TRANSPORTATION RESEARCH RECORD 795

Traffic Flow Theory and Characteristics

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1981

Transportation Research Record 795
Price $4.20
Edited for TRB by Mary McLaughlin

mode
1 highway transportation

subject area
55 traffic flow, capacity, and measurements

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
Traffic flow theory and characteristics.

(Transportation research record; 795)
Reports for the 60th annual meeting of the Transportation Research Board.
1. Traffic flow—Congresses. I. National Research Council (U.S.). Transportation Research Board. II. Series.
TE7.H5 no. 795 [HE332] 380.5s [388.3'143] 81-11325
ISBN 0-309-03209-1 ISSN 0361-1981 AACR2

Sponsorship of the Papers in This Transportation Research Record

GROUP 3—OPERATION AND MAINTENANCE OF TRANSPORTATION FACILITIES
Patricia F. Waller, University of North Carolina at Chapel Hill, chairman

Committee on Traffic Flow Theory and Characteristics
John J. Haynes, University of Texas at Arlington, chairman
Edmund A. Hodgkins, Federal Highway Administration, secretary
Charles R. Berger, Said M. Easa, John W. Erdman, Nathan H. Gartner, Richard L. Hollinger, Matthew J. Huber, John B. Kreer, Joseph K. Lam, Tenny N. Lam, Edward B. Lieberman, Roy C. Loutzenheiser, William R. McShane, Carroll J. Messer, Paul Ross, Richard Rothery, Joel Schesser, Steven R. Shapiro, Yosef Sheffi, Andrew St. John, William W. Wolman

David K. Witheford, Transportation Research Board staff

The organizational units, officers, and members are as of December 31, 1980.

Contents

Authors of the Papers in This Record

Ahmed, Samir A., School of Civil Engineering, Oklahoma State University, Stillwater, OK 74078
Benevelli, David A., Virginia Polytechnic Institute and State University, Blacksburg, VA 24060
Drenick, Rudolf F., Transportation Training and Research Center, Polytechnic Institute of New York, 333 Jay Street, Brooklyn, NY 11201
Garner, J.B., Department of Civil Engineering, University of Leeds, Leeds LS2 9JT, England
Gartner, Nathan H., Department of Civil Engineering, University of Lowell, 1 University Avenue, Lowell, MA 01854
Hauer, E., Department of Civil Engineering, University of Toronto, Toronto, Ontario M5S 1A4, Canada
Hurley, Jamie W., Jr., Department of Civil Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24060
Kelson, Mark D., Massachusetts Institute of Technology, E53-350, 50 Memorial Drive, Cambridge, MA 02139
Little, John D.C., Massachusetts Institute of Technology, E53-350, 50 Memorial Drive, Cambridge, MA 02139
Maze, Thomas H., Department of Civil Engineering, Wayne State University, Detroit, MI 48202
McShane, William R., Transportation Training and Research Center, Polytechnic Institute of New York, 333 Jay Street, Brooklyn, NY 11201
Michalopoulos, Panos G., Department of Civil and Mineral Engineering, University of Minnesota, Minneapolis, MN 55455
Mountain, L.J., Department of Civil Engineering, University of Liverpool, Brownlow Street, P.O. Box 147, Liverpool L69 3BX, England
Pagitsas, E., Department of Civil Engineering, University of Toronto, Toronto, Ontario M5S 1A4, Canada
Plum, Roger A., Department of Civil and Mineral Engineering, University of Minnesota, Minneapolis, MN 55455
Radwan, Ahmed E., Department of Civil Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24060
Shin, B.T., Department of Civil Engineering, University of Toronto, Toronto, Ontario M5S 1A4, Canada
Stephanedes, Yorgos J., Department of Civil and Mineral Engineering, University of Minnesota, Minneapolis, MN 55455

Estimation of Turning Flows from Automatic Counts

E. HAUER, E. PAGITSAS, AND B.T. SHIN

Traffic flows on intersection approaches can be obtained by using automatic counting machines. A method for the estimation of turning movements from approach counts is developed and tested. The problem is solved by identifying the most likely traffic-flow matrix that agrees with the given approach counts by using characteristic turning proportions. Traffic flows from 145 intersections in metropolitan Toronto have been coded. From this information, characteristic right- and left-turning proportions are estimated for five types of intersection approaches. Estimates of vehicle flows are derived by using these characteristic turning proportions. These estimates are then compared with the observed flows, and the accuracy of the estimation is explored. It appears that the method may in some cases serve as a useful tool.

Knowledge of traffic flows at intersections is needed for a variety of planning and management purposes. Obtaining an estimate of flows into and out of intersection approaches is relatively cheap. It can be done by automatic counting machines. However, for many purposes, the split of each traffic stream into left-turning, right-turning, and straight-through flows is of interest. The most common way of obtaining estimates of turning movements is from manual counts by observers, and such counts tend to be expensive.

Recently, alternative methods for the estimation of turning movements at intersections have been suggested (1-4). These methods use approach counts and other relevant information to obtain estimates of turning flows.

In this paper, we follow closely the method proposed by van Zuylen (3). The theoretical part in the following section contains no innovation and is included for completeness. The only difference is in the motivation of the method. Whereas van Zuylen relies on minimization of "information", our argument is based on the maximization of "likelihood".

The main purpose of the study is to explore the accuracy of the estimation obtained when the method is applied to a large number of urban intersections.

PROBLEM FORMULATION AND SOLUTION

Let i take on values 1,2,...,m and represent the label assigned to an intersection approach carrying traffic into the intersection. Similarly, let j take on values 1,2,...,n and represent the label assigned to an approach carrying traffic out of the intersection. With i and j as subscripts, let T_{ij} be the flow from approach i into approach j. When the T_{ij} are written in a table with m rows and n columns, a "flow matrix" is obtained. The task is to find estimates T_{ij}^* of T_{ij}--that is, to find an estimate of the flow matrix.

Two pieces of information are brought to bear on the estimation task:

1. It is assumed that the total flow into the intersection from approach i (O_i) and the total flow out of the intersection by approach j (D_j) are known for all i and j. This information is obtainable from, say, automatic traffic counters. By using this notation, $\sum_{j=1}^{n} T_{ij} = O_i$ and $\sum_{i=1}^{m} T_{ij} = D_j$. Barring errors of counting and negligible "end of count period" effects, the sum of entering flows (S) equals the sum of leaving flows ($\sum_{i=1}^{m} O_i = \sum_{j=1}^{n} D_j = S$).

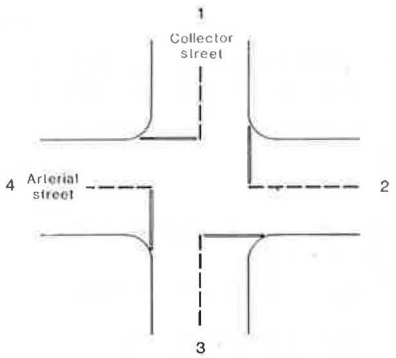
2. There exists some prior knowledge about the proportion of left and right turns at similar intersection approaches. Then let p_{ij} denote the proportion of the traffic emanating from approach i that at similar intersections turns in the direction of approach j. Naturally, $\sum_{j=1}^{n} p_{ij} = 1$.

To illustrate, consider the intersection shown in Figure 1 and the corresponding flow matrix. The cells on the diagonal are shaded to indicate that vehicles entering the intersection do not turn back by the same approach, and therefore these cells will have no entries. Any turn restrictions can be represented similarly. For example, if left turns were not permitted from the arterial street into the collector, the two cells (row 2, column 3 and row 4, column 1) would also be shaded. Assuming that the proportion of leftand right-turning traffic from an arterial street into a collector street is normally 0.02 each, and that the corresponding value for flow from the collector into the arterial is 0.30, the p_{ij} values can be entered into the matrix as shown. Finally, imagine that two automatic traffic counters were placed on each leg of the intersection, one counting entering traffic and the other counting traffic leaving the intersection. These eight counts are entered as row and column sums in the flow matrix. For example, the flow entering the intersection from approach 1 has been counted as 100 vehicles for the count period, whereas the flow leaving the intersection by approach 1 during the same period has been counted as 50 vehicles.

The stage is now set for analyzing the problem of finding an estimate of the flow matrix T.

There are many different combinations of numbers that, when listed in the flow matrix, will meet the

Figure 1. Intersection, flow matrix, and p_{ij} values.



From \ To	1	2	3	4	
1	(0)	(0.30)	(0.40)	(0.30)	100
2	(0.02)	(0)	(0.02)	(0.96)	600
3	(0.40)	(0.30)	(0)	(0.30)	200
4	(0.02)	(0.96)	(0.02)	(0)	700
	50	800	100	650	1600

Figure 2. Computations for example intersection: use of the algorithm.

p_{ij}							Iteration				
							1	2	3	4	5
To / From	1	2	3	4	0	A^0	A^1	A^2	A^3	A^4	A^5
1	0.00	0.30	0.40	0.30	100	2.5	2.06	1.94	1.91	1.90	1.90
2	0.02	0.00	0.02	0.96	600	15.0	15.37	15.86	16.29	16.65	16.95
3	0.40	0.30	0.00	0.30	200	5.0	6.28	6.54	6.59	6.60	6.62
4	0.02	0.96	0.02	0.00	700	17.5	16.70	16.24	15.89	15.71	15.46
D	50	800	100	650	1600						

Iteration 1	B^1	18.9	42.0	60.6	39.0
Iteration 2	B^2	15.86	43.16	68.24	37.66
Iteration 3	B^3	15.35	44.11	70.52	36.58
Iteration 4	B^4	15.24	44.93	71.04	35.74
Iteration 5	B^5	15.21	45.37	71.06	35.07

given row and column sums. The question is which of these many "feasible solutions" is most likely to have occurred during the period in which the automatic counts were taken.

Each of these feasible solutions could have arisen in a large number of ways. The number of ways in which flows $T_{i1}, T_{i2}, \ldots, T_{in}$ can be selected from a total of O_i is known to be

$$O_i! / \prod_1^n T_{ij}! \tag{1}$$

To keep the notation simple, all sum (Σ) and product (π) operators will be understood to skip over the "shaded" cells of the flow matrix.

In view of the specified proportions p_{ij}, each of the events included in the total given by Equation 1 arises with a probability of

$$\prod_1^n (p_{ij})^{T_{ij}} \tag{2}$$

It follows that the relative frequency w(T) with which some specific flow matrix T should be expected to arise is given by

$$w(T) = \prod_{i=1}^m \left[O_i! \prod_1^n (p_{ij})^{T_{ij}} / \prod_1^n T_{ij}! \right] \tag{3}$$

The task is to identify the flow matrix T* for which the value of w(T) is the largest.

Since the product $\prod_1^n O_i!$ is fixed, the task can be transformed into finding T that maximizes

$$\ln \prod_1^m \prod_1^n (p_{ij}^{T_{ij}} / T_{ij}!) = \sum_1^m \sum_1^n (T_{ij} \ln p_{ij} - \ln T_{ij}!) \tag{4}$$

According to Stirling's formula,

$$\ln T_{ij}! = T_{ij} \ln T_{ij} - T_{ij} + \tfrac{1}{2}(\ln T_{ij} + \ln 2\pi) + [r(T)_{ij}/12T_{ij}] \tag{5}$$

Since $T_{ij} >> 1$ and $r(T_{ij}) < 1$, all but the first two terms can be neglected. In preparation for the forthcoming minimization, note that, by using this approximation,

$$\ln T_{ij}! \cong T_{ij} \ln T_{ij} - T_{ij} \tag{6}$$

and

$$d \ln T_{ij}! / d T_{ij} \cong \ln T_{ij} \tag{7}$$

Making use of the approximation in Equation 6, we wish to find T that maximizes

$$\sum_1^m \sum_1^n T_{ij} \ln p_{ij} - T_{ij} \ln T_{ij} + T_{ij} \tag{8}$$

subject to

$$\sum_{j=1}^n T_{ij} = O_i \quad \text{for } i = 1, \ldots, m \tag{9}$$

and

$$\sum_{i=1}^m T_{ij} = D_j \quad \text{for } j = 1, \ldots, n \tag{10}$$

Let α_i be the Lagrange multiplier for the m-row constraints and β_i the Lagrange multiplier for the n-column constraints. Taking derivatives of the Lagrange expression, we find that

$$T_{ij}^* = p_{ij} e^{-\alpha_i} e^{-\beta_i} = p_{ij} A_i B_j \tag{11}$$

Equation 11 constitutes the solution of the problem as formulated. It remains to determine the value of the unknown multipliers A_i and B_j so as to satisfy the row and column sums as specified in Equation 10. The algorithm for doing so is described below. The underlying rationale for this estimation procedure and its weaknesses are discussed more fully by Hauer and Shin (5).

COMPUTATIONS

The problem formulated above and its solution belong to the general class of "biproportional" models. A brief description of the biproportional problem and its history, applications, and properties is given by Murchland (6).

Possibly the simplest solution algorithm consists of repeated balancing of the vectors A and B and is named after Kruithof (7). It consists of the following steps:

1. Set A_i (current) = $O_i/S^{1/2}$.
2. Find B_j from $B_j = D_j/\sum_{i=1}^{m} p_{ij}A_i$ (current).
3. Find A_i (new) from A_i (new) = $O_i/\sum_{j=1}^{m} p_{ij}B_j$.
4. Compare A_i (new) and A_i (current). If the largest difference is sufficiently small, use last A_i and B_j to find $T_{ij}^* = p_{ij}A_iB_j$. Otherwise, set A_i (current) = A_i (new) and return to step 2.

To illustrate, consider the intersection described in Figure 1. The proportions (p_{ij}), inflows (O_i), and outflows (D_j) are reproduced in Figure 2. Step 1 of the algorithm produces the column under A^0. Thus, for example, $A_1 = O_1/S^{1/2} = 100/1600^{1/2} = 2.5$. In step 2, by using A^0, the vector B^1 is calculated. For instance, $B_4^1 = 650/(2.5 \times 0.30 + 15.0 \times 0.96 + 5.0 \times 0.30 + 17.5 \times 0.00) = 39.0$. Going to step 3, vector A^1 is found by using B^1. For instance, $A_1^1 = 100/(18.9 \times 0.00 + 42.0 \times 0.30 + 60.6 \times 0.40 + 39.0 \times 0.30) = 2.06$. The first round of computations ends by comparing the new vector A with the previous one. Unless the desired closure is attained, a new round of computations is carried out. In this example, results of five iterations are listed. A gradual convergence of the multiplier vectors A and B is evident. Were one to calculate estimates T* on the basis of A^3, B^3, then Figure 3a would apply. Estimates of T* using A^5, B^5 are shown in Figure 3b.

PROPORTION OF TURNING MOVEMENTS AT INTERSECTIONS IN METROPOLITAN TORONTO

The procedure described and illustrated above produces estimates of vehicle flows by using data from automatic counters and prior knowledge about typical proportions of turning movements at intersections. The better the estimates of these typical proportions (p_{ij}), the more accurate the resulting estimate of the flow matrix will be. If no information is available about the intersection at hand or about the area in which it is situated, one may be inclined to rely on some gross average proportions. Thus, for example, the "average conditions" for signalized intersections on which the charts in the Highway Capacity Manual (8, p. 133) are based correspond to 10 percent left turns and 10 percent right turns. In an attempt to improve the quality of turning-flow estimation, the turning proportions at 145 intersections in metropolitan Toronto have been examined.

For each intersection, complete traffic-flow counts were available for four periods:

Period	Hours
Morning peak	7:00-9:00 a.m.
Evening peak	4:00-6:00 p.m.
Off-peak	11:00 a.m.-12:00 noon
Daytime	7:00 a.m.-6:00 p.m.

Figure 4 shows a typical traffic-flow diagram that served as a source of data. The information has been coded and keypunched as shown in Figure 5.

For coding, five types of intersections have been defined:

1. Central business district (CBD) (within rectangle bounded by Spadina, Dundas, Jarvis, and Front Streets),
2. Arterial with arterial,
3. North-south arterial with east-west collector,
4. East-west arterial with north-south collector, and
5. Collector with collector.

In this study, an arterial is a main thoroughfare with two or more lanes in one direction. In metro-

Figure 3. Computations for example intersection: (a) T^{*3} and (b) T^{*5}.

a

To / From	1	2	3	4	0
1		26	54	20	100
2	5		24	571	600
3	40	90		70	200
4	5	673	22		700
D	50	789	100	661	

b

To / From	1	2	3	4	0
1		25	54	21	100
2			23	572	600
3	40	87		72	199
4	5	673	22		700
D	50	785	99	665	

Figure 4. Typical traffic-flow diagram.

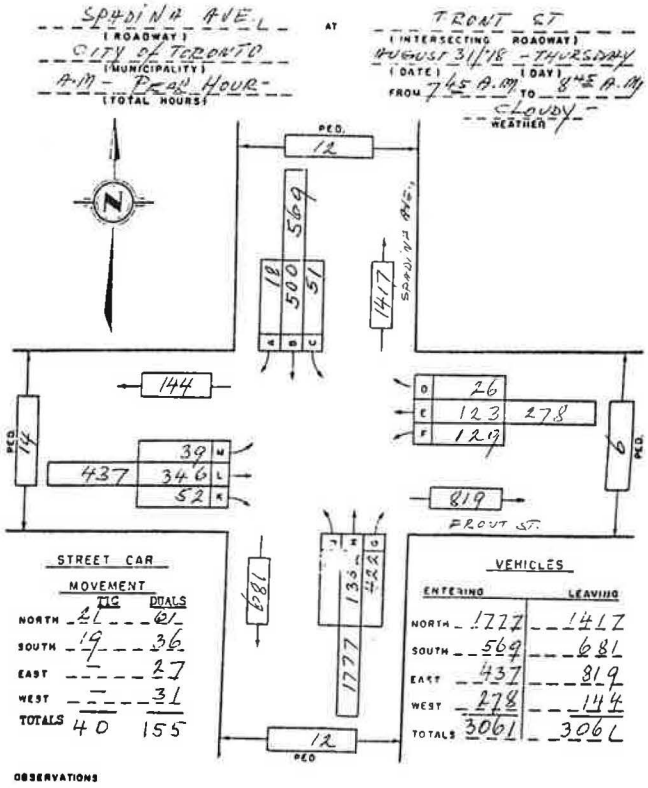
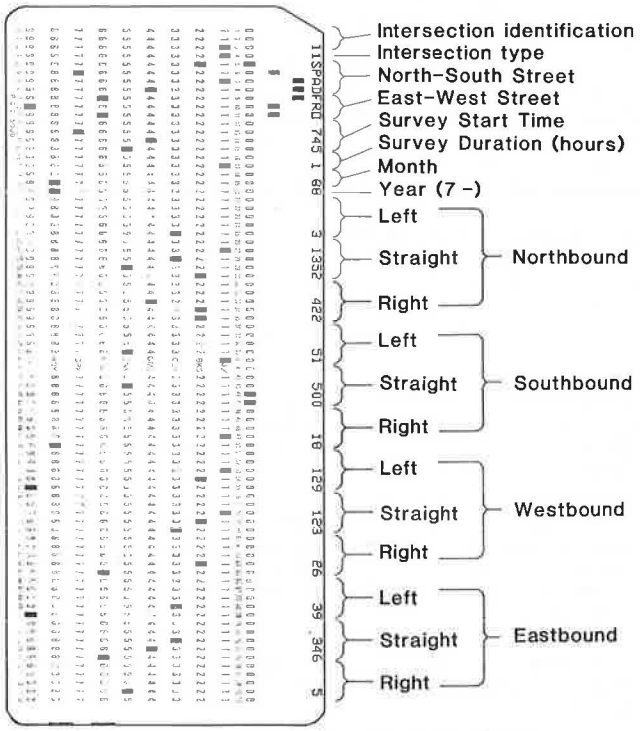


Figure 5. Data card for traffic-flow diagram in Figure 4.



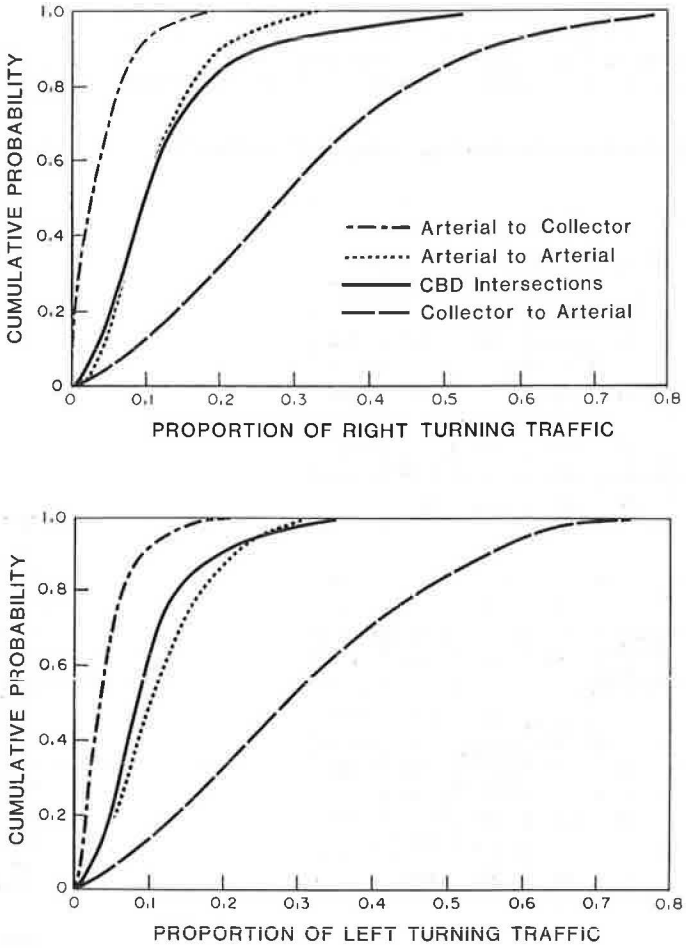
politan Toronto, arterials are approximately 1.25 miles apart. All lesser streets are included in the category "collector".

Many factors can serve to explain the differences between the turning proportions; these factors include the type of intersection or approach, the time of day, the direction of movement, the location in the urban area, nearby turn restrictions, and local land use. After detailed exploration of the various factors, it has been found that a large part of the difference can be attributed to the function of the road from which the vehicles enter the intersection and the function of the road by which they leave the intersection. The location in the urban area is another important factor. Accordingly, the estimates of average turning proportions given in Table 1 have been obtained. The empirical probability distribution of p_{ij} for approach types 1-4 is shown in Figure 6.

Table 1. Average turning proportions.

No.	Type of Approach	Proportion		No. of Approaches in Sample
		Turning Left	Turning Right	
1	CBD	0.10	0.12	92
2	Arterial to arterial	0.12	0.12	83
3	Arterial to collector	0.04	0.05	52
4	Collector to arterial	0.30	0.32	53
5	Collector to collector	0.10	0.20	3

Figure 6. Probability distribution of p_{ij}: left- and right-turning traffic.



COMPARISON OF OBSERVED FLOWS AND THEIR ESTIMATES

By using the average turning proportions in Table 1 and the observed approach flows, estimates of the entire flow matrix were obtained for all 145 intersections. This allows a comparison of observed flows and the estimates that could be obtained if turning flows were not counted.

To illustrate, the correspondence of flows and estimates for the intersection shown in Figure 4 is as follows:

Observed Flows	Estimated Flows	Observed Flows	Estimated Flows
18	12	3	0
500	506	1352	1353
51	50	422	420
26	43	39	19
123	128	346	348
129	105	52	68

The plotting of observed versus estimated flows for each movement of each type of approach defined in Table 1 for all 145 intersections results in Figures 7-11.

ACCURACY OF ESTIMATION

One of the main purposes of this study has been to explore the accuracy of estimates obtainable by this method. A first impression of such accuracy can be obtained by scanning Figures 7-11. Several observations can be made:

1. Points seem to be distributed approximately symmetrically around the bisector (except, of course, near the origin).
2. The band of points surrounding the bisector appears to be of constant width. This is attributable to the minimization process that provides the rationale to the estimation algorithm.
3. The accuracy of estimation varies by approach type. One reason for this variation is the distribution of p_{ij} (Figure 6). The wider the distribution of p_{ij}, the greater is the chance that the average value used in a specific case differs sub-

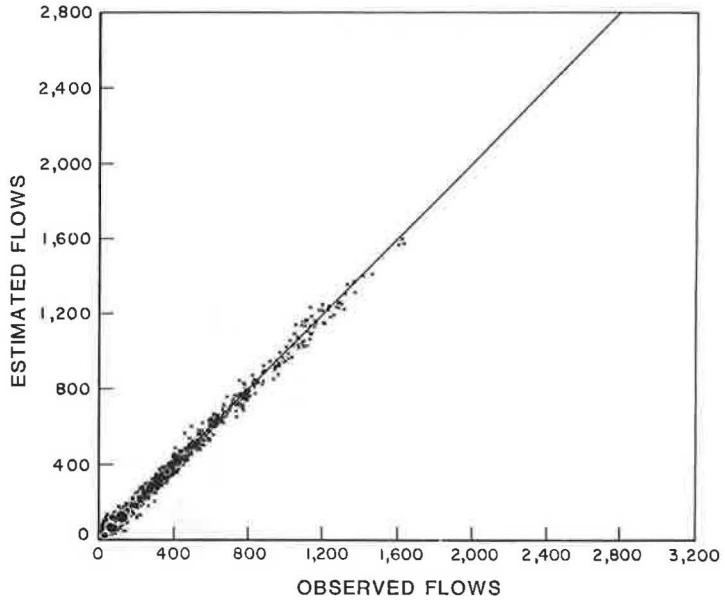


Figure 7. Correspondence of observed and estimated flows: CBD intersections.

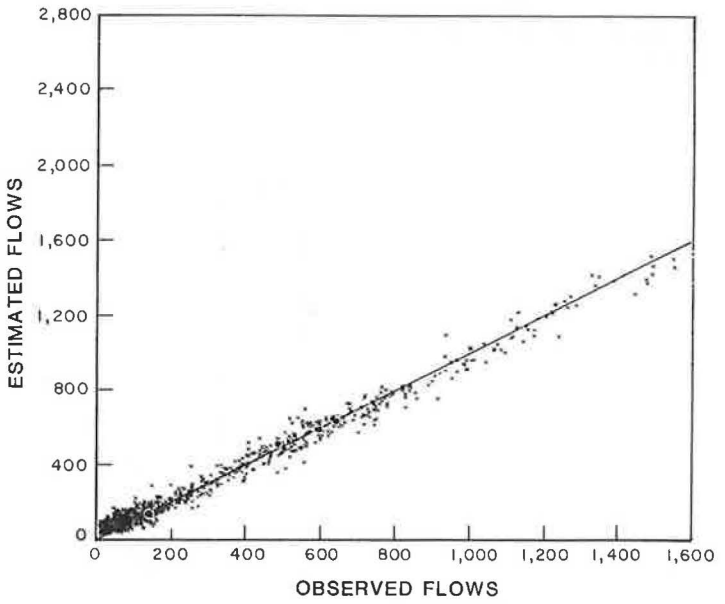


Figure 8. Correspondence of observed and estimated flows: arterial to arterial.

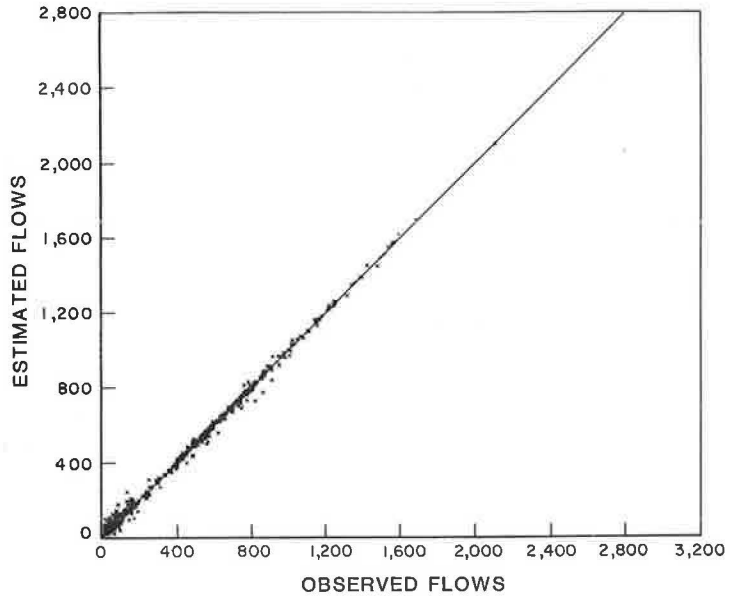


Figure 9. Correspondence of observed and estimated flows: arterial to collector.

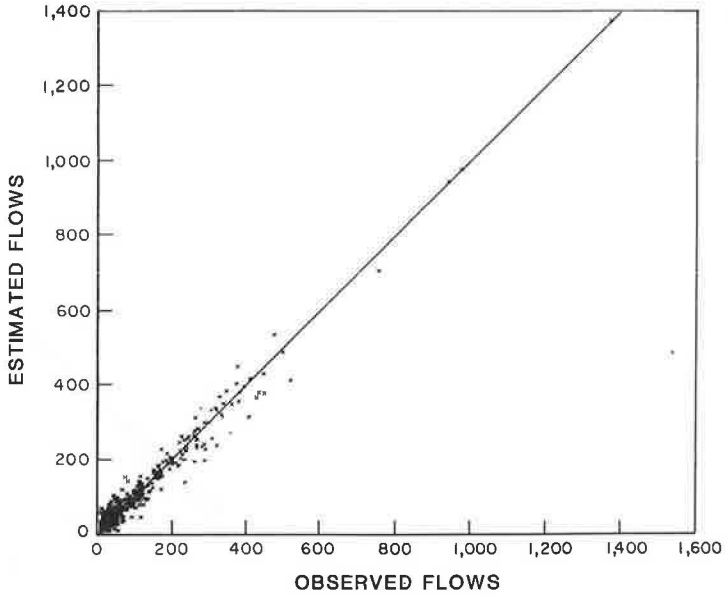


Figure 10. Correspondence of observed and estimated flows: collector to arterial.

