HIGHWAY RESEARCH RECORD

Number 456

Traffic Flow Properties and Theory

6 reports
prepared for the
52nd Annual Meeting

Subject Area

54 Traffic Flow

HIGHWAY RESEARCH BOARD
DIVISION OF ENGINEERING NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING
Washington, D.C. 1973

NOTICE

These papers report research work of the authors that was done at institutions named by the authors. The papers were offered to the Highway Research Board of the National Research Council for publication and are published here in the interest of the dissemination of information from research, one of the major functions of the Highway Research Board.

Before publication, each paper was reviewed by members of the HRB committee named as its sponsor and accepted as objective, useful, and suitable for publication by the National Research Council. The members of the review committee were chosen for recognized scholarly competence and with due consideration for the balance of disciplines appropriate to the subject concerned.

Responsibility for the publication of these reports rests with the sponsoring committee. However, the opinions and conclusions expressed in the reports are those of the individual authors and not necessarily those of the sponsoring committee, the Highway Research Board, or the National Research Council.

Each report is reviewed and processed according to the procedures established and monitored by the Report Review Committee of the National Academy of Sciences. Distribution of the report is approved by the President of the Academy upon satisfactory completion of the review process.

ISBN 0-309-02189-8
Library of Congress Catalog Card No. 73-16610
Price: $2.00

CONTENTS

FOREWORD

The six papers in this RECORD report on research designed to improve our basic understanding of the theory of different traffic flow situations. They will be of interest primarily to researchers and flow theorists and, to a lesser extent, traffic control specialists.

In the first paper, Garner and Parsons describe their work in England, which compared three methods of measuring journey speeds and flows: license matching, moving observer, and arrival output. The authors conclude that the arrival-output method gives more consistent and accurate results than the moving-observer method, is less costly than the license-matching method, and permits measurement of variations in flow over relatively short time intervals. They suggest that it should be much more widely used.

The work by Munjal and Hsu focuses on tests of the validity of three mathematical models developed as a means of describing lane-changing behavior. They use data from aerial photographs to evaluate a linear model proposed by Gazis, Herman, and Weiss, a nonlinear model by Oliver and Lam, and a stochastic model by Worrall, Bullen, and Gur. Validity measurements were made by statistical analysis.

Haefner and Warner present a technique for comparing traffic control alternatives and for measuring changes in performance of a complex traffic system. They analyze the traffic situation as a stochastic process and then present the elements of Markovian decision theory as a control evaluation format. Limited field examples from the merging section of the Baltimore Harbor Tunnel are also presented.

Flow on a two-lane hill was studied by Soyster and Wilson who extended the deterministic model by allowing a probabilistic distribution of concentrations for a given mean value of flow. Hence, instead of two concentrations corresponding to a mean flow rate, they generate a probability distribution that varies with time for a whole range of concentrations.

Seeking ways to further explore traffic dynamics, Lee and Yu set out to establish an acceptable parameter for the internal energy of traffic flow. They analyzed four vehicle-interaction related parameters by using data from aerial photographs and concluded that one of them, the coefficient of variation of speed, is a suitable measure of internal energy.

In the final paper, Athol and Bullen propose a freeway control strategy based on a two-state traffic flow pattern with the primary control parameter being the probability of transition from uncongested flow to congested flow. Their trials of the proposal suggest that it could have direct applicability to existing surveillance and control hardware.

MEASURING JOURNEY SPEEDS AND FLOWS

J. B. Garner, University of Leeds, England; and
D. R. Parsons, Durham County Council, England

The paper sets out to compare three methods of measuring journey speeds and flows. The methods studied are license matching, moving observer, and arrival output. In particular, the results obtained from the moving-observer and arrival-output methods were compared directly with those obtained from the license-matching method, which was taken as the standard method. Observations were taken simultaneously for each of the three methods at five locations. The locations chosen covered highways in urban, suburban, and rural areas and involved highways of varying design standards. A statistical analysis of the results showed that the arrival-output method, which is seldom if ever used, gives far more consistent and accurate results than the more conventional moving-observer method. Moreover, the arrival-output method can measure variations in flow over relatively short time intervals in addition to the usual hourly flows. The paper also shows that the moving-observer method and the arrival-output method involve almost identical cost, whereas the license-matching method is considerably more expensive. It concludes that there is no logical reason why the simple arrival-output method should not be used in preference to the moving-observer method.

•QUANTITATIVE information about road traffic is necessary in order to deal with problems of traffic congestion. For example, vehicle speeds and flows should be known so that an economic assessment of road improvement schemes can be made. Methods must therefore be available whereby these speeds and flows can be determined quickly and accurately on various types of roads for different traffic flow conditions.

The object of this paper is to examine in some detail the relative merits of three methods of measuring journey times (and hence speeds) and volumes over a given length of road. The three methods considered are the (standard) license-matching method, the moving-observer method, and the arrival-output method.

In particular, the paper is concerned with the effectiveness of the arrival-output method, which, as far as the authors are aware, has not previously been compared with the more conventional license-matching and moving-observer methods.

TEST SITES

Four sections of road were chosen for the study so that information could be collected and a comparison made of roads of varying lengths and types. The sites chosen were as follows:

1. City center route—part of the Headrow in the center of Leeds and 330 m in length (Fig. 1a);
2. Radial route—a section of Meanwood Road, Leeds, a little more than 1.5 km in length (Fig. 1b);
3. Ring road route—part of the Leeds Ring Road, 1.41 km in length (Fig. 1c); and
4. Rural route—a section of the M1 motorway from intersection 41 to intersection 43 (at the time of the study, intersection 42 was not open) (Fig. 1d).

MEASUREMENT METHODS

License-Matching Method

Table 1 gives the reported number of license matchings (as recommended by Sawhill and Berry) that have to be made on various types of facilities in order that the mean journey time and speed can be determined with an error of less than 5 percent with a 95 percent degree of confidence.

Because the heavy volumes of traffic at all sites in this study made it impracticable to record all licence numbers, samples were instead selected in accordance with the requirements given in Table 1. To be truly representative, a sample has to be distributed systematically throughout the periods of observation, during which time there is little change in traffic volumes. In this study, the method selected to ensure a systematic distribution was to record all vehicles whose registration number ended with an even digit.

Moving-Observer Method

Flows and speeds are obtained from the following formulas, which were derived by Wardrop and Charlesworth (2):

$$q = \frac{x + y}{t_w + t_a}$$

and

$$t = t_w - \frac{y}{q}$$

where

- q = flow in vehicles per unit time,
- t = mean journey time,
- x = number of vehicles met in the section when observer is traveling against the stream,
- y = number of vehicles that overtake the observer minus the number of vehicles that he overtakes when traveling with the stream,
- t_w = journey time of the observer when traveling with the stream, and
- t_a = journey time of the observer when traveling against the stream.

The mean journey time and, hence, the mean journey speed were first determined for each run of the test car; the overall mean journey speed was then determined.

Arrival-Output Method

This method of gathering speed-flow data is not at all well-known and so will be described here in some detail. It is somewhat similar to the moving-observer method in that a test car carrying an observer with a stopwatch is fed into the stream of traffic; roadside observers with stopwatches are also posted at the start and finish of the test section. The observers with stopwatches are also posted at the start and finish of the test section. The observer in the test car, observer A, records the time taken for the test car to cover the test section. Meanwhile, when the test vehicle enters the test section, the first roadside observer, observer B (Fig. 2), proceeds to count the number of vehicles that pass the starting point during each successive, say, 1-min interval. Immediately after the test vehicle passes the finishing point, the second roadside observer, observer C, proceeds to count the number of vehicles passing that point in each successive 1-min interval.

When the test vehicle has completed the test section, it makes its way back to the starting point. As it passes the beginning of the test section for the second time, traveling in the same direction as for the first run, observer B completes his first run

Figure 1. Test sites: (a) central city route, (b) radial route, (c) ring road route, and (d) rural route.

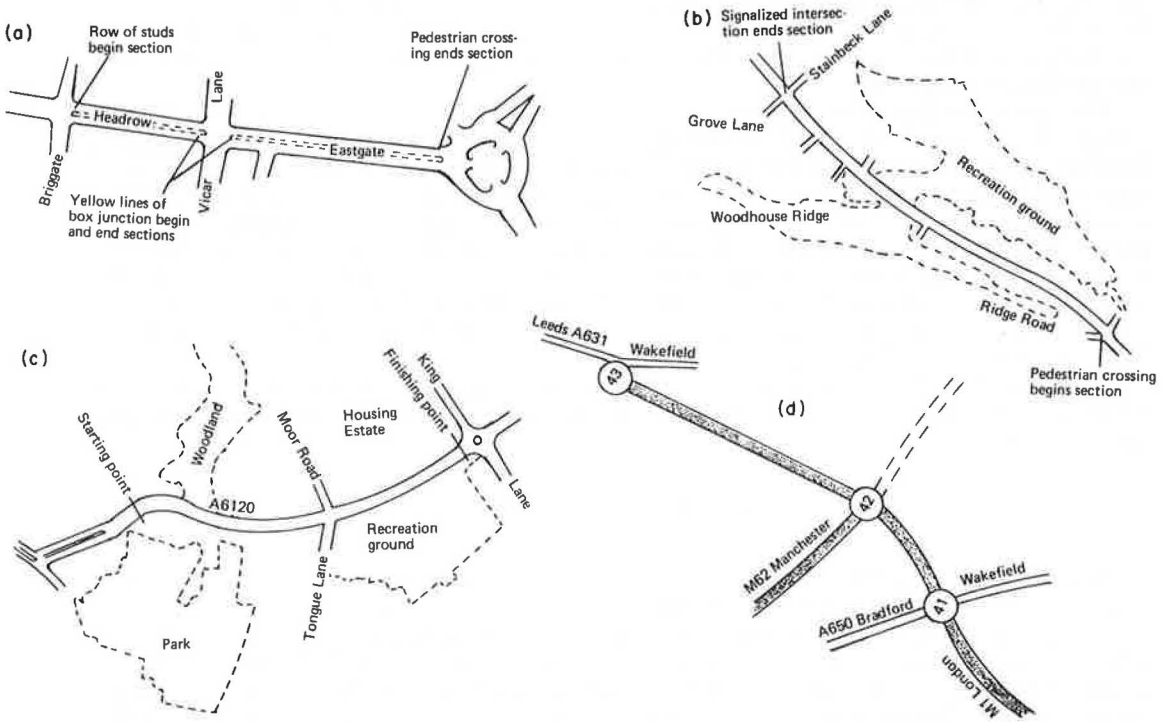
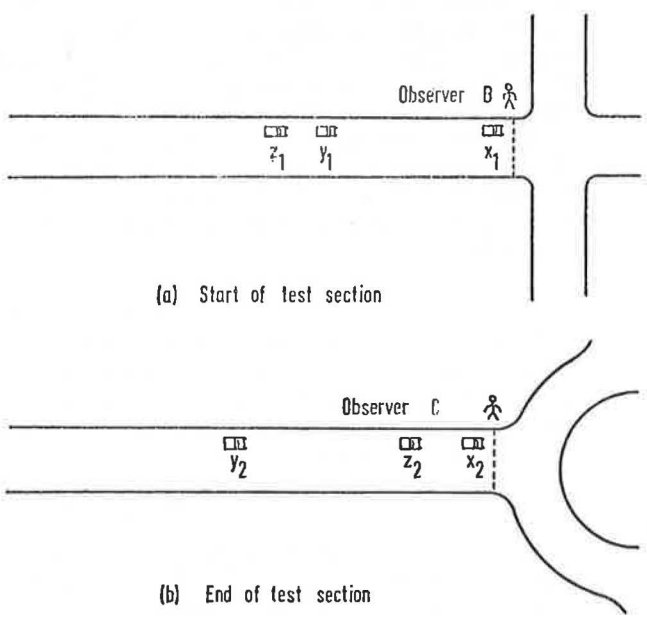


Table 1. Number of license matchings required for various facility types (1).

Location	Type of Facility	Number of Matchings Required
Urban	Signalized, two lane, uncongested	32
	Signalized, two lane, congested	36
	Multilane, uncongested	80
	Multilane, congested	102
Rural	Two lane, up to 1,130 vph	25
	Two lane, up to 1,440 vph	41
	Four lane, uncongested	30

Figure 2. Positioning of cars and observers for moving-observer method.



measurements and proceeds to record data for the second run. Observer A simply records the time taken for each test run, i.e., the time taken for the test vehicle to travel from the starting point to the finishing point of the test section. This he does for each run of the test vehicle. This procedure is repeated until the required number of runs has been completed.

The theory behind the method is as follows. Assume that x is the test vehicle and y and z are two following vehicles and that x_1, y_1, and z_1 in Figure 2a are their relative positions as they pass observer B and x_2, y_2, and z_2 are their positions as they pass observer C (Fig. 2b). Let ty_1 and tz_1 be the headway between the time that the test vehicle crosses the starting point and the time that vehicles y and z cross the same point. Similarly, ty_2 and tz_2 are the corresponding headway times at the finishing point. Let T be the time taken for the test vehicle to cover the test section. Then the time taken for y to cover the test section is $T + ty_2 - ty_1$ and the time taken for z to cover the test section is $T + tz_2 - tz_1$. Therefore, the mean time taken by y and z is $\frac{1}{2}(T + ty_2 - ty_1 + T + tz_2 - tz_1)$.

This may be extended to show that, when there is a continuous stream of n vehicles following the test vehicle, the mean time (T_n) for them to cover the test section is given by

$$T_n = T + \frac{1}{n} \Sigma\, t_2 - \frac{1}{n} \Sigma\, t_1$$

Timing every individual vehicle as it passed the roadside observers would be a very difficult operation, particularly when the traffic flow was heavy. Hence, vehicles are considered in groups that pass the observers in each successive 1-min time interval after the test vehicle. The distribution of these vehicles through the 1-min interval is assumed to be uniform, and hence they are all assumed to pass the observer midway through the 1-min time interval, i.e., the first group of vehicles is assumed to be 30 sec behind the test vehicle, the second group 1.5 min behind the test vehicle, and so on.

If a shorter time interval is chosen for recording the following vehicles, then it is reasonable to assume that more accurate results would be obtained. It would, however, lead to more work in the field and also to more computation in the office. A time interval of 1 min was chosen arbitrarily for this particular study and was found to give satisfactory results. Where there are moderate to heavy flows of traffic, the rate of flow over a short time period will tend to become more constant and any errors introduced will be of a compensating nature.

The mean journey time for the traffic stream is then determined, and, from this, the mean journey speed is easily calculated. In addition to this information, flows and variations in flows over the test period may be determined, inasmuch as the roadside observers are recording the numbers of vehicles passing them during the short time intervals.

DATA COLLECTED

In this study, data were collected simultaneously by each of the three methods to enable a direct comparison of vehicle speeds. In fact, the same test vehicle and test runs were used to gather the moving-observer and arrival-output data, and separate sets of results were recorded for the license-matching method to coincide with each run of the test vehicle.

