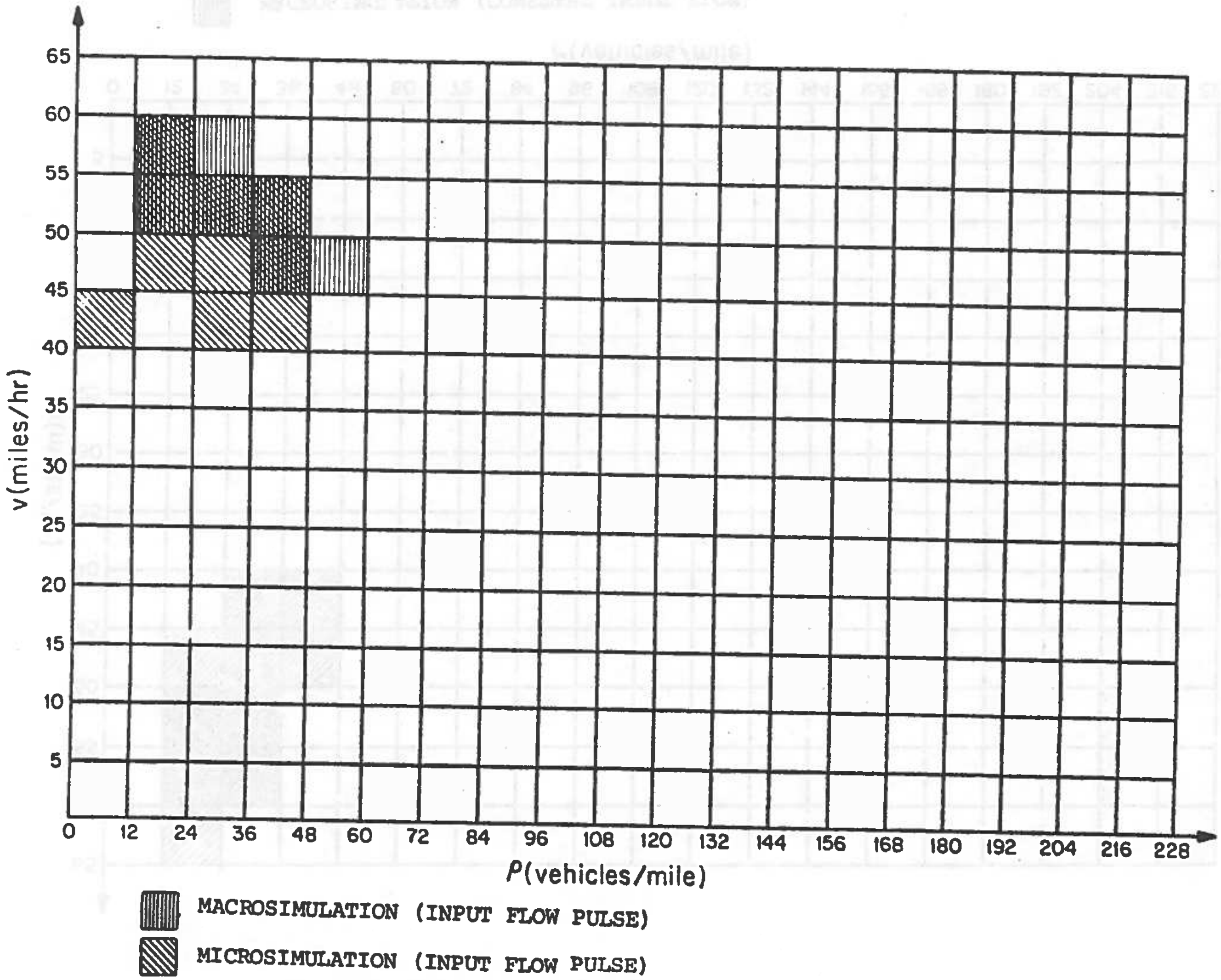


FIGURE 7.2a Discrete State-Space Trajectories at Low Density: Run A

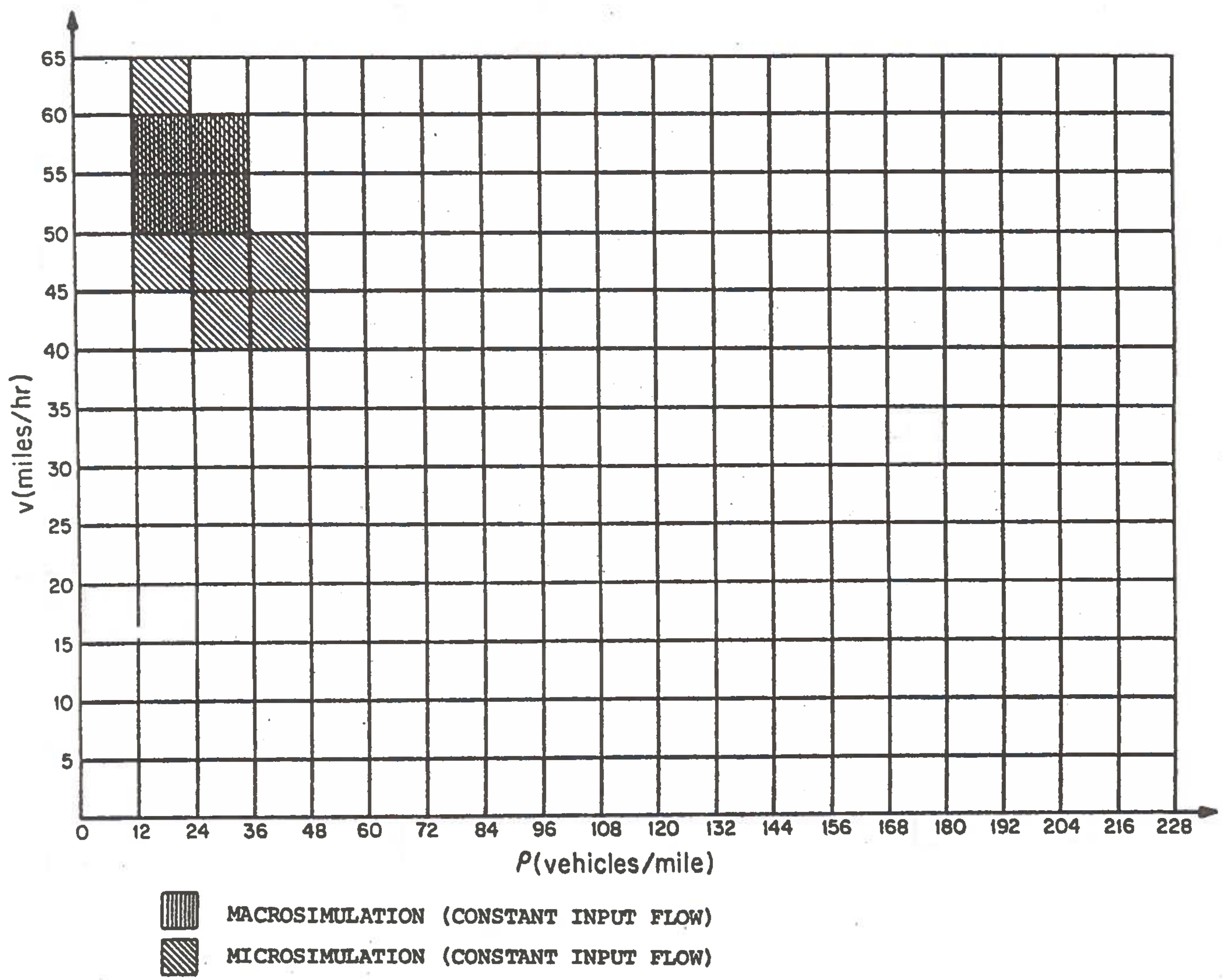


FIGURE 7.2b Discrete State-Space Trajectories at Low Density: Run B

that particular mark was compatible with the given Markov chain; else the mark was changed for that of the closest state with nonzero probability.

Results from the filter simulations were evaluated using the criteria described in section 5; namely, mean-squared error and fraction of time the right state was chosen (eqs. 5.7 and 5.8).