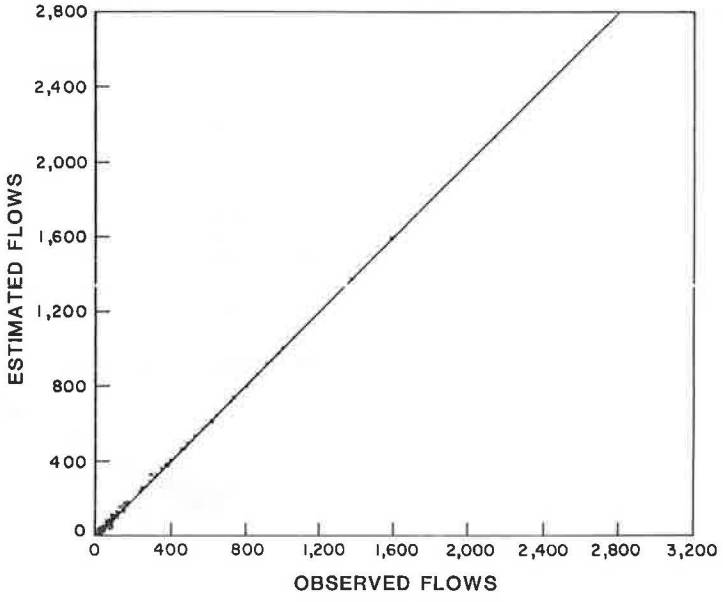


Figure 11. Correspondence of observed and estimated flows: collector to collector.

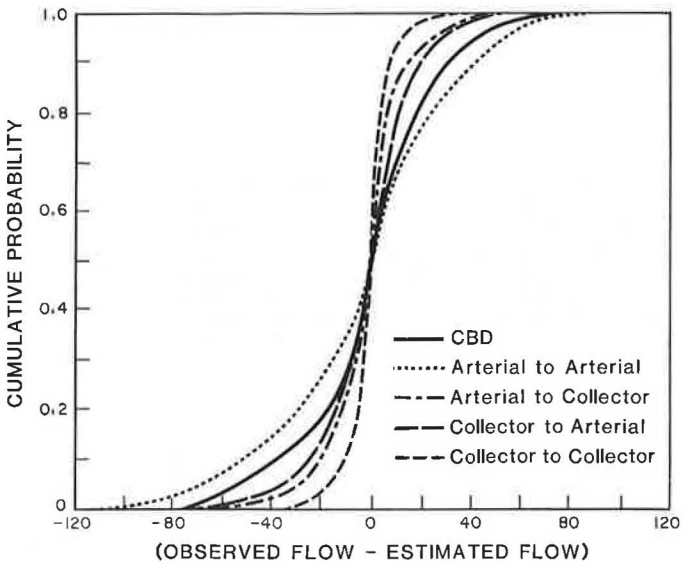


Figure 12. Distribution of difference between observed and estimated flows.

stantially from the value that actually prevails at that intersection.

4. The attainable accuracy is slightly exaggerated by the fact that the average turning proportions used (Table 1) have been derived from the same data that served for the plotting of Figures 7-11. However, limited sensitivity tests indicate that the accuracy of estimation is not affected significantly when p_{ij} are changed by a few percentage points.

The standard deviations of the difference between observed flows and their estimates are given below:

Type of Approach	Standard Deviation (no. of vehicles)
CBD	28.09
Arterial to arterial	40.83
Arterial to collector	17.67
Collector to arterial	20.07
Collector to collector	9.74

The probability distribution of this difference by approach type is shown in Figure 12. From Figure 12 one can read the probability of the difference between the observed and estimated flow to be in a certain range. To illustrate, for a CBD intersection the probability of an error less than ±60 vehicles is approximately 0.97 - 0.03 = 0.94.

SUMMARY

This paper presents a method for the estimation of vehicle turning movements from intersection approach flows. The estimation method identifies the most likely set of flows that agrees with the observed approach counts, taking also into account typical proportions of leftand right-turning flows. In principle, the method suggested by van Zuylen (3) is followed.

Flow estimates are obtained by iterative computations. The algorithm used in these computations is incorporated into a FORTRAN computer program.

To obtain realistic estimates of turning flows, prior information about characteristic leftand right-turning proportions is required. This has been obtained through the analysis of flows at 145 intersections in metropolitan Toronto. Results are summarized in the form of average turning proportions for five types of approaches.

By using the turning proportions so obtained and the observed approach flows, estimates of all flows on all 145 intersections for three periods of the day have been calculated. There appears to be a surprisingly close correspondence between the actual and the estimated flows. An empirical probability distribution curve for the difference between the actual and the observed flows is given for each of the five approach types. These can be used to anticipate the accuracy of estimation in similar circumstances.

When the obtainable accuracy is sufficient for the purpose at hand, the method described in this paper may be an attractive alternative to the conduct of a field survey by observers.

REFERENCES

1. M. Jeffreys and M. Norman. On Finding Realistic Turning Flows at Road Junctions. Traffic Engineering Control, Vol. 18, No. 1, Jan. 1977, pp. 19-21,25; Addendum, April 1977, p. 207.
2. A. Mekky. On Estimating Turning Flows at a Road Junction. Traffic Engineering Control, Vol. 20, No. 10, Oct. 1979, pp. 486-487.
3. H.J. van Zuylen. The Estimation of Turning Flows on a Junction. Traffic Engineering Control, Vol. 20, No. 11, Nov. 1979, pp. 539-541.
4. M. Norman and N. Hoffmann. Non-Iterative Methods for Generating a Realistic Turning Flow Matrix for a Junction. Traffic Engineering Control, Vol. 20, No. 12, Dec. 1979, pp. 587-589.
5. E. Hauer and B.T. Shin. Estimation of Origin-Destination Matrices from Observed Flows: Simple Systems. Toronto-York Joint Program in Transportation, Univ. of Toronto and York Univ., Toronto, Res. Rept. 65, Feb. 1980.
6. J.D. Murchland. Applications, History, and Properties of Biand Multi-Proportional Models. Transport Studies Group, University College, London, England, 1978.
7. J. Kruithof. Calculation of Telephone Traffic. De Ingenieur, Vol. 52, No. 8, 1937.
8. Highway Capacity Manual. HRB, Special Rept. 87, 1965.

A Probabilistic Model of Gap Acceptance Behavior

THOMAS H. MAZE

Although gap acceptance measures are often used in traffic engineering problems, there are no known, widely used techniques to account for the distribution of gap acceptance behavior at a specific site. The use of a simple-to-calibrate logit function to model the distribution of gap acceptance behavior is described. Because of the ease of modeling and the resulting good fit derived for this application of the model, it is believed that similar efforts could be conducted to model the actual distribution of gap acceptance behavior for use in specific design problems. However, further study with greater quantities of data and at other locations is recommended.

This paper describes an empirical study of the gap acceptance behavior of drivers merging from a stop on a one-way, single-lane street into the major flow on a one-way, single-lane arterial. The distribution of gap acceptance is a model with a logit function. By use of a logit function, the cumulative probability of accepting a gap of a specific length is modeled with a reasonably good fit.

Measures of gap acceptance enter into the calculation of the capacity of unsignalized intersections, warrants for stop signs, the capacity of weaving and merging areas, and other design problems. Although gap acceptance measures are important to design, there is no known, widely accepted technique that permits the design engineer to account for the distribution of gap acceptance behavior at a specific design study location. The Transportation Research Board's recent interim update of the Highway Capacity Manual (1) accounts for gap acceptance behavior with average, aggregate measures. These measures are used despite the fact that gap acceptance behavior is known to vary in different locations with respect to not only the mean length of gaps accepted but also the skew of the distribution of gap lengths accepted (2).

In cases where data on gap acceptance behavior have been collected for specific sites, they have traditionally been gathered by using cameras and pen recorders. Not only does this method imply great drudgery but also, as one recent gap acceptance study indicated, it cannot reliably measure gap lengths at smaller intervals than 0.5 s (3)--not to mention the possibility of error involved in transcribing data from films and pen recorders to a form usable for data processing.

Because of the cost, drudgery, and error involved in collecting data, and because there are no easy-to-use methods for synthesizing distributions of gap acceptance behavior even if data are collected, it is understandable that average, aggregate measures are often resorted to in design problems.

This paper presents a simple but robust means of synthesizing a distribution of gap acceptance behavior. The paper is divided into the following sections:

1. A description of the theoretical model that structures the inputs of the gap acceptance decision;
2. A description of the data collection site, the means by which data are collected, and the data themselves;
3. A description of the empirical model and its calibration; and
4. The conclusions derived from the empirical investigation.

CONCEPTUAL MODEL

To structure the model, a description of the behavior of drivers when they are confronted with a gap acceptance decision is theorized. The theory as defined is no more than fitting a rational decision process into a conceptual, mathematic framework.

Decision Inputs

In setting up a structure for a rational gap acceptance decision, two constraints are invoked:

1. No driver will accept a gap in the major stream that he or she believes will certainly lead to a collision.
2. No merging driver gains admittance to the major stream through intimidation of major-stream drivers.

These constraints are sometimes violated, but such types of behavior are considered irrational and are dropped from consideration. Given these two constraints, the inputs to the decision process can be postulated as follows.

Driver Risk

Risk is the value the driver places on the probability of collision during a merge with the major flow of traffic. All drivers are assumed to be adverse to accepting a gap in the major stream that implies a high degree of risk. The driver assigns a positive value to the risk of accepting all gaps, and the value of risk will become large as the gap length becomes small. A driver will decide to accept a gap only if the value of risk is less than the value assigned to the estimated delay of waiting for a larger gap.

Value of Delay Time

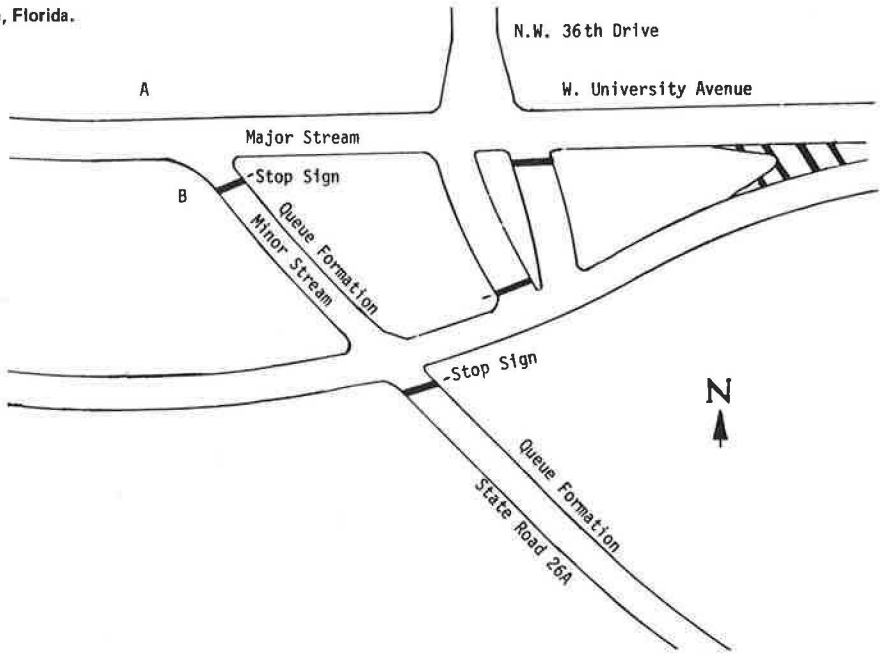
The value of delay due to gap refusal is the driver's estimated value of the time that will elapse until a suitable gap occurs. The driver's assessed time value of delay is always positive and is modified by (a) the length of time spent in the queue at the intersection and (b) the traffic volume in the major stream.

The length of time spent in the queue is a measure of exposure to irritation caused by the gap acceptance process. The degree of irritation the driver has been exposed to will modify the driver's weighting of the time value of delay. In other words, as the driver approaches the head of the queue, the estimated time value of delay will be marginally increasing with additional delay.

As traffic volumes in the mainstream increase, the merging driver understands that the delay caused by waiting for a larger gap will become greater. Therefore, the driver's estimated time value of delay due to gap refusal increases with increasing mainstream traffic volumes.

In another study (4), these modifications of driver behavior were loosely titled "pressure of traffic demand". Although the term "pressure" does conjure the correct image of what takes place as traffic demand increases, in reality pressure has nothing to do with gap acceptance behavior. What is actually being modified by increased traffic demand is the driver's estimate of the value of delay time.

Figure 1. Study intersection in Gainesville, Florida.



Decision Process

The two inputs into the gap acceptance decision process, the value of risk and the value of delay time, can be structured into a model of gap acceptance behavior. In the structure, risk is assumed to be independent of the length of the minor-stream queue and the major-stream traffic volume. If the value of risk assigned to accepting a gap is greater than the assigned value of delay due to not accepting a gap, then the gap is refused. If the value of risk assigned to accepting a gap is less than the assigned value of delay due to not accepting a gap, then the gap is accepted. This decision process is defined in Equations 1-3:

$$(VR_i) > (VT_i) \times f(Q_i, V_i)_{\text{gap refusal}} \quad (1)$$

$$(VR_i) < (VT_i) \times f(Q_i, V_i)_{\text{gap accepted}} \quad (2)$$

$$(VR_i) = (VT_i) \times f(Q_i, V_i)_{\text{undefined}} \quad (3)$$

where

VR_i = value assigned to the risk of accepting gap i,
VT_i = value assigned to the time penalty estimated for refusal of gap i,
$f(Q_i,V_i)$ = function that accounts for modifications in driver delay-time judgment,
Q_i = queue length in minor stream at time of gap i, and
V_i = mainstream traffic volume at time of gap i.

COLLECTION AND PREPARATION OF FIELD DATA

An unsignalized intersection in the western part of Gainesville, Florida, was chosen for the study of gap acceptance behavior (see Figure 1). This intersection was designed so that a single-lane, one-way, major-stream movement intersects a single-lane, one-way minor stream. The intersection was observed during the period of greatest congestion, the afternoon peak.

During the afternoon peak, the minor-stream traffic would back up and build a large queue. During the congested period, acceptable gaps in the mainstream became less frequent. Saturation of the intersection tested gap acceptance behavior under critical conditions.

The data collection was conducted during a Friday afternoon in the autumn over a period of 2-3 h. The equipment was set up on the north side of University Avenue, across the street from the merge area (point A in Figure 1). Three people were used to record gap acceptances. Two observers operated hand-held switches. One of these two observers recorded a signal whenever, in time, a mainstream vehicle entered the merge area, thus measuring the time length of all mainstream gaps. The second observer recorded, in time, the acceptance of a gap by a minor-stream vehicle. The data were stored by using the MEMODYNE system, which stores data on magnetic tape (cassette) and records the time lengths of gaps by means of an internal clock.

The third observer measured the length of the minor-stream queue at 1-min intervals. The queue-length data were later merged with the data stored by the MEMODYNE system.

The field data were transferred to disk storage at the University of Florida computer facilities. The data were then formatted for analysis with a standard statistical software package (5).

EMPIRICAL MODEL

The conceptual decision inputs included an assigned value of risk and the estimated value of delay due to gap refusal. To say the least, measurements of such driver perceptions would be difficult to collect. Thus, the weights drivers placed on inputs are estimated through regression. The driver-perceived values are hypothesized to take the following form:

$$VR_i = H(t_i) \quad (4)$$

$$VT_i \times f(Q_i V_i) = G(t_i, Q_i, V_i) \quad (5)$$

where t_i is the time length of gap i and where it is observed that, (a) if $H(t_i) > G(t_i, Q_i, V_i)$, then the gap is refused and (b) if

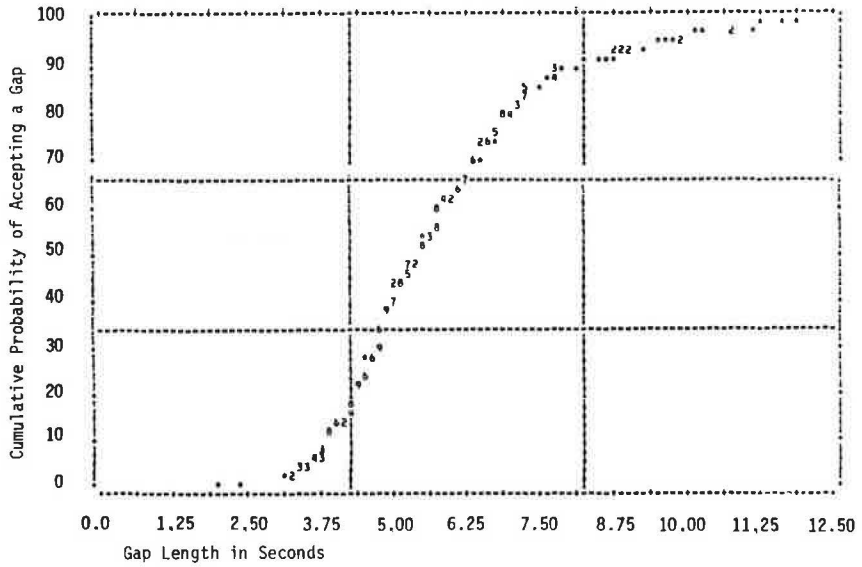


Figure 2. Gap length versus cumulative probability of gap acceptance.

$H(t_i) < G(t_i, Q_i, V_i)$, then the gap is accepted. Since the model of gap acceptance behavior is only interested in gaps accepted, only events where formula b held are examined here. In addition, because the variable used to model the value of risk is a subset of the variables used to model time value of delay and because the events examined are a homogeneous set (only gaps accepted), the model can be condensed to the following:

$$H(t_i) - G(t_i, Q_i, V_i) = F(t_i, Q_i, V_i) \quad (6)$$

Analysis Technique

Choice modeling is commonly done by using the cumulative probability of making a certain choice. In the case of gap acceptance, the choice is whether or not to accept a gap. Such choice phenomena are often modeled by use of either probit or logit analysis.

Probit analysis has been used in the past to synthesize distributions of gap acceptance behavior. Notable examples are Solberg's and Oppenlander's analysis of unsignalized intersections (2) and Drew's analysis of merging at freeway ramps (6). In these studies, a probit functional form was used to model the cumulative probability of accepting gaps of varying lengths. Probit analysis fits the dependent variable to a normal, cumulative probability distribution. To estimate a model of this functional form requires the use of maximum likelihood. Although maximum-likelihood procedures have been greatly improved, maximum likelihood is still cumbersome.

In a recent study of gap acceptance, Radwan and Sinha (3) modeled the cumulative probability of accepting gaps of varying lengths with logit analysis. They collected data by using time-lapse photography and a 20-pen recorder. By using data collected from a stop-controlled multilane intersection, Radwan and Sinha constructed a biased model by forcing a symmetrical logit function to fit what they admit is a skewed distribution (median is not equal to the mean). Although their model specification is biased, Radwan and Sinha have provided an example of the applicability of logit analysis in modeling gap acceptance behavior.

Because of the logit's simplicity of calibration, it is chosen for use in this study. The logit closely approximates the probit and may be linearly transformed to provide easy estimation of model parameters with linear regression. The simple, dichotomous choice logit functional form is

$$P = \{1/[1 + e^{F(x)}]\} \quad -\infty F(x) < \infty \quad (7)$$

and its linear transformation is

$$\ln[P/(1 - P)] = F(x) \quad (8)$$

where

P = cumulative probability of accepting a gap,
x = variables related to the gap acceptance decision, and
F(x) = linear function.

Dependent Variable

The dependent variable, the cumulative probability of accepting a gap of a specific length, is calculated by using the following equation:

$$P_i = (d_i/N) \; 0 < P < 1 \quad (9)$$

where

P_i = cumulative probability of accepting a gap of time length i,
d_i = number of gaps accepted of time length i or less, and
N = total number of events.

A plot of the values derived from the calculation of P_i (Equation 9) is shown in Figure 2. The uniform S-shape of the data points is the first clue that the study is on the right track in using a cumulative probability functional form to model gap acceptance.

The plot of the cumulative probability of accepting a gap, shown in Figure 2, is skewed. The mean length of the gaps accepted is greater than the value that coincides with the gap length that was accepted by 50 percent of the sample (median). The slope on the lower part of the curve is steep and then tends to flatten at the top. In other words, marginal change in the bottom of the curve is greatest, and the marginal change decreases as the curve is followed to the top.

In the data preparation, the problem of multiple

drivers accepting one gap arose. Because the majority of gap acceptance observations are single-vehicle acceptances (261 events), the analysis is limited to acceptances of gaps by one vehicle. Presumably, with more data the model could be expanded to include multiple acceptance. However, the objective of this study is to show how a simplistic technique can be used to efficiently model the gap acceptance decision. Therefore, it is assumed that only modeling single-vehicle gap acceptance would prove the case for logit analysis.

Independent Variables

Drew (6) modeled the cumulative probability of accepting a gap with the following specification:

$$P_i = F_p[\alpha + \beta(\log t_i)] \quad (10)$$

where

α = intercept,
β = slope coefficient, and
$F_p(\ )$ = probit functional form.

Drew skewed his independent variable by using the logarithm of t_i instead of t_i. The model specified here is similar to Drew's, but the skew is accounted for in a different manner. The conceptual model specified in Equation 6 is specified by using the linear form given below:

$$P_i = F_L(B_0 + B_1 X_i + B_2 Q_i + B_3 V_i) \quad (11)$$

where

B = slope coefficient,
$X_i = (\bar{T}/t_i) - 1$,
$\bar{T}$ = mean time length of all gaps accepted, and
$F_L(\ )$ = logit functional form.

X_i is used as the independent variable in Equation 11 instead of t_i for two reasons:

1. $\bar{T}/t_i$ is used instead of t_i because, when $t_i \leq \bar{T}$, the changes in the values of $\bar{T}/t_i$ range from one to infinity and, when $t_i \geq \bar{T}$, the values of $\bar{T}/t_i$ range from zero to one. Thus, the marginal changes of $\bar{T}/t_i$ are greatest when $t_i \leq \bar{T}$, which is consistent with the skew of the distribution of gap lengths accepted (Figure 2).
2. One is subtracted from $\bar{T}/t_i$ so that, when $t_i = \bar{T}$, the value of the independent variable would be zero. This made the regression parameters easier to interpret.

Estimations

Neither mainstream traffic volume nor queue length is found to have statistically significant slope coefficients. Because they add nothing to the model, they are dropped. Not including them does not seem to affect the strength of the final estimate. The fact that the mainstream volume does not modify gap acceptance behavior does not seem as surprising as queue length not having a significant impact on gap acceptance behavior. Sometimes the queue was as long as 20 cars or more, a delay that would seem long enough to modify behavior. The bias in the model of deleting volume queue length is examined later.

The following linear model was estimated by using ordinary least squares:

$$Y_i = \alpha + \beta(X_i) + E_i \quad (12)$$

where $Y_i = \ln[P_i/(1 - P_i)]$ and E_i = stochastic error. The resulting parameter estimates and regression statistics are as follows:

$$Y_i = 0.422 - 0.965[(\bar{T}/t_i) - 1] \quad (13)$$

t-statistic of $\beta_1 = 10.599$
$R^2 = 0.931$
$F = 3525.667$

Caution must be used in accepting the regression statistics as totally valid. In their calculation, it is assumed that the function estimated was linear. Still, the fit that is found by using $[(\bar{T}/t_i) - 1]$ is quite good, and the true statistics will be close to those formulated.

Bias Due to Deleted Variables

In this instance, the volume of mainstream traffic demand does not appear to have a significant impact on the length of gaps accepted. However, the lack of significance in this case does not mean that mainstream volume does not have an impact on gap acceptance in general. Study of other intersections

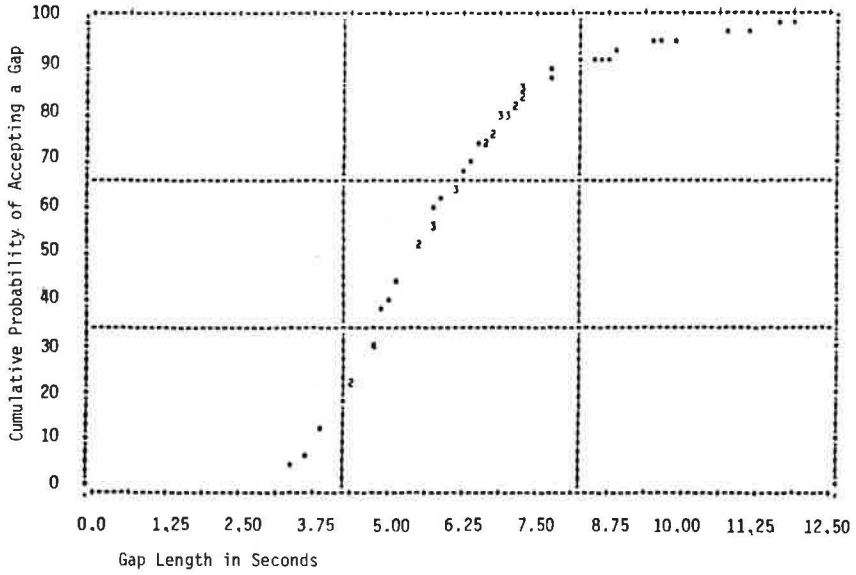


Figure 3. Gap length versus cumulative probability of gap acceptance for queue length of 1-5 automobiles.

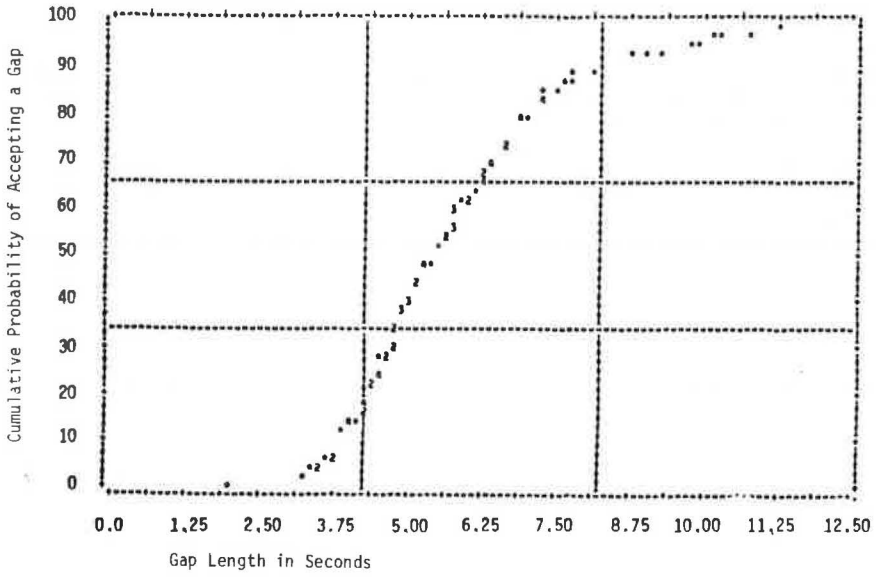


Figure 4. Gap length versus cumulative probability of gap acceptance for queue length of 6-10 automobiles.

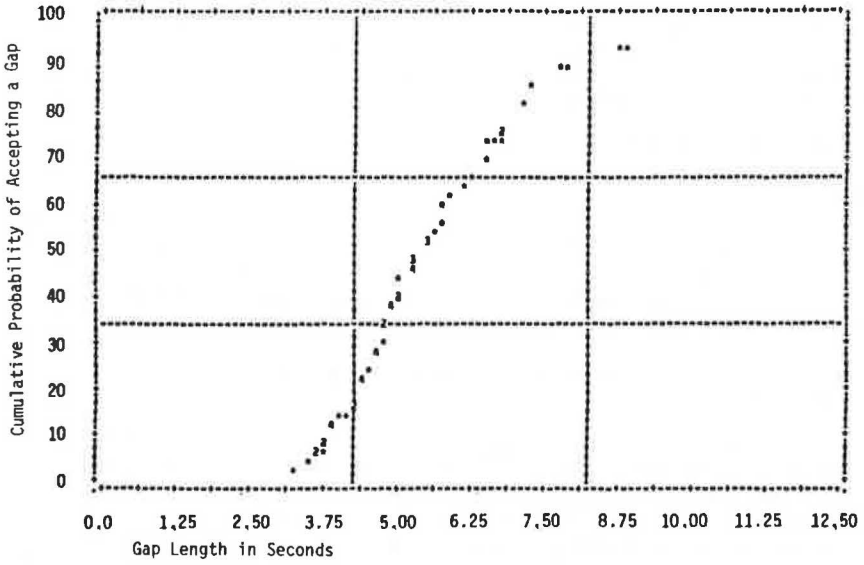


Figure 5. Gap length versus cumulative probability of gap acceptance for queue length of 11-15 automobiles.

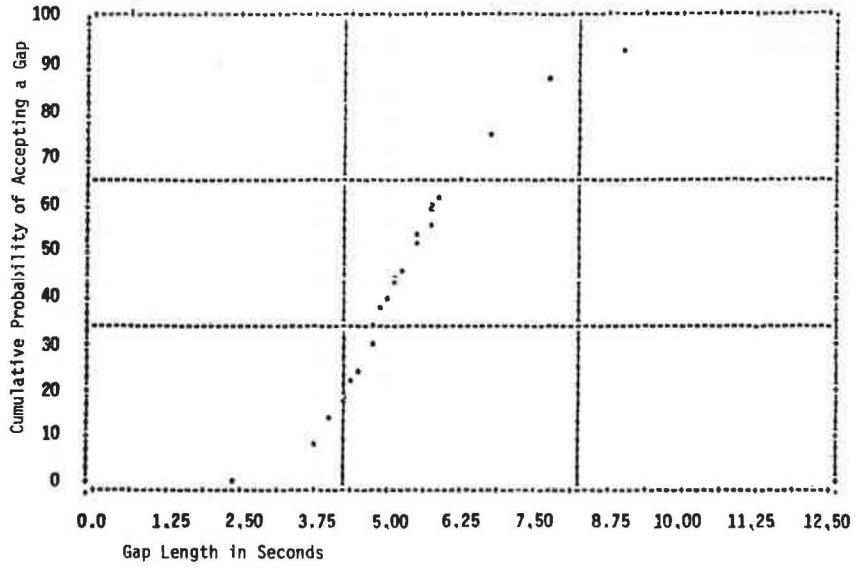


Figure 6. Gap length versus cumulative probability of gap acceptance for queue length of 16-20 automobiles.