DISCUSSION OF TESTS

If the license-matching speeds are considered as "standards" (they meet the statistical requirements specified in Table 1), it is interesting to compare them with the speeds determined by the moving-observer and arrival-output methods.

Accuracy

The mean speeds given in Tables 2 and 3 indicate that there is relatively little difference between the values obtained by the three methods. The arrival-output method

gives slightly better results than the moving-observer method, the averages of the differences between the mean speeds obtained from the license-matching method and the moving-observer and arrival-output methods being 2.22 km/h and 1.83 km/h respectively.

There is, however, a much more significant difference between the two sets of results when the individual runs are considered.

Data given in Table 4 show that there is a great reduction in the standard deviations of the differences between the speeds obtained from the individual runs on the license-matching and arrival-output methods as compared with those obtained by considering the license-matching and moving-observer methods.

It is possible to apply an "F-test" to these results and to determine the level of significance of the difference between the two sets of results. The results of the F-test, given in Table 5, show quite clearly that the arrival-output method gives results that are far more consistent than those from the moving-observer method.

Costs

A comparison was also made of the costs incurred in gathering the data by each of the three methods. It was found that the collection and analysis of data from the moving-observer and arrival-output methods involved almost exactly the same amount of work and the same cost.

In contrast, the license-matching method was more than twice as expensive in terms of man-hours of work to gather and analyze the data. It should be noted that this latter method only provides information about journey times and vehicle speeds, whereas the arrival-output and moving-observer methods can also be used to provide information about flows.

Each of the methods investigated measured journey times and vehicle speeds in one direction only. If speeds had been required in both directions on a particular stretch of road, the license-matching method necessarily would have involved a second completely independent survey. The moving-observer method would have required one extra observer in the test vehicle, and the arrival-output method would have required two extra roadside observers. Hence it is clear that, where vehicle speeds are required in both directions, the moving-observer and arrival-output methods involve only a relatively small increase in cost to obtain the information in the field, whereas the cost of collecting the same information using the license-matching method would be doubled.

SOME OTHER COMPARISONS

The arrival-output and moving-observer methods each facilitate measurement of flows as well as mean vehicle speeds. The moving-observer method, however, only gives a series of "spot" flows measured for the duration of each individual run. In the arrival-output method, the number of vehicles passing the roadside observers in successive short intervals of time is recorded; hence, not only is it possible to get a more accurate value of the mean flow over the period of the survey, but also any fluctuations that occur in that flow may be detected.

In a recent paper evaluating the moving-observer method of measuring traffic speeds and flows (3), it was concluded that the method was sensitive to minute-by-minute variations in the traffic stream. It was suggested that these variations would need to be overcome by increasing the length of the test run or by utilizing a greater number of test runs or both. In contrast, not only is the arrival-output method independent on the minute-by-minute variations in the traffic stream, but also it actually detects and measures them.

It was also concluded (3) that, where traffic volumes were low, the number of test runs required by the moving-observer method in order to achieve a given degree of accuracy might be so great as to render the method uneconomical and impractical to use. This is not the case with the arrival-output method: The time-intervals into which observed vehicles are classified need only be reduced in order to maintain the required degree of accuracy. If the volumes are exceptionally low, e.g., less than

Table 2. Speeds, in km/h, on the central city route.

Run No.	Briggate to Vicar Lane			Vicar Lane to Eastgate Roundabout			Briggate to Eastgate Roundabout		
	Moving-Observer	Arrival-Output	License-Matching[a]	Moving-Observer	Arrival-Output	License-Matching[a]	Moving-Observer	Arrival-Output	License-Matching[a]
1	16.09	15.08	16.87 (40)	21.28	27.00	26.70 (36)	17.38	18.12	19.83 (36)
2	—	12.07	15.56 (44)	13.79	18.97	24.46 (45)	7.56	15.06	18.51 (38)
3	8.01	12.38	14.85 (54)	9.64	17.56	18.96 (50)	9.29	15.26	16.90 (39)
4	7.44	7.10	10.91 (38)	27.00	25.07	26.04 (43)	13.60	12.76	13.20 (39)
5	10.36	13.41	13.55 (63)	26.20	19.51	25.83 (79)	16.77	17.12	17.40 (55)
6	11.02	10.06	11.54 (68)	29.02	8.16	12.42 (92)	17.98	8.63	13.24 (66)
7	7.68	7.10	9.72 (85)	4.02	5.94	7.31 (99)	5.17	6.65	8.14 (55)
Mean	10.10	11.03	13.29	18.71	17.46	20.25	12.54	13.37	15.32

[a]Figures in parentheses refer to the number of license matchings made on that particular run.

Table 3. Speeds, in km/h, on radial, ring road, and motorway routes.

Run No.	Radial Route			Ring Road Route			Motorway Route		
	Moving-Observer	Arrival-Output	License-Matching[a]	Moving-Observer	Arrival-Output	License-Matching[a]	Moving-Observer	Arrival-Output	License-Matching[a]
1	36.39	44.56	43.76 (40)	71.39	73.47	69.77 (23)	94.18	94.65	89.72 (44)
2	45.27	47.12	45.67 (45)	73.47	68.49	69.39 (21)	74.67	88.10	89.80 (30)
3	49.99	43.89	43.05 (67)	59.63	64.99	69.72 (23)	86.12	81.35	87.64 (38)
4	49.55	42.58	45.03 (48)	63.36	73.47	72.34 (24)	102.79	84.22	92.26 (38)
5	54.75	40.20	42.91 (41)	74.54	81.75	75.83 (19)	92.36	96.56	99.47 (36)
6	35.26	43.39	46.35 (43)	74.54	56.33	56.81 (45)	93.71	93.71	100.76 (35)
7	43.55	49.55	48.46 (43)	42.60	60.35	61.86 (35)	95.56	91.48	99.88 (30)
8	48.71	50.42	47.48 (39)	61.81	57.60	55.65 (47)			
9				75.08	58.26	57.32 (30)			
10				74.54	59.63	58.02 (38)			
11				69.44	80.47	75.22 (27)			
12				74.54	55.70	53.17 (29)			
Mean	45.43	45.21	45.34	67.91	65.88	64.59	91.34	90.01	94.22

[a]Figures in parentheses refer to the number of license matchings for that particular run.

Table 4. Standard deviations of the differences between speeds obtained on individual runs.

Site	License-Matching Versus Arrival-Output Method (km/h)	License-Matching Versus Moving-Observer Method (km/h)
Central city (Briggate-Vicar Lane)	1.24	2.32
Central city (Vicar Lane-Eastgate)	2.54	9.14
Central city (Briggate-Eastgate)	1.58	5.20
Radial route (Meanwood Road)	2.27	7.68
Ring road route (A6120)	3.04	12.89
Motorway route (M1)	4.76	8.40

Table 5. Results of F-test evaluation of arrival-output and moving-observer methods.

Site	Location	F-Test (percent)
Central city	Briggate to Vicar Lane	90.0
Central city	Vicar Lane to Eastgate	99.5
Central city	Briggate to Eastgate	99.0
Radial route	Meanwood Road	99.5
Ring road route	A6120	99.9
Motorway route	M1	90.0

about 100 vph in a given direction of travel, the actual times of individual vehicles behind the test vehicle may be recorded (rather than the numbers passing in a given time interval).

A final conclusion (3) was that the moving-observer method was best suited to medium and heavy traffic volumes, in which case the number of test runs required to achieve a given degree of accuracy can be relatively small. The tests described here suggest, however, that, even under such traffic conditions, the results obtained from the arrival-output method are far more consistent than those from the moving-observer method, and, hence, even fewer runs of the test vehicle are required to give the same degree of accuracy.

REFERENCES

1. Walker, W. P. Speed and Travel Time Measurement in Urban Areas. HRB Bull. 156, 1957, pp. 27-44.
2. Wardrop, J. C., and Charlesworth, G. A Method of Estimating Speed and Flow of Traffic From a Moving Vehicle. Jour. Institute of Civil Engineers, Vol. 3, No. 1, 1954, pp. 158-171.
3. O'Flaherty, C. A., and Simons, F. An Evaluation of the Moving Observer Method of Measuring Traffic Speeds and Flows. Proc. Fifth Conf. of the Australian Road Research Board, Vol. 5, No. 3, 1970, pp. 40-54.

EXPERIMENTAL VALIDATION OF LANE-CHANGING HYPOTHESES FROM AERIAL DATA

P. K. Munjal and Y. S. Hsu, System Development Corporation, Santa Monica

Lane changing is a very important component in highway traffic flow. Many researchers have recently presented mathematical models to describe lane-changing behavior. This paper focuses on the linear model by Gazis, Herman, and Weiss, the nonlinear model by Oliver and Lam, and the stochastic model by Worrall, Bullen, and Gur. Our objective is to evaluate the validity of these models by using aerial photographic data. Unknown parameters of the linear and nonlinear models, as well as the probability transition matrix of the stochastic model, are estimated by using the experimental data. Some statistical analyses are carried out to measure their validity.

•LANE CHANGING is a very common and complex phenomenon in highway travel. There may be a variety of reasons why a driver changes lanes: driver's lane preference, local traffic concentration, and average speed, to name just a few. It is impossible to model lane changing in mathematical forms that would take into account all causes for a lane change. Even if we could do that, the model would be too complex to have any practical value. This is one of the reasons why we want to study the lane-changing phenomenon in a macroscopic fashion. Another reason is that, even though traffic is a nondeterministic process, we cannot identify each individual driver's behavior. Thus, the best we can do is to study their average behavior.

The objective of this study is to validate and compare the available lane-changing models. There is a definite need to understand the relation between lane-change maneuvers and traffic flow conditions. The results we found may be directly applicable to the development of freeway traffic control strategies.

Several lane-changing studies have been made before. Oliver (7) proposed a theoretical model for lane changing on a two-lane, unidirectional roadway. In his paper, traffic was assumed to behave as a compressible fluid, obeying the equation of continuity.

$$\frac{\partial k_1}{\partial t} + \frac{\partial q_1}{\partial x} = P_{21}(x, t) - P_{12}(x, t)$$

$$\frac{\partial k_2}{\partial t} + \frac{\partial q_2}{\partial x} = P_{12}(x, t) - P_{21}(x, t) \qquad (1)$$

where

k_i = concentration of lane i, i = 1, 2;
q_i = flow of lane i, i = 1, 2;
$P_{12}(x, t)$ = lane-change function that describes transfer of vehicles from lane 1 to lane 2; and
$P_{21}(x, t)$ = lane-change function that describes transfer of vehicles from lane 2 to lane 1.

Furthermore, the lane-changing functions were assumed to satisfy

$$P_{12}(x, t) = \alpha k_1^2(x, t)[k_{2j} - k_2(x, t)]$$

$$P_{21}(x, t) = \beta k_2^2(x, t)[k_{1j} - k_1(x, t)] \tag{2}$$

where α and β are unknown constants to be estimated from experimental data, and k_{1j} and k_{2j} are jam concentrations of lanes 1 and 2 respectively. Experimental results were given in a later paper by Oliver and Lam (8).

A different approach was given by Worrall, Bullen, and Gur (12) where an elementary stochastic model was hypothesized. They made the following assumptions:

1. Lane changes were independent with an equal probability of occurrence for all vehicles; and
2. $X_{ij}^{mt} \geq 0$, if $|i-j| = 1$; $X_{ij}^{mt} = 0$, if $|i-j| \neq 1$ where X_{ij}^{mt} = number of lane changes observed between lanes i and j within subsection m during time t, and

$$Pr(X_{ij}^{mt} = N) = \frac{\exp(-\lambda_{ij}^{m} \times t)(\lambda_{ij}^{m} \times t)^N}{N!} \tag{3}$$

(N = 0, 1, 2, ...) as a Poisson process where λ_{ij}^{m} equals average number of lane changes between lanes i and j within subsection m during unit time and may depend on the flow or density.

It is assumed that the probability of a vehicle changing lanes in section m is a function only of its position in section m-1 and of the lane into which the change is made. The position of the vehicle is proposed as an outcome of a finite Markov process that defines a probability transition matrix T within section m. Specifically,

$$T(m) = \begin{bmatrix} t_{11}(m), & t_{12}(m) \ldots, & t_{1r}(m) \\ t_{21}(m), & t_{22}(m) \ldots, & t_{2r}(m) \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ t_{r1}(m), & t_{r2}(m) \ldots, & t_{rr}(m) \end{bmatrix}$$

for an r-lane highway, where $t_{ij}(m)$ is the probability that a vehicle in lane i in section m-1 will make a lane change to lane j in section m. For simplicity, T(m) is further assumed to be independent of m. The probability transition matrix is to be estimated from experimental data.

The compressible fluid approach was also applied by Gazis, Herman, and Weiss (1) and later extended by Munjal and Pipes (5) to multilane freeway on-ramp perturbation studies. In these studies, the rate of lane changes was hypothesized as

$$\frac{\partial q_1}{\partial x} + \frac{\partial K_1}{\partial t} = a(K_2 - K_1)$$

$$\frac{\partial q_2}{\partial x} + \frac{\partial K_2}{\partial t} = a(K_1 - K_2) \tag{4}$$

for a two-lane uniform unidirectional freeway and as

$$\frac{\partial q_1}{\partial x} + \frac{\partial K_1}{\partial t} = aK_2 - bK_1$$

$$\frac{\partial q_2}{\partial x} + \frac{\partial K_2}{\partial t} = bK_1 - aK_2 \tag{5}$$

for a two-lane non-uniform unidirectional freeway, where K_i is the deviation from the equilibrium concentration in lane i, i = 1, 2. No experimental studies were made for these on- and off-ramp models. However, a similar study was carried out for a freeway lane drop by Munjal and Pipes (6) in which aerial photographic data were used for experimental validation and gave encouraging results.

The work by Levin (3) is concerned with the mathematical modeling of the delay and distance experienced by a vehicle making a lane change by using gap-acceptance concepts.

Although we have mentioned four lane-changing hypotheses, Levin's work (3) is not considered further here because of the complexity of his delay and distance models and the excessive data required for validation. Therefore, we have reduced our study to three models, the linear lane-changing model (Eq. 5), the nonlinear lane-changing model (Eqs. 1 and 2), and the stochastic model (Eq. 3).

The aerial photographic data available from the Federal Highway Administration are of the three-lane Long Island Expressway. Our first task is to extend the two-lane linear and nonlinear models to three-lane models. The unknown parameters of the linear and nonlinear models, as well as the probability transition matrix of the stochastic model, are estimated by using the aerial data. Some statistical analysis is also carried out to provide a quantitative measure of the validity of each model.

DATA ACQUISITION AND REDUCTION

Data were supplied from two sites that were selected to study the traffic flow on grade- and curvature-free multilane freeway sections with no nearby on- and off-ramps. These were the Long Island Expressway (three lanes wide) in New York and the Palisades Interstate Parkway (two lanes wide) in New Jersey. Traffic count studies showed both sites to carry a medium-to-high flow of traffic. The present analysis is carried out for the Long Island Expressway only because it provides more accurately reduced data. Daily 5-min traffic counts were taken for a week to determine a reasonable estimate of different time periods for various constant traffic-flow levels. The Long Island Expressway site is free of access for a distance of 3.2 miles; the westbound direction was chosen for data collection. This section is between the interchanges at Guinea Woods Road and Jericho Turnpike (Fig. 1).