7.4 SUMMARY OF RESULTS

Table 7.1 shows results of the filtering simulations. Vertical entries refer to three different cases from which the Markov chain (Q matrix) was estimated. They are (see section 7.1):

a) Microscopic simulation with an input flow pulse,

b) Macroscopic simulation with a square wave input flow, and

c) Microscopic simulation for high-density conditions .

For each of these cases, the three types of filters -- unmarked, velocity marks, and velocity plus headway marks -- were simulated.

Horizontal entries are the vehicle-arrival point processes used as input to the filters. Beside, cases (a) and (c) already mentioned, (recall that no point process is available fromthe macrosimulations), the following were also considered:

d) Microsimulation with constant input flow, and

e) Microsimulation in very high-density conditions. This case, as compared with (c), was produced not only by starting the system at a high-density initial condition but also by applying a larger input flow mean.

Two performance criteria -- equations (5.7) and (5.8) -- are given for each case in table 7.1, the first being the mean-squared error. This table was obtained from a single realization of every case of interest, and thus, no attempt is made to draw quantitative conclusions regarding the statistics of the errors. Our main interest at this point of the development is to obtain qualitative understanding of the point process filters when applied to traffic data.

For quick reference to table 7.1, we will use the notation (i,j) to describe the element on ith row and j<u>th</u> column. For instance, (6.3) is the case in which Q was estimated from a microsimulation with high-density

TABLE 7.1 FILTERING RESULTS

Filter Type		a)	b)	d)	e)
No Marks	a)	0.0069 40.5	.0222 0	0.0051 0	0.0443 0
	b)	0.0173 22.0	0.0223 0	0.0196 0	0.0469 0
	c)	0.0144 0	0.0262 32.5	0.0212 0	0.0171 0
Velocity Marks	a)	0.0113 33.0	0.0258 2	0.0076 19.5	0.0452 4.0
	b)	0.073 4.5	0.0658 0	0.0201 0	0.082 0
	c)	0.071 0	0.0259 17.5	0.0583 0.5	0.0162 0
Velocity and Headway Marks	a)	0.0163 23.5	0.0237 0	0.0155 16.0	0.0667 6.0
	b)	0.092 0	0.0946 0	0.0432 0	0.304 0
	c)	0.071 0	0.0246 12	0.0583 0.5	0.0165 1.0

conditions (case c); a filter with velocity marks was used, and the arrival point process was generated from a microsimulation with constant input flow (case d). Also, the performance criteria are

$e_1 = 0.0583$, and

$e_2 - 0.5\%$.

From the table, we observe the following:

1) Performance criteria

The percentage of correctly estimated states (e_2) seems to bear little relationship to the mean-squared error e_1. Furthermore, e_2 is somewhat random, and due to this, we will focus on criterion e_1 solely.

2) Mismatches

Table 7.1 contains a number of mismatched cases such as (3,3) -- estimating Q from a high-density condition and using it to filter a low-density point process. These experiments were performed with the idea of studying the sensitivity of the filters to the matrix Q. Notice that when Q is estimated from low-density processes, the filters do better if the point process is also of low density. Compare (1,1), (1,2), (1,3) and (1,4), as well as the second row. On the other hand, filters with Q estimated from a high-density simulation od not perform better with high-density point process as can be seen in row three of table 7.1. This behavior may be due to modeling errors since for congested situations, Poisson process is less accurate a model of the arrival-time sequence.

3. Marked filters

Use of the mark information causes the filter performance to deteriorate. Compare, for instance, (1,1), (4,1), and (7,1). This implies that homogeneity was too simplistic an assumption, and that speed and density of an individual vehicle, and aggregate speed and density on a section of the roadway, are not correlated enough. The only cases in which use of marks did not dteriorate, the filter performance (though it did not make it better either) was estimation of high-density point processes. Compare (3,2), (6,2), and (9,2).

It is clear that in such cases, most of the cars travel at the same speed and have the same headway; thus, an individual vehicle's velocity does provide some information about spatial mean state of the traffic flow (aggregate speed and density).