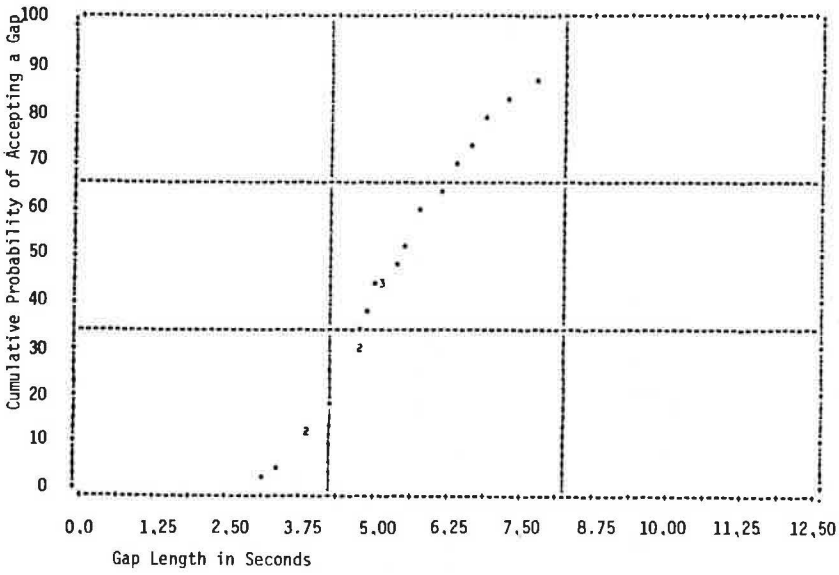


Figure 7. Gap length versus cumulative probability of gap acceptance for queue length of ≥20 automobiles.

is needed before concrete evidence will be available.

More surprising is that the length of the queue does not seem to affect the length of the gap accepted. The median gap lengths accepted as queue lengths increased (minimum gap lengths accepted by 50 percent of the sample) are given below:

Queue Length (no. of cars)	Median Gap Length (s)
1-5	5.50
6-10	5.55
11-15	5.58
16-20	5.60
≥21	5.58

There appear to be no significant differences with increasing queues. In Figures 3-7, the cumulative probabilities of accepting gaps of various lengths stratified with respect to queue length are plotted. All plots appear to have relatively the same distributions. If the observations are consistent with the theory of the marginally increasing value of additional delay time, then the plots with longer queues should appear steeper and closer to the left-hand side. This does not appear to be the case. Thus, it is assumed that there is minimal bias due to the deletion of queue length from the model. However, the real proof of bias would only come with a more data-intensive effort.

CONCLUSIONS

This paper describes an empirical study of the gap acceptance behavior of drivers merging from a stop on a one-way, single-lane street into the major flow on a one-way, single-lane arterial. When a logit function is used, the cumulative probability of accepting a gap of a specific length is modeled with a reasonably good fit.

This study shows the effectiveness of using a simplistic logit form to model gap acceptance behavior. However, more work should be done to account for variables deleted in this study. In addition, the acceptance of gaps by multiple drivers and under other circumstances, such as in freeway weaving and merging areas, should be investigated. However, the simplistic method described here of accounting for the distribution of gap acceptance behavior would permit the design engineer to have better knowledge of driver behavior at the design location with a minimum of effort.

ACKNOWLEDGMENT

The use of K. Courage's MEMODYNE system for data collection and his assistance are greatly appreciated.

REFERENCES

1. Interim Materials on Highway Capacity. TRB, Transportation Research Circular 212, Jan. 1980.
2. P. Solberg and J.C. Oppenlander. Lag and Gap Acceptance at Stop-Controlled Intersections. HRB, Highway Research Record 118, 1965, pp. 48-67.
3. A.E. Radwan and K.C. Sinha. Gap Acceptance and Delay at Stop-Controlled Intersections on Multi-Lane Divided Highways. Institute of Transportation Engineers Journal, March 1980, pp. 38-44.
4. F.A. Wagner, Jr. An Evaluation of Fundamental Driver Decisions and Reactions at an Intersection. HRB, Highway Research Record 118, 1965, pp. 68-84.
5. N.H. Nie and others. Statistical Package for the Social Sciences, 2nd ed. McGraw-Hill, New York, 1975.
6. D.R. Drew. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968, pp. 193-196.

Sensitivity of Fuel-Consumption and Delay Values from Traffic Simulation

JAMIE W. HURLEY, JR., AHMED E. RADWAN, AND DAVID A. BENEVELLI

The use of a fuel-consumption model developed by using field measurements obtained previously by Claffey is described. The model, called the modified fuel-consumption model, was derived in a form that is suitable for insertion into the NETSIM computer simulation program. Sensitivity analyses were performed by using the existing NETSIM fuel model and the modified fuel-consumption model. The effect of two headway distributions (uniform and shifted negative exponential) on fuel consumption and delay was tested for a hypothetical isolated intersection and an existing small, open network. The impact of various saturation headway values on delay and fuel consumption was investigated, and the incorporation of grade effects in NETSIM through changes in the saturation headway was evaluated. Significant differences in results were found between the two fuel models, and significant differences in both delay and fuel-consumption estimates were found between the two headway models. It was found that delay and fuel consumption were insensitive to saturation headway values between 2.0 and 2.2 s, but significant differences resulted for low to moderate volumes when saturation headway was increased to 2.4 s. Although gradient effects per se have yet to be developed for inclusion in NETSIM, these effects on saturation headway values and the consequent impacts on delay and fuel consumption were found to be significant only at high volumes for grades between -4 and +4 percent.

Heavy dependence on imported oil and the recent extreme increases in the price of gasoline are responsible for the initiation of plans to retime traffic signals in urban areas so as to minimize fuel consumption rather than delay. Currently, there are two schools of thought on how this may be done. Some investigators have used macroscopic approaches to illustrate that extremely long cycle lengths are necessary to minimize fuel consumption (1,2). Naturally, a delay penalty is incurred with this strategy. Other researchers, using either macroscopic or microscopic approaches, have observed that the cycle length that minimizes delay also minimizes fuel consumption (3-5). Any concerted effort to retime traffic signals to minimize fuel consumption should logically be delayed until basic questions such as this have been resolved. Although this paper does not address this particular problem, it does attempt to provide some degree of insight into the state of the art regarding traffic operations and fuel consumption.

Use of microscopic traffic simulation computer programs has long been considered to be a viable and practical technique for evaluating traffic flow. The alternative--collecting and analyzing field data--is time consuming, expensive, and may in the case of traffic networks be impracticable. A detailed simulation program that predicts fuel consumption with reasonable accuracy could be valuable not only for resolving the basic questions concerning minimization of fuel consumption but also for validating other computer programs more likely to be used by engineers in the field. The network flow simulation (NETSIM) program (6) is an existing program that shows considerable promise for serving as such a baseline. NETSIM was developed primarily for closed-network applications, but it has been validated for isolated intersections (6). The program is capable of simulating both pretimed and actuated control and includes fuel-consumption and emissions data in its output. NETSIM has, in fact, been used in some of the investigations concerning cycle lengths that minimize fuel consumption (3,5). Although the program is broad and comprehensive, the making of general statements based on its output might be premature. Questions exist concerning the impact of certain input variable values, the internal program logic, and fuel consumption and delay. The concerns addressed in this paper are

1. The choice of fuel-consumption model,
2. The effect of the headway-distribution model on fuel consumption and delay,
3. The impact of saturation-headway values on fuel consumption and delay, and
4. The effect of grade on fuel consumption through changes in saturation headway.

The remainder of the paper describes program modifications and analyses that provide insight into these concerns.

DEVELOPMENT OF FUEL-CONSUMPTION MODEL

The fuel-consumption tables used in the current version of the NETSIM program were developed from laboratory-based data obtained for 1971-model-year vehicles (7). These tables provide instantaneous rates of fuel consumption in gallons per 100 000 s as a function of vehicle speed and acceleration. Program flexibility permits, at the user's option, the inclusion of alternative fuel-consumption tables. One set of fuel-consumption data frequently used by analysts is that of Claffey (8). Claffey's data may be an attractive alternative to some, since they were measured in the field rather than in a laboratory. These fuel-consumption data were obtained over a wide range of operating conditions for vehicles of the mid-1960s. One might question whether either the fuel-consumption data currently stored in NETSIM or Claffey's data are appropriate for today's fleet-mix and fuel-consumption characteristics. Although nothing can be said with certainty, two earlier studies indicate that these impacts may be small (9,10).

Claffey's data are not presented in a form that is directly suitable for insertion into the NETSIM program. The data that pertain to this study are given in terms of fuel consumption at constant speed, fuel consumption while idling, and excess fuel consumed during speed-change cycles. Excess fuel consumption during a speed-change cycle is defined as the additional fuel consumed beyond that which would have been required had the vehicle continued at constant speed (without performing the speed-change maneuver).

Direct comparisons between Claffey's data base and that of the NETSIM program can be made by converting the NETSIM data to the form used by Claffey. Claffey measured idling fuel consumption at 2.2 L/h (0.58 gal/h) as opposed to the 3.4 L/h (0.90 gal/h) used in NETSIM. The differences in the Claffey and NETSIM data for constant speed at zero grade are shown in Figure 1 for speeds up to 64 km/h (40 miles/h). In order to compare the excess fuel consumed during speed-change cycles, the NETSIM instantaneous consumption rates were integrated over time. This was done for the entire speed-change cycle by using the speed profile that would be generated internally by NETSIM for an isolated vehicle.

A typical NETSIM speed-change profile is shown in

Figure 2. Deceleration takes place at -0.3 m/s² (-1 ft/s²) from some initial speed to 90 percent of the initial speed, followed by deceleration at -2.1 m/s² (-7 ft/s²) to all lower speeds. An acceleration rate of 2.4 m/s² (8 ft/s²) is used for speeds between zero and 3.4 m/s (11 ft/s) and 1.8 m/s² (6 ft/s²) is used for speeds between 3.4 and 5.2 m/s (11 and 17 ft/s). For all speeds greater than 5.2 m/s, the program uses an acceleration rate of 0.9 m/s² (3 ft/s²). The results of these calculations are shown with Claffey's speed-change data in Figure 3. The two data sets exhibit the same general trends, although the differences in magnitudes are relatively large.

To convert Claffey's data to instantaneous consumption rates as required by NETSIM, it was necessary to develop an analytic model that would not only reasonably duplicate Claffey's data but would do so for the NETSIM speed profile. In other words, the task was to develop a model that would enable the NETSIM program to give the "correct" fuel consumption, even though Claffey's tests almost certainly used a somewhat different speed profile. Stepwise regression was used to develop a number of candidate models. Of these candidates, only one model was found that would predict Claffey's data with reasonable accuracy and at the same time show reasonable values and trends for instantaneous fuel consumption. This model is based on the following assumptions:

1. The instantaneous fuel-consumption rate is never less than the idling rate.
2. For decelerations less than -0.3 m/s (-1 ft/s), the engine operates at a no-load/no-throttle condition, which results in the instantaneous idling-consumption rate.

For all conditions other than these, the model is given in terms of the NETSIM program units by

$$\dot{g} = 16.117 + 0.1658V + 0.007\,252V^2 + 9.626a - 0.009\,577aV^2 + 0.208\,45a^2V \quad (1)$$

where

$\dot{g}$ = instantaneous fuel-consumption rate (gal/100 000 s),
a = instantaneous acceleration rate (ft/s²), and
V = instantaneous velocity (ft/s).

The implemented model, valid for typical urban speeds only, estimates Claffey's data well for idling, all constant-speed conditions, and all speed-change cycles except for the stop-go speed-change cycle from an initial speed of 16 km/h (10 miles/h). With that exception, the model estimates Claffey's values within 4 percent. A considerably larger error for the 16-km/h stop-go speed-change cycle is possibly caused by the fact that the NETSIM speed profile assumes a higher acceleration rate over most of the acceleration portion of the maneuver than would likely be encountered in practice. In reality, the only time a vehicle is likely to accelerate to a final speed of 16 km/h is when in a queue, and one would not expect

Figure 1. Fuel-consumption rates at constant speed.

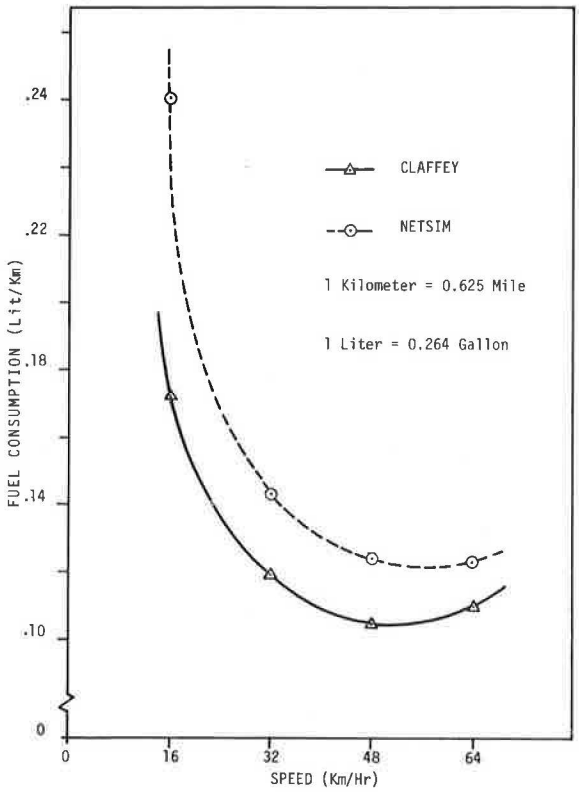
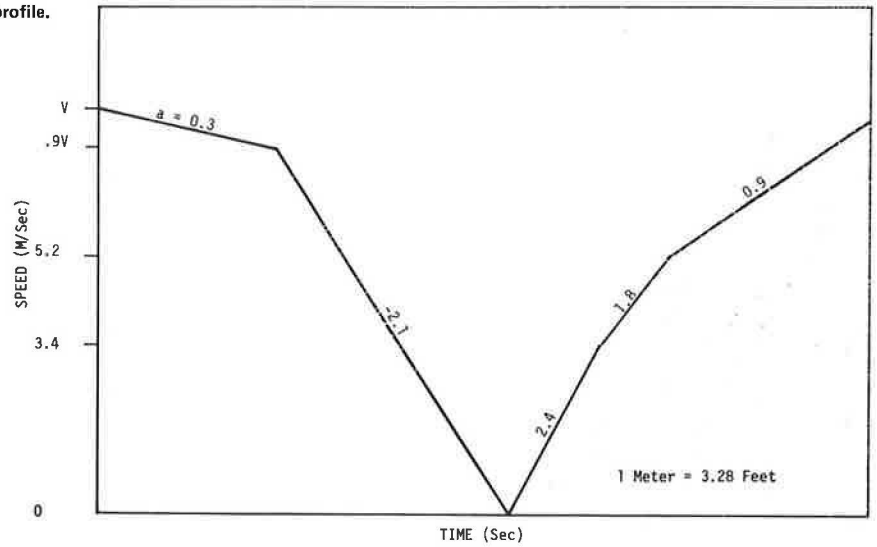


Figure 2. NETSIM speed-change profile.



to observe high acceleration rates under this condition. It should be noted that NETSIM's car-following logic would probably preclude such high values of acceleration being attained in a queue.

HEADWAY-DISTRIBUTION MODEL

The NETSIM program uses car-following logic to trace individual vehicles through the system. Before car following can take place, however, vehicles are "generated" according to a uniform statistical distribution. This logic for vehicle generation may be reasonable for a large network with high traffic volumes, but it could give inaccurate results for lowto moderate-volume isolated intersections and small networks. NETSIM was modified to generate vehicles according to a shifted-negative-exponential distribution with an assumed minimum headway of 1 s. It is noted that vehicles were generated in this manner in the original version of NETSIM (called UTCS-1).

ANALYSES

Sensitivity analyses were performed with respect to the fuel-consumption model, the headway-distribution model, and values of saturation headway to assess their impact on fuel consumption and delay (saturation headway is defined as the headway corresponding to a saturation flow rate). A hypothetical isolated intersection, shown in Figure 4, was analyzed for a stream that consisted of passenger cars only with no turning movements. Intersection control was assumed to be simple two-phase, pretimed control with a 60-s cycle length and 40 s of green plus amber allocated to the main street for all test cases. The link-node diagram for this hypothetical intersection is shown in Figure 5. To illustrate the magnitudes of the impacts and their variation with demand for an isolated intersection, fixed cross-street demand volumes of 400 vehicles/h and main-street volumes of 600, 1200, and 1800 vehicles/h on each approach were considered. This implies near-saturation conditions on the cross street and approximately 31, 63, and 94 percent saturation on the main street, respectively. It is noted that the degree of saturation for the respective approaches and demands will vary with the saturation headway assumed.

The impact of the intersection approach gradient on traffic performance has not been extensively studied. An early study by Conley (11) suggested that both negative and positive intersection approach grades decrease starting-time delay and thus increase the saturation flow rate. Dick (12) concluded that the relation between grade and rate of flow was essentially linear and that, for every 1 percent increase (decrease) in grade, saturation flow was reduced (increased) by 3 percent for gradients between -5 and +10 percent. The effects of roadway gradient on traffic operation cannot be directly obtained from a NETSIM simulation exercise, but they can be incorporated in part by changing the

Figure 3. Excess fuel consumed during speed-change cycles.

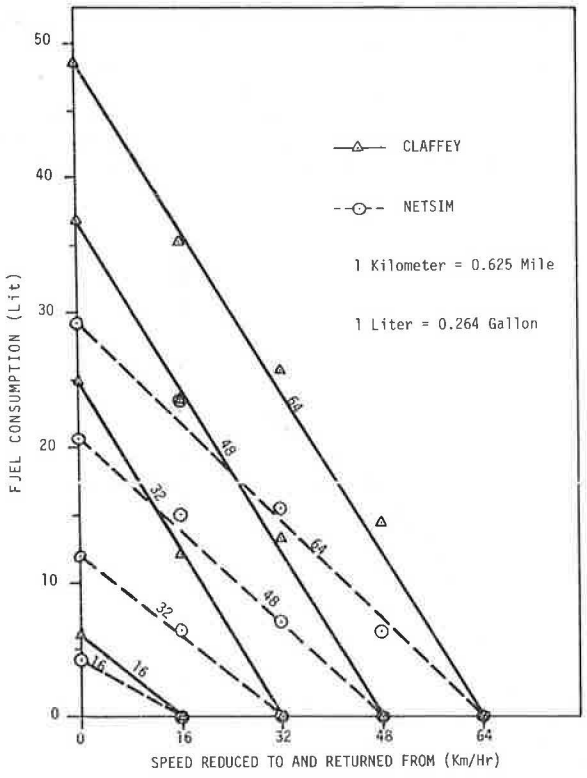


Figure 4. Isolated intersection.

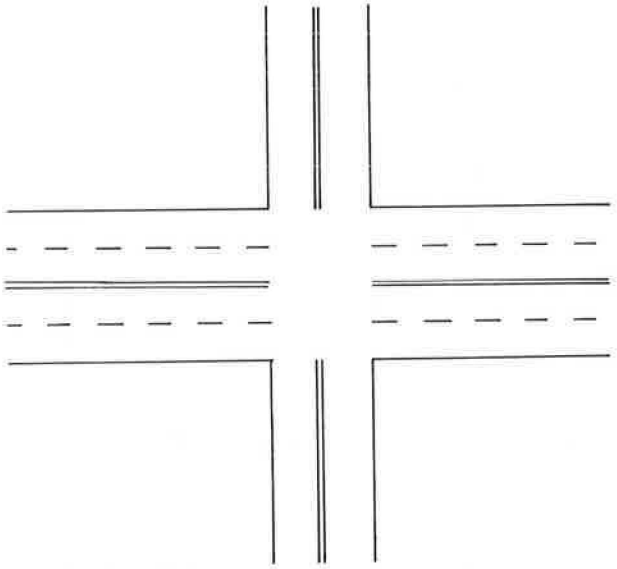


Figure 5. Link-node diagram for intersection.

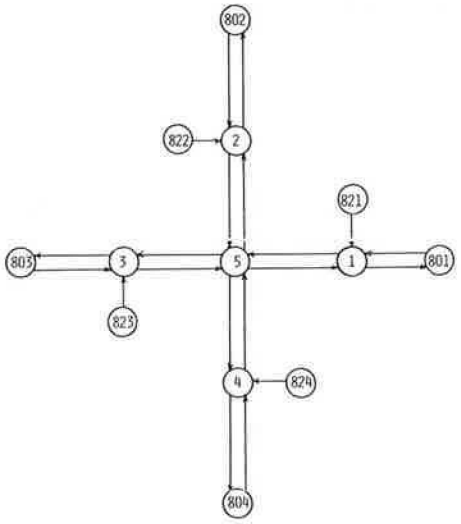
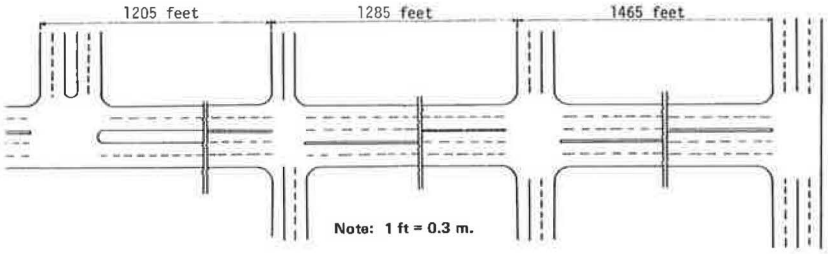


Figure 6. Prices Fork Road network.



average intersection discharge rate. For a base saturation headway of 2.2 s, and by using the adjustment factor suggested by Dick (12), saturation headways of 1.9 and 2.5 s were used to simulate the effect of -4 and +4 percent gradients, respectively. These values were input to NETSIM for the major-street approaches; a 2.2-s saturation headway was assumed for the minor-street approaches. The average fuel-consumption values (gallons per vehicle) estimated by NETSIM are attributed to traffic interaction only, and the possible effect of grade per se on fuel-consumption rates is not taken into account. Components for uniform speed and speed-change cycles as functions of grade have yet to be developed in a form suitable for NETSIM.

To illustrate the impacts of fuel-consumption and headway-distribution models on an actual network, a second set of analyses was performed for the Prices Fork Road arterial in Blacksburg, Virginia. The Prices Fork Road system is a small, open network that consists of four signalized intersections. The analyses were performed by using volumes measured in the morning peak period and a cycle length of 70 s at all intersections. The cycle splits and offsets used were obtained from the TRANSYT computer program (13). Cross-street volumes ranged from 10 to 940 vehicles/h, and link volumes on Prices Fork Road ranged from 260 to 1290 vehicles/h. The system as a whole, then, could be described as a moderate-volume operation. A physical description of the Prices Fork Road system is shown in Figure 6, and the link-node diagram is shown in Figure 7.

For the isolated intersection, the analyses are logically broken down into the four sets of simulation runs described in Table 1. Several replicates were simulated for each run (a minimum of four), each of 900-s duration. Several trials were attempted in order to select an optimum link length for both the major and minor streets. The optimum length should be long enough to prevent traffic spillback on the intersection approaches but should not be too long to negate the speed-change-cycle effect on fuel-consumption values at low volumes.

For the Prices Fork Road network, four cases were simulated for 1 h each:

1. Uniform headway distribution and NETSIM fuel models,
2. Uniform headway distribution and the modified fuel model,
3. Shifted-negative-exponential headway distribution and the NETSIM fuel model, and
4. Shifted-negative-exponential headway distribution and the modified fuel model.

Figure 7. NETSIM link-node diagram for Prices Fork Road.

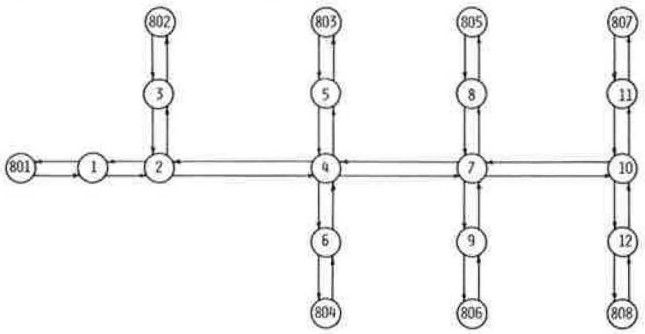


Table 1. Layout of four simulation runs for hypothetical isolated intersection.

Run Group	Run Number	Headway Distribution	Saturation Headway (s)	Grade (%)	Fuel Model	Main-Street Volume (vehicles/h)
1	1	U	2.2	0	NETSIM	600
	2	U				1200
	3	U				1800
	4	SNE			NETSIM	600
	5	SNE				800
	6	SNE				1200
2	7	U	2.2	0	Modified	600
	8	U				1200
	9	U				1800
	10	SNE			Modified	600
	11	SNE				1200
	12	SNE				1800
3	13	SNE	2.0	0	Modified	600
	14	SNE				1200
	15	SNE				1800
	16	SNE				600
	17	SNE	2.4			1200
	18	SNE				1800
4	19	SNE	1.9[a]	4	Modified	600
	20	SNE	2.5[b]			1200
	21	SNE	2.2[c]			1800

Note: U = uniform; SNE = shifted negative exponential.
[a]Downgrade. [b]Upgrade. [c]Cross street.

Figure 8. Effect of fuel model: uniform distribution.

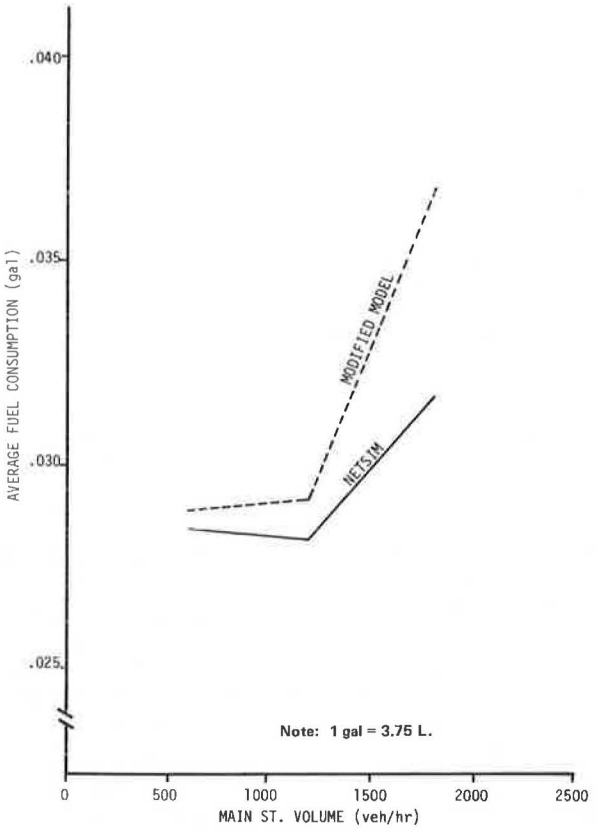
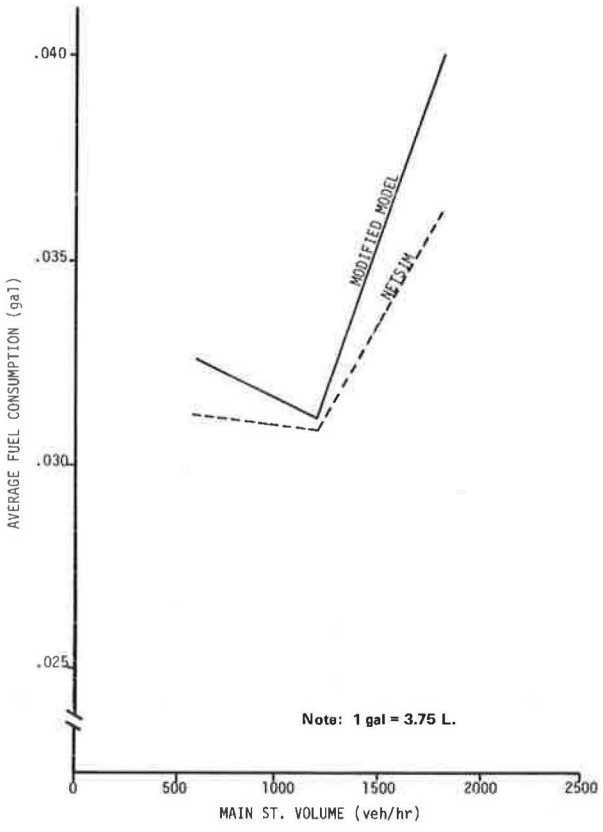


Figure 9. Effect of fuel model: shifted negative exponential distribution.



RESULTS

Isolated Intersections

Fuel Models

By using the uniform headway distribution, it was found that the modified fuel model produced significantly higher fuel-consumption values at the 95 percent level than the NETSIM model. The difference between the two models, shown in Figure 8, increases with increasing main-street volume (or degree of saturation). A similar comparison in which the shifted-negative-exponential headway distribution was used indicated that there was no significant difference in the two fuel models (see Figure 9), although the modified fuel model gave consistently higher consumption values.