The data were collected by aerial photography. A sequence of 70-mm color photographs was taken at 2-sec intervals with a Maurer 220 pulsed-sequence camera, a lightweight camera designed for aerial reconaissance. A 38-mm Zeiss Biogen wide-angle lens with a relative aperture of f/4.5 was used, allowing filming of about 1 mile of freeway from the helicopter, a Bell 47G3B1, hovering at a constant altitude of 4,000 ft. Magazines of 225 ft were used, which allowed continuous filming of about 30 min of traffic by using the 2-sec frame. Photographs obtained in this manner were projected on a film reader on-line with a computer, and this system permitted an accurate measure of the coordinates of vehicles in the photographic image.

The Benson-Lehner 29E film-reading system containing two crosswires and 10x magnification optics was used to read the x-y coordinates of the vehicles. This information was stored in electronic accumulators in the 282E Telecordex, which was connected to an IBM 1800 computer through a special interface. The data-reduction processing immediately followed the reading of data. There was a real-time feedback to the operator if any rereading of data was required.

Details of the film-reading technique and the associated computer software to develop trajectories of vehicles relative to an actual ground-based coordinate are given by Tashjian and Knobel (9).

Some of the important features are

1. All photographic image points of interest corrected for the optical distortion produced by the combined effect of the aerial camera lens and film magazine,
2. Position and orientation of the aerial camera in ground coordinates determined from camera and ground reference points,
3. Reference points transformed from film coordinates to ground coordinates,
4. Automobile coordinates transformed from film coordinates to ground coordinates,

5. Automobile coordinate points translated into distance and lane position relative to a prespecified ground point,
6. Position and speed of a car in its previous frame predicted from the position and speed of that car in any subsequent frame,
7. Automobile position matched in previous frame to the trajectory (i.e., the predicted position computed during processing of every frame),
8. Car trajectories smoothed for improved estimate of position and speeds, and
9. Data tape generated that contains a car number, associated distances along the road, and corresponding times, speeds, and lane numbers of the car.

The concentration k and the number of lane changes P_{ij} from lane i to lane j are needed for model validations. For reasons to be explained in the next section, we subdivide the film into 3-min time periods. Within each time period, we calculate the flow q, space-mean speed $\bar{v}$, and concentration k. Because the aerial data were taken at an interval of 2 sec, these parameters can be obtained from the following procedures.

Let $R_1, R_2, \ldots R_k$ denote a set of points along the roadway. For each point R_j that a car passes during the filmed period, we will have the following situation. We observe that the car at time t_o is at a distance x_o and at time $t_o + \Delta t$ is at a distance x_1; R_j is between x_o and x_1. Δt is the time interval of the photograph taken, say 2 sec. The relation is shown in Figure 2. Then the time that the vehicle passes R_j is, by linear interpolation, given by

$$t = \frac{R_j - x_o}{x_1 - x_o} \Delta t + t_o \text{ sec} \tag{6}$$

The velocity at R_j is estimated by

$$v = \left(\frac{R_j - x_o}{x_1 - x_o}\right) v_1 + \left(\frac{x_1 - R_j}{x_1 - x_o}\right) v_o \tag{7}$$

where v_1 and v_o are the velocities at x_1 and x_o respectively. Thus, when R_j tends to x_o, v would tend to v_o.

If the number of cars passing R_j in the photographed time period (3 min) is n, then the average velocity (space-mean speed) is

$$\bar{v} = \frac{n}{\sum_{i=1}^{n} \frac{1}{v_i}} \tag{8}$$

and the density is

$$k = q/\bar{v} \tag{9}$$

where $q = n \times 20$ is the hourly flow.

We shall make use of these statistics in the next section.

MODEL VALIDATIONS

Because the aerial data were taken from a three-lane site, we need first to extend the two-lane model to three lanes for a non-uniform roadway. The non-uniformity means that the three lanes have different q-k relations. The model as extended by Munjal and Pipes (5) is

$$\begin{aligned}
(D_t + c_1 D_x)K_1 &= aK_2 - bK_1 \\
(D_t + c_2 D_x)K_2 &= bK_1 - aK_2 + cK_3 - aK_2 \\
(D_t + c_3 D_x)K_3 &= aK_2 - cK_3
\end{aligned} \tag{10}$$

The time interval of 3 min is chosen for calculating concentrations because

1. It is long enough to average out fluctuations over a constant flow period,
2. It contains enough cars to be statistically meaningful, and
3. It is short enough that a sufficient number of intervals are available from the constant flow period.

The first algorithm is used to estimate a, b, and c in Eq. 10 and is as follows:

1. Find $k_1(i)$, $k_2(i)$, $k_3(i)$ for each 3-min interval i, i = 1, 2, . . . , n by using Eqs. 8 and 9, where the subscripts stand for the lane number. Next, compute $K_1(i) = k_1(i) - \bar{k}_1$, where $\bar{k}_1$ is the mean of $k_1(i)$. $K_2(i)$ and $K_3(i)$ are similarly computed.
2. Find the number of net lane changes for each lane and for each time period i, over a 3,200-ft stretch of road section and denote them as $\ell_1(i)$, $\ell_2(i)$, and $\ell_3(i)$.
3. Use a least squares procedure to obtain estimates of a, b, and c. That is, we find a, b, and c that minimize

$$f = \sum_{i=1}^{n} \{[aK_2(i) - bK_1(i) - \ell_1(i)]^2 + [bK_1(i) - aK_2(i) + cK_3(i) - aK_2(i) - \ell_2(i)]^2 + [aK_2(i) - cK_3(i) - \ell_3(i)]^2\}$$

The minimizing values of a, b, and c are denoted by a_o, b_o, and c_o respectively.

The aerial data used are from two films, the first of 849 frames, starting at 9:50 a.m. on August 21, 1969, and the second of 821 frames starting at 5:55 p.m. on August 22, 1969. The numbers of car trajectories are 271 for lane 1, 538 for lane 2, and 584 for lane 3 on the first film and 249 for lane 1, 496 for lane 2, and 562 for lane 3 on the second film. The flow and space-mean speed are 2,953 cars/hour and 84.8 ft/sec for the first period and 2,852 cars/hour and 86.7 ft/sec for the second period. We feel that these two films can be considered to have the same constant flow rate. They both belong to service level B [according to the Highway Capacity Manual (2)]. Net lane changes for lanes 1, 2, and 3 are obtained and are given in Table 1 under the "experimental" columns. Estimates of a, b, and c by using the above data are $a_o = 15.04$, $b_o = -15.84$, and $c_o = -0.39$. These parameters are used in the linear model, and the theoretical net lane changes are computed by

$$L_1 = a_oK_2(i) - b_oK_1(i)$$

$$L_2 = b_oK_1(i) - a_oK_2(i) + c_oK_3(i) - a_oK_2(i)$$

$$L_3 = a_oK_2(i) - c_oK_3(i), \quad i = 1, 2, \ldots, n \tag{11}$$

The theoretical values of L_j (j = 1, 2, 3) are those under the "model" columns in Table 1.

The two-lane model by Oliver and Lam (8) is now extended to a three-lane model resulting in the following form

$$D_tk_1 + D_xq_1 = P_{21} - P_{12}$$

$$D_tk_2 + D_xq_2 = P_{12} - P_{21} + P_{32} - P_{23}$$

$$D_tk_3 + D_xq_3 = P_{23} - P_{32} \tag{12}$$

with

$$P_{12}(i) = \alpha k_1^2(i)[k_{2_j} - k_2(i)]$$

$$P_{21}(i) = \beta k_2^2(i)[k_{1_j} - k_1(i)]$$

Figure 1. Long Island Expressway test site.

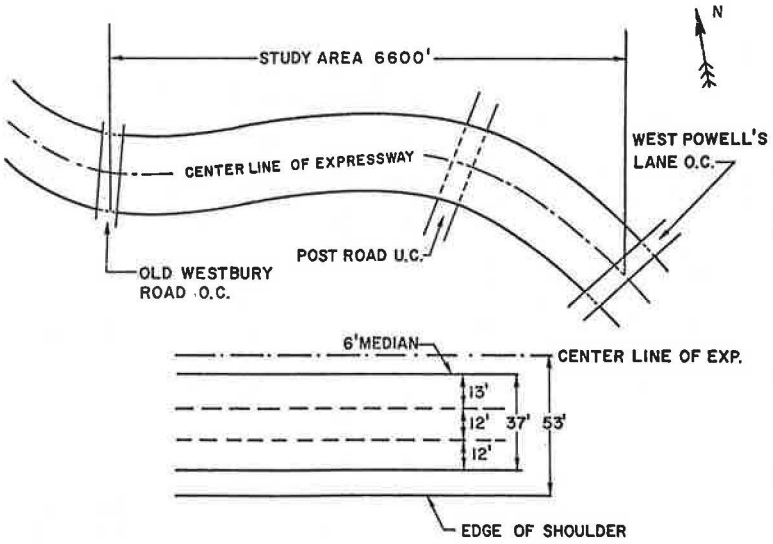


Figure 2. Relation between vehicle position and time.

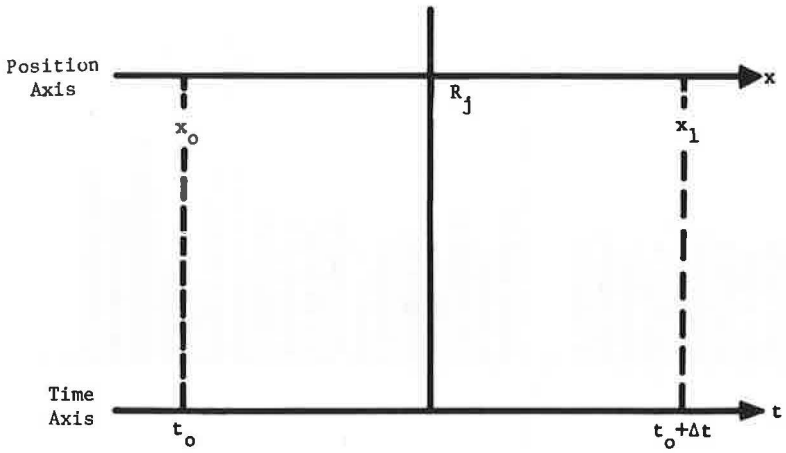


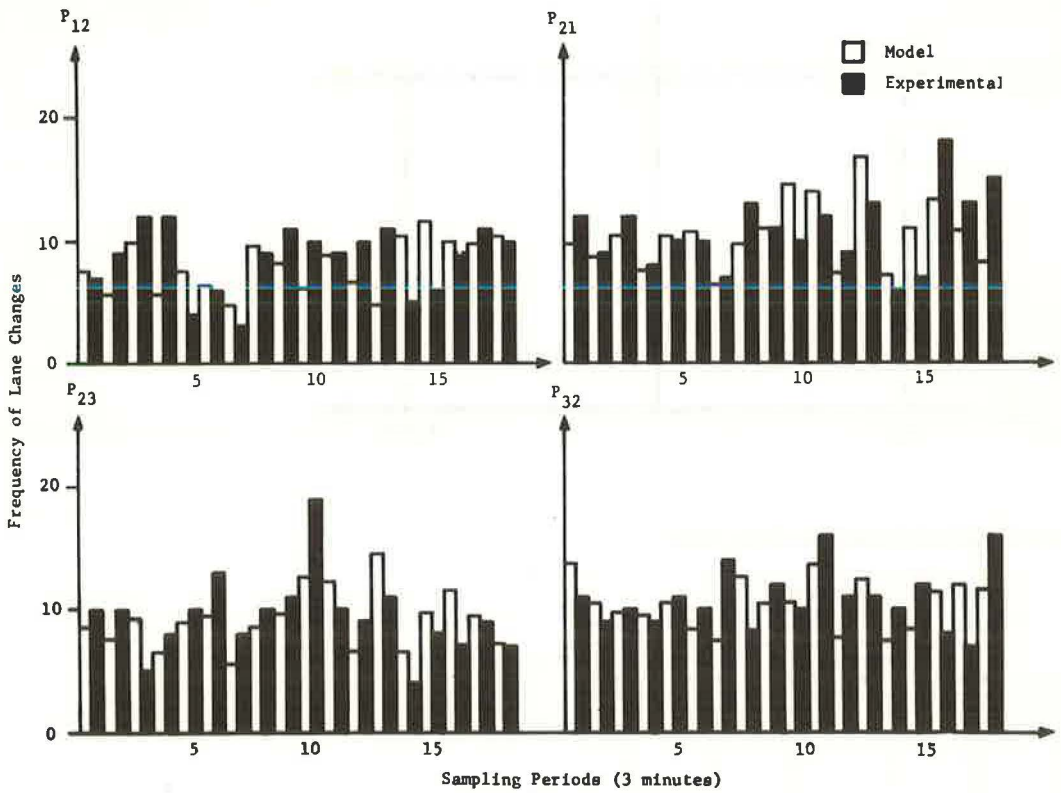
Table 1. Net lane changes from linear model.

	Number of Net Lane Changes					
	Lane 1		Lane 2		Lane 3	
Interval	Model	Experimental	Model	Experimental	Model	Experimental
1	-0.2	5	0.3	-4	-0.1	-1
2	1.5	0	-0.8	-1	-0.7	1
3	0.7	-1	-0.8	6	0.1	-5
4	-2.2	-4	3.5	5	-1.3	-1
5	-0.1	6	0.1	-5	0	-1
6	-0.4	4	0.2	-7	0.2	3
7	-3.1	4	1.1	2	2.0	-6
8	0.4	4	-0.3	-6	-0.1	2
9	0.4	0	-0.7	1	0.3	-1
10	1.1	0	-2.8	-9	1.7	9
11	1.9	3	-3.5	3	1.6	-6
12	-1.8	-1	3.2	3	-1.4	-2
13	1.4	2	-3.9	-2	2.5	0
14	-0.7	1	2.3	5	-1.6	-6
15	1.5	1	-1.8	3	0.3	-4
16	2.0	9	-3.3	-8	1.3	-1
17	0.9	2	-1.2	-4	0.3	2
18	-0.2	5	1.1	4	-0.9	-9
Mean	0.173	2.28	-0.406	-0.83	0.233	-1.45

Table 2. Results of nonlinear model.