It should also be noted that headway information is already imbedded into the filtering scheme and need not be explicity included as a mark. This is so because the filter acts over the arrival point process, and estimates are updated each time an arrival occurrs, which implictly takes into account the separation time between arrivals, or headway. Figures 7.3, 7.4, and 7.5 illustrate cases (1,1), (4,1), and (7,1), respectively.

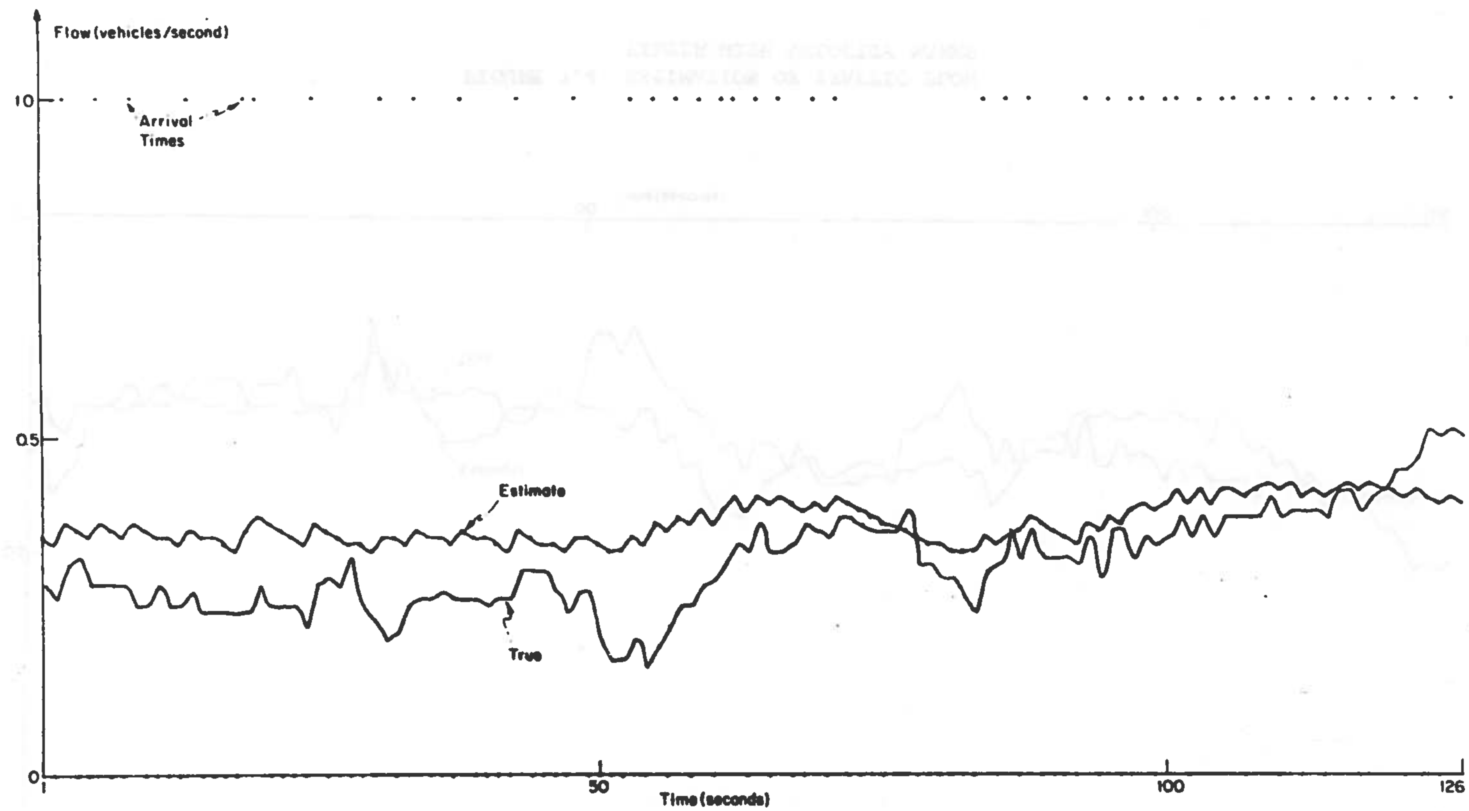


FIGURE 7.3 ESTIMATION OF TRAFFIC FLOW
FILTER WITHOUT MARKS

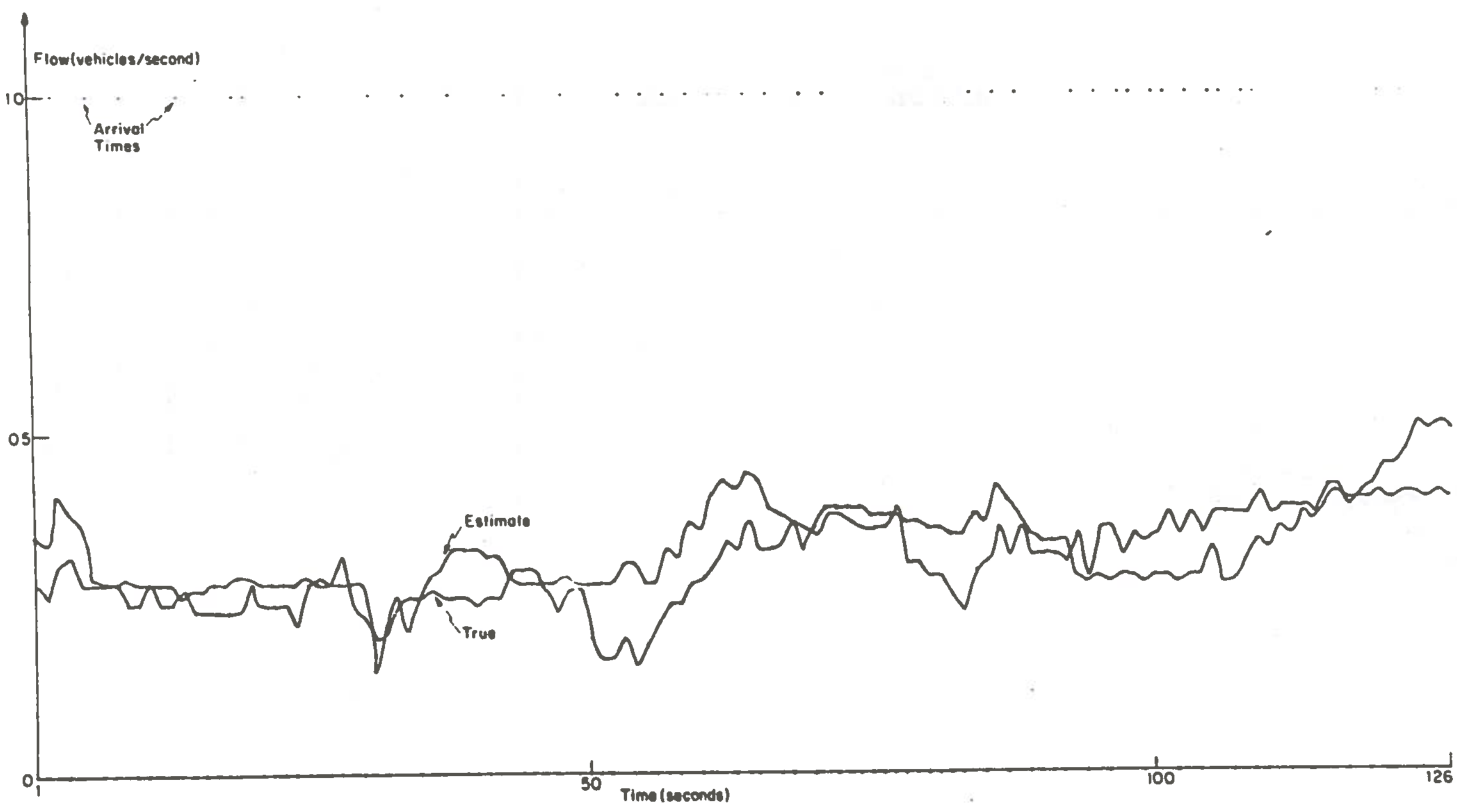


FIGURE 7.4 ESTIMATION OF TRAFFIC FLOW FILTER WITH VELOCITY MARKS

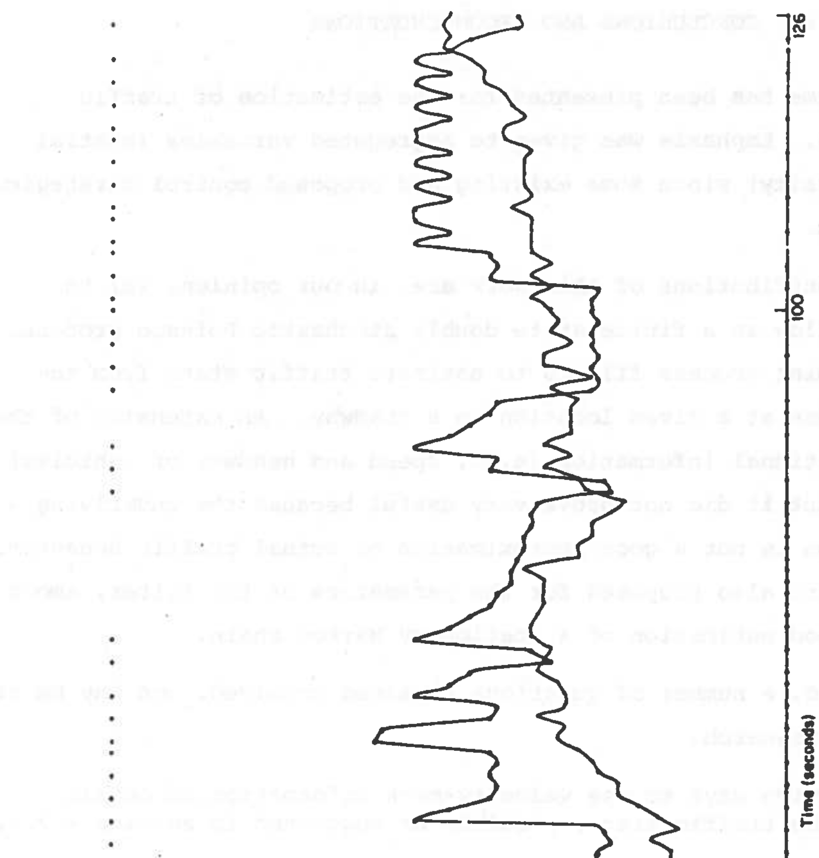


FIGURE 7.5 ESTIMATION OF TRAFFIC FLOW FILTER WITH VELOCITY AND HEADWAY MARKS

8. CONCLUSIONS AND RECOMMENDATIONS

A filtering scheme has been presented for the estimation of traffic variables on freeways. Emphasis was given to aggregated variables (spatial mean velocity and density) since some existing and proposed control strategies deal mainly with them.

The principal contributions of this work are, in our opinion, (a) the modeling of traffic flow as a finite-state doubly stochastic Poisson process, and (b) the use of point process filters to estimate traffic state from the vehicle-arrival process at a given location in a roadway. An extension of the model to include additional information (e.g., speed and headway of vehicles) was also presented, but it did not prove very useful because the underlying homogeneity assumption is not a good approximation of actual traffic behavior. Estimation schemes were also proposed for the parameters of the filter, among them maximum likelihood estimation of a stationary Markov chain.

On the other hand, a number of questions remained unsolved, and may be of interest for further research:

1) Find alternative ways to use velocity-mark information to obtain better estimates of the traffic state, possibly as suggested in section 4.2.1,

2) Estimate filter parameters from actual data and determine their characteristics,

3) Incorporate information from several presence detectors placed along the roadway,

4) Test filter algorithms against actual freeway data, and

5) Control of roadway traffic. Ideas presented in this work could be used for future research to develop alternative control strategies based upon the finite-state characterization of traffic.