Headway Models

The average delay values for the shifted-negative-exponential headway distribution were found to be significantly higher than those of the uniform headway model (see Figure 10). For the range of main street volumes considered, the shifted-negative-exponential distribution produced values that were, on the average, 52 percent higher than those of the uniform distribution. This result suggests using the shifted-negative-exponential headway distribution instead of the uniform distribution when using NETSIM to simulate traffic through an isolated intersection. This is suggested because a "bunching" of vehicles can be represented through the use of the shifted-negative-exponential headway distribution, whereas the uniform headway distribution can-

Figure 10. Effect of headway model on delay.

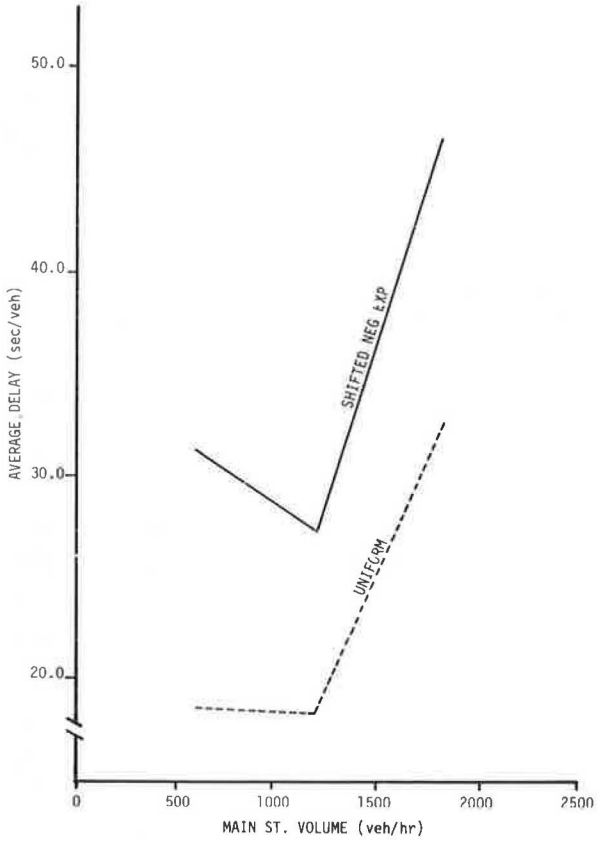


Figure 11. Overall comparison of fuel-consumption values.

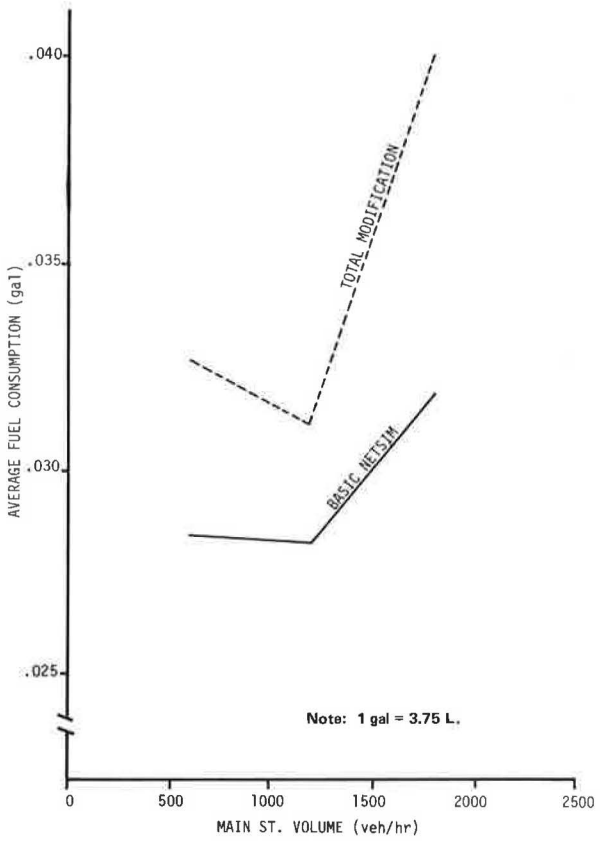
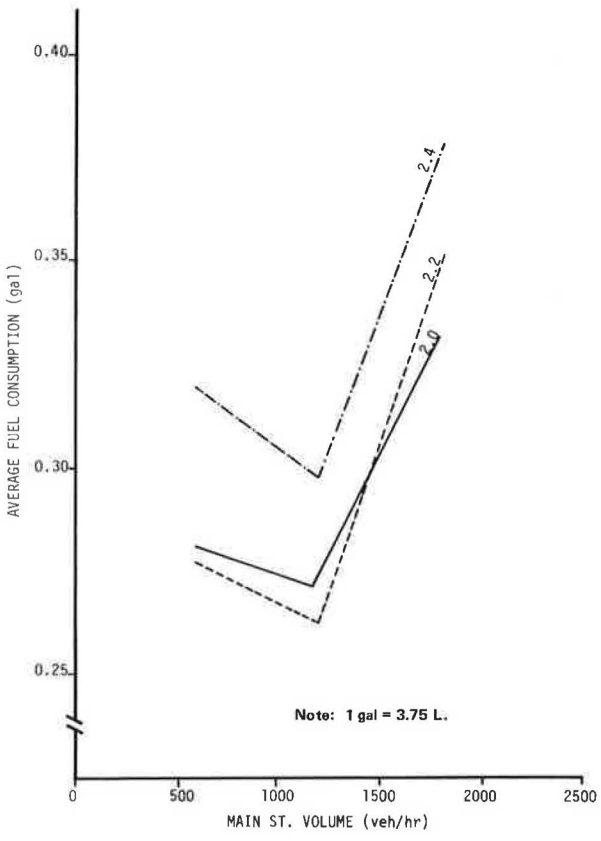


Figure 12. Effect of saturation headway on fuel consumption.



not reproduce this real-world traffic pattern.

In addition, fuel-consumption values were also found to be significantly different between the two headway distributions.

Overall Comparison

Figure 11 shows a comparison between the fuel-consumption values of the basic NETSIM model and a program modified to incorporate both the modified fuel model and the shifted-negative-exponential headway distribution. The results, which are significant, indicate that fuel-consumption values from the "totally modified" model average 17.4 percent higher than those of the basic NETSIM program. The delay differences would be the same as those shown in Figure 10.

Saturation Headway

The effects of saturation headway on fuel consumption and delay are shown in Figures 12 and 13, respectively. Increasing saturation headway from 2.0 to 2.2 s has no significant effect on either delay or fuel consumption. However, an increase in saturation headway from 2.2 to 2.4 s produces a significant effect on both delay and fuel consumption only for major-street volumes of 600 and 1200 vehicles/h.

Gradient Effects

The effects of grade on delay and fuel consumption on the main street were estimated by using changes in saturation-headway values (again, it should be noted that these changes do not include all of the effects of grade). As Figures 14 and 15 show, the

Figure 13. Effect of saturation headway on delay.

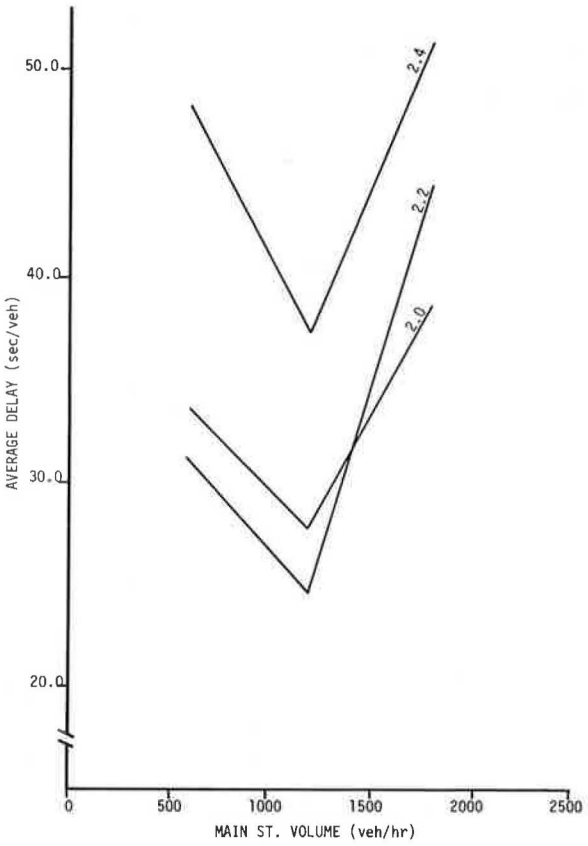


Figure 14. Effect of grade on fuel consumption.

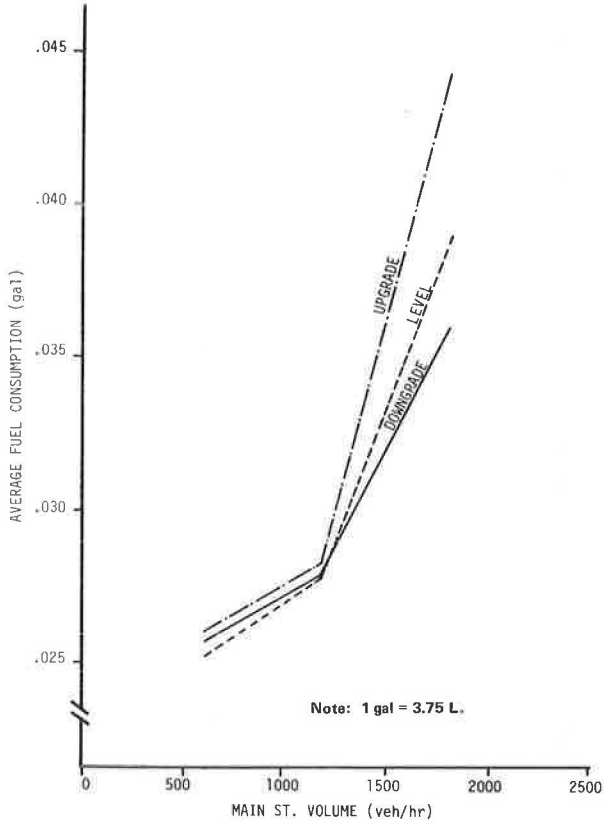
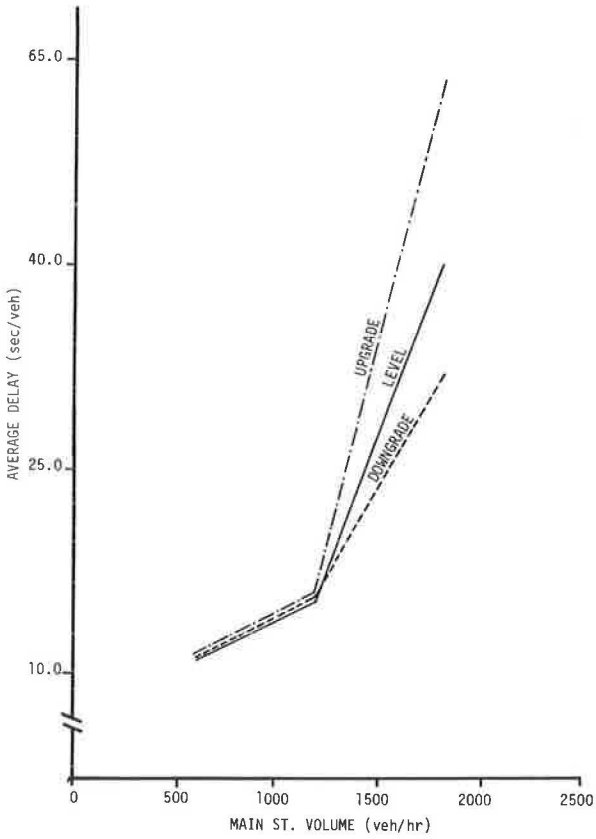


Figure 15. Effect of grade on delay.



results for both fuel consumption and delay showed no difference between zero and +4 or -4 percent grades at major-street volumes of 600 and 1200 vehicles/h. However, it was found that both delay and fuel-consumption values for the +4 and -4 percent grades were significantly different from those for zero percent grade for a major-street approach volume of 1800 vehicles/h. It appears that incorporating the effect of grade through changes in saturation headway can be useful at high volumes.

Prices Fork Road Network

The results of the Prices Fork Road analyses are summarized below (1 gal = 3.75 L):

	Fuel Consumption (gal)		
Headway Model	NETSIM	Modified Model	Delay (s/vehicle)
Uniform	155	169	53.4
Shifted negative exponential	192	199	92.5

The shifted-negative-exponential headway model shows an average delay per vehicle that is 73 percent greater than that for the uniform-headway-distribution model. Similarly, a change in headway models from the uniform to the shifted negative exponential causes increases in fuel-consumption estimates of 24 percent for the NETSIM fuel model and 18 percent for the modified model. A change from the basic NETSIM program to one that uses both the modified fuel model and the shifted-negative-exponential headway distribution results in a 28 percent increase in the estimate of fuel consumption.

CONCLUSIONS

The modified-fuel-consumption model developed in this study produced fuel-consumption values higher than those of the NETSIM model for both the uniform and shifted-negative-exponential headway distributions. NETSIM has proved to be an efficient tool in simulating traffic through large networks, especially at high volumes. The results of the headway-distribution model runs for the isolated intersection and the Prices Fork Road network suggest that the shifted-negative-exponential distribution should be used instead of the uniform distribution for simulating traffic at isolated intersections and small, open networks, especially those with low to moderate traffic volumes.

NETSIM results proved to be sensitive to saturation headways greater than 2.2 s. Reducing the saturation headway to less than 2.2 s does not significantly affect the results from the simulation model. Except for high-volume conditions, both delay and fuel consumption were found to be insensitive to the effect of grade on saturation flow rate.

The results of this research do not answer all significant questions concerning the universality of the conclusions drawn from the NETSIM simulations. For example, the results presented show that there is a significant difference between fuel models, although no statement is made as to which of the two is best. The fact is that both models are based on old vehicles and an entirely new model needs to be developed based on field tests with newer vehicles. Modeling of field-based truck fuel consumption should also be included in NETSIM, and sensitivity analyses should be performed with respect to traffic mix. The present inability of the program to consider the effect of grade on fuel consumption is of

concern. Such a capability needs to be developed and sensitivities analyzed. Other headway distributions such as the Erlang, Lognormal, Pearson, and Composite models should be investigated. Only when there is a thorough understanding of the significant variables and their impact on fuel consumption can any reliable conclusions be drawn concerning signal timing for the purpose of minimizing fuel consumption. It is felt that, of the available computer programs, NETSIM offers the greatest potential for determining these requirements.

REFERENCES

1. C.S. Bauer. Some Energy Considerations in Traffic Signal Timing. Traffic Engineering, Feb. 1975.
2. K.G. Courage and S.M. Parapar. Delay and Fuel Consumption of Traffic Signals. Traffic Engineering, Nov. 1975.
3. S.L. Cohen and G. Euler. Signal Cycle Length and Fuel Consumption and Emissions. TRB, Transportation Research Record 667, 1978, pp. 41-48.
4. J.W. Hurley and R.P. Ball. Energy Savings for Fixed Time Traffic Control. Modeling and Simulation, Vol. 9, Proc., 9th Annual Pittsburgh Conference, 1978.
5. J.W. Hurley and R.P. Ball. Evaluation of Energy-Based Signal Settings for Traffic-Actuated Control. Modeling and Simulation, Vol. 10, Proc., 10th Annual Pittsburgh Conference, 1979.
6. Network Flow Simulation for Urban Traffic Control Systems: Phase II. Federal Highway Administration, U.S. Department of Transportation, Vols. 1-5, 1977.
7. E.B. Lieberman and S.L. Cohen. New Technique for Evaluating Urban Traffic Energy Consumption and Emissions. TRB, Transportation Research Record 599, 1976, pp. 41-45.
8. P.J. Claffey. Running Costs of Motor Vehicles as Affected by Road Design and Traffic. NCHRP, Rept. 111, 1971.
9. J.W. Hurley. Feasibility of Transportation Projects: An Energy-Based Methodology. Univ. of Florida, Gainesville, Doctoral dissertation, March 1975.
10. G.E. Gonet. A Preliminary Investigation of the Fuel Consumption Characteristics of Freeway Ramp Metering. Virginia Polytechnic Institute and State Univ., Blacksburg, M.Sc. thesis, March 1979.
11. R. Conley. Effect of Grade on Starting Headway Time. Traffic Engineering, Vol. 22, No. 4, Jan. 1952.
12. A. Dick. Effect of Gradients on Saturation Flow at Traffic Signals. Traffic Engineering and Control, Vol. 5, No. 5, Sept. 1963.
13. The TRANSYT Signal Timing Reference Book. Federal Highway Administration, U.S. Department of Transportation, 1978.

Abridgment

Traffic Data Acquisition from Small-Format Photography

L.J. MOUNTAIN AND J.B. GARNER

A simple and economical method for collecting traffic data at complex urban intersections is described. The technique, developed at the University of Leeds, involves the collection of data in digital form by use of small-format photography taken from a hovering helicopter. The data are obtained in this form by means of a coordinate reader and processed by means of a computer that, using a two-dimensional coordinate transformation system, transforms the data to ground data and outputs information on a range of traffic-flow parameters. These parameters include approach volumes, the origins and destinations of all vehicles followed through the intersection, and the mean journey time and speed for each route. The potential of 16- and 35-mm photography to provide suitable photographic coverage is evaluated. The accuracy of the coordinate transformation systems is determined, and a computer-based coordinate-matching technique is developed. Finally, the accuracy and costs of obtaining traffic data by using the technique are compared with those associated with more conventional ground-survey methods. It was found that, where comprehensive traffic data are required, the technique can provide a simple, accurate, and economical method of traffic data collection and could be a workable alternative to conventional ground-survey methods.

While conventional ground-survey methods seem incapable of keeping abreast of the expanding data needs of today's traffic engineers, photographic techniques appear to have the potential to do so, particularly when combined with modern coordinate reader and computer technologies (1). The major problem associated with photographic techniques has, in the past, always been the difficulties associated with extracting and analyzing the vast quantities of information available from photographs (2), a problem that is particularly acute at intersections because of the complex nature of traffic movements at such locations. The basic objective of the research described in this paper was to find a solution to this problem.

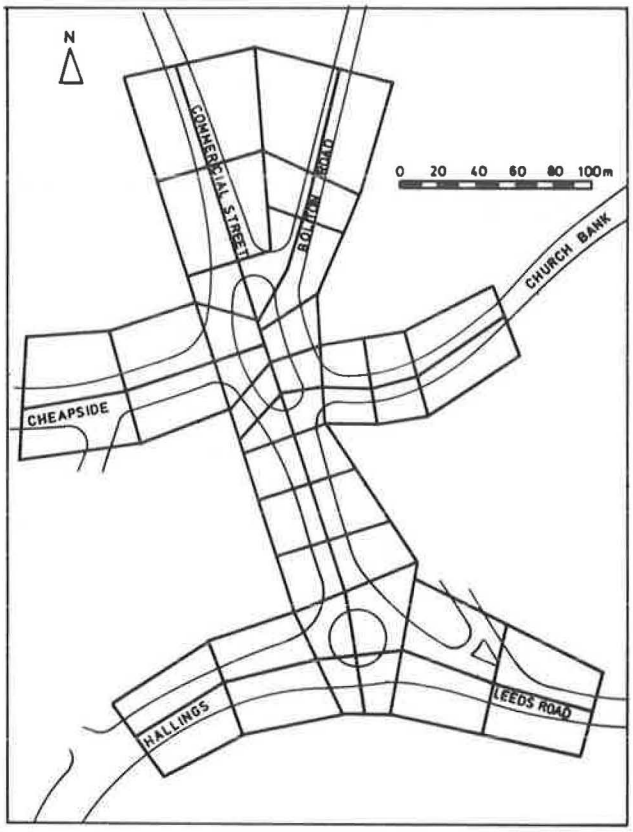
STUDY AREA

The site selected for study in this project was a complex double roundabout system near the center of Bradford in northern England. One portion of the system has five legs and the other has three, which gives a total of 36 possible routes through the study area. In order to include any queues on the approaches, it was decided that it would be necessary to include a minimum distance of 50 m on each approach to the roundabouts. Thus, the minimum area of coverage necessary was approximately 0.15x0.31 km.

DATA COLLECTION

To obtain comprehensive traffic data at such a location, it is essential that the photography used should be capable of providing continuous coverage of the entire intersection for a period of at least 1 h. Time-lapse photography is immediately suggested, but using any type of ground camera position would have restricted the field of view too much. A better vantage point could be obtained from the air,

Figure 1. System of zones used for vehicle tracking: Forster Square, Petergate, Bradford.



but there is then a problem in keeping the study area continuously in the field of view (3,4).

On balance, it was concluded that a time-lapse camera mounted in a hovering helicopter offered the best possibilities for providing the required photographic coverage. It was found that problems caused by helicopter vibration could be, if not totally eliminated, reduced to an acceptable level simply by the use of a fast shutter speed. Two cameras were used: the 16-mm Vinten Mark III and the 35-mm Robot Motor Recorder 24C. Color-reversal films were chosen for use in both cameras because small-format color photography is relatively inexpensive and the aid to vehicle tracking afforded by color justifies the additional costs over black-and-white emulsions (5).

In this study, filming was done at a rate of 1 frame/s. Analysis could then be carried out both by using every frame and by missing some frames so that a variety of time-lapse intervals could be simulated and compared.

DATA EXTRACTION

In the past, the reduction of traffic data has been done by various means that span the full range from manual to fully automatic (2). Although automatic methods will probably offer the ultimate solution to the problems encountered in data extraction, the need for further research and development of both hardware and software, together with a considerable reduction in the costs involved, renders semiautomatic methods the most feasible solution for the immediate future (6,7).

Semiautomatic methods are based on the identification of vehicle positions in terms of photographic coordinates, expressed relative to any arbitrary pair of rectangular axes. This is achieved by means of machines known as coordinate readers. These machines are semiautomatic in that a human operator is required to identify the points of interest and activate the machines, which then automatically output the rectangular coordinates of the points of interest on recording media such as paper tape. Semiautomatic techniques are simple to apply and less tedious than manual methods, and the equipment required is readily available and inexpensive in comparison with that required for fully automatic data reduction.

COORDINATE TRANSFORMATION

Photographic coordinates obtained by using a coordinate reader do not make any allowance for the distortions inherent in single photographs, and the axes and origin will not necessarily be the same from frame to frame. It is thus necessary to convert photographic coordinates to a common system of coordinates, the most convenient of which in Britain is the National Grid.

Conversion can be accomplished by using mathematical transformation techniques. There are two such techniques available: linear transformation (8) and projective transformation (9). Both techniques were thoroughly investigated. It was found that the photographic coordinates could be converted to National Grid coordinates quickly and conveniently by using either linear or projective transformation techniques and at a level of accuracy adequate for the majority of traffic engineering purposes. Although it was found that both methods were capable of similar levels of accuracy in the case of the study area used, projective transformation was preferred for its greater flexibility (10).

VEHICLE TRACKING

Having obtained suitable photographic coverage of the study area and having found a method of relating vehicle positions on the photographs to their actual positions on the ground, the next, and perhaps most important, stage in the project was to devise a technique that would enable the coordinates of individual vehicles to be matched from frame to frame. In order to do this, each set of coordinates had initially to be given an identifying code to relate the coordinates to the appropriate vehicle. The best way to do this is to base these codes on the observable characteristics of the vehicles. The photography obtained in this study provided sufficient detail to determine the color, type, and lane of travel of each vehicle. Numerical identification codes were devised based on these characteristics.

Vehicle identification codes obtained in this way are, of course, by no means unique, and it is necessary to devise a method by which the mismatching of the coordinates of apparently identical vehicles can be avoided. Attempts have been made to solve this problem in the past (8,11), but the techniques devised have in all cases been applicable to links on which the direction of vehicle travel is constant. A rather different approach is needed for the study of intersections where not only a vehicle's direction can change but also each vehicle has a choice among a number of alternative directions.

The approach adopted in this study was to use a system of zones. The Bradford study area was divided into zones (see Figure 1) so that a vehicle in any particular zone in one frame could only be in one, or perhaps two, other zones in the next frame. In order to follow vehicles, a computer program was

written so that, for a vehicle in a particular zone, a search was made for an identical vehicle in the next frame in the most likely zone first, then in the next most likely zone, and so on until a match was found.

In this way, the coordinates of each vehicle at discrete intervals of time were determined. A number of traffic-flow parameters were then determined, including the flow on each leg of the roundabout classified according to direction, the origins and destinations of all vehicles followed through the roundabout system, and the journey time and speed for each route.

RELATIVE ACCURACY OF THE DATA OBTAINED

One essential requirement of any useful new technique for traffic data collection is that it should be capable of providing levels of accuracy similar to those currently achieved by using conventional methods, at similar or reduced costs. In addition to comparing the accuracy of the extracted data and the costs associated with obtaining these data, by using both 16- and 35-mm photography, the efficiency of using 1-, 2-, 3-, and 4-s time-lapse intervals was also investigated.

The levels of accuracy attainable by using the photographic technique are given below:

Parameter	Error Limit for 95 Percent of Data (%)
Link volumes	±11.6
Origin-destination volumes	±17.0
Journey times	±10.9

The levels of accuracy were not found to vary significantly with the type of photography or the time-lapse interval used.

Although these levels of accuracy are lower than those that might be achieved by using conventional methods, they were within limits acceptable for the majority of engineering and planning purposes. The slightly reduced accuracy levels are, in fact, offset to some extent by the additional data that can be readily obtained from the photography. These include traffic density, vehicle paths, and gap acceptance, all of which could not be easily obtained by using conventional methods.

RELATIVE COSTS OF THE PHOTOGRAPHIC TECHNIQUE

The cost of a traffic survey can be evaluated in terms of two basic units, time and money, and it is desirable that the costs in terms of both should be minimized. Relative costs in terms of these two basic units are summarized below [1 £ = U.S. $2.32 (1980 average)]:

Method	Cost	
	Time (man-hours)	Money (£)
Photographic		
1-s time-lapse interval	1100	2700
4-s time-lapse interval	300	1000
Conventional	550	1700

Investigations showed that an overall reduction in the number of man-hours required for data collection and analysis could be achieved by using the photographic technique with a 2-, 3-, or 4-s time-lapse interval. The photographic technique, however, requires that the majority of man-hours be worked consecutively and thus, although the photographic technique requires considerably fewer personnel, the overall time period required to produce data in a useful form tends to be rather longer than that required for conventional surveys. If a 3- or 4-s time-lapse interval is used, however, an overall time period of less than two months should generally be adequate to obtain comprehensive peak-hour traffic-flow data at an intersection similar to the one studied in this project (assuming that it is not necessary to wait very long for suitable flying conditions). Although the overall time period is about twice as long as the overall time required to obtain similar data from a ground count, the time scale is not excessive and should be acceptable in the majority of cases.

Monetary costs do not vary significantly with either the 16- or 35-mm formats, and from this point of view either could be used equally well. The time-lapse interval used is, however, one of the major determinants of costs. If it were necessary for the user to pay full agency rates for computer processor time, the photographic technique would be rendered uneconomical irrespective of the time-lapse interval used. If it is necessary to pay only overhead costs for computing time, then the use of a 3- or 4-s time-lapse interval can render the photographic technique economical irrespective of the traffic density. The use of a 1-s time-lapse interval, on the other hand, would make the technique both too lengthy and too expensive.

CONCLUSIONS

Small-format, time-lapse photography taken from a hovering helicopter provides a relatively simple and economical technique for collecting comprehensive traffic data at complex intersections. Extraction of the data from the photography by using a coordinate reader to obtain the data in digital form, combined with computer analysis of the data, can form the basis for a simple and relatively rapid data-extraction technique. The technique can be used at any location to provide comprehensive traffic data, if suitable photographic coverage is available, and can provide data with a reasonable level of accuracy at costs similar to or lower than those encountered in using conventional ground-survey methods.

REFERENCES

1. D. Bayliss. Traffic Data for Transport Policies and Programmes. Proc., Conference on Traffic Data Collection, Univ. of Leeds, Leeds, England, Jan. 1979, pp. 1-25.
2. The Practicality of Aerial Traffic Data Collection. Institute of Traffic Engineers, Washington, DC, Informational Rept., 1975, pp. 5-24.
3. J.B. Garner and L.J. Mountain. The Potential of 16mm and 35mm Time Lapse Photography Taken from a Helicopter for Traffic Studies at Complex Intersections. Photogrammetric Record, Vol. 9, No. 52, 1978, pp. 523-535.
4. J.B. Garner and L.J. Mountain. Traffic Data Collection: An Alternative Method? Traffic Engineering and Control, Vol. 19, No. 10, 1978, pp. 451-454.
5. D. Berry, G. Ross, and R. Pfefer. A Study of Left-Hand Exit Ramps on Freeways. HRB, Highway Research Record 21, 1963, pp. 1-16.
6. L.J. Mountain. Traffic Data Collection. Traffic Engineering and Control, Vol. 29, No. 3, 1979, pp. 129-131.
7. C.M.G. Francis. The Microdensitometer as an Aid to the Interpretation of Multispectral Photographs for Mineral Exploration Purposes. Photogrammetric Record, Vol. 8, No. 45, 1975, pp. 309-316.
8. J. Uren and J.B. Garner. Integrated System for Urban Traffic Data Collection. TRB, Transpor-

tation Research Record 699, 1979, pp. 41-49.
9. J. Taylor and R. Carter. Photogrammetric Data Acquisition for a Freeway Ramp Operations Study. HRB, Highway Research Record 319, 1970, pp. 78-87.
10. L.J. Mountain. The Potential of Small-Format Photography for Traffic Data Collection. Univ. of Leeds, Leeds, England, Ph.D. thesis, 1979.
11. J. Uren. Urban Traffic Data Collection from Conventional Aerial Photography. Proc., Conference on Traffic Data Collection, Univ. of Leeds, Leeds, England, Jan. 1979, pp. 143-166.