	Number of Lane Changes From Lane i to j								Net Gain (Cars) of Lane i					
	P_{12}		P_{21}		P_{23}		P_{32}		L_1		L_2		L_3	
Interval	Model	Experimental	Model	Experimental	Model	Experimental	Model	Experimental	Model	Experimental	Model	Experimental	Model	Experimental
1	7.6	7	9.9	12	8.6	10	13.7	11	2.3	5	2.8	-4	-5.1	-1
2	5.8	9	8.7	9	7.6	10	10.4	9	2.9	0	-0.1	-1	-2.8	1
3	9.9	12	10.4	12	9.2	5	9.9	10	0.5	0	0.2	5	-0.7	-5
4	5.3	12	7.7	8	6.7	8	9.6	9	2.4	-4	0.5	5	-2.9	-1
5	7.5	4	10.3	10	9.0	10	10.6	11	2.8	6	-1.2	-5	-1.6	-1
6	6.3	6	10.8	10	9.5	13	8.3	10	4.5	4	-5.7	-7	1.2	3
7	4.9	3	6.4	7	5.7	8	7.5	14	1.5	4	0.3	2	-1.8	-6
8	9.6	9	9.8	13	8.6	10	12.7	8	0.2	4	3.9	-6	-4.1	2
9	8.2	11	11.0	11	9.7	11	10.3	12	2.8	0	-2.2	1	-0.6	-1
10	6.1	10	14.6	10	12.7	19	10.6	10	8.5	0	-10.6	-9	2.1	9
11	8.9	9	14.0	12	12.2	10	13.8	16	5.1	3	-3.5	3	-1.6	-6
12	6.7	10	7.4	9	6.6	9	7.7	11	0.7	-1	0.4	3	-1.1	-2
13	4.8	11	16.8	13	14.5	11	12.4	11	12.0	2	-14.1	-2	2.1	0
14	10.5	5	7.1	6	6.4	4	7.4	10	-3.4	1	4.4	5	-1.0	-6
15	11.7	6	10.8	7	9.8	8	8.4	12	-0.9	1	-0.5	3	1.4	-4
16	9.9	9	13.3	18	11.7	7	11.4	8	3.4	9	-3.7	-8	0.3	-1
17	9.8	11	10.8	13	9.5	9	12.0	7	1.0	2	1.5	-4	-2.5	2
18	10.3	10	8.1	15	7.2	7	11.6	16	-2.2	5	6.6	4	-4.4	-9
Mean	7.88	8.55	10.44	10.83	9.18	9.38	10.46	10.83	2.56	2.28	-1.28	-0.83	-1.28	-1.45

Figure 3. Lane-change frequencies (nonlinear model).



$$P_{23}(i) = \gamma k_2^2(i)[k_{3j} - k_3(i)]$$

$$P_{32}(i) = \delta k_3^2(i)[k_{2j} - k_2(i)] \tag{13}$$

(i = 1, 2, . . . , n) where

P_{ij} = number of lane changes from lane i to lane j,
k_{ij} = jam concentration, and
α, β, γ, and δ = parameters to be estimated.

The second algorithm that is used in this paper estimates α, β, γ, and δ of Eq. 13 and compares the theoretical and experimental statistics in the following manner:

1. Use the values $k_1(i)$, $k_2(i)$, and $k_3(i)$ calculated from the first algorithm;
2. Find α, β, γ, and δ such that all of the following

$$S_1 = \sum_{i=1}^{n} \{P_{12}(i) - \alpha k_1^2(i)[k_{2j} - k_2(i)]\}^2$$

$$S_2 = \sum_{i=1}^{n} \{P_{21}(i) - \beta k_2^2(i)[k_{1j} - k_1(i)]\}^2$$

$$S_3 = \sum_{i=1}^{n} \{P_{23}(i) - \gamma k_2^2(i)[k_{3j} - k_3(i)]\}^2$$

$$S_4 = \sum_{i=1}^{n} \{P_{32}(i) - \delta k_3^2(i)[k_{2j} - k_2(i)]\}^2 \tag{14}$$

are minimized, and let α_o, β_o, γ_o, and δ_o be the minimizing values of α, β, γ, and δ respectively; and

3. Find theoretical values of P_{ij} by using α_o, β_o, γ_o, and δ_o in Eq. 13, and compare the differences

$$L_1(i) = P_{21}(i) - P_{12}(i)$$

$$L_2(i) = P_{32}(i) - P_{23}(i) + P_{12}(i) - P_{21}(i)$$

$$L_3(i) = P_{23}(i) - P_{32}(i) \tag{15}$$

for i = 1, 2, . . . , n, which are net gains or losses for each lane due to lane changes.

Using the same data as used for the linear model, we summarized the experimental and computed results by using the nonlinear model as given in Table 2 and shown in Figure 3. More information is provided by Table 2 than by Table 1 in that not only is the net gain for each lane due to lane changes recorded (L_1, L_2, L_3) but also each individual lane-changing flow is recorded (P_{12}, P_{23}, P_{32}). If we consider the average behavior of lane changers, i. e., if we look at the mean values of the samples we collected, the nonlinear lane-changing hypothesis seems to be superior to the linear lane-changing hypothesis. This is not surprising because the nonlinear model has more mechanisms than the linear model.

To validate the stochastic model, we divide the road stretch into 16 sections of 200 ft each and calculate the probability distribution of cars in each lane of each road section for the entire filmed period (821 + 849 = 1,670 frames). The probability transition matrix T for sections 1 and 16 is also calculated:

$$T = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix}$$

where T_{ij} is the ratio of the number of lane changes from lane i to lane j to the total flow in lane i. The 200-ft road section is chosen mainly to ensure that $T_{ij} = 0$ for $|i-j| > 1$.

The algorithm proposed by Worrall and Bullen (10) was used to calculate the approximate T, say $\hat{T}$, for the entire road stretch. This $\hat{T}$ is employed for each 3-min interval to obtain the theoretical value of lane changes. That is, we want to find $\hat{T}$ such that

$$\hat{a}_{16} = a_1 \times \hat{T}^{15}$$

is as close to the experimental value of a_{16} as possible, where a_1 and a_{16} are the distributions of vehicles by lane in section 1 and section 16 respectively, and $\hat{a}_{16}$ is the distribution, by lane, of vehicles in section 16 estimated by using the transition matrix $\hat{T}$. Therefore, the probability transition matrix of the entire road stretch (16 sections) is just $R = \hat{T}^{15}$. The estimated R from the algorithm is

$$R = \begin{bmatrix} 0.7235 & 0.2478 & 0.0287 \\ 0.1513 & 0.7038 & 0.1449 \\ 0.0211 & 0.1398 & 0.8391 \end{bmatrix}$$

This transition matrix, R, is used for each 3-min time period for estimating the number of lane changes. Results are given in Table 3 and shown in Figure 4.

STATISTICAL COMPARISONS

A better comparison can be made if we employ some quantitative measure of the validity of each model. The approximate normal statistic u can provide such a measure. We outline the procedure in the following.

1. Calculate

$$e_j(i) = L_j(i) - \hat{L}_j(i), \quad j = 1, 2, 3, \quad i = 1, 2, \ldots, n \tag{16}$$

where j is the lane number, n is the total sample size, and $L_j(i)$ and $\hat{L}_j(i)$ indicate the number of net lane changes for lane j in sample i from the experimental data and computed data, in turn.

2. Obtain

$$\bar{e}_j = \frac{1}{n} \sum_{i=1}^{n} e_j(i) \tag{17}$$

3. Compute

$$s_j = \left\{ \sum_{i=1}^{n} [e_j(i) - \bar{e}_j]^2 \Big/ (n-1) \right\}^{1/2} \tag{18}$$

4. The approximate u-statistic is

Table 3. Results of stochastic model.

	Number of Lane Changes From Lane i to j								Net Gain (Cars) of Lane i					
	P_{12}		P_{21}		P_{23}		P_{32}		L_1		L_2		L_3	
Interval	Model	Experimental	Model	Experimental	Model	Experimental	Model	Experimental	Model	Experimental	Model	Experimental	Model	Experimental
1	7.8	7	9.8	12	9.2	10	11.2	11	2.0	5	0	-4	-2.0	-1
2	6.7	9	9.3	9	8.8	10	10.1	9	2.6	0	-1.3	-1	1.3	1
3	9.0	12	9.9	12	9.7	5	9.6	10	0.9	0	-1.0	5	0.1	-5
4	6.4	12	8.5	8	7.8	8	9.4	9	2.1	-4	-0.5	5	-1.6	-1
5	8.1	4	8.8	10	9.4	10	10.2	11	0.7	6	0.1	-5	-0.8	-1
6	8.0	6	9.7	10	8.8	13	9.1	10	1.7	4	-1.4	-7	-0.3	3
7	6.4	3	8.0	7	7.6	8	8.6	14	1.6	4	-0.6	2	-1.0	-6
8	9.0	9	9.8	13	9.4	10	10.9	8	0.8	4	0.7	-6	-1.5	2
9	8.1	11	9.9	11	9.7	11	9.9	12	1.8	0	-1.6	1	-0.2	-1
10	6.4	10	11.0	10	10.4	19	9.8	10	4.6	0	-5.2	-9	0.6	9
11	8.1	9	11.0	12	10.5	10	11.4	16	2.9	3	-2.0	3	-0.9	-6
12	7.0	10	8.1	9	7.8	9	8.5	11	1.1	-1	-0.4	3	-0.7	-2
13	6.2	11	11.7	13	11.0	11	10.9	11	5.5	2	-5.6	-2	0.1	0
14	9.0	5	8.1	6	8.0	4	8.5	10	-0.9	1	1.4	5	-0.5	-6
15	9.5	6	9.5	7	9.4	8	8.6	12	0	1	-0.8	3	0.8	-4
16	9.0	9	10.5	18	10.3	7	10.3	8	1.5	9	-1.5	-8	0	-1
17	8.7	11	9.9	13	9.5	9	10.8	7	1.2	2	0.1	-4	-1.3	2
18	8.7	10	8.8	15	8.4	7	10.3	16	0.1	5	1.8	4	-1.9	-9
Mean	7.92	8.55	9.6	10.83	9.25	9.38	10.07	10.83	1.68	2.28	-0.99	-0.83	-0.69	-1.45

Figure 4. Lane-change frequencies (stochastic model).

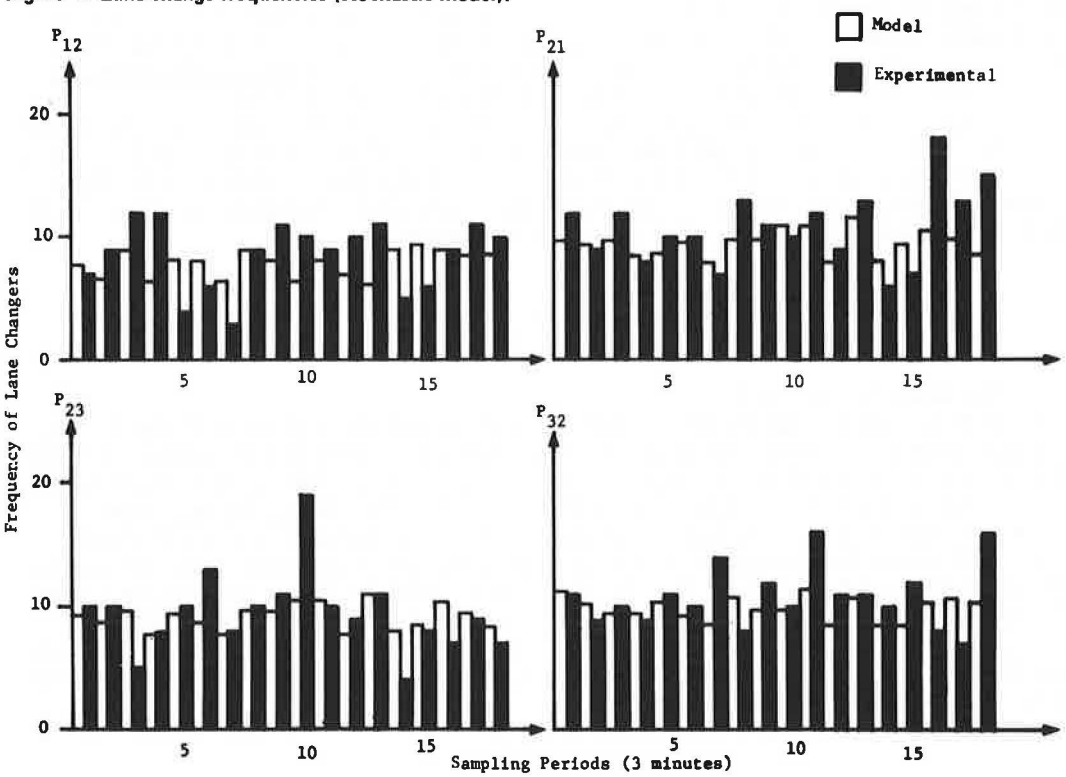


Table 4. u-statistic of lane-changing models.

	Lane 1		Lane 2		Lane 3		Composite Tail
Model	$\|u\|$	α^a	$\|u\|$	α^a	$\|u\|$	α^a	Statistic[b]
Linear	3.09	0.002	0.36	0.719	1.79	0.073	14.325
Nonlinear	0.168	0.867	0.295	0.768	0.182	0.856	1.124
Stochastic	0.677	0.498	0.147	0.883	0.807	0.420	3.378

[a] Represents the exceedance (tail) probability given by the formula $\alpha = \text{prob}\{|u| > u_\alpha\}$.
[b] Using Fisher's Combination-of-Tests Statistic (11)

$$\Lambda = -2 \sum_{1}^{k} \ln \alpha_j \sim \text{chi-square } (2k)$$

where k is the degree of freedom.

$$u_j = \frac{\sqrt{n-1}\,\bar{e}_j}{s_j} \tag{19}$$

The summary of the u-statistics for both models is given in Table 4.

The entries in the last column of Table 4 obey the chi-square distribution with 6 degrees of freedom in accordance with Fisher's Combination-of-Tests Statistic where we have combined the results over all three lanes for each model. Referring to the tabulation of chi-square values, it is seen that an observed value of Λ or greater for the linear model has occurrence probability of less than 0.05, whereas the corresponding values for the nonlinear and stochastic models are approximately 0.98 and 0.75. These results are only approximate, but they do indicate relative ranking. On this basis we rank the three models as the nonlinear model, the stochastic model, and the linear model in this order according to the u-statistics.

The superiority of a model over others can also be viewed by the complexity of the model or the amount of information used in parameter estimation. The nonlinear model clearly has more mechanisms than the linear model. Moreover, the value of jam concentration is used in the nonlinear model, whereas it is not used in the linear model. The stochastic model is a completely different approach and does not hypothesize a density oscillation between lanes. It estimates the frequency of lane changes by fixing the boundary conditions at both ends of the road stretch. Because of this complexity, it is very difficult to evaluate this model against the other two models. However, the extensiveness of data preparation is about the same for all three models.

One shortcoming of the stochastic model is that the algorithm suggested by Worrall and Bullen only gives an approximate solution of the probability transition matrix. This probably explains why, with more parameters to estimate (six for a three-lane highway), it is still not better than the nonlinear model according to the validation results. (It is noted that, although T has nine unknowns, the constraints

$$\sum_{j=1}^{3} t_{ij}, \quad i = 1, 2, 3$$

reduce the unknowns to six.)