The use of finite-state models to characterize traffic behavior has many consequences for both filtering and control-system design. In many ways, we see how this framework is more descriptive and natural to the discrete nature of traffic. On the other hand, there are frew tools available to decide in a systematic way responses to issues such as the minimum number of states required in the underlying Markov-chain model. While it can be argued that this is a modeling question, it is also closely tied to the formulation of the control problemand remains a challenge in Markov-decision theory.

REFERENCES

[1] Athans, M., Houpt, P.K., et al. Stochastic Control of Freeway Corridor Systems, ESL-P-620, MIT, August 1975.

[2] Chang, D., Gazis, D.C., Traffic Density Estimation with Consideration of Line Changing, RC 5250, IBM. February 1975.

[3] Gazis, D.C., Knapp, R. On Line Estimation of Traffic Densities from Time Series of Flow and Speed Data. Transportation Science. Vol. 5, No. 3 (pp. 283-301). August 1971.

[4] Gazis, D.C., Szeto, M. Surveillance and Control of Traffic Systems by Application of Kalman Filtering Techniques. RC 3690, IBM. January 1972.

[5] Gazis, D.C., Szeto, M. Estimation of Traffic Densities at the Lincoln Tunnel from Time Series of Flow and Speed Data. Stochastic Point Process (P. Lewis, ed.); pp. 151-165. Wiley, 1972.

[6] Gazis, D.C., Szeto, M. Design of Density Measuring Systems for Roadways. IBM, Transportation Research Record No. 495; pp. 44-52. 1974.

[7] Isaksen, L.; Payne, H.J., Regulation of Freeway Traffic. JACC Proceedings. 1972.

[8] Isaksen, L., Payne, H.J. A Systematic Approach to Freeway Traffic Control and Surveillance. Highway Research Board. January, 1972.

[9] Karlin, S. A First Course in Stochastic Processes. Academic Press, New York and London. 1968.

[10] Looze, D.P., Houpt, P.K., Athans, M. Dynamic Centralized and Decentralized Control Strategies. ESL-R-610, MIT. August, 1975.

[11] Mitchell, W.J., The Estimation and Simulation of Traffic Flow Using Car-Following and Fluid-Analog Models. S.M. Thesis, Electronic Systems Laboratory, MIT. 1977.

[12] Mikhalkin, B., Payne, H.J. Estimation of Speed from Presence Detectors. Highway Research Board Record TE7, no. 388, pp. 73-83. 1972.

[13] Orlhac, D.L., Athans, M., Speyer, J., Houpt. P. Estimation of Traffic Variables via Extended Kalman Filter. ESL-R-611, MIT. September 1975.

[14] Rudemo, M. Doubly Stochastic Poisson Processes and Process Control. Adv. Applied Probability 4, pp. 318-338. 1972.

[15] Segall, A. Stochastic Processes in Estimation Theory. ESL-P-588, MIT. January 1975.

[16] Segall, A. Recursive Estimation from Discrete Time Point Processes. ESL-P-591, MIT. February, 1975.

[17] Snyder, D. Filtering and Detection for Doubly Stochastic Poisson Processes. IEEE Trans. on Information Theory, Vol. IT-18, No. 1, January, 1972.

[18] Snyder, D. Random Point Processes. Wiley 1975.

[19] Sparks, W., Newell, G. Statistical Properties of Traffic Counts. Published in Stochastic Point Processes: Statistical Analysis, Theory and Applications. P. Lewis, Editor. Wiley Interscience. 1972.

[20] Van Trees, H.L. Detection, Estimationa dn Modulation Theory, Part I. Wiley 1975.

APPENDIX

Report of Inventions

Although there are no inventions reported here, the report represents an advance in that an alternative approach to estimation of traffic variables (mean speed and density) is developed by modeling the traffic as a finite-state Market Chain as described in detail in section 2. A sequential point process filter has been designed to estimate the state from observations on the vehicle arrival time sequence as described in detail in Section 3.

110 copies

U. S. DEPARTMENT OF TRANSPORTATION
TRANSPORTATION SYSTEMS CENTER
KENDALL SQUARE, CAMBRIDGE, MA. 02142