Decentralized Control of Congested Street Networks

RUDOLF F. DRENICK, SAMIR A. AHMED, AND WILLIAM R. McSHANE

A mathematical model for traffic flow in city streets and its control is presented. The model is thought to be appropriate to the kind of control systems anticipated over the next 10-20 years. These systems are expected to rely to a considerable extent on communications that are less binding on driver behavior than the traffic signals that are now virtually the sole communication-control devices. The model is accordingly probabilistic. It is not limited to the problem of optimal signalization. On the contrary, optimal as well as satisfactory traffic control can be based on it. In either case, the problem develops into one of a special kind of nonlinear programming and of very large scale. A scheme is described for its decomposition into decentralized control, and several algorithms for its computational execution are outlined.

The most important technological development during the past decade may have been that of the very-large-scale-integration (VLSI) circuit chip. By all present indications, its impact will be widespread and profound. The control of traffic in city streets is likely to be affected. One can perhaps anticipate that current control systems will be supplemented and perhaps even supplanted by others that use the new technology. The new generation of such systems might seek to induce desirable traffic patterns by greater flexibility in its adjustment of red-green splits and in its use of turn signals at intersections. It might also combine the conventional signals with traffic advisories broadcast over general, citizens band, and perhaps even dedicated radio transmission channels, all in an effort to create a more satisfactory traffic-flow pattern.

The common characteristic of most of these strategies is that they are not binding on the driver in the way that conventional traffic signals now are. One can accordingly expect drivers' reactions to them to be of an even less deterministic nature than their reactions to the present ones. A model of the traffic system under these conditions seems most appropriate if formulated stochastically, and this is what has in fact been done in this study. The control variables of the system are, roughly speaking, the probabilities with which vehicles can be induced to make right turns, left turns, or no turns at the intersections of the street network in response to the various signals to which their drivers are exposed. The traffic signals, of course, remain as a set of control variables as well.

The problem of designing a traffic control system can then be viewed as that of making the best, or at least a satisfactory, choice of those probabilities. Formulated in this way, it develops into a constrained mathematical programming problem whose solution is made difficult partly by its nonlinearity and partly by its large scale. The nonlinearity is admittedly of a very special kind. The constraints, as well as the objective functions, are typically multilinear in the control variables. This is a feature that should be exploited in the solution procedure, and the several solutions that have been considered do so.

The large scale of the problem, on the other hand, suggests decentralized control schemes. Such schemes are at least intuitively most appropriate when the controlled system is made up of many subsystems geographically distributed over a wide area. It is then an appealing idea to exercise control over each subsystem separately, perhaps based mainly on inputs obtained locally, and to perform the necessary coordination through a hierarchy of supervisory controllers. Decentralized control schemes have the potential of reducing the cost of data communications, providing a high level of fail-safe capability, and allowing greater flexibility in the design and implementation of control strategies.

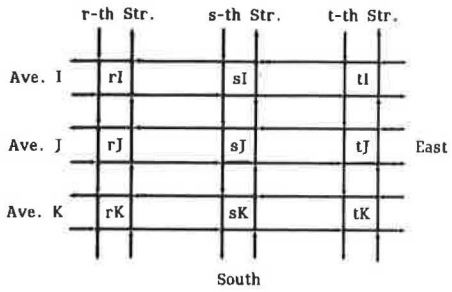
Surprisingly, however, little work has been reported on the application of decentralized control concepts to optimize the operations of large-scale urban transportation systems and even less work that treats these systems stochastically (1). The currently accepted versions of these concepts are explained in a recent article by Barry (2). Chu (3) explored the optimal decentralized control of a string of high-speed, densely packed vehicles using on-board controllers. Looze and others (4) and Kumar and others (5) proposed decentralized control schemes for regulating traffic on urban freeway corridors. Saridis and Lee (6) discussed the general problem of hierarchical control and management of traffic systems, while Chu (7), Sarachik (8), Singh and Titli (9), and Gershwin and others (10), among others, suggested decentralized control algorithms for large street networks, in nonstochastic formulations. A comprehensive survey of decentralized control methods and their applications to large-scale systems has recently appeared (11).

In this paper, a traffic network model based on probabilistic concepts is developed, and the problem of traffic control in such a network is formulated. The reasoning that leads to decentralization as a technique for the solution of that problem is outlined, and, finally, solution algorithms are briefly discussed, those that are already available as well as those that have been developed in this study.

TRAFFIC NETWORK MODEL

A traffic network is treated in this study as a stochastic system that is controlled by influencing the probability that a vehicle arriving at an intersection will make a right or left turn or no turn at all. The discussion presented here is based on a rather simple network of streets. Its extensions to

Figure 1. Segment of a street network.



more general and complex configurations will, however, be readily apparent.

Consider a network that is a rectangular grid of streets numbered 1,2,3, etc., and avenues labeled A,B,C, etc. Figure 1 shows the segment involving the streets numbered r, s, and t and avenues I, J, and K. During every red-green cycle, intersection sJ processes an input quadruple i of arriving platoons into an output quadruple k of departing ones, and it does so with certain "processing" probability p_{ik}^{sJ}. The indices i and k represent the sizes, compositions, and directions of the arriving and departing platoon quadruples at intersection sJ, respectively. A platoon is said to have departed from an intersection when it arrives at the next one. The processing probabilities at the different intersections are the control variables of the system. Signalization is present as a concern but is derived, rather than explicit, in this formulation.

Since every platoon quadruple i is processed into some quadruple k, one must have

$$\Sigma_k p_{ik}^{sJ} = 1 \qquad p_{ik}^{sJ} \geq 0 \tag{1}$$

where i and k range over all of their possible values. In addition, if p_i^{sJ} is the joint probability of arrival of quadruple i and q_k^{sJ} is the joint probability of departure of quadruple k, then

$$q_k^{sJ} = \Sigma_i p_{ik}^{sJ} p_i^{sJ} \tag{2}$$

This equation relates the input and output probabilities at the same intersection. A second relation prevails between inputs and outputs at different intersections. It expresses the fact that input quadruple i at a particular intersection consists of portions of the output quadruples released from the four intersections immediately connected to it. In Figure 1, the input quadruple arriving at intersection sJ is made up of the platoons coming from intersections sI, sK, rJ, and tJ. The joint arrival probability could thus be expressed as

$$p_i^{sJ} = (q_{k_1}^{sI})(q_{k_2}^{sK})(q_{k_3}^{rJ})(q_{k_4}^{tJ}) \tag{3}$$

where the four factors at the right-hand side represent the probabilities of having output platoon patterns from intersections sI, sK, rJ, and tJ that match the input quadruple pattern to intersection sJ.

The processing of quadruple i into k by intersection sJ presumably takes a certain known processing time t_{ik}^{sJ}, which depends on both the input i and the output k. One way of assessing the performance of intersection sJ, therefore, would be by its mean processing time:

$$\tau^{sJ} = \Sigma_{ik} t_{ik}^{sJ} p_{ik}^{sJ} p_i^{sJ} \tag{4}$$

Equations 1-4 written down for all intersections describe the state of traffic flow in the entire network. Control is exercised by adjusting the processing probabilities at the different intersections in such a way that a satisfactory flow pattern results. In this study, a pattern is considered satisfactory if it avoids "overload" throughout the network. The term overload can be given various interpretations; the simplest (though not obviously the most realistic) is that overload is avoided at intersection sJ if

$$\tau^{sJ} \leq \Delta^{sJ} \tag{5}$$

where Δ^{sJ} is the cycle length there. The reason for adapting this definition is that a violation of Equation 5 at any one intersection will eventually lead to unbounded waiting times and queue lengths there. These would entail the same phenomena at neighboring intersections and thus represent an expanding nucleus of congestion in the network.

The problem of achieving a satisfactory flow pattern, as formulated here, is seen to be a mathematical programming problem, though one without an obvious objective function. A set of processing probabilities p_{ik}^{sJ} that obeys the constraints of Equations 1-5 is to be determined, if it exists. In other words, any feasible solution to the problem is regarded to be satisfactory. It is customary in mathematical programming to prescribe a suitable objective function and to seek a solution that is optimal relative to it. In this case, one can take the same approach on the basis of the argument that overload at even only one intersection, as just explained, represents a highly undesirable traffic phenomenon and that it accordingly should be avoided if at all possible. This view suggests a rather natural choice of the objective function, namely

$$\tau = (1/N)\Sigma\tau^{sJ} \tag{6}$$

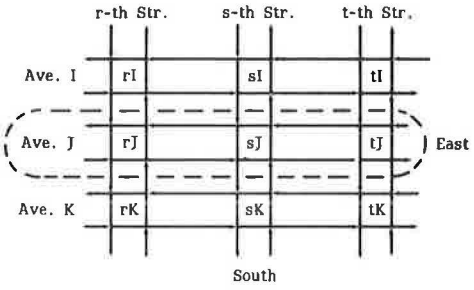
in which the sum extends over all intersections in the network (and the factor 1/N is there mainly for cosmetic reasons). The minimization of τ would of course be subject to the constraints of Equations 1-5. Additional constraints may be needed to impose origin-destination (O-D) specifications, queue-length limits, and other traffic restrictions. These are not mentioned here for simplicity of exposition.

If the Equation 6 choice of an objective function is considered inappropriate, others can be substituted for it, and several that have been considered by traffic engineers are natural candidates. The nature of the problem, and to some extent also that of its solution, are not greatly affected by such modifications.

The problem in any case is of very large scale in general, regardless of whether scale is measured in terms of the number of variables or the number of constraints. The number depends on what variables or constraints are counted. In terms of variables, if one considers only the processing probabilities p_{ik}^{sJ}, there will be as many as there are platoon quadruplets per intersection, squared (to allow for arrivals and departures), and multiplied by the number N of intersections in the network. In terms of constraints, if one considers only the probability and overload constraints (Equation 5 and Figure 1), there will be N(N + 1).

The problem is also nonlinear in its variables, as one readily recognizes. If Equation 6 is used as the objective function, the nonlinearities are of a special nature. The variables enter into the constraints and the objective function in sums of products, but none are raised to powers other than 0 or 1. They are, in other words, multilinear func-

Figure 2. Decentralization by avenue.



tions of those variables. The same is true of the problem if objective functions other than that of Equation 6 are used, provided they are expected values of cost-benefit parameters attached to the traffic flow.

CONTROL DECENTRALIZATION

The large scale of the programming problem described in the preceding section makes it highly desirable to seek a solution by way of a decomposition algorithm or, which is saying the same thing, to effect control by decentralization. Of the various schemes that have been considered so far in this study, one that is patterned on the hierarchical decomposition procedures developed by Mesarovic and his students (9, Chapter 4) currently appears the most promising. A brief description of these procedures and their use in the study is given here.

The idea of the hierarchical decomposition schemes is to resolve the Lagrangian L that corresponds to a large-scale programming problem into a number of terms

$$L = \Sigma_J L^J \tag{7}$$

each of which involves variables associated mainly with a single subsystem rather than with the system as a whole. There evidently is a good deal of latitude in the interpretation of the word "mainly" and, for that matter, the word "subsystem". In fact, success with the approach is often tantamount to a judicious exploitation of that latitude.

The approach that has been used in this study so far is the following. It has been assumed that the traffic in the network is characterized by a "dominant direction of flow," e.g., from north to south in Figure 1. One can then assign the control of traffic along and across each avenue to a subordinate controller, as indicated in Figure 2. The coordination of the signaling among avenues can be assigned to one supervisory controller or to several.

To do so, one collects in each of the terms L^J all those making up the Lagrangian L that can be associated with the traffic on and across avenue J. With the objective function of Equation 6, this makes

$$L^J = (1/N)\Sigma_s \tau^{sJ} + \Sigma_s \lambda^{sJ}(\tau^{sJ} - \Delta) \tag{8}$$

disregarding any additional terms attributable to O-D constraints and others. The mean processing times τ^{sJ} in Equation 8 are explicitly multilinear functions of the processing probabilities $p_{ik}{}^{sJ}$ of intersection sJ but implicitly also of those of other intersections. Strictly speaking, τ^{sJ} would be a function of the processing probabilities of all other intersections but, with the dominant direction of flow, only of those lying north of avenue J.

One can now seek an optimal set of processing probabilities in the usual way--namely, among those that satisfy the following equations:

$$\partial L/\partial p_{ik}{}^{sA} = \partial L^A/\partial p_{ik}{}^{sA} + \partial L^B/\partial p_{ik}{}^{sA} + \ldots + \partial L^Z/\partial p_{ik}{}^{sA} = 0 \tag{9}$$

$$\partial L/\partial p_{ik}{}^{sB} = \partial L^B/\partial p_{ik}{}^{sB} + \ldots + \partial L^Z/\partial p_{ik}{}^{sB} = 0 \tag{10}$$

...

$$\partial L/\partial p_{ik}{}^{sZ} = \partial L^Z/p_{ik}{}^{Z} = 0 \tag{11}$$

and so on, along with the overload and other constraints. The form of these equations is due to the assumption of a dominant direction of flow that implies that L^A depends only on the $p_{ik}{}^{sA}$, L^B only on $p_{ik}{}^{sA}$ and $p_{ik}{}^{sB}$, etc.

The form of Equations 9-11 further suggests that the controller for avenue J be assigned the solution of the optimization problem represented by the equation

$$\partial L^J/\partial p_{ik}{}^{sJ} + \alpha_{ik}{}^{sJ} = 0 \quad (J = A, B, \ldots, Z) \tag{12}$$

along with overload and other constraints pertinent to avenue J. This optimization would be for the processing probabilities $p_{ik}{}^{sJ}$ of the intersections along that avenue. The quantity

$$\alpha_{ik}{}^{sJ} = \partial L^K/\partial p_{ik}{}^{sK} + \ldots + \partial L^Z/\partial p_{ik}{}^{sZ} \tag{13}$$

would be supplied to the avenue J controller by a superior to be treated as an additive constant during the optimization. It would be updated by the superior and resupplied to the subordinates for a new optimization.

The actual computational procedure, as in most nonlinear programming problems, would not, however, seek a direct simultaneous solution of Equations 12 and 13 but would use a search algorithm selected from among the several candidates briefly discussed in the next section of this paper.

SOLUTION TECHNIQUES

It has been pointed out that the problem of devising satisfactory or optimal traffic control, as formulated in this study, is a constrained nonlinear programming problem. It is, in fact, of a special kind that has been referred to here as "multilinear". This is true regardless of whether or not the control scheme is decentralized.

Several existing solution techniques suggest themselves. Among them are, first of all, a number of well-tested algorithms for the solution of constrained nonlinear programming in general (12,13). Their very generality, however, tends to be a disadvantage in that their convergence may be slow and uncertain in practice. Another group of techniques goes under the term of geometric programming (14,15). They apply to a more restricted class of problems but one that includes many of the multilinear ones. Unfortunately, they do not readily accommodate equality constraints of the kind that are inherent in the traffic-control problem formulation discussed here and thus are not particularly appropriate.

An effort was accordingly made in this study to investigate solution algorithms that are tailored to multilinear problems. One such algorithm exploits the fact that the multilinear programming problem is a natural generalization of the linear ones and hence is a rather direct analog to the well-known simplex algorithm (16).

As of this writing, however, a number of specialized techniques patterned on the gradient method are being favored because they seem most amenable to the

kind of extensions in the traffic-control problem that are expected to become necessary in the near future.

DISCUSSION OF FORMULATION

This paper has described the present status of a study that seeks to formulate the problem of traffic control in city streets, with a view to the way in which the control might be executed in another decade or two. It may be of interest to add some brief remarks on the thinking that led to the current formulation, the features of it that are now thought to be undesirable, and the developments in it that are anticipated for the near future.

The idea of the formulation arose from a recent effort at developing a mathematical approach to organization theory (17). The parallel between the control of traffic in city streets and organizations may seem rather remote, but there are in fact a number of important analogies. For one, both can be designed with the aim of avoiding overload among junctions--i.e., among the intersections of city streets and among the members of an organization. Moreover, in both the control variables develop into probabilities of the kind that have here been called processing probabilities. There are, however, substantial distinctions as well. Most important may be the fact that the flow of traffic in a street network is a much more involved phenomenon than the flow of information and material in a well-functioning organization. In fact, it may be safe to say that an organization that had as disorderly a flow pattern as city traffic would be virtually unmanageable.

The main shortcoming of the current formulation of the control problem is felt to be its nondynamic character. It is a tacit assumption in a solution by mathematical programming that, once obtained, the solution will also be promptly adopted. In traffic control, however, and especially in the kind of control scheme envisioned here, this is unlikely to be so. The control system will thus have to monitor its own success with the traffic pattern by means of suitably placed sensors and adjust its control signals accordingly. The dynamics of the resulting feedback loops will have to be combined with the driver characteristics in order to achieve satisfactory operation. At this time, the gradient-like methods mentioned in this paper seem the most amenable of those considered or developed so far.

REFERENCES

1. Traffic Control Systems Handbook. Federal Highway Administration, U.S. Department of Transportation, 1976.
2. A.J. Barry. Distributed Intelligence in Traffic Control Systems. Proc., Engineering Foundation Conference on Research Directions in Computer Control of Urban Traffic Systems, Asilomar, CA, ASCE, New York, 1979, pp. 58-69.
3. K.C. Chu. Decentralized Control of High-Speed Vehicular Strings. Transportation Science, Vol. 8, No. 4, 1974, pp. 361-384.
4. D.P. Looze and others. On Decentralized Estimation and Control with Application to Freeway Ramp Metering. IEEE Trans. on Automatic Control, Vol. AC-23, No. 2, 1978, pp. 268-275.
5. K.S. Kumar and others. Reliable Decentralized Control Strategies for Freeway Regulation. U.S. Department of Transportation, April 1980.
6. G.N. Saridis and C.S. Lee. On Hierarchically Intelligent Control and Management of Traffic Systems. Proc., Engineering Foundation Conference on Research Directions in Computer Control of Urban Traffic Systems, Asilomar, CA, ASCE, New York, 1979, pp. 209-218.
7. K.C. Chu. Decentralized Real-Time Control of Congested Traffic Networks. International Business Machines Corp., Armonk, NY, Res. Rept. RC-6337, Dec. 1976.
8. P.E. Sarachik. Clearing of Congested Multi-Destination Networks. Proc., Engineering Foundation Conference on Research Directions in Computer Control of Urban Traffic Systems, Asilomar, CA, ASCE, New York, 1979, pp. 231-250.
9. M.G. Singh and A. Titli. Systems Decomposition, Optimization, and Control. Pergamon Press, Oxford, England, 1978.
10. S.B. Gershwin and others. Hybrid Optimization in Urban Traffic Networks. U.S. Department of Transportation, Rept. DOT-TSC-RSPA-79-7, 1979.
11. N.R. Sandell and others. Survey of Decentralized Control Methods for Large-Scale Systems. IEEE Trans. on Automatic Control, Vol. AC-23, No. 2, 1978, pp. 108-128.
12. D.G. Luenberger. Introduction to Linear and Nonlinear Programming. Addison-Wesley, Reading, MA, 1965.
13. E. Polak. Computational Methods in Optimization: A Unified Approach. Academic Press, New York, 1971.
14. R.J. Duffin, E.L. Peterson, and C. Zener. Geometric Programming. Wiley, New York, 1967.
15. C.S. Beightler and D.T. Philips. Applied Geometric Programming. Wiley, New York, 1976.
16. R.F. Drenick. Decentralized Control of Street Traffic. U.S. Department of Transportation, 1980.
17. R.F. Drenick. Organization and Control. In Directions in Large-Scale Systems (Y.C. Ho and S.K. Mitter, eds.), Plenum Press, New York, 1976, pp. 279-302.

Improved Estimation of Traffic Flow for Real-Time Control

YORGOS J. STEPHANEDES, PANOS G. MICHALOPOULOS, AND ROGER A. PLUM

A critical review of the most widely accepted demand prediction algorithms is presented. Based on data collected at four intersections, sensitivity analysis of the best existing algorithms indicates that very little improvement in their performance could be achieved. A new, simpler algorithm, which requires considerably less information and fewer computations, is subsequently proposed and compared with the best of the existing algorithms. The results suggest that for 5-min prediction the second-generation Urban Traffic Control System predictor (UTCS-2) is usually better. However, in cycle-by-cycle prediction the proposed algorithm is considerably (as much as 41 percent) better than the best of the existing algorithms.

The problems associated with computerized signal control are numerous, ranging from demand prediction algorithms to reliability analysis, detector placement, and safelock design. In recent publications (1-3), traffic models and signal control strategies have been developed.

The major objective of this study is to determine the most reliable prediction algorithm suitable for implementing a recently developed (3) real-time control policy for critical intersections. This determination depends on two basic criteria: (a) algorithm performance and (b) the ability of the selected algorithm to estimate average arrival flow rates on a cycle-by-cycle basis. The second criterion is required for the implementation of a policy such as that mentioned above.

A critical review of the most widely accepted demand prediction algorithms is performed first. This review includes a summary of performance characteristics in which emphasis is placed on the effectiveness and drawbacks of each algorithm from the limited tests found in the literature. Potential improvements to the best existing algorithms, suggested in the literature, are discussed, and sensitivity tests are performed that indicate the extent of improvement in algorithm performance that could be expected to result from such changes.

Subsequently, a new demand prediction algorithm is proposed and compared with the second- and third-generation Urban Traffic Control System (UTCS) predictors (4,5), which were found to be the best (for the purposes of this study) among the existing algorithms. The current-measurement and historical-average predictors are also included in the comparisons. The comparison tests are based on 10 data sets collected at four intersections over a three-month period. These tests are more extensive and detailed than previous ones in that they include both isolated and coordinated intersections controlled by pretimed or actuated signals.

The major findings can be summarized in three parts:

1. The test results of a comparison of the performance of UTCS-2, UTCS-3, historical average, and current measurement are in agreement with previous studies (6,7). More specifically, the 5-min predictions of both UTCS-2 and UTCS-3 track the trend of the actual values of the volume measurements, and both improve the prediction in comparison with using the current measurement as the predicted value. However, in both cases the predicted values time-lag the actual measurements. The test results also show that UTCS-2 performs consistently better than UTCS-3. However, UTCS-2 provides less information or very little additional information over historical averages.
2. The sensitivity analysis performed here on UTCS-2 and UTCS-3 parameters indicates that not much improvement in expected UTCS performance could be achieved by varying the parameters away from the values recommended in the literature (6,8). These observations reinforce the need for the development of simpler and more accurate demand prediction algorithms.
3. An algorithm is proposed that, in its simplest form, degenerates to a moving average. The tests show that, when few data are available, the moving average is the most accurate method. When more detailed data are available, the complete algorithm performs at least as well as or better than a moving average.

BACKGROUND

Review of Demand Prediction Algorithms

The existing demand prediction algorithms fall into three general categories: (a) the second generation, (b) the third generation, and (c) algorithms developed after the third generation. The second generation is designed for control intervals on the order of 5-15 min, and the third generation is designed on a cycle-by-cycle basis.

Second-generation algorithms are older and typically require extensive historical data as reference. They use current traffic measurements to correct for the traffic deviations from the average historical pattern. Second-generation UTCS (UTCS-2) (4), ASCOT (9), and ASCOT-RTOP (10) all belong to this category.

Third-generation algorithms (5,7), generally more recent than the second generation, were developed with the objective of making predictions based on current traffic measurements only. However, the third-generation UTCS (5), the best-known algorithm in this category, requires a "representative" data set for estimating prediction coefficients. This assumption is in conflict with the idea of "highly responsive control software" (i.e., the third-generation control software) for which the predictor was designed (7).

The Baras-Levine algorithms (11-13), which constitute the most recent approach to demand prediction, fall in the third category. These algorithms are based on the hypothesis that, in contrast with previous assumptions, the data from traffic sensors represent a point process that is not Poisson. They therefore use point-process techniques to develop improved filter-predictors for use in traffic-responsive (nearly real-time) computer control of urban traffic. Their algorithm, F/P I (13), is aimed primarily at critical intersection control and is based on a time-varying Markov chain model that represents a linearization and discretization of nonlinear traffic dynamics. F/P I was found to be more accurate and more informative than ASCOT by its authors (13). It needs, however, more computation time. In addition, unlike algorithms in the previous two categories, the Baras-Levine algorithm

Table 1. Algorithms tested and evaluated.

Algorithm	Description
UTCS-2	$\hat{v}_t = m_t + \gamma(m_{t-1} - f_{t-1}) + (1-\alpha)\sum_{s=0}^{t-1}\alpha^s(f_{t-s-1} - m_{t-s-1}) + \gamma(1-\alpha)\sum_{s=0}^{t-2}\alpha^s(f_{t-s-2} - m_{t-s-2})$
UTCS-3	$\hat{v}_t = \gamma_1 f_t + (1-\gamma_1)\left[\hat{\mu}_0\alpha^t + (1-\alpha)\sum_{s=0}^{t-1}\alpha^s f_{t-s-1}\right]$
Historical average	$\hat{v}_t = m_t$
Current measurement	$\hat{v}_t = f_t$
Proposed	$\hat{v}_t = a_1 f_t + a_2(f_t - f_{t-1}) + a_3\left(\sum_{k=1}^{N} f_{t-k}/N\right) + a_0$

predicts queue size at an intersection rather than demand.

Both second-generation and third-generation demand prediction algorithms are the results of extensive research. However, the elaborate formulas they offer leave much to be desired. Other researchers have postulated a number of factors that, in their opinion, have apparently hampered the success of demand prediction algorithms. One factor brought forward by Kreer (8) is that the vehicles that are measured should be the same ones that are affected by the resulting change in control action; the only types of control that might satisfy this requirement (but might also result in increased computation costs) are critical intersection control and vehicle-actuated modes of control (8). To be sure, a major reason for the apparent failure in solving the traffic-control problem is that no predictor built to date is adaptive to the changing, underlying traffic-flow process. At coordinated networks, treating demand prediction as an open-loop process is another reason, intimately related to the first one, for this failure. In recent research by Mengert, Brown, and Yuan (7), it was proposed to use the Trigg and Leach method (14), a smoothing algorithm, and another technique that they developed to make UTCS-2 and UTCS-3 adaptive. Significantly, preliminary tests did not reveal substantial improvements. Box-Jenkins-type analyses and other estimation techniques such as Kalman filtering may, however, promise future improvements (7).

Of the five algorithms presented above, only UTCS-2 and UTCS-3 have been chosen for further analysis and testing. The Baras-Levine algorithm (13) could not be included, since it cannot be used for demand prediction. Neither version of the ASCOT algorithm (1,10) has been chosen for three basic reasons:

1. ASCOT is a second-generation technique and as such cannot respond to traffic on a cycle-by-cycle basis; testing two similar techniques (i.e., UTCS-2 and ASCOT) could not be justified unless one were significantly different from the other.
2. ASCOT has data requirements that are significantly greater than those of any of the algorithms reviewed.
3. ASCOT requires extensive instrumentation that is not available in most real systems.

Two more algorithms have been included in comparison tests and performance evaluation. The historical average, one of the two algorithms, assumes that the volume during any specified time period equals the smoothed historical volume for that period as obtained from earlier observations. The second algorithm, the current measurement used as the predicted value, assumes that the volume during any given time period (in this case, 5 min or one cycle) is the same as that during the previous time period. As a result of this assumption, prediction inherently lags behind observation by at least one time period. This method has the simplest data requirements.

Finally, the proposed algorithm assumes prediction to be a linear function of the current volume, the difference between the current and previous volume, and an average volume during the previous three to five time periods. All algorithms tested and evaluated in this work are presented in Table 1 and explained in the following sections.

UTCS-2

The second-generation UTCS, UTCS-2, predicts the next-control-interval (on the order of 5-15 min) traffic volume at each detector location in real time based on the measurements from the same location only. The algorithm makes use of both smoothed historical traffic data and current traffic-volume measurements from the vehicle detector.

The UTCS-2 set of equations has been presented elsewhere (4). The complexity of the solution, however, which is of major interest to this study, is usually not shown. By solving the difference equations of UTCS-2 (4), it can be shown that UTCS-2 results in the following demand prediction equation:

$$\hat{v}_t = m_t + \gamma(m_{t-1} - \gamma f_{t-1}) + (1-\alpha)\sum_{s=0}^{t-1}\alpha^s(f_{t-s-1} - m_{t-s-1}) + \gamma(1-\alpha)\sum_{s=0}^{t-2}\alpha^s(f_{t-s-2} - m_{t-s-2}) \quad (1)$$

in which

$$m_t = a_0 + \sum_{i-1}^{k}[a_i\cos(2\text{nit}/N) + b_i\sin(2\text{nit}/N)] \quad (2)$$

and

$$\gamma = \left\{(n-1)\sum_{s=1}^{n-1}\left[f_s - m_s - (1-\alpha)\sum_{p=0}^{s-1}\alpha^p(f_{s-p-1} - m_{s-p-1})\right] \times \left[f_{s-1} - m_{s-1} - (1-\alpha)\sum_{p=0}^{s-2}\alpha^p(f_{s-p-2} - m_{s-p-2})\right]\right\} \div (n-2)\sum_{i=1}^{n}\left[f_s - m_s - (1-\alpha)\sum_{p=0}^{s-1}\alpha^p(f_{s-p-1} - m_{s-p-1})\right]^2 \quad (3)$$

where

$\hat{v}_t$ = predicted volume at time t;
m_t = historical volume at time t;
f_t = measured volume at time t;
d_t = empirical adjustment at time t;
α = constant computed off-line from representative volume data of the location in question (e.g., for the UTCS system in Washington, D.C., α was 0.2);
γ = smoothing coefficient (e.g., for the UTCS system in Washington, D.C., γ was 0.9);
a_0, a_i, b_i = coefficients (computed off-line) of Fourier series approximation of historical traffic patterns for each measurement location;
k = user input parameter determining the fidelity of Fourier series approximation, usually the result of a trade-off between Fourier series accuracy and storage space and computation effort (in general, for more rapidly varying functions, higher values of

k should be used; k-values from 6 to 20 have been used in past applications);
n = number of sample points of the representative data set; and
N = total number of time intervals in the representative data set (e.g., for 15-min intervals, the data for a 24-h day will consist of 96 intervals).