We should be more careful here to interpret the u-statistics given in Table 4. Any deviation from the assumptions in which the u-statistic is derived can result in a large value of u, e.g., the non-zero mean of $\bar{e}_j$ in Eq. 17, the mutual dependence of the samples. If corrections can be made, the u-statistics will, in general, be improved.

Both models assume the unknown parameters to be independent of concentration. These parameters have been estimated by first fitting the experimental data by using a constant flow of traffic. How valid this assumption is can be tested by using data from a different constant flow level of traffic and computing the lane changes by using the same parametric values. Available aerial data do not provide enough samples for this kind of study. However, some preliminary inspection suggests that the unknown parameters are concentration-dependent.

CONCLUSIONS

Three lane-changing models, the linear lane-changing model by Gazis, Herman, and Weiss, the nonlinear model by Oliver and Lam as extended here, and the stochastic model by Worrall and Bullen, were selected for experimental validation with aerial photographic data supplied by the Federal Highway Administration. The unknown parameters of the linear and nonlinear models, as well as the probability transition matrix of the stochastic model, were estimated from the data. The number of lane changes was then calculated by using models with estimated parameters. It was found, statistically, that the nonlinear model gave excellent validation results for every lane of the three-lane Long Island Expressway, and the linear and stochastic models gave excellent results only for lane 2. Generally speaking, the three models are ranked as the nonlinear model, the stochastic model, and the linear model.

REFERENCES

1. Gazis, D. C., Herman, R., and Weiss, G. H. Density Oscillations Between Lanes of Multilane Highways. Operations Research, Vol. 10, 1962, pp. 658-667.
2. Highway Capacity Manual—1965. HRB Spec. Rept. 87, 1965.
3. Levin, M. Some Investigations of the Freeway Lane Changing Processes. Civil Engineering Dept., Texas A&M Univ., PhD dissertation, 1970.
4. Munjal, P. K., Hsu, Y. S., and Lawrence, R. L. Analysis and Validation of Lane-Drop Effects on Multi-Lane Freeways. Transportation Research, Vol. 5, 1971, pp. 257-266.
5. Munjal, P. K., and Pipes, L. A. Propagation of On-Ramp Density Waves on Non-Uniform Unidirectional Two-Lane Freeways. System Development Corp., Tech. Memo. TM-3858/017/00, 1969.
6. Munjal, P. K., and Pipes, L. A. Propagation of On-Ramp Density Waves on Uniform Unidirectional Multi-Lane Freeways. Transportation Science, Vol. 5, 1971, pp. 390-402.
7. Oliver, R. M. A Two-Lane Traffic Model. Operations Research Center, Univ. of California, Berkeley, Rept. 64-34, 1965.
8. Oliver, R. M., and Lam, T. Statistical Experiments With a Two-Lane Flow Model. Proc. Third International Symposium on Theory of Traffic Flow, New York, 1965, pp. 170-180.
9. Tashjian, Z. S., and Knobel, H. C. On-Line Processing of Aerial Photographic Data. Univ. of California, Los Angeles, Eng. 7144, 1971.
10. Worrall, R. D., and Bullen, A. G. R. Lane Changing on Multilane Highways. Northwestern Univ., Evanston, Ill., final rept., 1969.
11. Kendall, M. G. The Advanced Theory of Statistics, Vol. II. Hafner Publ. Co., 1951.
12. Worrall, R. D., Bullen, A. G. R., and Gur, Y. An Elementary Stochastic Model of Lane Changing on a Multilane Highway. Highway Research Record 308, 1970, pp. 1-12.

TOWARD A MARKOVIAN TRAFFIC CONTROL EVALUATION SYSTEM

Lonnie E. Haefner*, Washington University; and
John A. Warner III*, R. L. Banks and Associates

The objective of this paper is to present a technique for evaluating traffic control alternatives and resultant performance changes in a complex traffic system. The system can be described as existing in a finite number of states, with some known probability of transition from one state to another in a given time period. Limited field examples from the merging section of the Baltimore Harbor Tunnel are presented and analyzed through a multistage optimization process, termed Markovian decision theory. The technique prescribes an optimal traffic control alternative for each possible state of the system. Changes in flow-density relationships, employment of traffic cones, and a hypothesized metering example are discussed as preliminary tests of the technique.

•MANY TRAFFIC SITUATIONS exist as very complex entities, and there are several ways to improve operation. Control alternatives may be of a permanent type (striping, pretimed versus actuated signalization) or an immediate option within a given control system (metering rate). Such a choice is often based on the criterion of change of flow rate and related to long-term gains of the system and society through amelioration of congestion. The ability to relate an alternative to recognizable long-term gains can be difficult, particularly if the traffic system can operate in a variety of ways at different times.

The objective of this paper is to illustrate preliminary application of a decision algorithm designed to account for these aspects of a complex traffic system by using examples from the merging area of the Baltimore Harbor Tunnel. Through a multistage optimization process over a finite number of recognizable states of the system, the technique prescribes an optimal traffic control alternative for each state. The collection of such optimal alternatives, termed the optimal strategy, maximizes long-run gains to the system through induced changes in the flow-density relationship.

CONCEPTUAL FORM OF THE MODEL

Consider an area of controlled-access highway that requires vehicular traffic to execute merging maneuvers. The state i of the merging area is defined by the level of vehicular density k. There is a probability p_{ij} that can be associated with the transition from one state i to another j. The set of all possible transition probabilities (transition matrix) then describes the behavior of the system. For example, in the four-state (i = 4) case, the transition matrix might be

$$p_{ij} = \begin{bmatrix} 0.6 & 0.3 & 0.1 & 0 \\ 0.2 & 0.6 & 0.2 & 0 \\ 0 & 0.1 & 0.4 & 0.5 \\ 0 & 0.1 & 0.3 & 0.6 \end{bmatrix}$$

*This work was performed while the authors were with the Civil Engineering Department, University of Maryland.

Gain of an Ergodic Process

The gain g of an ergodic process can be found from

$$g = \sum_{i=1}^{N} \pi_i q_i$$

where q_i is the expected immediate return in state i and π_i is the steady-state probability of state i. The gain can be visualized as the return per transition of the process.

For the Baltimore Harbor Tunnel case, we get $\pi = (0.33,\ 0.67)$ and the gain of the process

$$g = \sum_{i=1}^{2} \pi_i q_i$$

which equals 13.55 vehicles/stage. This result yields total rewards through use of the formula

$$v_i(n) = ng + v_i$$

that are in close accordance with those obtained from the recurrence relationship.

It has been demonstrated how the behavior of the merging area (in its present configuration) can be analyzed. A methodology will now be explored for determining the relative worth of alternate configurations.

ALTERNATE MERGING AREA CONFIGURATIONS

Consider, as an example, the addition of a cone line between lanes 2 and 3 (Fig. 3). Assume that, for this alternative, the area's descriptive parameters become

$$p_{ij} = \begin{bmatrix} 0.55 & 0.45 \\ 0.25 & 0.75 \end{bmatrix} \qquad R_{ij} = \begin{bmatrix} 23 & 22 \\ 20 & 9 \end{bmatrix}$$

which yields

$$q = \begin{bmatrix} 22.55 \\ 11.75 \end{bmatrix}$$

and $\pi = (0.357,\ 0.643)$, which gives

$$g = \sum_{i=1}^{2} \pi_i q_i$$

which equals 15.61 vehicles/stage. In this example, we have shown that the hypothetical alternative cone placement would allow a higher rate of vehicular flow.

Up to this point the analysis of permanent changes in the operation of the merging area has been discussed. A methodology for optimizing area traffic behavior through a real-time control procedure is now developed.

ALTERNATE REAL-TIME OPERATING PROCEDURES

Consider, again, the geometric alignment of the area described for the Baltimore Harbor Tunnel. Assume that, in addition to operating the system in this "uncontrolled"

Figure 1. General flow-density relationship.

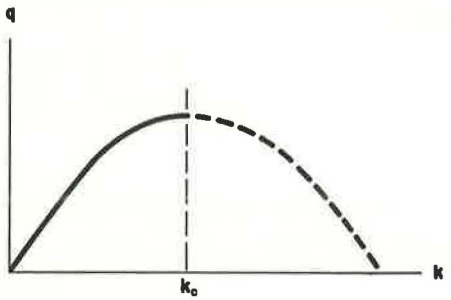


Figure 2. Existing layout of Baltimore Harbor Tunnel merging area.

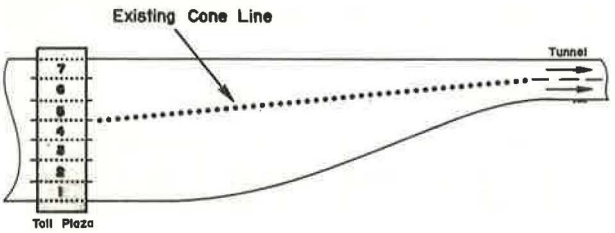
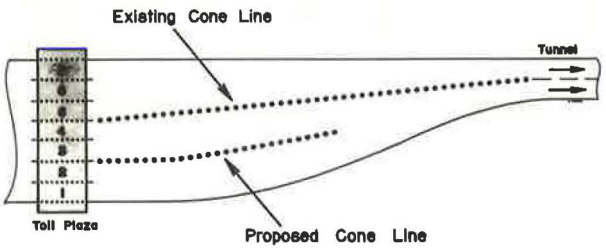


Figure 3. Hypothetical alternative layout of Baltimore Harbor Tunnel merging area.



These transitions may be defined in discrete time (e.g., 15 sec, 30 sec, or 1 min) intervals or as a continuous process (in which case they become transition rates).

Associated with each transition there is a "reward." For purposes of study, the reward is defined as the vehicular flow q measured at the outlet of the merging area. Then a reward matrix (similar in form to the transition matrix) can be constructed, as shown below, in which the elements are changes in q resulting from the system transitions from i to j.

$$R_{ij} = \begin{bmatrix} 10 & 15 & 6 & 3 \\ 12 & 20 & 8 & 5 \\ 8 & 12 & 5 & 4 \\ 6 & 8 & 4 & 2 \end{bmatrix}$$

The general relationship between density and flow is shown in Figure 1 (1, 2).

An objective is to analyze the behavior of this system over time, measured by the number of stages, or state transitions n, that occur. If we assume that the transition matrix is dependent only on the present state of the system, it can be analyzed as a Markov process (3):

$$\sum_{j=1}^{N} p_{ij} = 1$$

and $0 \leq p_{ij} \leq 1$.

Using Markovian decision theory, we can compute various characteristics of the process (4). One characteristic of particular interest is the expected reward from a set of staged decisions, given a starting point in time. In a more sophisticated analysis, it is possible to change both the transition and reward matrices by specifying a set of traffic control alternatives. Then each control alternative has transition and reward matrices associated with it. Given a performance objective (e.g., maximize flow over a period of time), it is possible to define an optimum policy, i.e., the set of optimal traffic control decisions at each stage of the process for each possible state of the system.

If one further assumes that the process is completely ergodic (i.e., after it has been operational for a long time, the probability of the system being in any given state is independent of its starting state), then the long-term average earnings per unit time, defined as the gain, can be found. The optimum policy (set of decisions) is that that maximizes gain.

Application of this optimization technique to a real-time control system in other than the trivial case requires that a decision optimizing immediate return does not necessarily optimize long-range return. It is obvious, if this condition is not satisfied, that the stream of rewards from a series of decision stages could be optimized by simply selecting the decision that optimizes the return at each immediate stage (5).

APPLICATION OF THE MODEL TO THE BALTIMORE HARBOR TUNNEL MERGING AREA

The merging area of the Baltimore Harbor Tunnel (Fig. 2) requires cars to merge into two lanes within a distance of approximately 500 ft. Under present operating discipline, vehicles entering the area in lanes 5, 6, and 7 are separated from those in lanes 1 through 4 by a cone line.

Preliminary data indicate that the behavior of the area can be approximated by the relationship shown in Figure 1, where q is the traffic flow in vehicles per transition stage interval (taken to be 30 sec) and k is the vehicular density in vehicles/ft.

For an elementary two-stage Markovian model, state one exists when k is less than k_c and state two exists when k is greater than or equal to k_c, which results in the following hypothetical matrices:

$$p_{ij} = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix} \qquad R_{ij} = \begin{bmatrix} 24 & 21 \\ 21 & 6 \end{bmatrix}$$

Expected Reward of a Policy

The expected reward $v_i(n)$ from a set of staged decisions (policy), given a starting point i, is defined by the recurrence relationship

$$v_i(n) = \sum_{j=1}^{N} p_{ij} [r_{ij} + v_j(n-1)] \qquad i = 1, 2, \ldots N, \text{ and } n = 1, 2, \ldots$$

By defining q_i, the expected reward from the next stage transition, given the starting state i,

$$q_i = \sum_{j=1}^{N} p_{ij} r_{ij} \qquad i = 1, 2, \ldots N$$

we can write the recurrence relationship in the form

$$v_i(n) = q_i + \sum_{j=1}^{N} p_{ij} v_j (n-1) \qquad i = 1, 2, \ldots N, \text{ and } n = 1, 2, \ldots$$

As an example, suppose our problem contained two states, with matrices

$$R = \begin{bmatrix} 9 & 3 \\ 3 & -7 \end{bmatrix} \qquad p = \begin{bmatrix} 0.5 & 0.5 \\ 0.4 & 0.5 \end{bmatrix}$$

Then, after we have computed

$$q = \begin{bmatrix} 6 \\ -3 \end{bmatrix}$$

the recurrence relationship can be used to derive the following values:

n	$v_1(n)$	$v_2(n)$
0	0	0
1	6	-3
2	7.5	-2.4
3	8.55	-1.44
4	9.555	-0.444
5	10.5555	0.5556

Therefore, using this recurrence method for the Baltimore Harbor Tunnel case gives the following total expected reward:

n	$v_1(n)$	$v_2(n)$
0	0	0
1	22.8	9.0
2	40.08	20.76
3	55.15	33.63
4	69.34	46.93
5	83.17	60.41
6	96.86	73.96

mode, the option exists at each transition point in time to control density by some traffic operations technique (e.g., metering of traffic input). The set of control decisions at each stage is desired.

For example, assume that behavior under the "control" alternative (denoted by the superscript 2) is described by

$$p_{ij} = \begin{bmatrix} 0.9 & 0.1 \\ 0.6 & 0.4 \end{bmatrix} \qquad R_{ij} = \begin{bmatrix} 22 & 21 \\ 21 & 4 \end{bmatrix}$$

The following table then summarizes the pertinent information:

i	a	p^a_{i1}	p^a_{i2}	R^a_{i1}	R^a_{i2}	q^a_i
1 ($k < k_c$)	1	0.6	0.4	24	21	22.8
	2	0.9	0.1	22	21	21.9
2 ($k > k_c$)	1	0.2	0.8	21	6	9.0
	2	0.6	0.4	21	4	14.2

By using the Markovian policy iteration method (see Appendix),[1] we can determine the set of staged decisions that will maximize flow through the area (4). This computation is presented below.

Step 1: Set $v_1 = v_2 = 0$ and enter the policy improvement routine.

Step 2: It will choose the decision that maximizes immediate returns, giving

$$\bar{d} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \bar{p} = \begin{bmatrix} 0.6 & 0.4 \\ 0.6 & 0.4 \end{bmatrix} \qquad \bar{q} = \begin{bmatrix} 22.8 \\ 14.2 \end{bmatrix}$$

Step 3: Entering the value determination routine gives

$$g + v_1 = 22.8 + 0.6v_1 + 0.4v_2$$

$$g + v_2 = 14.2 + 0.6v_1 + 0.4v_2$$

By setting $v_2 = 0$, we get $g = 19.36$, $v_1 = 8.6$, and $v_2 = 0$.

Step 4: Now, applying the policy improvement routine gives

i	a	$q^k_i + \sum_{j=1}^{2} p^k_{ij} v_j$
1	1	22.8 + 0.6(8.6) + 0.4(0) = 27.96
	2	21.9 + 0.9(8.6) + 0.1(0) = 29.64*
2	1	9.0 + 0.2(8.6) + 0.8(0) = 10.72
	2	14.2 + 0.6(8.6) + 0.4(0) = 19.36*

Step 5: This yields

$$\bar{d} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \qquad \bar{p} = \begin{bmatrix} 0.9 & 0.1 \\ 0.6 & 0.4 \end{bmatrix} \qquad \bar{q} = \begin{bmatrix} 21.9 \\ 14.2 \end{bmatrix}$$

Step 6: The value determinations are

$$g + v_1 = 21.9 + 0.9v_1 + 0.1v_2$$

$$g + v_2 = 14.2 + 0.6v_1 + 0.4v_2$$

[1] The original manuscript included an appendix entitled Markov Policy Iteration Method available in Xerox form at the cost of reproduction. When ordering, refer to XS-46, Highway Research Record 456.

Step 7: This yields $g = 20.8$, $v_1 = 11.00$, and $v_2 = 0$.
Step 8: Again, by applying policy improvement, we get

i	a	$q_i^k + \sum_{j=1}^{2} p_{ij}^k v_j$
1	1	$22.8 + 0.6(11.00) + 0.4(0.0) = 29.4$
	2	$21.9 + 0.9(11.00) + 0.1(0.0) = 31.8^*$
2	1	$9.0 + 0.2(11.00) + 0.8(0.0) = 11.2$
	2	$14.2 + 0.6(11.00) + 0.4(0.0) = 20.8^*$

Step 9: This gives

$$\bar{d} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

for the second consecutive time, identifying it as the optimal decision. The optimal policy, then, is to choose alternative 2 (employ the control alternative) at each decision point in time.

CONCLUSIONS

Efficient use of this technique requires data sufficient to permit formation of the transition and reward matrices. Preliminary efforts with twin time-lapse cameras (one recording flow and the other density) indicate that this method will provide suitable results.

One shortcoming of the technique is the fundamental assumption that the system under study operates as an ergodic Markov process, with present-state operation purely a function of the state of the system immediately prior to it and long-run operation independent of initial state. In addition, the appropriate definition of states is worthy of detailed study for each individual problem. An excessive number of states will yield a cumbersome and costly algorithm. Likewise, too few states will obscure the traffic flow phenomena at levels relevant for control considerations. Definition of states should be closely related to the sensitivity of the alternatives' ability to meaningfully alter the flow-density relationship. The choice of transition stage time is also an important consideration. Too lengthy a period might allow several important state changes to occur. Too short a period would result in increased cost of data acquisition and system operation. Preliminary operation of the traffic system discussed indicates that a stage length roughly equal to the time required to travel through the merging area at periods of moderate flow, that of 30 sec, is best. In general, the transition stage time chosen for study should relate to realistic needs for monitoring the system and allow for stabilization of short-term perturbations resulting from employment of a control alternative.

In any system where use of this evaluation approach is considered, the data collection phase should include checking to see how closely the system under study operates as a Markov process. Once this assumption is met, a wide range of geometric and control alternatives can be tested for any traffic situation that can be characterized by a flow-density relationship.

ACKNOWLEDGMENTS

The authors wish to acknowledge E. C. Carter, J. W. Hall, and S. Palaniswamy for their helpful discussion and suggestions on analytic approaches to the problem and F. Arzt for data reduction efforts.

Support from the Federal Highway Administration and the Maryland State Department of Transportation is gratefully acknowledged. The opinions, findings, and conclusions expressed in this paper are those of the authors and not necessarily those of the Maryland State Department of Transportation or the Federal Highway Administration.

REFERENCES

1. An Introduction to Traffic Flow Theory. HRB Spec. Rept. 79, 1964.
2. Drew, D. R. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968.
3. Feller, W. An Introduction to Probability Theory and Its Application. John Wiley and Sons, New York, 1957.
4. Howard, R. A. Dynamic Programming and Markov Processes. M.I.T. Press, 1960.
5. Bellman, R. Dynamic Programming. Princeton Univ. Press, 1957.

A STOCHASTIC MODEL OF FLOW VERSUS CONCENTRATION APPLIED TO TRAFFIC ON HILLS

A. L. Soyster and G. R. Wilson, Pennsylvania State University

In the fundamental relationship between flow and concentration, flow (in vph) increases with concentration (in vehicles/mile) until a critical point is reached. After this critical point, flow decreases to zero as concentration increases to saturation. This is a deterministic model relating flow rate q to concentration k. In this paper this deterministic model is extended by allowing a probabilistic distribution of concentrations for a given mean value of flow. The specific application is to traffic proceeding up a two-lane hill. In this stochastic model, platoons arrive at the bottom of the hill in a Poisson fashion with parameter λ and at the top of the hill in a Poisson fashion with parameter μ. Because the size of platoon at the top of the hill is generally considerably larger than that at the bottom, $\lambda > \mu$. The distributions of platoon sizes at both the bottom and the top of the hill are additional parameters in the formulation. Vehicles on a hill represent a birth and death process where arrival of vehicles at the bottom corresponds to births and arrival of vehicles at the top of the hill corresponds to deaths. Because the lower bound on the number of vehicles is zero and the upper bound is determined from the length of the hill and the length of vehicles, there are a finite number of possible states. These states are incorporated into a finite Markov chain with a transition matrix determined by λ, μ, and the distribution of platoon sizes at the bottom and top of the hill. The transition matrix generates the probability of various concentrations on the hill as a function of the input parameters and time t. Hence, instead of two concentrations corresponding to a mean flow rate, we generate a probability distribution that varies with time for a whole range of concentrations. The Markov process also generates certain dynamic properties of the system such as relative stability. These and other stochastic properties of the Markov process are included to provide an extension of the classical flow-concentration deterministic model.

•ONE of the most complicating features in the analysis of traffic flow is its probabilistic nature. Two identical roadways may have an average flow of 500 vph, but the various parameters that measure the performance of the roadway could be significantly different in any given time interval. Some of these parameters are speed, concentration, number of passes completed and aborted, number of accidents, and percentage of slow-moving vehicles. The actual state of these traffic parameters fluctuates over time, so we must usually be satisfied with measuring an average value and perhaps some extreme values.

Solomon (12) observed that variation in speed from the normal flow of traffic was a leading cause of accidents. Very slow or very fast vehicles are involved in an abnormally high percentage of accidents. This study emphasizes the role and need for a more detailed analysis of traffic flow, in particular the need for a probabilistic model that treats random fluctuations in traffic behavior as a function of time.

It is the purpose of this paper to analyze the stochastic nature of one important traffic parameter, concentration. The classical relationship between flow and concentration is a deterministic one (6, 8): The flow rate increases with concentration until a critical point is reached, after which the flow rate decreases to zero. In this paper the deterministic model embodied in the fundamental diagram of flow and concentration is ex-

tended by allowing a probabilistic distribution of concentrations for a given average value of flow, which is itself a random and changing quantity. Furthermore, the relative stability of traffic flow will be measured by observing the rapidity with which a low concentration is transformed into a high concentration and vice versa. Hence, certain dynamic characteristics of traffic behavior will be presented.

The model to be presented is a Markov birth and death process. Along with the requisite mathematical development, data and results of a field study that tested the feasibility and utility of the model are included. It should be noted, though, that the model developed in this paper is but a tool and not an end in itself. The model will produce, with appropriate input, information about probabilities of various concentrations as a function of time and other factors.

The two-lane hill is a frequently encountered configuration that causes disruption and turbulence in the normal flow on level roadways. It is a physical setting in which concentration of vehicles obviously and directly affects flow rate. Every driver has experienced the agony of heavy traffic proceeding uphill; the speed of the platoon is controlled by the slowest moving vehicle when passing is not permitted or is too risky. As the concentration increases, so does the probability of encountering a slow-moving truck. The two-lane hill will be the physical setting for the stochastic extension of the fundamental diagram of flow and concentration.

BIRTH AND DEATH STOCHASTIC MODELS

The birth and death process is one type of stochastic process in which the time parameter is continuous and state space is discrete. Usually a population of individuals (or things) is considered where the size of the population at time t is X(t). During the interval t to t + Δt, the population may increase (birth) or decrease (death). We will consider the number of vehicles on the hill as our population of individuals; when a vehicle arrives at the bottom of the hill it is a birth, and as a vehicle reaches the crest of the hill it is a death. As t varies, vehicles will enter and depart the hill, i.e., births and deaths, and X(t) will then denote the number of vehicles on the hill at time t. If we then can obtain the probabilistic description of X(t) we have a probabilistic description of the number of vehicles on the hill for any time t.

One of the simplest birth-death processes is the one used to derive the stochastic nature of a single-server queuing system with Poisson arrivals and an exponential service time. In this system the probability of a birth in a small interval Δt is assumed to be proportional to the length of Δt and likewise for a death. Usually λ is the average birth rate, and μ is the average death rate. If we denote the probability of n individuals in the system at time t as $P_n(t)$, then the forward Kolmogorov difference equations are (making standard assumptions)

$$P_n(t + \Delta t) = (1 - \lambda\Delta t - \mu\Delta t)\, P_n(t) + (1 - \lambda\Delta t)\, \mu\Delta t\, P_{n+1}(t) + (1 - \mu\Delta t)\, \lambda\Delta t\, P_{n-1}(t) \qquad n \geq 1$$

$$P_0(t + \Delta t) = (1 - \lambda\Delta t)\, P_0(t) + (1 - \lambda\Delta t)\, \mu\Delta t\, P_1(t)$$

What this system means is that we can be in state n at time t + Δt in three possible ways:

1. Be in state n at time t and have no births and no deaths in the interval Δt,
2. Be in state n + 1 at time t and have no births but one death in the interval Δt, or
3. Be in state n - 1 at time t and have no deaths but one birth in the interval Δt.

The usual procedure is to let Δt become very small and obtain a system of differential-difference equations whose solution determines various properties of the process X(t). A closed-form solution for the time-varying X(t) is very difficult in this example, as in most other models, but the so-called steady-state solution is readily attained and widely known. The steady-state solution is appropriate for large values of t wherein initial transient influences become dampened. In this example the steady-state solution would be

$$P[X = n] = (1 - \lambda/\mu)\, (\lambda/\mu)^n$$

This steady-state distribution is very simple, and as such we can calculate important properties of the birth-death process in the steady state; e.g., the average number of individuals in the system is $\lambda/(\mu - \lambda)$, whereas the expected time in the system for each individual is $1/(\mu - \lambda)$.

Modeling this single-server queuing system mathematically permits both analysis and synthesis of the system. For example, we can predict the change in system characteristics by varying the parameters λ and μ, or we might have an optimization problem in which we want to minimize the expected number of individuals in the system subject to constraints on λ and μ. These same comments of course apply to a valid model of traffic proceeding up a hill. We would like to have a model with which we could perform the following operations (however, we do not deal with such applications specifically in this paper):

1. Predict changes in system characteristics if the passenger or transport arrival rate or both change,
2. Predict changes in system characteristics if physical changes in the roadway are made, and
3. Optimize various objective functions subject to constraints on the parameters.

For example, we might want to predict changes in the probability distribution of concentrations of a hill if heavy transport vehicles increased in density by 25 percent or if the speed capabilities of trucks decreased 15 percent because of heavier loads.

CHARACTERISTICS OF THE TRAFFIC MODEL

In describing a multiple birth and death process that approximates the flow of traffic proceeding uphill on a two-lane highway, we should first see how this model differs from the single-server queuing model. In the single-server queuing model we assume that individuals arrive at a service station, form a waiting line, and wait their turn for service. Only one individual is served at a time. This is not the case in the traffic system, for as soon as a vehicle arrives at the foot of the hill it begins service, i.e., climbing the hill. The traffic model is then a self-service model; we let X(t) be the number of vehicles on the hill at time t. Another difference is that in the single-server queuing model we assume that all arriving customers are homogeneous in the sense that their service distributions are all the same. In the traffic model we have two types of customers or arrivals: transport vehicles and passenger vehicles. The performance of each type of vehicle on the hill is considerably different. A third crucial difference lies in the nature of the arrivals and services: In the queuing model we assumed that individuals arrive in a Poisson fashion and are serviced in an exponential fashion. In the traffic model that is simply not true, for vehicles do not flow freely or randomly on the highway, especially on hills where platoons are formed behind slow-moving vehicles. What happens is vehicles arrive at the foot of the hill in bulk and leave the hill in bulk.

The multiple birth and death process for traffic flow on hills includes features not present in the ordinary queuing model, i.e., self-service, nonhomogeneous vehicles, bulk arrivals, and bulk finishes. $P_{ij}(t)$ is defined as the probability of i cars and j trucks on the hill at time t; the possible states of our system will be pairs of nonnegative integers (i, j). We make two assumptions.

First, the distribution of times between platoon arrivals at both the bottom and the top of the hill is exponential with parameters $1/\lambda$ and $1/\mu$ respectively. Equivalently, platoon arrivals at the bottom and the top of the hill can be shown to be Poisson with parameters λ and μ respectively. The measurement of time between platoon arrivals by our own convention shall be from the front of the lead car of a platoon to the front of the lead car of the next platoon.