It can be seen that the UTCS-2 prediction equation (1) is a function of

$$\hat{v} = \hat{v}[m(t), f(t), n, \alpha] \tag{4}$$

where

$$m = m(a_0, a_i, b_i, k, N, t; i = 1 \text{ to } k) \tag{5}$$

UTCS-3

The predictor for the third-generation UTCS software, UTCS-3, predicts traffic volume two control intervals into the future. Like UTCS-2, UTCS-3 forecasts the volume at each location in real time based on measurements from the same location. However, it is different from UTCS-2 in that the prediction process relies solely on current-day measurements (no historical traffic pattern is required for prediction). By solving the difference equations of UTCS-3 (5), it can be shown that UTCS-3 results in the following demand prediction equation:

$$\hat{v}_{t+j} = \gamma_j f_t + (1 - \gamma_j)\left[\hat{\mu}_0 \alpha^t + (1 - \alpha) \sum_{s=0}^{t-1} \alpha^s f_{t-s-1}\right] \tag{6}$$

in which

$$\gamma_j = \left\{(n - 1) \sum_{s=1}^{n-j} \left[f_s - \hat{\mu}_0 \alpha^s - (1 - \alpha) \sum_{p=0}^{s-1} \alpha^p f_{s-p-1}\right] \times \left[f_{s+j} - \hat{\mu}_0 \alpha^{s+j} - (1 - \alpha) \sum_{p=0}^{s+j-1} \alpha^p f_{s+j-p-1}\right]\right\} \div (n - 1 - j) \sum_{s=1}^{n} \left[f_s - \hat{\mu}_0 \alpha^s - (1 - \alpha) \sum_{p=0}^{s-1} \alpha^p f_{s-p-1}\right] \tag{7}$$

where

$\hat{v}_{t+j}$ = predicted volume for time (t + j) at time t;
γ_j = extrapolation constant computed off-line from representative volume data of the location in question;
f_t = measured volume at time t;
μ_t = exponentially smoothed volume measurement, also referred to as "coarse prediction of volume"; and
α = smoothing coefficient [a value of 0.95 has been used in past applications (6)].

It can be seen that the UTCS-3 prediction equation (5) is a function of

$$v = \hat{v}[\hat{\mu}_0, \hat{i}(t), n, \alpha] \tag{8}$$

Comparative Evaluation and Drawbacks of UTCS-2 and UTCS-3

Both UTCS predictors are based on single-location traffic measurements. Both use the linear combination of residues (differences between traffic measurements and either historical data or smoothed traffic data) as the basic feature for prediction. The second-generation predictor requires historical data as the reference. The third-generation predictor does not require historical data, makes predictions based on current traffic measurements (7), and can be applied to undersaturated links only (5).

The drawbacks of UTCS-2 (7) led to the development of UTCS-3. Major UTCS-2 drawbacks are related to its high reliance on historical data: Traffic volume can vary substantially, depending on various external (with respect to algorithm) factors (e.g., weather conditions, special events, developments in other modes of transportation, and even the traffic-control change itself). UTCS-2 is not responsive to such changes. Because of this reliance on historical data, UTCS-2 is not readily transferable across systems and therefore is not practical. A large data base is required for the historical data. This data base consumes computer storage space and must be updated periodically off-line. Furthermore, an analysis conducted early in the UTCS project in which "simulated" traffic data were used indicated that historical data were not always necessary to achieve good prediction.

As can be seen from a comparison of Equations 1-3 and Equations 6-8, UTCS-2 prediction is indeed a function of two time-dependent functions (more accurately, a function of the difference of two time-dependent functions): measured volume f(t) and historical volume m(t). UTCS-3 prediction, on the other hand, is a function of f(t) only. The constants to be predetermined are n, the number of sample points of the representative data set, and α, the smoothing coefficient; these constants are needed irrespective of the UTCS predictor specification chosen.

A small number of past performance tests comparing UTCS-2 and UTCS-3 (6,8) have indicated that UTCS-3 is not capable of achieving a performance as high as that of UTCS-2, which was consistently better--i.e., had both a lower mean square and a lower mean absolute value error. In addition, UTCS-2 had a larger portion of small-magnitude errors than UTCS-3 (6). Time-lagging is a serious drawback of both algorithms, but it is especially obvious with UTCS-3, where it is inherently two time intervals long and cannot be compensated for. One result of this inherent time lag is that, when there is a detector outage, UTCS-2 provides reasonably good values during the outage and is available as soon as vehicle detector operation is restored whereas UTCS-3 will not provide predicted volumes until two time intervals after the vehicle detector is restored.

RESEARCH APPROACH

For testing the demand prediction algorithms, 10 data sets were collected at six locations in the Minneapolis-St. Paul metropolitan area during October, November, and December 1979. The details are summarized in Table 2, which shows that the selected locations include both coordinated and isolated intersections under pretimed or actuated control.

Two error measurements were computed for each data set and algorithm: (a) mean square error (MSE), which penalizes large prediction errors, and (b) mean absolute error (MAE), which indicates the expected typical error for an individual prediction. These error measurements, which have been established in the literature for comparing prediction performance (6,7), are defined as follows:

$$MSE = [\Sigma(\text{measured volume} - \text{predicted volume})^2]/N \tag{9}$$

$$MAE = [\Sigma|\text{measured volume} - \text{predicted volume}|]/N \tag{10}$$

where N is the total number of predictions.

In using current UTCS algorithms, all of the

Table 2. Summary of data-set characteristics.

Location	Data Set	Approach Classification	Control Policy	Date and Duration
Oak Street, S.E., and Delaware Avenue, northbound	1	Coordinated	Semiactuated	5-min intervals, 3:00-6:30 p.m., 15 days in Oct.-Nov. 1979
Oak Street, S.E., and Washington Avenue, S.E., westbound	2	Isolated	Pretimed	5-min intervals, 3:00-6:30 p.m., 15 days in Oct.-Nov. 1979
Oak Street, S.E., and Washington Avenue, S.E., eastbound	3	Coordinated	Pretimed	5-min intervals, 3:00-6:30 p.m., 15 days in Oct.-Nov. 1979
Oak Street, S.E., and Washington Avenue, S.E., southbound	4	Coordinated	Pretimed	5-min intervals, 3:00-6:30 p.m., 15 days in Oct.-Nov. 1979
Union Street, S.E., and Washington Avenue, S.E., westbound	5	Coordinated	Pretimed	Nov. 13, 2:45-3:50 p.m., 40 cycles
	6	Coordinated	Pretimed	Nov. 14, 8:30-9:35 a.m., 39 cycles
	7	Coordinated	Pretimed	Nov. 14, 3:55-5:00 p.m., 33 cycles
	8	Coordinated	Pretimed	Nov. 15, 7:30-8:35 a.m., 38 cycles
Fifth and Excelsior, Hopkins	9	Isolated	Pretimed	Dec. 3, 3:50-4:55 p.m., 45 cycles
	10	Isolated	Pretimed	Dec. 4, 3:40-4:55 p.m., 45 cycles

necessary constants were obtained from the literature (6,7) except the smoothing constant, α, in which case the steady-state α value was found and used. Historical averages were formed by using data from seven days. The Statistical Package for the Social Sciences regression package (15) was used to determine constants where needed by the formulas.

All tests for which results are presented later in this paper are valid for comparison among prediction algorithms. Since the data used for testing were collected at a number of locations, nothing can be concluded regarding the relation between control policy and prediction accuracy.

PROPOSED ALGORITHM

Because of the disadvantages of the existing algorithms already described and some additional problems discussed later in this paper, it was decided that a new algorithm should be developed. In addition to meeting the objectives set forth in the introduction, the new algorithm should avoid arbitrary methods of treating the data and should instead be based on theory in as straightforward a manner as possible. It should be flexible enough to become part of closed-loop traffic control once the demand prediction was incorporated into the traffic process. It should also be simpler and at least as accurate as UTCS-2 or, in cycle-by-cycle prediction (of most interest to the project objectives), UTCS-3.

The proposed algorithm uses the volume during the next time period as the predicted variable; the current volume, the difference between current volume and previous volume, and the average volume during the previous three, four, or five time periods are the independent variables. The algorithm can be derived by using data from one or more previous days and used for demand prediction on the following day. On-line derivation is also possible.

The prediction equation is

$$\hat{v}_t = a_0 + a_1 f_t + a_2 (f_t - f_{t-1}) + a_3 \left(\sum_{k=1}^{N} f_{t-k} \Big/ N \right) \quad (11)$$

where N is the number of time periods considered (N = 3, 4, and 5 are suggested values) and a_1, a_2, and a_3 are control coefficients that can be found by using standard regressive techniques to best fit the measured data for the location in question.

From the above it can be seen that the proposed algorithm is a function of

$$\hat{v} = \hat{v}\,[f(t), a_i, N; i = 0 \text{ to } 3] \quad (12)$$

It can also be seen that the algorithm is simpler in form than either UTCS-2 (Equations 1 and 2) or UTCS-3 (Equations 6 and 7). It should also be noted that the prediction equation (Equation 11) is of the general form

$$\hat{v}(t) = a_0 + a_1 f(t) + a_2 [df(t)/dt] + a_3 \int f(t)\,dt \quad (13)$$

and could be treated as a proportional-plus, derivative-plus integral control problem if the traffic system were closed-loop--i.e., if the demand estimates were used to set the traffic signals controlling the flow f(t). The benefits expected from such a general treatment are known in system analysis and well documented in the literature (16,17).

In an open-loop application, such as at isolated intersections, the accuracy of the proposed algorithm depends directly on the regularity of the traffic data trend--i.e., on the similarity between the trend of the data used for the determination of the control coefficients and that of the actual measurements. This implies that the algorithm should be used during periods compatible with those during which data were collected. Updating algorithm coefficients will further improve accuracy. However, extensive historical data are not required by the algorithm. In a closed-loop application, the algorithm would keep its present form. The coefficients a_i could then be determined through analytic methods well established in the theory of dynamic optimal control (16-18).

A simpler version of the algorithm introduced above is obtained for $a_0 = 0$, $a_1 = 0$, $a_2 = 0$, $a_3 = 1$ in Equation 11:

$$\hat{v}_t = \sum_{k=1}^{N} f_{t-k} \Big/ N \quad (14)$$

which is the moving average. As demonstrated later, this version can achieve high prediction accuracy, as good as or slightly worse than that of Equation 11, but it is inherently slow in responding to abrupt demand changes. In the absence of such changes, and in the absence of any previous information on a particular intersection, this most simple version would be preferable to all others.

TEST RESULTS

In this section, sensitivity tests on certain parameters of the two best existing demand prediction algorithms are first performed. The tests indicate that the performance improvement achieved by varying these parameters is negligible. The proposed algorithm is then tested, evaluated, and compared with the other algorithms given in Table 1. Two prediction intervals--5 min and cycle-by-cycle--are considered by using the data set given in Table 2.

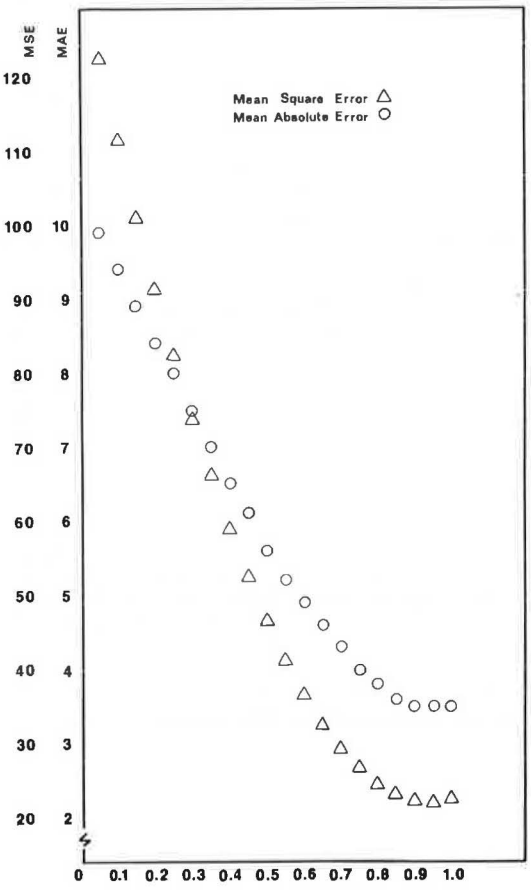
Sensitivity Analysis and UTCS-2 and UTCS-3 Possible Improvements

It has been proposed elsewhere (7) that a potential improvement to UTCS-3 performance lies in computing the parameter γ_j on-line by using a formula that

Table 3. Prediction errors for range of α values in UTCS-3 (data set 1).

α	Error (vehicles/5 min/lane) MSE	MAE
0.05	123.3	10.0
0.10	112.3	9.5
0.15	102.0	9.0
0.20	92.3	8.5
0.25	83.2	8.1
0.30	74.7	7.6
0.35	66.9	7.1
0.40	59.8	6.6
0.45	53.3	6.2
0.50	47.4	5.7
0.55	42.1	5.3
0.60	37.5	5.0
0.65	33.5	4.7
0.70	30.2	4.4
0.75	27.5	4.1
0.80	25.4	3.9
0.85	24.0	3.7
0.90	23.2	3.6
0.95	23.0	3.6
1.00	23.5	3.6

Figure 1. Prediction errors for range of α values in UTCS-3: data set 1.



is different from Equation 7. Such a change would make γ_j a time-varying function adaptive to the latest trend of the traffic deviations. By following similar reasoning, α, the smoothing constant in both UTCS-2 (Equation 3) and UTCS-3 (Equation 8), can be made adaptive by using the Trigg and Leach method (14).

To obtain an indication of the extent to which UTCS performance could be improved through such changes, a sensitivity analysis was performed. First, error sensitivity with respect to changes in α was analyzed. For each α value, a total number of sample points n was chosen so that α achieved a steady-state level. For any data set, errors were then recorded for a complete range of α values.

The tests led to two major conclusions:

1. Around the best α values (i.e., when errors are lowest), errors are not very sensitive to changes in α (see Table 3 and Figure 1). This may be verified, for example, from Table 3, where the MSE elasticity with respect to α can be found to be very low: e_α(MSE) = 0.16; in the same table, it can be seen that e_α(MAE) = 0.
2. Values of α and γ previously recommended for UTCS-2 and UTCS-3 (6,7) were used, and they performed quite well. Performance was worse when any α or γ values other than the ones recommended from the Washington, D.C., application (6,7) were used with UTCS-2 (see Table 4). Performance improved by, at most, 2.8 percent for "best" α values in the UTCS-3 application (see Table 5).

These sensitivity-analysis results indicate that, if methods for calculating α and γ were improved, the improvement of UTCS performance would be insignificant. This can be concluded since, for the locations examined, error sensitivity to α and γ changes around values recommended in the literature was very low. These results are strengthened when it is observed that the locations examined had quite different properties and were controlled by different policies. These results also support the need for a better demand prediction algorithm.

Five-Minute Prediction

Data sets 1-4 (Table 2) were used in these tests. Five algorithms were tested: UTCS-2, UTCS-3, current measurement, historical average, and the proposed algorithm. Use of UTCS-3 for one-time-step, 5-min prediction was possible (6) by considering 5 min as one time step and by setting j = 1 in Equation 6. The coefficients developed for the proposed algorithm and the test results are summarized in Tables 6 and 7, respectively.

The proposed versions in Table 6 follow the form of the prediction introduced previously by Equation 11. The two errors, MSE and MAE, for the five algorithms are presented in Table 7 in two ways. The value for each error is given so that conclusions on algorithm performance can easily be drawn; evi-

Table 4. Prediction errors for varying α and γ values in UTCS-2.

Data Set	MSE (vehicles/5 min/lane) α = 0.9[a] γ = 0.2[a]	γ = 0.4	α = 0.8 γ = 0.2	γ = 0.4	MAE (vehicles/5 min/lane) α = 0.9[a] γ = 0.2[a]	γ = 0.4	α = 0.8 γ = 0.2	γ = 0.4
1	13.1	14.3	14.0	15.6	3.0	3.0	3.0	3.0
2	61.8	61.9	64.2	65.9	6.4	6.4	6.4	6.5
3	51.4	51.1	52.6	59.4	6.1	6.3	6.2	6.5
4	19.1	21.8	21.0	24.3	3.4	3.7	3.6	3.9

[a]Value recommended from past applications (6, 7).

Table 5. Prediction errors for varying α values in UTCS-3.

Data Set	MSE (vehicles/5 min/lane)		MAE (vehicles/5 min/lane)	
	$\alpha = 0.95$[a]	$\alpha = 0.85$	$\alpha = 0.95$[a]	$\alpha = 0.85$
1	23.0	22.5	3.6	3.5
2	89.5	88.2	7.7	7.6
3	84.8	83.3	7.5	7.5
4	30.8	30.1	4.3	4.3

[a]Value recommended from past applications (6, 7).

Table 6. Proposed algorithm versions (5-min prediction).

Proposed Version	Control Coefficient				No. of Time Periods (N)
	a_0	a_1	a_2	a_3	
1	11.748	0.521	0	0	-
2	16.713	0.517	0	0	-
3	20.061	0	0.067	0	-
4	18.375	0	0	0.457	3
5	16.962	0	0	0.497	4
6	14.254	0	0	0.573	5
7[a]	0	0	0	1	3
8[a]	0	0	0	1	4
9[a]	0	0	0	1	5

[a]Corresponds to the moving average.

Table 7. Prediction errors of five types of prediction algorithms for 5-min prediction.

Data Set[a]	Algorithm	MSE		MAE	
		Value	Difference from UTCS-3 (%)	Value	Difference from UTCS-3 (%)
1	Current	23.5	0.9	3.6	0.0
	UTCS-3	23.3	-	3.6	-
	UTCS-2	13.1	-44	3.0	-17
	Historical	12.8	-45	3.0	-17
	Version 1	19.0	-18.5	3.2	-11.1
	Version 7	22.5	-3.4	3.7	2.8
	Version 8	23.1	-0.9	3.8	5.6
	Version 9	24.5	5.2	4.0	11.1
2	Current	92.9	3.8	7.6	-1.3
	UTCS-3	89.5	-	7.7	-
	UTCS-2	61.8	-31	6.4	-17
	Historical	65.0	-27	6.7	-13
	Version 2	89.4	-0.1	8.0	3.9
	Version 7	117.3	31.1	8.5	10.4
	Version 8	126.8	41.7	8.6	11.7
	Version 9	122.9	37.3	8.3	7.8
3	Current	87.1	2.7	7.6	1.3
	UTCS-3	84.8	-	7.5	-
	UTCS-2	51.4	-39	6.1	-19
	Historical	55.8	-34	6.2	-17
	Version 4	71.0	-16.3	7.3	-2.7
	Version 5	71.3	-15.9	7.2	-4.0
	Version 6	73.4	-13.4	7.3	-2.7
	Version 7	67.5	-20.4	6.8	-9.3
	Version 8	61.5	-27.5	6.3	-16.0
	Version 9	63.5	-25.1	6.4	-14.7
4	Current	34.8	13.0	4.7	9.3
	UTCS-3	30.8	-	4.3	-
	UTCS-2	19.1	-38	3.4	-21
	Historical	16.9	-45	3.4	-21
	Version 3	24.8	-19.5	3.8	-11.6
	Version 7	26.8	-13.0	4.0	-7.0
	Version 8	24.8	-19.5	3.7	-14.0
	Version 9	24.3	-21.1	3.8	-11.6

[a]As in Table 2.

dently,lower errors indicate better algorithm performance. In addition, each algorithm is compared with UTCS-3, and the deviation of its error with respect to that of UTCS-3 is presented. A positive deviation means that the algorithm in question has an error greater than that of UTCS-3 and is therefore less desirable than UTCS-3. A negative deviation implies that the algorithm has an error smaller than that of UTCS-3 and is therefore more desirable. The best-performing algorithm in 5-min prediction, UTCS-2, was not chosen as a basis for comparison since it could not be used in cycle-by-cycle prediction.

The following conclusions can be drawn from the test results and the relative performance comparisons given in Table 7:

1. For any data set, at least one version of the proposed algorithm performs substantially (as much as 21 percent) better than UTCS-3.
2. The proposed algorithm does not always perform as well as UTCS-2 or the historical average. This is especially true for the isolated location (data set 2), where the mean absolute error is 25 percent higher than that of UTCS-2. A probable reason for the superior performance of UTCS-2 in this case is the importance of historical data for 5-min prediction in isolated locations.
3. Versions 7-9 of the proposed algorithm, which degenerate to a moving average of three to five previous periods, exhibit performance very similar to that of the complete algorithm.
4. By increasing N (the number of time periods considered for an average) from three to five in the proposed algorithm, performance is not significantly affected.
5. For all locations, UTCS-2 errors are lower than UTCS-3 errors.
6. For all locations, the errors of the historical algorithm are lower than the errors of UTCS-3. At two of the four locations, the historical algorithm also performs better than UTCS-2; however, for all locations, the difference in performance between the historical algorithm and UTCS-2 is not significant.
7. For all locations, prediction from the current measurement alone is worse than that of either UTCS-2 or UTCS-3.

Figures 2-4 show the performance of four of the algorithms tested for data sets 2, 3, and 4, respectively. The historical average has not been plotted together with the rest of the algorithms, since it exhibits behavior very similar to that of UTCS-2. The conclusions cited above can also be drawn from these figures.

Cycle-by-Cycle Prediction

Both coordinated and isolated intersections, corresponding to data sets 5-10 in Table 2, were used in the cycle-by-cycle prediction tests. Three algorithms were tested: UTCS-3, current measurement, and the proposed algorithm. The two historically based algorithms investigated for 5-min prediction--i.e., historical and UTCS-2--could not be used for cycle-by-cycle prediction since signal cycles did not begin and end at the same times each day.

The coefficients developed for the proposed algorithm and the test results are summarized in Tables 8 and 9, respectively. Coefficients for versions 7-9 (Table 6), which correspond to the moving average, are derived on-line. Coefficients for versions 10-12 (Table 8) are derived from data set 6--i.e., a period adjacent to the prediction period of data set 8. Coefficients for versions 13-15 (Table 8) are derived from the combined set of data sets 5-7--i.e., from periods not compatible with the prediction period of data set 8; for this and the previous test only, the deviation and prediction periods were chosen to be different so that the performance of the algorithm under less than ideal conditions could be examined. Finally, coefficients

Figure 2. Five-minute volume: data set 2.

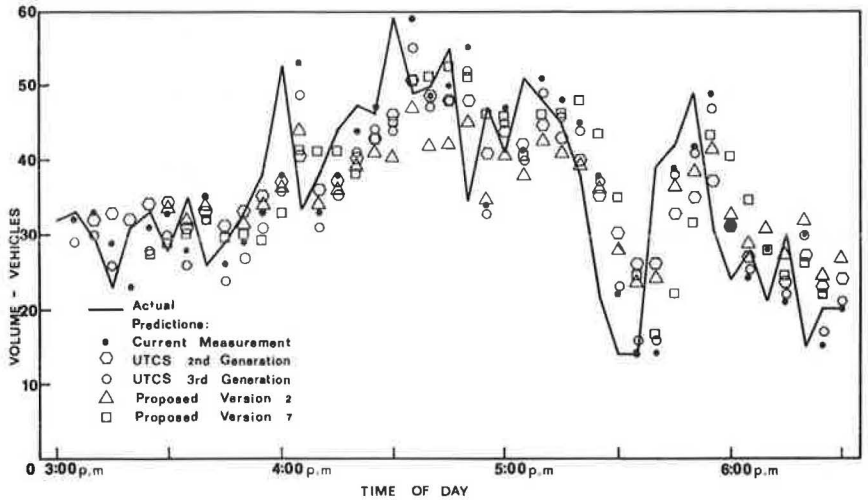


Figure 3. Five-minute volume: data set 3.

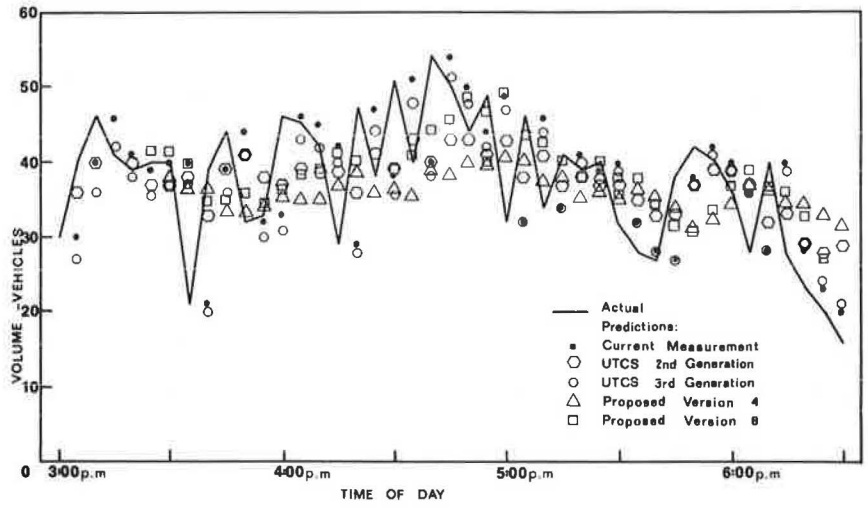
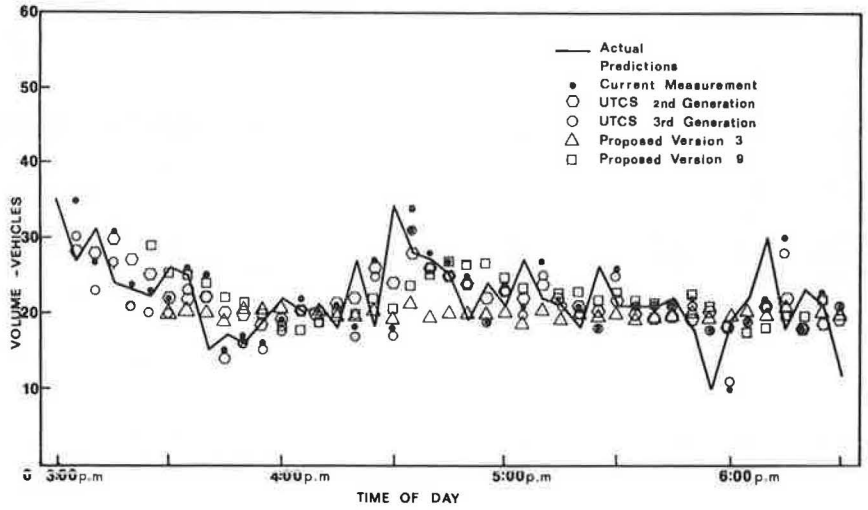


Figure 4. Five-minute volume: data set 4.



for versions 16-18 (Table 8) are derived from data set 9--i.e., a period almost identical to the prediction period. Here, the algorithm was expected to perform best.

The following conclusions can be drawn from the test results and the relative performance comparisons given in Table 9:

1. At all times, at least three versions of the proposed algorithm are substantially (as much as 41 percent) better than UTCS-3. These versions correspond to the moving average. As in the case of 5-min prediction, the unsatisfactory UTCS-3 performance can be attributed, at least in part, to the occasional congestion at the test sites.

Table 8. Proposed algorithm versions (cycle-by-cycle prediction).

Version	Control Coefficient				No. of Time Periods (N)
	a_0	a_1	a_2	a_3	
10	7.435	0	0	-0.359	3
11	9.759	0	0	-0.781	4
12	11.053	0	0	-1.014	5
13	2.044	0	0	0.762	3
14	1.597	0	0	0.818	4
15	1.265	0	0	0.861	5
16	6.839	0	0	0.277	3
17	8.030	0	0	0.149	4
18	8.098	0	0	0.143	5

Table 9. Prediction errors of three types of prediction algorithms for cycle-by-cycle prediction.

Data Set[a]	Algorithm	MSE		MAE	
		Value	Difference from UTCS-3 (%)	Value	Difference from UTCS-3 (%)
5	Current	13.14	26	2.98	21
	UTCS-3	10.44	-	2.47	-
	Version 7	7.77	-26	2.22	-10
	Version 8	7.19	-31	2.13	-14
	Version 9	6.94	-34	2.12	-14
6	Current	4.70	10	1.71	9
	UTCS-3	4.27	-	1.57	-
	Version 7	3.74	-12	1.55	-1
	Version 8	3.20	-25	1.44	-8
	Version 9	3.20	-25	1.43	-9
7	Current	7.44	2	2.04	-2
	UTCS-3	7.33	-	2.09	-
	Version 7	6.01	-18	2.03	-3
	Version 8	6.12	-17	2.03	-3
	Version 9	6.04	-18	2.07	-1
8	Current	9.43	22	2.51	15
	UTCS-3	7.74	-	2.18	-
	Version 7	5.07	-34	1.78	-18
	Version 8	5.21	-33	1.76	-19
	Version 9	4.60	-41	1.68	-23
	Version 10[b]	6.20	-20	1.99	-9
	Version 11[b]	8.30	7	2.37	9
	Version 12[b]	9.37	21	2.55	17
	Version 13[c]	6.52	-16	1.98	-9
	Version 14[c]	5.90	-24	1.83	-16
	Version 15[c]	5.79	-25	1.84	-16
5-8	Current	8.75	17	2.32	12
	UTCS-3	7.47	-	2.08	-
9	Current	17.02	15	3.19	7
	UTCS-3	14.82	-	2.97	-
	Version 7	9.22	-38	2.47	-17
	Version 8	8.70	-41	2.48	-16
	Version 9	9.22	-38	2.53	-15
10	Current	8.53	7	2.44	7
	UTCS-3	7.99	-	2.29	-
	Version 16[d]	5.43	-32	2.00	-13
	Version 17[d]	5.31	-34	1.97	-14
	Version 18[d]	5.31	-34	1.97	-14
	Version 7	6.20	-22	1.99	-13
	Version 8	6.30	-21	2.06	-10
	Version 9	6.02	-25	2.02	-12

[a] As in Table 2.
[b] Equations derived from data set 6.
[c] Equations derived from data sets 5-7.
[d] Equations derived from data set 9.