Second, given a platoon arrival, the change in the state of the hill can be by more than one vehicle. Thus, a discrete distribution A_{ij} gives the probability of i cars and j trucks arriving in a platoon at the bottom of the hill for all combinations of i and j.

$$\sum_{ij} A_{ij} = 1 \qquad A_{ij} \geq 0$$

Similarly, a discrete distribution f_{ij} gives the probability of i cars and j trucks arriving in a platoon at the top of the hill and departing the system for all combinations of i and j.

$$\sum_{ij} f_{ij} = 1 \qquad f_{ij} \geq 0$$

For example we may have

$$f_{10} = 0.1$$
$$f_{01} = 0.1$$
$$f_{11} = 0.2$$
$$f_{2,1} = 0.3$$
$$f_{3,1} = 0.3$$

Given that a platoon arrives at the crest and that the state of the hill is (1,1), we must use conditional probability to find the correct probabilities of f_{10}, f_{01}, f_{11}; the modified distribution would be

$$f_{10} = \frac{0.1}{0.1 + 0.1 + 0.2} = 0.25$$
$$f_{01} = 0.25$$
$$f_{11} = 0.50$$

The aforementioned process should really not be called a birth and death process, for the standard terminology of a birth and death process requires that, in a small interval of time Δt, only a single birth or death has positive probability. (Rosenshine of Pennsylvania State University suggested the name multiple birth and death.) Finally we must observe that in the queuing model presented earlier it was implicitly assumed that one could have any number of individuals in the system. Certainly this is not the case on the hill where there are physical limitations due to the actual length of the roadway and corresponding lengths of vehicles. Assume then that N and M are upper limits to the number of cars and trucks on the hill; if the state of the system is at some point N and M, then no more vehicles can enter the hill until some vehicles in the system leave. Hence the process under consideration will have two reflecting barriers: the state (0,0) and the state (N,M) where N + M = Q. Q is the maximum number of vehicles that can be physically present on the hill at any one time when we consider the average length of cars and trucks and make plausible assumptions about the proportion of each present. However, for the sake of illustration we simplify the problem by calling N and M the respective upper limits for cars and trucks present on the hill.

MODEL FORMULATION

We now consider the Kolmogorov differential-difference equations that describe this multiple birth-death process with reflecting barriers. Because the required notation is a bit abstruse for a general model let us first set N = 4 and M = 2. $P_{ij}(t)$ is the probability of i cars and j trucks on the hill at time t where of course for each $t \in (0, \infty)$

$$\sum_{(i,j)} P_{ij}(t) = 1$$

We shall also specify the conditional probability distribution of arrivals and finishes given that an arrival or finish has occurred. Let A_{ij} be the conditional probabilities of i cars and j trucks arriving, given that an arrival has occurred, and f_{ij} be the conditional probabilities of i cars and j trucks finishing the hill (i.e., reaching the crest), given that a finish has occurred. In our example set

$$A_{10} = 0.5 \qquad f_{10} = 0.6$$
$$A_{01} = 0.3 \qquad f_{01} = 0.17$$
$$A_{11} = 0.1 \qquad f_{11} = 0.17$$
$$A_{21} = 0.1 \qquad f_{21} = 0.50$$

(The fact that the possible sets of bulk arrivals and bulk finishes are the same is only coincidental in this example.)

The usual procedure in birth-death processes is to write $P_{ij}(t + \Delta t)$ where Δt is some very small interval in terms of $P_{ij}(t)$. For example, we set i = 2 and j = 2 and consider $P_{22}(t + \Delta t)$; i.e., we want to write an expression for the probability of being in state (2,2) at time $t + \Delta t$. There are the three mutually exclusive and exhaustive ways of being in state (2,2) at time $t + \Delta t$:

1. Be in state (2,2) at time t and have no arrivals or no finishes in Δt,
2. Be in state (i,j) at time t and have (2-i, 2-j) arrivals and no finishes in Δt, or
3. Be in state (i,j) at time t and have (i-2, j-2) finishes and no arrivals in Δt.

When t is very small we cannot have both an arrival and a finish in t time so that no other possibilities are available. Hence, when t is sufficiently small $P_{22}(t + \Delta t)$ is approximately equal to the sum of the following three expressions:

1. Prob—no arrivals or finishes in Δt and system in state (2,2) at time t,
2. $\sum_{(i,j)}$ Prob—(2-i, 2-j) vehicles arrive and no finishes in Δt and the system is in state (i,j) at time t, and
3. $\sum_{(i,j)}$ Prob—(i-2, j-2) vehicles finish and no arrivals in Δt and the system is in state (i,j) at time t.

We can now write in more classical terminology the Kolmogorov equations where we have utilized the independence of the probability of being in a given state and the event of arrivals and finishes (except at boundaries) plus the fact that the Prob [2-i, 2-j) arrivals] = Prob [(2-i, 2-j) arrivals | an arrival] × Prob [an arrival]. $P_{22}(t + \Delta t) = (1 - \lambda\Delta t)(1 - \mu\Delta t)P_{22}(t) + (1 - \mu\Delta t)\lambda\Delta t\,[A_{21}P_{01}(t) + A_{11}P_{11}(t) + A_{01}P_{21}(t) + A_{10} \times P_{12}(t)] + (1 - \lambda\Delta t)\,\mu\Delta t\,[f_{10}P_{32}(t)]$. Rearranging terms, dividing by Δt, and neglecting terms on the order $(\Delta t)^2$ yield

$$\frac{P_{22}(t + \Delta t) - P_{22}(t)}{\Delta t} = (-\lambda - \mu)\,P_{22}(t) + \lambda\,[A_{21}\,P_{01}(t) + A_{11}\,P_{11}(t) + A_{01}\,P_{21}(t) + A_{10} \times P_{12}(t)] + \mu f_{10}\,P_{32}(t)$$

Now we let $\Delta t \to 0$ and on the L.H.S. we have by definition $P'_{22}(t)$, i.e., the derivative of $P'_{22}(t)$. In this example we have 15 possible states so that employing the same limiting procedure to each of the 15 possible states would yield a linear system of 15 homogeneous differential equations of the first degree; the system would have the following simple form:

$$P'(t) = A\,P(t)$$

where A is a 15 × 15 matrix and not a function of t.

One of the easiest methods of obtaining the 15 forward Kolmogorov equations is to write the approximate probabilities of moving from state (i,j) at time t to state (i',j') at time $t + \Delta t$. This is shown in detail in Figure 1, where for brevity we have omitted the Δt associated with each λ and μ and terms on the order $(\Delta t)^2$. The first line in Figure 1 indicates that, if the process is in state (0,0) at time t, it can conceivably be in states [(0,0), (1,0), (0,1), (1,1), (2,1)] at time $t + \Delta t$. In particular the probability that the process will be in one of these states is 1. The forward Kolmogorov equations for state (2,2) were obtained by setting $P_{22}(t + \Delta t)$ equal to the column entries below (2,2) multiplied by their respective states. In fact this is how all the forward Kolmogorov equations could be obtained. Note that in the row associated with state (3,0) at time t are the conditional probabilities with superscripts attached. If the process is in state (3,0) at time t and an arrival occurs, then it is impossible for the arrival to contain two cars and one truck; hence, the probabilities of the other three possibilities must be amended. In our example $A^{(3)}_{10}$, $A^{(3)}_{01}$, and $A^{(3)}_{11}$ are the conditional distributions of arrivals, given that an arrival occurs and that the arrival can be in only one of three states, (1,0), (0,1), or (1,1). In this example

$$A_{10}^{(3)} = \frac{0.5}{0.5 + 0.3 + 0.1}$$

$$A_{01}^{(3)} = \frac{0.3}{0.5 + 0.3 + 0.1}$$

$$A_{11}^{(3)} = \frac{0.1}{0.5 + 0.3 + 0.1}$$

The Kolmogorov equations for this example produce a system of 15 linear differential, first-degree, and homogeneous equations of the form

$$P'(t) = A\,P(t)$$

At this juncture we can proceed in one of two directions:

1. Find the solution of this system of linear differential equations with various initial boundary conditions, or
2. Find the steady-state probabilities by making t very large, i.e., set $P'_{ij}(t) = 0$ and solve the system of 15 linear equations, subject to $\sum_{ij} P_{ij} = 1$.

The form of the solution to the linear system of differential equations in the first direction is

$$P(t) = e^{At}\,P(0)$$

Methods for obtaining this solution and solutions for the second direction are discussed in the next section.

SOLUTION PROCEDURE

For the example in the previous section we had 15 possible concentrations, i.e., N = 4 and M = 2. Suppose that the arrival of platoons at the bottom of the hill is 5/min and at the top of the hill is 3/min so that we set $\lambda = 5$ and $\mu = 3$ in Figure 1. The conditional distributions $\{A_{ij}\}$ and $\{f_{ij}\}$ will be the same as those given earlier. The average flow rate of platoons at the bottom of the hill is 5/min, but of course in some minutes there may be 0 platoons and in other minutes 10 platoons moving through the system in such a way that the limits N and M are not violated.

$B_{(0,0)}(t)$ is a 15-component vector that represents the probability of being in each of the 15 states at time t given that at time 0 the state was (0,0). Of course, for different values of t these probabilities are different, but the sum of the 15 components is 1. To obtain $B_{(0,0)}(t)$ we must solve the following set of 15 linear differential equations:

$$B'_{(0,0)}(t) = A\,B_{(0,0)}(t)$$

where A is the 15 × 15 matrix in Figure 1 with $\lambda = 5$ and $\mu = 3$. The general form of such systems is given by

$$B_{(0,0)}(t) = e^{At}\,B_{(0,0)}(0)$$

where $B_{(0,0)}(0)$ is the boundary condition and represents a starting state of (0,0) at time 0. The solution procedure is to obtain the matrix e^{At}; there are two well-known procedures for deriving this matrix (2, 10). The first relies on obtaining 15 distinct eigenvalues of the matrix A so that A can be diagonalized, and the second is simply a series expansion of the matrix e^{At}. We have chosen the second method for computational expedience.

The series expansion for the matrix e^{At} is

$$e^{At} = I + At + \frac{(At)^2}{2!} + \frac{(At)^3}{3!} + \ldots$$

Figure 1. Transition matrix at t + Δt.

t \ t + Δt	0,0	1,0	2,0	3,0	4,0	0,1	1,1	2,1	3,1	4,1	0,2	1,2	2,2	3,2	4,2
0,0	$1-\lambda$	λA_{10}				λA_{01}	λA_{11}	λA_{21}							
1,0	μ	$1-\lambda-\mu$	λA_{10}				λA_{01}	λA_{11}	λA_{21}						
2,0		μ	$1-\lambda-\mu$	λA_{10}				λA_{01}	λA_{11}	λA_{21}					
3,0			μ	$1-\lambda-\mu$	$\lambda A_{10}^{(3)}$				$\lambda A_{01}^{(3)}$	$\lambda A_{11}^{(3)}$					
4,0				μ	$1-\lambda-\mu$					λ					
0,1						$1-\lambda-\mu$	λA_{10}				λA_{01}	λA_{11}	λA_{21}		
1,1	$\mu f_{11}^{(3)}$	$\mu f_{01}^{(3)}$				$\mu f_{10}^{(3)}$	$1-\lambda-\mu$	λA_{10}				λA_{01}	λA_{11}	λA_{21}	
2,1	μf_{21}	μf_{11}	μf_{01}				μf_{10}	$1-\lambda-\mu$	λA_{10}				λA_{01}	λA_{11}	λA_{21}
3,1		μf_{21}	μf_{11}	μf_{01}				μf_{10}	$1-\lambda-\mu$	$\lambda A_{10}^{(3)}$				$\lambda A_{01}^{(3)}$	$\lambda A_{11}^{(3)}$
4,1			μf_{21}	μf_{11}	μf_{01}				μf_{10}	$1-\lambda-\mu$					λ
0,2						μ					$1-\lambda-\mu$	λ			
1,2						$\mu f_{11}^{(3)}$	$\mu f_{01}^{(3)}$				$\mu f_{10}^{(3)}$	$1-\lambda-\mu$	λ		
2,2						μf_{21}	μf_{11}	μf_{01}				μf_{10}	$1-\lambda-\mu$	λ	
3,2							μf_{21}	μf_{11}	μf_{01}				μf_{10}	$1-\lambda-\mu$	λ
4,2								μf_{21}	μf_{11}	μf_{01}				μf_{10}	$1-\mu$

Figure 2. Transition matrix for illustrative example.

(a)

t \ t + Δt	0,0	1,0	2,0	3,0	4,0	0,1	1,1	2,1	3,1	4,1	0,2	1,2	2,2	3,2	4,2
0,0	.158	.103	.060	.029	.012	.063	.077	.088	.057	.040	.017	.034	.051	.062	.148
1,0	.143	.099	.064	.034	.015	.055	.070	.090	.062	.047	.014	.029	.046	.06;	.171
2,0	.121	.093	.069	.041	.020	.043	.060	.092	.068	.058	.010	.022	.039	.059	.204
3,0	.101	.085	.072	.046	.024	.033	.051	.094	.073	.068	.007	.016	.034	.058	.238
4,0	.091	.078	.068	.043	.023	.029	.047	.097	.075	.070	.006	.014	.033	.060	.265
0,1	.143	.093	.052	.024	.009	.065	.078	.090	.056	.037	.019	.038	.057	.069	.164
1,1	.135	.089	.054	.027	.011	.056	.071	.093	.061	.043	.015	.032	.052	.069	.192
2,1	.123	.087	.059	.031	.014	.047	.063	.096	.066	.051	.012	.023	.044	.065	.217
3,1	.107	.082	.062	.035	.016	.039	.055	.098	.071	.059	.009	.019	.039	.063	.246
4,1	.097	.077	.061	.035	.017	.033	.051	.100	.074	.062	.007	.017	.037	.064	.270
0,2	.134	.082	.044	.020	.008	.066	.078	.091	.055	.035	.021	.092	.064	.077	.184
1,2	.123	.079	.046	.022	.009	.056	.071	.095	.061	.041	.016	.034	.057	.076	.214
2,2	.114	.078	.050	.025	.011	.047	.064	.098	.066	.048	.013	.027	.049	.072	.239
3,2	.104	.076	.053	.027	.012	.039	.056	.101	.071	.054	.009	.021	.042	.069	.264
4,2	.098	.075	.055	.029	.013	.035	.052	.102	.073	.058	.008	.018	.039	.067	.278

(b)

0,0	1,0	2,0	3,0	4,0	0,1	1,1	2,1	3,1	4,1	0,2	1,2	2,2	3,2	4,2
.120	.086	.058	.031	.014	.046	.062	.096	.067	.052	.012	.025	.044	.065	.223

The number of terms required to approximate e^{At} of course depends on the size of t; for large values of t a larger number of terms are required.

When t = 1 min, the probabilities of being in each of the 15 states, given one starts in a particular state, are given in Figure 2. If at t = 0 the state of the hill is (0,0), then the probability that the state of the hill is (0,0) at t = 1 min is 0.1589, whereas the probability that the state of the hill is (4,2) at t = 1 is 0.1481. The probabilities of various concentrations at t = 1 depend quite naturally on the state of the hill at t = 0 (each row in Fig. 2 is different). The probability of being in state (0,0) at t = 1, given that the hill is in state (0,1) at t = 0, is 0.1433. If we want to know the probabilities of various concentrations when t = 2, the matrix e^{At} would be needed where t = 2.

It might be expected that for large values of t the probabilities of various concentrations are independent of the starting states. This indeed is true and is a well-known fact about Markov processes. In this small illustrative example, this steady state was reached when t = 6 min. Figure 2b shows the steady-state probabilities; the probability of being in state (0,0) in the long run is 0.120 regardless of the starting state. The time to reach steady state is determined largely by the number of states possible, and in this example the number is only 15. In a later section a field study is described where the time to reach steady state is nearly 2 hours.