2. From data sets 9 and 10, it can be verified that the more sophisticated versions of the proposed algorithm (versions 16-18) perform even better than the moving average. Therefore, the best performance is obtained when the algorithm is derived from data collected, on an earlier day, during a period identical to the prediction period.

3. When the algorithm is derived from data collected during periods incompatible with the prediction period, in most cases performance is still considerably (as much as 25 percent) better than UTCS-3. In such cases, however, the moving average offers a slightly better prediction than the more sophisticated versions and is therefore preferable.

4. The performance of the proposed algorithm is not affected by the approach type (i.e., coordinated or isolated).

5. For all data sets, the current measurement predictor is least desirable.

Figures 5-8 illustrate the performance of the three algorithms in cycle-by-cycle prediction of data sets 5, 8, 9, and 10, respectively. For data set 8, Figure 6 shows three versions of the proposed algorithm--one that uses a data set from the previous day (version 10), one that uses three data sets from the previous and the current day (version 15), and one that is equivalent to a moving average (version 9) using four data sets. For data set 9, Figure 7 shows the proposed version equivalent to a moving average whereas, for data set 10, Figure 8 shows both that version and the complete proposed algorithm. The conclusions derived from Table 9 and cited above could also be drawn from these figures.

As Figure 8 shows, the proposed algorithm achieves superior performance by weighing a constant average, describing volume trend during the same period on a previous day, much more heavily than cycle-by-cycle traffic fluctuations. This suggests that the assumption of average arrivals frequently used in practice for isolated intersections is not unreasonable. In contrast, in Figures 3 and 6 and in Table 8 it is seen that this assumption is unreasonable for coordinated intersections.

Finally, it should be noted that, as Tables 7 and 9 indicate, the improved performance of the proposed algorithm over UTCS-3 is much more noticeable in cycle-by-cycle than in 5-min prediction.

CONCLUSIONS

The test results of the previous section suggest that in 5-min prediction, for all locations, UTCS-2 performs better than UTCS-3 and that both are superior to the current measurement for prediction. The results also indicate that predictions based on historical data alone are as good as and frequently better than UTCS-2 predictions. The results are consistent with previous findings and lead to the conclusion that computations and data requirements can be significantly reduced by choosing the historical algorithm over either UTCS-2 or UTCS-3 for 5-min prediction. The results also indicate that, in 5-min prediction, at least one version of the proposed algorithm performs better than UTCS-3 in all cases; however, it does not always perform as well as UTCS-2.

In cycle-by-cycle prediction, which was of the greatest interest in this study, UTCS-2 and the historical algorithm cannot be used. From the remaining three algorithms, again the current measurement performs worse than UTCS-3 while the moving-average version of the proposed algorithm performs better than UTCS-3 in all cases. When the proposed algorithm is derived from data collected on an earlier day during a period identical to the prediction period, it performs better than all algorithms examined. However, the complete version of the proposed algorithm has certain disadvantages in comparison with the moving-average version: (a) It may need to be updated frequently, (b) it requires that data be collected on at least one previous day, and (c) it performs best when used during a specified time of day, which makes it necessary to develop more than one equation for each day (a minimum of

Figure 5. Flow rate: data set 5.

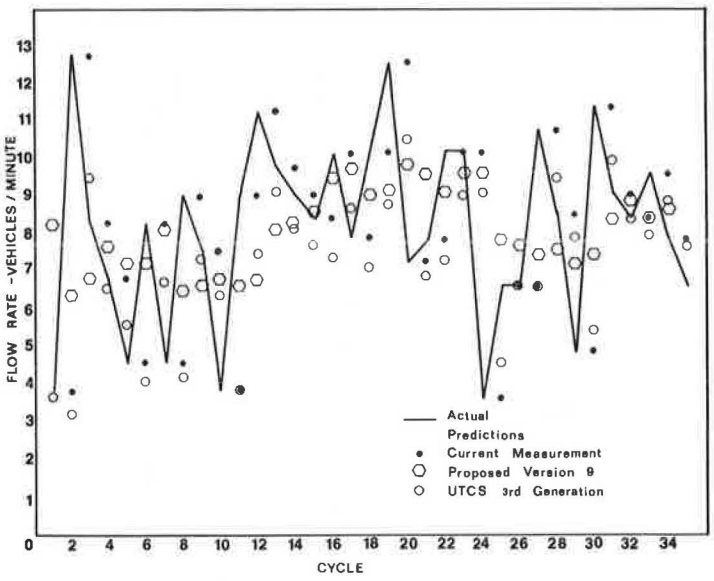


Figure 6. Flow rate: data set 8.

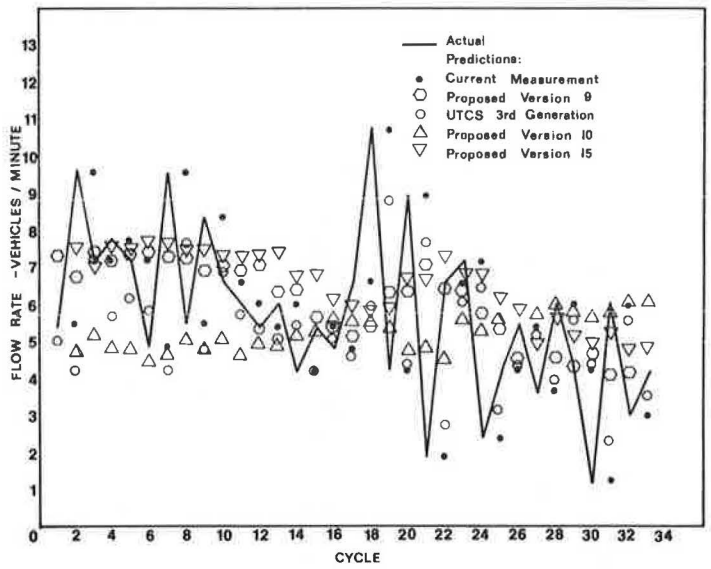


Figure 7. Flow rate: data set 9.

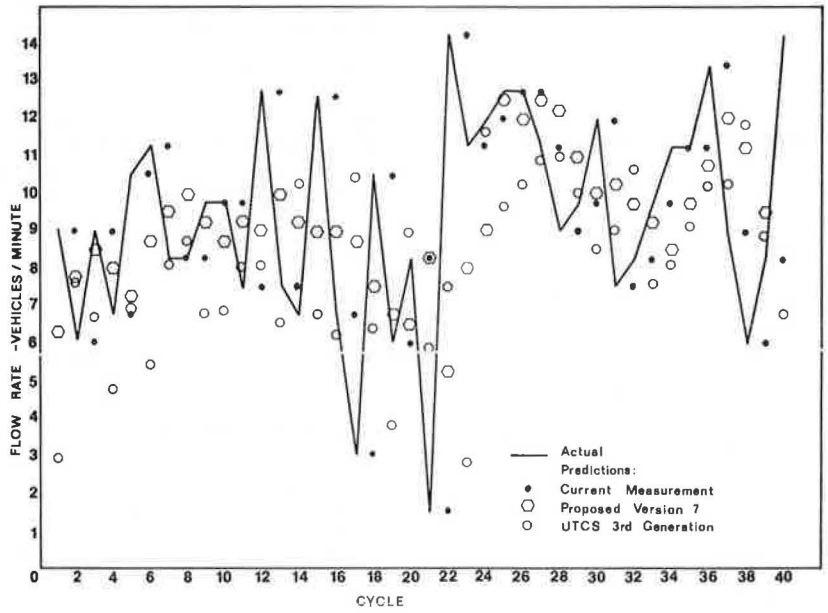
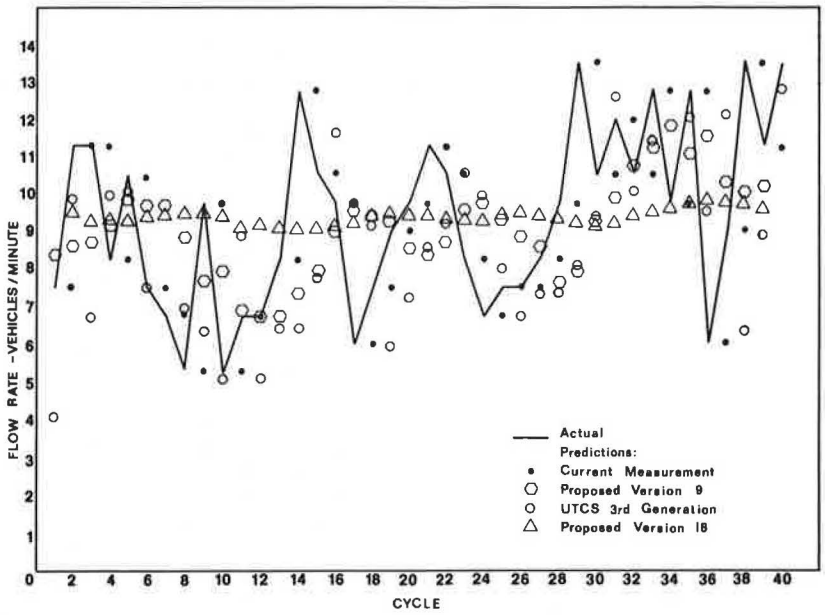


Figure 8. Flow rate: data set 10.



three equations for each peak and off-peak period would be required).

Despite these criticisms, the proposed algorithm has significant advantages over existing algorithms. It does not need extensive historical data as UTCS-2 and the historical average do, and it can be applied in cycle-by-cycle prediction whereas UTCS-2 and the historical average cannot. Furthermore, it performs better (as much as 41 percent better for the versions examined) than the UTCS-3. It should also be pointed out that it could easily be optimized at a later date by using established optimal-control-theory methodologies.

The moving-average version of the proposed algorithm will usually offer prediction that is more accurate than that offered by the best existing algorithms. It achieves such performance with minimal data, since volume or flow-rate measurements from the previous three intervals are sufficient. It is therefore recommended that the moving-average version of the proposed algorithm be used initially, especially if little is known about the demand characteristics of a particular intersection. If more information on demand is available, it is desirable to use the complete proposed algorithm.

ACKNOWLEDGMENT

We would like to acknowledge financial support from the Program of University Research of the U.S. Department of Transportation.

REFERENCES

1. G. Stephanopoulos, P.G. Michalopoulos, and G. Stephanopoulos. Modelling and Analysis of Traffic Queue Dynamics at Signalized Intersections. Transportation Research, Vol. 13A, 1979, pp. 295-307.
2. P.G. Michalopoulos, G. Stephanopoulos, and V.B. Pisharody. Modelling of Traffic Flow at Signalized Links. Transportation Science, Vol. 14, No. 1, 1980, pp. 9-41.
3. P.G. Michalopoulos, G. Stephanopoulos, and G. Stephanopoulos. An Application of Shock Wave Theory to Traffic Signal Control. Transportation Research, 1980.
4. Urban Traffic Control System and Bus Priority System Traffic Adaptive Network Signal Timing Program: Software Description. Federal Highway Administration, U.S. Department of Transportation, Aug. 1973.
5. E.B. Lieberman and others. Variable Cycle Signal Timing Program: Volume 4--Prediction Algorithms, Software and Hardware Requirements, and Logical Flow Diagrams. Federal Highway Administration, U.S. Department of Transportation, May 1974. NTIS: PB 241 720.
6. J.B. Kreer. A Comparison of Predictor Algorithms for Computerized Traffic Control Systems. Traffic Engineering, Vol. 45, No. 4, April 1975, pp. 51-56.
7. P. Mengert, P. Brown, and L. Yuan. Prediction Algorithms for Urban Traffic Control. Transportation Systems Center, U.S. Department of Transportation, Cambridge, MA, Internal Project Memorandum, Feb. 1979.
8. J.B. Kreer. Factors Affecting the Relative Performance of Traffic Responsive and Time-of-Day Traffic Signal Control. Transportation Research, Vol. 10, 1976, pp. 75-81.
9. Improved Control Logic for Use with Computer-Controlled Traffic. Stanford Research Institute, Menlo Park, CA, NCHRP Project 3-18(1), Final Rept., June 1977.
10. Improved Operation of Urban Transportation Systems: Volume 3--The Development and Evaluation of a Real-Time Computerized Traffic Control Strategy. Canada Ministry of Transport, Toronto, Ontario, 1976.
11. J.S. Baras and W.S. Levine. Estimation of Traffic Flow Parameters in Urban Traffic Networks. Proc., IEEE Annual Conference on Decision and Control, San Francisco, 1977, pp. 428-433.
12. J.S. Baras, W.S. Levine, and T.L. Lin. Discrete Time Point Processes in Urban Traffic Queue Estimation. Proc., IEEE Conference on Decision and Control, San Diego, 1978, pp. 1025-1031.
13. J.S. Baras and others. Advanced Filtering and Prediction Software for Urban Traffic Control Systems. Transportation Studies Center, Univ. of Maryland, College Park, Draft Final Rept., 1979.
14. D.W. Trigg and A.G. Leach. Exponential Smooth-

ing with an Adaptive Response Rate. Operational Research Quarterly, Vol. 18, No. 1, 1967, pp. 53-59.
15. N.H. Nie and others. Statistical Package for the Social Sciences, 2nd ed. McGraw-Hill, New York, 1975.
16. J.J. D'Azzo and C.H. Houpis. Feedback Control System Analysis and Synthesis, 2nd ed. McGraw-Hill, New York, 1966.
17. R.C. Dorf. Modern Control Systems, 2nd ed. Addison-Wesley, Reading, MA, 1974.
18. S.J. Citron. Elements of Optimal Control. Holt, Rinehart, and Winston, New York, 1969.

Discussion

Nathan H. Gartner

The authors have conducted a comprehensive review of algorithms for short-range traffic-volume prediction of the type that were used in second- and third-generation UTCS control strategies. They also investigate a newly proposed, rather simple algorithm that in its general form calculates the predicted volume as a linear combination of the current (measured) value, the difference with respect to the previous value, and the moving average (analogous to the proportional-plus, derivative-plus integral controller). The authors tested this algorithm on a number of data sets covering 5-min intervals and single-cycle intervals (presumably, in the 1- to 2-min range). Judging by two forms of mean error criteria--MSE and MAE (Equations 9 and 10)--results indicate that, for the shorter, cycle-by-cycle predictions, the proposed algorithms perform better than the other algorithms that were tested.

The results and conclusions of the paper are veritable and confirm ideas obtained from previous studies on the use of prediction algorithms in traffic control. This discussion is only peripherally concerned with the particular results and methodology of the paper; it is primarily directed toward the ulterior objective of the study, that of implementing real-time traffic-control strategies.

Real-time traffic control is designed to provide an increased degree of responsiveness to changing traffic flows. The expectation is that intersection performance can be improved by capitalizing on this variability. Yet extensive field tests with the UTCS system and elsewhere show that such expectations did not materialize (19,20). The more responsive the strategy that was tested, the less effective was its performance. In analyzing these results, one may erroneously conclude that a library of signal-control plans generated off-line by using historical data (from another day, perhaps another year, but for the same daily period) is more effective than controls generated on-line by using very recent data (the past few minutes). However, a closer examination of these studies reveals that it is not the rationale that has failed (i.e., that traffic-responsive control should provide benefits over fixed-time control) but the models and procedures that were implemented that failed to produce the desired results.

Real-time traffic-control strategies that rely on predicted volumes are not truly responsive: They do not respond to actual traffic conditions but to hypothetical conditions. The traffic-flow process and the optimization procedure used in deriving the control plans form an inseparable closed-loop control system. The signal controls can only be effective if an accurate model is used in the optimization. Yet these strategies use an abstract model that is calibrated by the predicted volumes. Predictions are inherently inaccurate, and therefore the models cannot take account of the short-term fluctuations to which they are supposed to respond. In essence, by aggregating and averaging the data, the prediction algorithms destroy the information content that is most important for real-time control.

A good demonstration of the inadequacy of short-term volume predictions for their intended use is provided by the extensive data sets analyzed by the authors and shown in Figures 2-8. In all cases, the discrepancies between predicted and actual values are very substantial. The shorter the prediction interval, the larger are the relative discrepancies. [The authors should also calculate the mean relative error values of the type $1/N \Sigma$ (error/measured value).] A most telling example is shown in Figure 8, where the best predictor turns out to be an almost fixed value, notwithstanding the highly variable cycle-by-cycle flow rates. It is clear that one cannot conceivably derive a responsive strategy from such a prediction.

Furthermore, suppose one could predict the flows in each cycle with complete accuracy (i.e., with a zero mean error value). Even then the resulting real-time control strategy might be ineffective. For example, the following numbers represent vehicle arrivals for two cycles, grouped into 5-s intervals, on a signal-controlled approach with a 60-s cycle time:

Cycle	Vehicle Arrivals
1	1 1 2 1 1 2 0 2 0 0 1 1
2	0 1 0 0 1 1 2 1 2 1 1 2

During both cycles, the flow is the same (12 vehicles), yet the optimal control strategy for each should be entirely different because of the different distribution of the arrivals within the cycle.

To summarize, I offer the following conclusions:

1. Reliable estimates of future traffic volumes can only be obtained for lengthier time periods (of several minutes). These estimates can then be used to derive steady-state-type control strategies (e.g., first-generation traffic-responsive control).
2. The quality of predictions should not be judged merely by their average closeness to the actual values; rather, they should be evaluated in terms of the ultimate objective--the effectiveness of the control strategies that they produce.
3. Estimates of volumes for very short periods (e.g., cycles) are unreliable and cannot be used to provide effective real-time control. Therefore, there is a need to develop real-time traffic-control strategies that move away from the use of predicted average volumes and rely mostly on actual flows.

REFERENCES

19. J. MacGowan and I.J. Fullerton. Development and Testing of Advanced Control Strategies in the Urban Traffic Control System. Public Roads, Vol. 43, Nos. 2-4, 1979-1980.
20. P.J. Tarnoff. Concepts and Strategies: Urban Street Systems. Proc., International Symposium on Traffic Control Systems, Univ. of California, Berkeley, Aug. 1979.

Samir A. Ahmed

The authors have provided a needed critique of several predictor models that have been proposed for

real-time control of street intersections. It is true, as the authors have pointed out, that one of the major reasons for the unsatisfactory performance of real-time signal-control systems is the lack of a predictor model that describes the dynamic behavior of the traffic-flow process. The paper does not, however, substantiate the important points made on how a predictor model should avoid arbitrary methods of treating the data and instead be based on a solid theoretical foundation.

Predictor models can be categorized into two general categories: ad hoc models and point-process models. Ad hoc models (e.g., moving-average models, exponential smoothing models, and adaptive exponential smoothing models) propose arbitrary weighting schemes to be assigned to the current and previous observations on the variable of interest. The primary advantage of these models is their ease of implementation, and their major weakness is their inherent lack of a general approach for choosing among alternative weighting schemes. The proposed models described in this paper (Equation 11) and the rest of the models given in Table 1 are in this category. In using these models, much is left to the personal judgment of the engineer or investigator who must assume any special knowledge of the control system under consideration.

The other category of models--point-process models--forms a broad class of potential models for representing the stochastic nature of many physical processes. The eventual form of the predictor model is determined by the properties of the observations on how a process actually behaves and, among other linear predictor models, the resulting forecasts are optimal in terms of having minimum MSE (21). In addition, in view of the recent developments in the joint analysis of interrelated processes, it is possible to construct point-process models for predicting several processes observed in different parts of the control system.

To compare the prediction errors resulting from ad hoc models and point-process models, I applied three ad hoc models and a point-process model that has been developed for freeway traffic to a total of 166 data sets obtained from three freeway surveillance systems (22). Forecasts were made for 1 min ahead in time. By using the same error measures described in the paper, the point-process model has been found to be superior to the ad hoc models.

In conclusion, I believe that it is time that some serious research was devoted to the problem of predicting traffic-flow variables for the real-time control of street intersections. Such research should be based on the well-established theory of point-process models.

REFERENCES

21. G.E.P. Box and G.M. Jenkins. Time Series Analysis Forecasting and Control. Holden-Day, San Francisco, 1976.
22. S.A. Ahmed and A.R. Cook. Analysis of Freeway Traffic Time-Series Data by Using Box-Jenkins Techniques. TRB, Transportation Research Record 722, 1979, pp. 1-9.

Authors' Closure

We would like to thank Gartner for his constructive remarks, which are in close agreement with our views. With regard to Ahmed's comments, we observe the following.

As pointed out from the outset, the objective of this study was to test the existing models in order to select the most appropriate one for easy field implementation subject to the stated criteria. The development of the models was incidental and resulted from the need to simplify the prediction process. The simple predictors emerged as at least as accurate as and frequently better than the most complex ones.

With respect to the argument that prediction models fall into the two categories suggested by Ahmed, one can argue that, under a more comprehensive classification, models can be grouped as (a) descriptive or correlative, (b) causal, and (c) normative. Most efforts to date belong to the first category. Their inherent weakness is lack of attention to causal analysis and use of rather arbitrary and complex statistical methods to best describe the system modeled. Our attempt at capturing the basic causal relations within the system analyzed represents only a first modest step in entering the second category. We insisted on keeping our models simple in form, yet we included all major terms that would make sense to a traffic practitioner. Furthermore, the models can be simply transformed to become part of the complete feedback system analysis in either the time or the frequency domain, once the complete traffic system is modeled. Since an understanding of the causal relations governing the whole traffic system is currently lacking, it is our belief that a jump into normative interpretations would be premature at this stage. Since such an understanding is lacking, results from normative analysis would not offer substantial improvements and therefore would not be cost effective. To be sure, Ahmed's findings (22) are, in effect, consistent with this statement. For example, they indicate that a simple moving average could be as accurate as the method he proposed.

As for Ahmed's final conclusion, we again feel that the emphasis should not be on the exact statistical treatment of the problem. We believe that demand prediction should not be treated as a point process without using more valuable information that is currently ignored, information related to traffic dynamics. More specifically, a detailed traffic model should be developed for analyzing traffic flow in signalized links. Demand prediction would, then, be greatly simplified, since information from upstream detectors would be used. In addition, a traffic model is needed to identify the tail end of the queue and predict arrivals there. In short, complete dynamic traffic-behavior analysis is more crucial than exact statistical treatment of the prediction method.

Finally, the methods proposed by Ahmed and Cook (22) do not appear to be overwhelmingly superior, and their successful and cost-effective application to interrupted-flow conditions remains to be tested.

MAXBAND: A Program for Setting Signals on Arteries and Triangular Networks

JOHN D. C. LITTLE, MARK D. KELSON, AND NATHAN H. GARTNER

MAXBAND is a portable, off-line, FORTRAN IV computer program for setting arterial signals to achieve maximal bandwidth. Special features of the program include (a) automatically choosing cycle time from a given range, (b) permitting the design speed to vary within given tolerances, (c) selecting the best lead or lag pattern for left-turn phases from a specified set, (d) allowing a queue clearance time for secondary flow accumulated during red, (e) accepting user-specified weights for the green bands in each direction, and (f) handling a simple network in the form of a three-artery triangular loop. Green splits can be provided or, alternatively, flows and capacities can be given and splits calculated by using Webster's theory. The program produces cycle time, offsets, speeds, and order of left-turn phases to maximize the weighted combination of bandwidths. The optimization uses Land and Powell's MPCODE branch and bound algorithm. As many as 12 signals can be handled efficiently. The program is available from the Federal Highway Administration.

Signal-setting methods for fixed-time traffic-control systems separate broadly into two classes. The first, and historically oldest, consists of methods that maximize bandwidth and progression. This group includes, among others, those of Little and Morgan (1), Little (2) and Messer and others (3). The second group contains methods that seek to minimize delay, stops, or other measures of disutility. Examples are Hillier's combination method (4), Traffic Research Corporation's SIGOP (5), Robertson's TRANSYT (6), Gartner, Little, and Gabbay's MITROP (7), and Lieberman and Woo's SIGOP II (8).

Although disutility-oriented methods have now been available for some time, many traffic engineers continue to prefer maximal bandwidth settings because they have certain inherent advantages. For one thing, bandwidth methods use relatively little input; the basic requirements are street geometry, traffic speeds, and green splits. In addition, progression systems are operationally robust. Space-time diagrams let the traffic engineer visualize easily the quality of the results. Through accumulated experience, engineers who have knowledge of the specific streets can spot problems and, if necessary, make modifications to the settings. Furthermore, various studies [for example, the study by Wagner, Gerlough, and Barnes (9)] and much practical experience have shown that bandwidth systems give good results in the field. It may even be that drivers expect signal progression and take it as a measure of setting quality. In any case we take the position here that, if engineers are going to use bandwidth systems, they should have the best.

Morgan and Little (1) first computerized the setting of arterial signals for maximal bandwidth. The widely distributed program of Little, Martin, and Morgan (10) efficiently finds offsets for maximal bandwidth given cycle time, red times, signal distances, and street speed. The total bandwidth attained can be allocated between directions on the basis of flow.

Little (2) subsequently generalized the computation in several ways: The cycle time could be automatically selected from a given range, and so could speed. Networks could be solved. These and several further extensions became possible through a mixed-integer formulation of the problem.

The flexibility thereby introduced has several advantages. For example, maximal bandwidth calculations frequently have a disconcerting feature. On a long street, the signals that constrict bandwidth may turn out to be far apart. A small change in design speed or cycle time can produce quite different signal settings and bandwidths. Yet drivers do not hold their speeds exactly constant and, as shown by Desrosiers and Leighty (11), they tend to adjust their speeds to the signals. Therefore, it makes sense to let design speed between signals be a variable, at least within certain limits. Similarly, it is helpful to be able to consider a range of possible cycle times automatically and determine which one combines best with the street geometry to yield maximal bandwidth. The mixed-integer formulation permits this.

The approach has not become popular, however, for two principal reasons:

1. A person must invest substantial effort in learning how to formulate and solve problems in this way.
2. At the time the paper by Little (2) appeared, the solving of mixed-integer problems was inefficient and expensive. Since then, however, methods for solving mixed-integer problems have become better and large-scale computations have become cheaper.

Further research described here reveals that the mixed-integer formulation extends to multiphase signals. For example, asymmetric reds that occur when the green is delayed for left turns can easily be introduced into the formulation. The decision whether to put a left-turn arrow at the beginning or end of the green can be assigned to the optimization and resolved in whichever way maximizes total bandwidth.

Messer and others (3) have also developed a program, PASSER II, to consider this last issue. An advantage of the present approach is that the multiphase feature is combined with the flexibilities mentioned earlier into a single formulation that can be pursued to a mathematically proved optimum. A further extension is of potential value when secondary flows are significant. Turn traffic entering the artery at a previous intersection may build a queue that interferes with the progression. In this case, a time advance can be put into the through band to permit the queue to clear the intersection before the through platoon arrives.

A portable FORTRAN IV computer program has been developed that maximizes bandwidth according to methods described by Little (2) and extended here. The program, called MAXBAND, is designed to handle arteries and simple three-artery networks that contain as many as 17 signals. The program has been documented in a series of reports for the Federal Highway Administration (FHWA) (12).

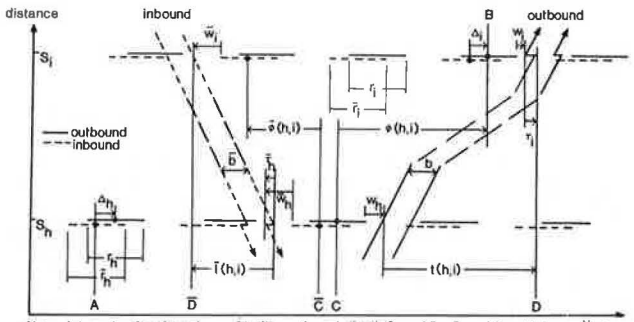
METHODOLOGY

Optimization

The optimization formulation draws and generalizes on Little (2). The basic geometry is shown in Figure 1. Let

$b(\bar{b})$ = outbound (inbound) bandwidth (cycles);

Figure 1. Space-time diagram showing green bands.



Note: Inbound and outbound green bands pass through signals S_h and S_i. Quantities with bars refer to inbound direction; those without, to outbound. Outbound reds are drawn solid and above inbound reds, which are dashed. In the general case shown, inbound and outbound reds need not coincide.

S_i = ith signal (i = 1,...,n);
r_i ($\bar{r}_i$) = outbound (inbound) red time at S_i (cycles);
w_i ($\bar{w}_i$) = time from right (left) side of red at S_i to left (right) edge of outbound (inbound) green band (cycles);
$t(h,i)$ [$\bar{t}(h,i)$] = travel time from S_h to S_i outbound (S_i to S_h inbound) (cycles);
$\phi(h,i)$ [$\bar{\phi}(h,i)$] = time from center of an outbound (inbound) red at S_h to the center of a particular outbound (inbound) red at S_i (cycles) {the two reds are chosen so that each is immediately to the left (right) of the same outbound (inbound) green band; $\phi(h,i)$ [$\bar{\phi}(h,i)$] is positive if the S_i center of red lies to the to the right (left) of the S_h center of red};
Δ_i = time from center of $\bar{r}_i$ to nearest center of r_i (cycles) (positive if center of r_i is to the right of center of $\bar{r}_i$); and
τ_i ($\bar{\tau}_i$) = queue clearance time, an advance of the outbound (inbound) bandwidth upon leaving S_i (cycles).

The fundamental equation for formulating the arterial problem arises from a physical constraint. It is derived with the help of Figure 1 by expressing the difference in time from A to B in two different ways: First, by using outbound-defined quantities, time A to B = Δ_h + integer number of cycles + $\phi(h,i)$. Then, by using inbound-defined quantities, time A to B = integer number of cycles - $\bar{\phi}(h,i)$ + another integer number of cycles + Δ_i. Setting these times equal, rearranging, and coalescing the integers into a single variable, m(h,i),

$$\phi(h,i) + \bar{\phi}(h,i) + \Delta_h - \Delta_i = m(h,i) \tag{1}$$

We call m(h,i) the loop integer in recognition of the more general case of networks. The terminology applies in the present case because the links S_h to S_i and S_i to S_h form a loop and Equation 1 states that the sum of times around the loop is an integer number of cycles.

From Figure 1, we also read from C to D:

$$\phi(h,i) + (1/2)r_i + w_i + \tau_i = (1/2)r_h + w_h + t(h,i) \tag{2a}$$

and from $\bar{C}$ to $\bar{D}$:

$$\bar{\phi}(h,i) + (1/2)\bar{r}_i + \bar{w}_i = (1/2)\bar{r}_h + \bar{w}_h - \bar{\tau}_h + \bar{t}(h,i) \tag{2b}$$

Substituting Equation 2 into Equation 1 to eliminate ϕ and $\bar{\phi}$ gives

$$t(h,i) + \bar{t}(h,i) + (1/2)(r_h + \bar{r}_h) + (w_h + \bar{w}_h) - (1/2)(r_i + \bar{r}_i) - (w_i + \bar{w}_i) - (\tau_i + \bar{\tau}_h) + \Delta_h - \Delta_i = m(h,i) \tag{3}$$

So far we have required that S_i follow S_h in the outbound direction, but this restriction is not necessary. For physical reasons we wish t(.) to satisfy t(h,j) = t(h,i) + t(i,j); by reason of which, setting h = j, we shall require t(i,h) = -t(h,i) and, by a similar argument, $\bar{t}(i,h) = -\bar{t}(h,i)$. With these relations, Equations 2 and 3 hold for arbitrary S_h and S_i, and

$$\phi(h,j) = \phi(h,i) + \phi(i,j) \qquad \phi(h,i) = -\phi(i,h) \tag{4}$$

$$m(h,j) = m(h,i) + m(i,j) \qquad m(h,i) = -m(i,h) \tag{5}$$

along with corresponding expressions for $\bar{\phi}$.

Notation becomes simpler if the signals are numbered sequentially from 1 to n in the outbound direction. Then define $x_i = x(i,i+1)$ for x = t, $\bar{t}$, m, ϕ, $\bar{\phi}$. Now Equation 3 gives

$$t_i + \bar{t}_i + (w_i + \bar{w}_i) - (w_{i+1} + \bar{w}_{i+1}) + \Delta_i - \Delta_{i+1} = -(1/2)(r_i + \bar{r}_i) + (1/2)(r_{i+1} + \bar{r}_{i+1}) + (\bar{\tau}_i + \tau_{i+1}) + m_i \tag{6}$$

From Figure 1 we also see that

$$w_i + b \leq 1 - r_i \tag{7a}$$

$$\bar{w}_i + \bar{b} \leq 1 - \bar{r}_i \tag{7b}$$

If for the moment we also require $b = \bar{b}$, we can collect Equation 6 and Equation 7 into a basic mixed-integer linear program (LP1) for setting arterial signals.

LP1: Find b, $\bar{b}$, w_i, $\bar{w}_i$, and m_i to max b, subject to

$$\left.\begin{aligned} \bar{b} &= b \\ w_i + b &\leq 1 - r_i \\ \bar{w}_i + \bar{b} &\leq 1 - \bar{r}_i \end{aligned}\right\} \quad i = 1, \ldots, n$$

$$(w_i + \bar{w}_i) - (w_{i+1} + \bar{w}_{i+1}) + (t_i + \bar{t}_i) + \Delta_i - \Delta_{i+1} = -(1/2)(r_i + \bar{r}_i) + (1/2)(r_{i+1} + \bar{r}_{i+1}) + (\bar{\tau}_i + \tau_{i+1}) + m_i \qquad i = 1, \ldots, n-1$$

$$m_i = \text{integer}$$

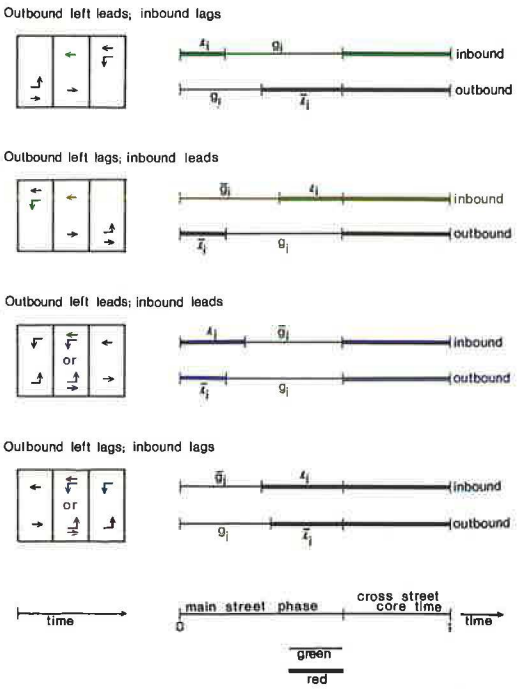
$$b, \bar{b}, w_i, \bar{w}_i \geq 0 \qquad i = 1, \ldots, n$$

LP1 has 3n constraints, 2n + 2 continuous variables, and n - 1 unrestricted integer variables.

In the formulation of LP1, the green band is defined on departure from the signal. Therefore, when queue clearance times are introduced, the jog put into the band will, under some circumstances, cause the tail of the arriving band to hit red. If desired, the green band can be defined so as to require room for both the arriving green band and the queue clearance jog. This can be done by adding t_i and $\bar{t}_i$, respectively, to the left-hand sides of Equations 7a and 7b (or 13a and 13b below). This change may reduce the bandwidth somewhat.

We next introduce a generalization that permits the optimization program to decide when the left-turn phase (if one is present) will occur with respect to the through green at any signal. The

Figure 2. The four possible patterns of left-turn phases.



left-turn green can be picked to lead or lag, whichever gives the most total bandwidth. At the same time, however, the traffic engineer must be able to specify which of the possible combinations of leads or lags will be permitted in a given instance.

Figure 2 shows the four possible patterns of left-turn green phases. Let

g_i ($\bar{g}_i$) = outbound (inbound) green time for through traffic at S_i (cycles),
ℓ_i ($\bar{\ell}_i$) = time allocated for outbound (inbound) left-turn green at S_i (cycles), and
R = common red time in both directions to provide for cross-street movements (cycles).

Since the time allocated to outbound left-turn green is inbound red time, we have (see Figure 2) $r_i = R + \bar{\ell}_i$, $\bar{r}_i = R + \ell_i$ and $r_i + g_i = 1$, $\bar{r}_i + \bar{g}_i = 1$. Moreover, we can express Δ_i, the time from the center of r_i to the next center of r_i, in terms of ℓ_i and $\bar{\ell}_i$ for each case as follows:

Pattern	Δ_i
1	$-(1/2)\ (\ell_i + \bar{\ell}_i)$
2	$(1/2)\ (\ell_i + \bar{\ell}_i)$
3	$-(1/2)\ (\ell_i - \bar{\ell}_i)$
4	$(1/2)\ (\ell_i - \bar{\ell}_i)$

All of these can be expressed in the following form:

$$\Delta_i = (1/2)\,[(2\delta_i - 1)\ell_i - (2\bar{\delta}_i - 1)\bar{\ell}_i] \tag{8}$$

where δ_i and $\bar{\delta}_i$ are 0 - 1 variables and the previous cases are now picked out by

Pattern	δ_i	$\bar{\delta}_i$
1	0	1
2	1	0
3	0	0
4	1	1

Therefore, we can use the mixed-integer program to select the pattern that will maximize bandwidth. If only certain patterns are to be permitted, restriction can be placed on the δ_i and $\bar{\delta}_i$ to enforce the requirements. For example, if only patterns 1 and 2 are permitted, the constraint $\delta_i + \bar{\delta}_i = 1$ is added.

We often wish to let the user favor one direction--say, by manipulating the ratio of inbound to outbound bandwidth. For example, this ratio might be set to that of the two flows. Such a requirement is easily built into the LP as a constraint. However, in making one green band larger than the other, we can never make it larger than the smallest green in that direction. Once this is achieved, it is foolish to restrict the opposite direction further just to satisfy the ratio. Therefore, we speak of a target ratio.

Let k = target ratio of inbound to outbound bandwidth. For the case of $k < 1$ (outbound favored), we can set up the objective function and ratio constraint as follows: $\max(b + k\bar{b})$, subject to $\bar{b} \geq kb$. The $k > 1$ case is also accommodated if we change the formulation to $\max(b + k\bar{b})$, subject to

$$(1 - k)\bar{b} \geq (1 - k)kb \tag{9}$$

For $k = 1$, the last inequality must be replaced by $b = \bar{b}$.

A further set of generalizations is possible. One of the most important is to let signal period (cycle length) and speed be variables. Each will be constrained by upper and lower limits. In addition, changes in speed from one street segment to the next can be limited. Let

T = cycle length (signal period) (s);
z = 1/T = signal frequency (cycles/s);
T_1, T_2 = lower and upper limits on cycle length, i.e., $T_1 \leq T \leq T_2$ (s);
$d(h,i)$ [$\bar{d}(h,i)$] = distance between S_h and S_i outbound (inbound) (m);
d_i = $d(i,i+1)$, $\bar{d}_i$ = $\bar{d}(i,i+1)$;
e_i, f_i ($\bar{e}_i, \bar{f}_i$) = lower and upper limits on outbound (inbound) speed (m/s); and
$1/h_i, 1/g_i$ ($1/\bar{h}_i, 1/\bar{g}_i$) = lower and upper limits on on change in outbound (inbound) reciprocal speed, i.e., $1/h_i \leq (1/v_{i+1}) - (1/v_i) \leq 1/g_i$ $(m/s)^{-1}$.

We are constraining change in speed by putting upper and lower limits on change in reciprocal speed. Although the two are not quite the same, either serves to prevent large, abrupt speed changes. Reciprocal speed is used because it enters linearly in the constraints and can be transformed into t_i. Thus,

$$t_i = (d_i/v_i)z$$
$$\bar{t}_i = (\bar{d}_i/\bar{v}_i)z \tag{10}$$

In the expanded formulation, t_i, $\bar{t}_i$, and z are decision variables that, once known, determine progression speeds.

We add to LP1 all of these generalizations to yield a more versatile mixed-integer linear program.

LP2: find b, $\bar{b}$, z, w_i, t_i, $\bar{t}_i$, δ_i, $\bar{\delta}_i$, and m_i to max $(b + k\bar{b})$, subject to

$$(1 - k)\bar{b} \geq (1 - k)kb \tag{11}$$

$$1/T_2 \leq z \leq 1/T_1 \tag{12}$$

$$\left.\begin{aligned} w_i + b &\leq 1 - r_i \\ \bar{w}_i + \bar{b} &\leq 1 - \bar{r}_i \end{aligned}\right\} \quad i = 1, \ldots, n \tag{13a, 13b}$$

$$(w_i + \bar{w}_i) - (w_{i+1} + \bar{w}_{i+1}) + (t_i + \bar{t}_i) + \delta_i \ell_i - \bar{\delta}_i \bar{\ell}_i - \delta_{i+1}\ell_{i+1} + \bar{\delta}_{i+1}\bar{\ell}_{i+1} - m_i = (r_{i+1} - r_i) + (\bar{\tau}_i + \tau_{i+1}) \quad i = 1, \ldots, n-1 \tag{14}$$

$$\left.\begin{aligned} (d_i/f_i)z &\leq t_i \leq (d_i/e_i)z \\ (\bar{d}_i/\bar{f}_i)z &\leq \bar{t}_i \leq (\bar{d}_i/\bar{e}_i)z \end{aligned}\right\} \quad i = 1, \ldots, n-1 \tag{15a, 15b}$$

$$\left.\begin{aligned} (d_i/h_i)z &\leq (d_i/d_{i+1})t_{i+1} - t_i \leq (d_i/g_i)z \\ (\bar{d}_i/\bar{h}_i)z &\leq (\bar{d}_i/\bar{d}_{i+1})\bar{t}_{i+1} - \bar{t}_i \leq (\bar{d}_i/\bar{g}_i)z \end{aligned}\right\} \quad i = 1, \ldots, n-2 \tag{16a, 16b}$$

$$b, \bar{b}, z, w_i, \bar{w}_i, t_i, \bar{t}_i \geq 0$$

$$m_i = \text{integer}$$

$$\delta_i, \bar{\delta}_i = 0, 1$$

where, if k = 1, Equation 11 is replaced by $\bar{b} = b$.

LP2 involves (11n - 10) constraints and (4n + 1) continuous variables, up to 2n (0 - 1) variables and n - 1 unrestricted integer variables, not counting slack variables. In addition, if the user decides to require or prohibit certain left-turn patterns, constraints on δ_i and $\bar{\delta}_i$ are added up to a maximum of 2n.

LP2 describes how the arterial case is solved. The triangular loop consists of three arteries. Its mathematical program consists of (a) an objective function that is a weighted combination of the objective functions of the individual LP2's (the weights being set by the user to express the relative importance of bandwidth on the three arteries), (b) all the constraints of the individual LP2's, and (c) the loop constraint. The loop constraint is

$$\sum_{(i,j)\in L} \psi_{ij} = n_L \tag{17}$$

where ψ_{ij} is the offset (in cycles) for link (i,j) in loop L and n_L is the loop integer.

Determining Green Splits

One option in MAXBAND is for the user to supply the green splits. As an alternative, the user can provide traffic volume and capacity information and the program will calculate the splits. This is done essentially by using the theory of Webster (13), who has shown that under certain circumstances total delay at an intersection is minimized by dividing the available cycle time among competing streams of traffic proportional to their volumes divided by their capacities.

In MAXBAND, the user who wishes to use this option provides volume and capacity information for the traffic, classified into four through movements and four left-turn movements for each intersection. Let

TRAT(i) = through-traffic ratio of volume to capacity in direction i,
LRAT(i) = left-turn traffic ratio of volume to capacity in direction i, and
i = OUT, IN, OUTC, INC = outbound main, inbound main, outbound cross street, and inbound cross street.

The procedure calculates

MAIN = max[TRAT(OUT) + LRAT(IN), TRAT(IN) + LRAT(OUT)] = the larger of through volume/capacity plus opposite left-turn volume/capacity for the two directions on the main street.
CROSS = max[TRAT(OUTC) + LRAT(OUTC), TRAT(INC) + LRAT(OUTC)] = the same quantity for the cross street.

The basic split between main street and cross street is

MM = MAIN/(MAIN + CROSS) = total time allocated to main street (fraction of cycle).
CC = CROSS/(MAIN + CROSS) = total time allocated to cross street (fraction of cycle).
Let L(OUT) [L(IN)] = outbound (inbound) left split and G(OUT) [G(IN)] = outbound (inbound) through split. Then,
L(OUT) = {LRAT(OUT)/[TRAT(OUT) + LRAT(IN)] x MM.
L(IN) = {LRAT(IN)/[TRAT(OUT) + LRAT(IN)]} x MM.
G(OUT) = MM - L(IN).
G(IN) = MM - L(OUT).

If necessary, these splits are then modified slightly to meet minimum green requirements.

COMPUTER PROGRAM

The computer program consists of (a) a control section, (b) an input section that accepts the problem data from the user, (c) a matrix generator that transforms the input into a form usable by the mixed-integer linear program, (d) a mathematical programming package, and (e) an output routine that interprets the mathematical programming results and prints them out in a form usable by a traffic engineer. The program contains approximately 11 000 lines of FORTRAN IV code, broken down as follows:

Item	No. of Lines
Control program	100
Input program	3200
Matrix generator	3200
MPCODE	2500
Output program	2000

Control Program

The control program manages the overall computation, calling each of the other programs as needed.

Input

Input to the program is on IBM cards or a corresponding card image file on another medium such as magnetic tape or disk. The basic inputs are as follows:

1. Overall problem information--The overall problem information includes a name for the run, an indicator for whether it is a loop problem or an artery problem, an indicator for whether metric or customary units are used, the acceptable range of cycle lengths, and the target ratio for the bandwidths on each artery and their weights, unless these are to be computed from volume information. Usually, default values are used in the mathematical programming package for certain parameters, such as the maximum number of iterations and reinversions, but, if the user wishes, these can be supplied as part of the overall problem information.

2. Network geometry--The order of signals on each artery is given, along with the distances between them (which may be different in each direction) and the names of their intersections. In the case of a loop, the intersection numbers at artery meetings are specified.

3. Green splits or traffic flows and capacities--The user may specify the green splits at each signal as a fraction of the overall cycle time. Alternatively, traffic flows and capacities can be given for each link, including cross streets and turning movements, and the program will calculate green splits by using Webster's formula.

4. Left-turn patterns--Left-turn phases can occur at the beginning or end of the green in either direction, creating four possible patterns for the through direction at each intersection. The user can specify which of the patterns are acceptable, and the program will choose among them to maximize bandwidth.

5. Queue clearance time--Queues may build up during red time as a result of turning movements onto the artery at previous intersections. Such queues may impede the flow of vehicles in the through band. The user may therefore specify at any intersection in either direction a queue clearance time as a fraction of the cycle length. The program will adjust the through band to arrive at the intersection after the queue has cleared and leave the intersection with the queue included as part of the band. In effect, this puts a jog into the through band, advancing it upon leaving the intersection by an amount equal to the queue clearance time.

6. Range of speed--For each link or, if preferred, for the artery as a whole, the user specifies a design speed and a speed tolerance. The program then chooses speeds for each link from this range so as to maximize bandwidth. In addition, the user can constrain the change in speed from one link to the next. If the user does not set limits on speeds and speed changes, default values of 10 percent are used.

More detailed specification of inputs is found in the MAXBAND User's Manual (12).

Matrix Generator

The mathematical program LP2 is a special case of the general linear mixed-integer problem: max cx, subject to

$$Ax \lessgtr b$$

$$x \geq 0$$

$$x_j \text{ integer } j \in J$$

where x is an n-vector of variables whose values are sought, c is an n-vector of objective functions coefficients, A is an (m x n) matrix of coefficients, b is an m-vector of right-hand-side constants, and J is a set of subscripts identifying the variables required to be integer.

The traffic problem described earlier as LP2 requires specific values for A, b, c, and J. These will change as the traffic problem changes. It is necessary to have a generalized program, called the matrix generator, that will take the traffic input data and convert them into the appropriate vectors and matrices for input to the mathematical programming package.

Mathematical Programming Package

A key part of the computer program is the routine for solving the mixed-integer linear program. After several linear programming codes were examined and comparative runs were made on two mixed-integer packages, MPCODE was selected. MPCODE is available in FORTRAN IV source code and is superbly documented in a hard-cover book by Land and Powell (14).

Output

The output of the program is divided into three parts: input cards, a data summary, and a solution report.

The input-card section is a simple printing of the input cards. The data-summary report contains the following information:

1. MPCODE values used by the Land and Powell system;
2. For an artery, (a) general information such as the name of the artery, the number of signals, limits on cycle length, units, target bandwidth ratio, and bandwidth weights; (b) arterywide values such as design speed, tolerances, and limits on changes between links; (c) intersection values, including splits with an indication of their origin, queue clearance times, minimum greens, and the permitted patterns for left turns; (d) link values as actually used, including length, design speed, and speed tolerances; and (e) volumes and capacities on all approaches, when provided.
3. For a loop, (a) general loop information, including upper and lower limits on cycle time and where the arteries meet, and (b) the same artery information described above but for each of the three arteries.

The solution report presents the following data:

1. An indicator for whether the problem has been solved successfully;
2. MPCODE statistics describing the number of iterations, etc., used by the Land and Powell algorithm to solve the problem ("number of solutions" is the number of integer solutions, including the optimal integer solution, found for the problem);
3. For an artery, (a) general information, including name of artery, number of signals, and type of units; (b) cycle time and bandwidths; (c) left patterns selected as optimal; (d) duration and offsets of splits in both fractions of a cycle and seconds; and (e) traversal times and speeds on links;
4. For a loop, (a) loop information, including chosen cycle time, bandwidths, and objective function; (b) same information as for a single artery for each of the three arteries; and (c) repeat of duration and offsets of splits for signals at artery meetings.

Examples of outputs for several test problems are given elsewhere (12).

TESTING

The testing of the program has included runs on a wide variety of problems and operation on several computers.

Table 1 gives run statistics for 10 arterial problems and 3 loop problems. The number of variables, constraints, and integer variables (total and free) relate to the mixed-integer program and represent measures of program difficulty. The number of branch and bound iterations is another measure of how much computation the program required. The number of solutions is the number of feasible integer solutions discovered, up to and including the optimal one. Input data for several of the problems can be found in Little and Kelson (12).

Table 1. MAXBAND performance statistics.

Problem Type	Problem	Problem Characteristics: No. of Signals	No. of Variables	No. of Constraints	No. of Integer Variables: Total	Free	Solution Characteristics: No. of Iterations[a]	No. of Solutions[b]	CPU Time (s)	Cost ($)
Arterial	Broadway, Cambridge, MA	5	25	45	4	4	36	1	2.17	0.73
	Voorhees scenario 1[c]	6	30	56	5	5	668	2	5.81	1.07
	Short version, Waltham artery	6	33	56	8	8	197	3	3.23	0.76
	Voorhees scenario 2[c]	6	36	56	11	11	957	6	7.01	1.23
	Voorhees scenario 2[c], computed splits	6	36	56	11	11	1 457	5	9.74	1.52
	Modified Voorhees scenario 2[c], computed splits	6	42	56	17	17	607	7	6.26	1.35
	Voorhees scenario 1[c], computed splits	6	42	56	17	17	2 089	6	14.08	2.03
	Modified Waltham artery	11	60	116	15	10	1 296	4	17.16	2.44
	Waltham artery	11	60	111	15	15	3 781	7	44.79	5.69
	Wisconsin Avenue, Washington, D.C.	17	88	177	19	19	8 700	12	~210.00	~25.00
Loops	Modified Attleboro loop	4	36	44	5	5	286	3	3.69	0.84
	Attleboro loop	4	37	44	5	5	432	4	4.44	0.92
	FHWA test network	15	93	168	19	16	32 885	15	628.81	74.87

[a]Total number of simplex iterations used. [b]Total number of feasible integer solutions found. [c]Test arterials provided by FHWA.

Figure 3. Eleven-signal test problem: Main Street, Waltham, Massachusetts.

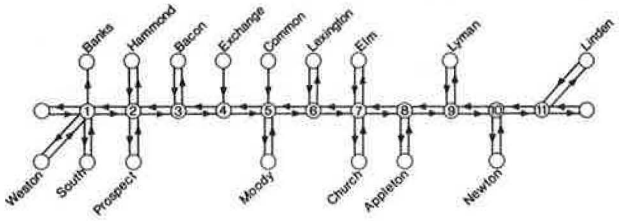


Figure 4. Main Street, Waltham: outbound green band.

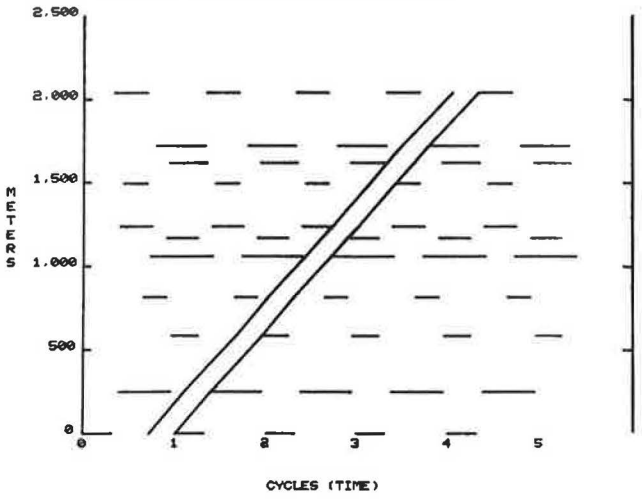
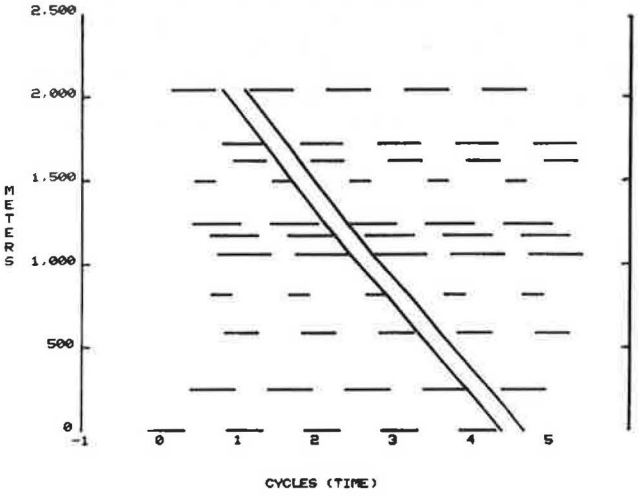


Figure 5. Main Street, Waltham: inbound green band.



The central processing unit (cpu) time is the number of seconds required on the Massachusetts Institute of Technology (MIT) IBM 370-168 and is the primary performance measure. As can be seen, most problems have been solved in a few seconds, although one problem, the 15-signal FHWA test network loop problem, took 10.5 min. The cost shown is the cost charged by the MIT Information Processing Service.

The settings developed by MAXBAND have been put into a traffic simulation program (NETSIM) for several of the test networks with good results. In addition, the MAXBAND output can be used to construct a space-time plot. Figure 3 shows Main Street, Waltham, Massachusetts. Figures 4 and 5 show space-time diagrams for its MAXBAND-generated settings. In Figure 4, note the slight changes in slope resulting from 10 percent permitted variation in design speed. In Figure 5, inbound red times differ from outbound because of left-turn phases.

The MAXBAND program has been transmitted to FHWA on magnetic tape and has been operated on computers to which FHWA has access.

DOCUMENTATION

Little and Kelson (12) provide documentation for MAXBAND in three volumes. Volume 1, the Summary Report, provides an overview of MAXBAND, including complete input and run data on three test networks. Volume 2, the User's Manual, describes the MAXBAND system and how to use it in detail, including worked-out examples of a basic symmetric artery, a basic asymmetric artery, a general artery, and a loop. Volume 3, the Programmer's Manual, gives computer program documentation: First, an overall description of the program is provided, organized by subroutine, and then a listing is given for each subroutine along with a description of what it does. In addition, major variables used in the subroutine are listed. No attempt is made to document MPCODE, since excellent documentation has already been provided by Land and Powell (14).

ACKNOWLEDGMENT

The development of the FORTRAN IV MAXBAND program was supported by a contract from FHWA. S.H. Cohen was contract manager. We also wish to acknowledge the assistance of Matthew Steele during several phases of the effort.

REFERENCES

1. J.T. Morgan and J.D.C. Little. Synchronizing Traffic Signals for Maximal Bandwidth. Operations Research, Vol. 12, 1964, pp. 896-912.
2. J.D.C. Little. The Synchronization of Traffic Signals by Mixed-Integer Linear Programming. Operations Research, Vol. 14, 1966, pp. 568-594.
3. C.J. Messer, R.H. Whitson, C.L. Dudek, and E.J. Romano. A Variable-Sequence Multiphase Progression Optimization Program. HRB, Highway Research Record 445, 1974, pp. 24-33.
4. J.A. Hillier. Appendix to Glasgow's Experiment in Area Traffic Control. Traffic Engineering and Control, Vol. 7, 1966, pp. 569-571.
5. SIGOP: Traffic Signal Optimization Program--A Computer Program to Calculate Optimum Coordination in a Grid Network of Synchronized Traffic Signals. Traffic Research Corp., New York, Sept. 1966. NTIS: PB 173 78.
6. D.I. Robertson. TRANSYT: A Traffic Network Study Tool. Road Research Laboratory, Crowthorne, Berkshire, England, Rept. LR 253, 1969.
7. N.H. Gartner, J.D.C. Little, and H. Gabbay. Optimization of Traffic Signal Settings by Mixed-Integer Linear Programming: Parts I and II. Transportation Science, Vol. 9, Nov. 1975, pp. 321-363.
8. E.B. Lieberman and J.L. Woo. SIGOP II: A New Computer Program to Calculate Optimal Signal Timing Patterns. KLD Associates, Inc., Huntington Station, NY, Jan. 1976.
9. F.A. Wagner, D.L. Gerlough, and F.C. Barnes. Improved Criteria for Traffic Signal Systems on Urban Arterials. NCHRP, Rept. 73, 1969.
10. J.D.C. Little, B.V. Martin, and J.T. Morgan. Synchronizing Traffic Signals for Maximum Bandwidth. HRB, Highway Research Record 118, 1966, pp. 21-47.
11. R.D. Desrosiers and C.H. Leighty. Traffic Flow Responses to Unannounced Increases in Progression Speeds of Signal Systems. Public Roads, Vol. 34, No. 1, April 1966, pp. 1-15.
12. J.D.C. Little and M.D. Kelson. Optimal Signal Timing for Arterial Signal Systems: Volume 1--Summary Report, Volume 2--User's Manual, Volume 3--Programmer's Manual. Federal Highway Administration, U.S. Department of Transportation, April 1980.
13. F.V. Webster. Traffic Signal Settings. Her Majesty's Stationery Office, London, Road Research Tech. Paper 39, 1958.
14. A. Land and S. Powell. FORTRAN Codes for Mathematical Programming. Wiley, London, 1973.