The steady-state probabilities of the various concentrations shown in Figure 2b represent a significant departure from the deterministic information given by an ordinary flow-concentration diagram. Instead of assigning a fixed value of concentration for a fixed flow, the stochastic model assigns a certain probability of various concentrations corresponding to a certain fixed mean flow of vehicles. Of course an average concentration could be computed, but the knowledge of probabilities of certain extreme conditions is available with this model, along with the time-varying behavior of these concentrations.

MEAN FIRST PASSAGE TIMES

The dynamic properties of the multiple birth and death model are in some ways illustrated in the previous section wherein it was shown that the probability distribution of concentrations changed with time until a steady state was reached. In this section a more natural and useful dynamic property is described: How long does it take for a road jam to dissipate to ordinary conditions, or, put another way, how long does it take to move from a high concentration state to a low state or vice versa? How quickly a hill can become clogged with vehicles is a measure of its relative stability.

The stability of traffic concentration on a hill can be analyzed by finding what is called the mean first passage time. The mean first passage time is simply the average number of minutes required to pass from a particular state to some other state. If the mean first passage time from state (0,0) to state (20,10) is relatively short, then the hill can become clogged in a very short span of time.

The mean first passage time from state j to state k m_{jk} is calculated from the following system of linear equations (4):

$$m_{jk} = 1 + \sum_{i \neq k} P_{ji}\, m_{ik} \qquad j \neq k$$

P_{ji} is the probability of a single transition from state j to some intermediate state i. But in our continuous-time Markov process, $P_{ji}(t)$ involves an unknown number of intermediate transitions before state i is reached. However, if we consider a sufficiently small t, the number of transitions between j and i approaches 1, and we can use the above system of equations in a valid fashion to get a good approximation of the transitions from state i to state k. If we multiply the number of transitions, m_{jk} by the time one transition on the average occurs, we get a valid approximation of the mean first passage times. In the interest of brevity, the mathematical development supporting these comments is omitted (Fig. 3).

Figure 3 shows the mean first passage times for the illustrative example when the time interval used in discretizing the Markov process was $\Delta t = 0.0001$. (It should be

noted that we found negligible differences for Δt as large as 0.05.) The mean first passage time from state (0,0) to state (4,2) was 2.09 min, whereas the mean first passage time from (2,1) to (0,0) was 2.00 min. Of course, with a larger number of states, the mean first passage times between the extreme pairs of state would be larger.

FIELD STUDY

In this section we shall describe some results of field studies where the feasibility and the validity of the assumptions for the stochastic model were tested.

The crucial mathematical assumption is that platoons of vehicles arrive at both the bottom and top of the hill in a Poisson fashion (8, 9); nonetheless data were collected from four hills in Centre and Blair Counties in central Pennsylvania. In all cases the hypothesis that platoons of vehicles follow a Poisson flow cannot be rejected at the 0.05 significance level when the classical chi-square goodness-of-fit test is used. It should be noted though that these tests were carried out during daylight hours wherein intercity traffic was only moderate; certain peak periods, such as the 5:00 p. m. rush hour, were not tested.

One particular two-lane hill in Centre County was chosen to be studied extensively. Here we collected data to estimate λ, μ, $\{A_{ij}\}$, and $\{f_{ij}\}$. The particular hill was Penn-144 between the towns of Centre Hall and Pleasant Gap; this hill is about $\frac{1}{2}$ mile long, and the bottom of the hill is in Centre Hall. The data were always collected during clear dry weather and between the hours of 9:00 and 11:30 a. m. Our estimates of the necessary parameters are given in Table 1 and are based on five observations during the months of July and August 1972. The tactical procedures used to collect, transcribe, and analyze the data can be found elsewhere (13). There were some difficult problems such as determining whether vehicles were platooned, and in many instances it was not entirely clear whether a vehicle should be classified as passenger or transport.

The estimates for λ and μ were 1.80 platoons/min and 1.28 platoons/min for this hill. The conditional distribution of platoon sizes is given in Table 1. The upper limits N and M were eventually set at 18 and 5 respectively; this permitted a total of 114 states (19 × 6). These limits do not represent upper bounds on the capacity of the hill; with a $\frac{1}{2}$-mile hill, these limits should be four or five times as large. The difficulty encountered was that, with matrices of the size 500 × 500, certain computational procedures become prohibitively costly; hence, practical limits of 13 cars and five trucks were used initially. Later a simple procedure for avoiding the dimensionality difficulty is discussed.

The probability transition matrix for the 114-state model was computed for increasing values of t. The steady-state distribution was reached for t = 2 hours (Table 2). This compares with the illustrative example where after 6 min the steady state was reached. Note that Table 2 shows that, as the number of trucks increases, correspondingly the number of cars increases. For example, the probability that there are no trucks and 18 cars on the hill at any one time is 0.001; however, the probability of five trucks and 18 cars is much larger at 0.070. The conditional probability of 18 cars given that there are five trucks would be 0.070/0.288 = 0.240.

Table 3 gives a sample of the mean first passage times. Note that the mean first passage time from state (0,0) to (18,5) is 63.84 min but from (18.5) to (0,0) the time is 109.47 min. Hence on the average the time to reach saturation from zero concentration is over twice as long as to go from saturation to zero concentration.

This type of information is, of course, not available from the classical flow-concentration diagram, which specifies the flow rate for a given concentration. Conversely a given flow rate would correspond to two distinct concentrations. If the flow rate were x, then concentrations y and z could generate the flow shown in Figure 4. The stochastic flow-concentration model generates a probability distribution for the various concentrations rather than just two points y and z corresponding to a particular flow x. One might expect that this probability distribution should be some type of bimodal distribution with modes at y and z. To some degree this was in fact found to be true.

Figure 3. Mean first passage times for illustrative example.

	0,0	1,0	2,0	3,0	4,0	0,1	1,1	2,1	3,1	4,1	0,2	1,2	2,2	3,2	4,2
0,0	--	1.08	2.71	6.01	12.7	2.64	1.77	1.35	2.27	3.41	14.1	6.07	3.05	2.31	2.09
1,0	1.43	--	2.02	5.45	12.2	3.50	1.87	1.27	2.01	3.09	15.0	6.48	3.22	2.29	1.93
2,0	2.18	1.32	--	3.91	10.8	4.14	2.49	1.18	1.75	2.48	15.7	7.12	3.49	2.32	1.69
3,0	2.64	1.94	1.62	--	7.54	4.58	2.88	1.39	1.52	1.74	16.3	7.57	3.80	2.38	1.44
4,0	2.78	2.20	2.24	3.34	--	4.67	2.98	1.37	1.76	0.77	16.3	7.62	3.82	2.37	1.08
0,1	1.36	1.75	3.10	6.31	13.0	--	1.36	1.38	2.36	3.50	11.6	4.78	2.39	2.01	1.99
1,1	1.82	1.73	2.97	6.14	12.8	3.09	--	1.01	2.09	3.27	14.2	5.22	2.45	1.32	1.76
2,1	2.00	1.79	2.71	5.81	12.4	3.82	2.25	--	1.52	2.84	15.3	6.77	2.79	1.81	1.49
3,1	2.48	1.84	2.51	5.31	11.8	4.37	2.64	1.22	--	1.94	16.0	7.31	3.56	1.86	1.16
4,1	2.66	2.16	2.41	5.15	11.2	4.52	2.85	1.16	1.70	--	16.1	7.45	3.63	2.17	0.67
0,2	2.03	2.16	3.35	6.50	13.1	1.90	1.76	1.48	2.42	3.54	--	1.92	1.60	1.71	1.87
1,2	2.24	2.21	3.31	6.42	13.0	2.85	1.80	1.34	2.27	3.38	12.5	--	.093	1.32	1.60
2,2	2.34	2.20	3.20	6.28	12.9	3.36	2.06	1.17	2.05	3.14	14.7	6.11	--	0.81	1.24
3,2	2.52	2.23	3.10	6.11	12.7	4.19	2.28	1.02	1.79	2.83	15.6	6.95	3.13	--	0.70
4,2	2.61	2.26	3.02	6.00	12.5	4.43	2.75	0.89	1.67	2.55	16.0	7.34	3.45	1.91	--

Table 1. Field study parameters.

Bottom of Hill		Top of Hill	
Arrivals	Value	Arrivals	Value
A_{10}	0.734	f_{10}	0.635
A_{20}	0.098	f_{20}	0.146
A_{30}	0.016	f_{30}	0.018
A_{40}	0.009	f_{40}	0.018
A_{50}	0.003	f_{70}	0.004
A_{01}	0.095	f_{01}	0.041
A_{11}	0.032	f_{11}	0.055
A_{21}	0.006	f_{21}	0.037
A_{02}	0.006	f_{31}	0.009
		f_{61}	0.009
		f_{71}	0.014
		$f_{2,2}$	0.005
		$f_{6,2}$	0.005
		$f_{12,2}$	0.004

Note: $\lambda = 1.80$, and $\mu = 1.28$.

Table 2. Steady-state probabilities for field study.

Cars	0	1	2	3	4	5
0	0.021	0.009	0.006	0.005	0.003	0.001
1	0.016	0.010	0.007	0.006	0.004	0.002
2	0.013	0.010	0.008	0.007	0.005	0.003
3	0.012	0.010	0.009	0.008	0.005	0.004
4	0.010	0.010	0.009	0.008	0.007	0.005
5	0.009	0.009	0.009	0.009	0.008	0.006
6	0.008	0.009	0.009	0.009	0.008	0.007
7	0.007	0.008	0.009	0.009	0.009	0.008
8	0.006	0.008	0.008	0.009	0.010	0.009
9	0.006	0.007	0.008	0.009	0.010	0.010
10	0.005	0.006	0.008	0.009	0.011	0.011
11	0.005	0.006	0.007	0.009	0.012	0.012
12	0.004	0.005	0.007	0.009	0.012	0.013
13	0.003	0.005	0.006	0.009	0.012	0.015
14	0.003	0.004	0.005	0.008	0.013	0.018
15	0.002	0.003	0.005	0.008	0.013	0.021
16	0.002	0.003	0.004	0.007	0.013	0.027
17	0.001	0.002	0.003	0.006	0.013	0.036
18	0.001	0.002	0.003	0.005	0.011	0.070
Total	0.128	0.126	0.130	0.149	0.179	0.288

Table 3. Mean first passage times for field study, in min.

From	To (0, 0)	To (6, 2)	To (18, 5)
(0, 0)	0.48	59.71	63.84
(8, 0)	84.27	67.61	53.80
(18, 0)	107.5	82.34	26.28
(0, 1)	25.30	57.55	63.51
(8, 1)	84.57	61.85	53.04
(18, 1)	107.53	82.00	22.86
(8, 2)	87.80	49.85	51.46
(18, 2)	107.87	81.90	18.37
(0, 3)	67.78	61.44	60.76
(8, 3)	94.86	65.61	46.77
(18, 3)	108.52	82.25	13.00
(0, 4)	70.60	65.07	60.57
(8, 4)	96.70	72.93	46.78
(18, 4)	109.1	82.88	07.05
(0, 5)	79.06	69.56	59.24
(8, 5)	99.84	77.87	44.26
(18, 5)	109.47	83.51	00.14

Figure 4. Flow-density diagram.

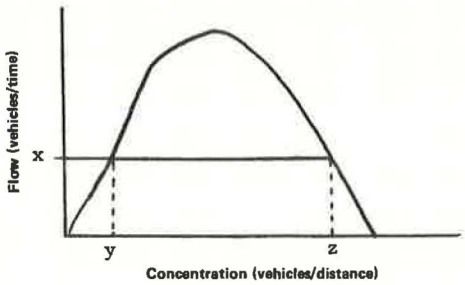
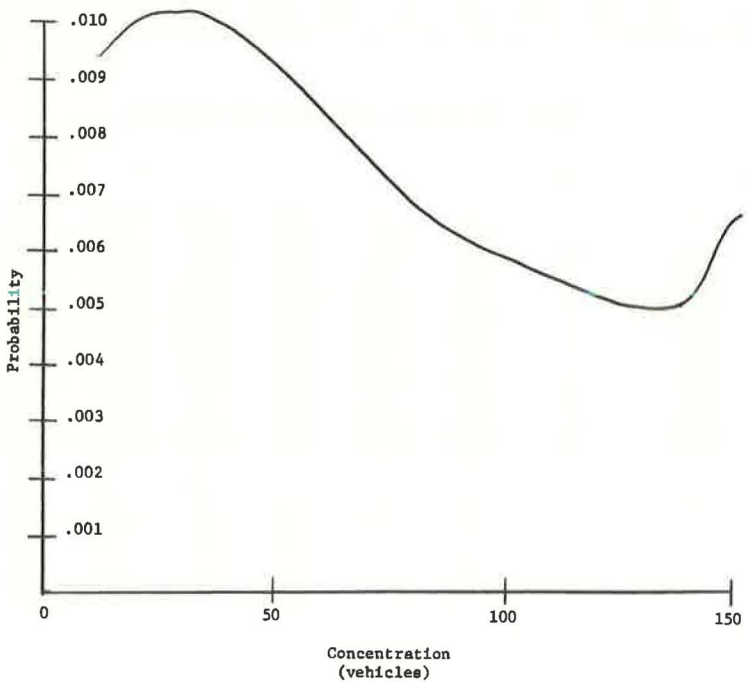


Figure 5. Field study distribution of concentration for total vehicles.



As a demonstration of this phenomenon, the stochastic model used for the Centre Hall Mountain field study was adapted so that cars and trucks were treated identically. Previously, the upper bounds on cars and trucks were 18 and five respectively, which allowed only 23 vehicles on the hill, but this set of bounds requires 114 states. If, instead, all vehicles are treated identically then the model could use 114 as an upper bound. The actual upper bound on the number of vehicles allowed was set at 150. Figure 5 shows the steady-state probabilities in graphical form; the two modes are at state 7 and state 150.

FURTHER CONSIDERATIONS

The stochastic model described in this paper appears to be a valuable tool in the analysis of traffic flow and concentration, although no attempt has been made in this study to apply the methodology to the design and control of roadways. But consideration is now under way for applications in passing safety and speeds and certain other theoretical extensions.

One shortcoming of the model is that the parameters λ and μ are time-homogeneous. Obviously over long enough time spans we should treat λ and μ as functions of time, i.e., $\lambda(t)$ and $\mu(t)$. This extension is certainly feasible, for the only additional difficulty is that the linear differential equations now become time-dependent. The transition matrices should then become even more useful when time intervals over rush hours are included. Furthermore mean first passage times become very important.

ACKNOWLEDGMENTS

We would like to acknowledge the Pennsylvania Transportation and Traffic Safety Center, Pennsylvania State University, for support of this project.

REFERENCES

1. Almond, J., ed. Proceedings, Second International Symposium on the Theory of Road Traffic Flow, London, 1963. Organization for Economic Co-operation and Development, 1965.
2. Bharucha-Reid, A. Elements of the Theory of Markov Processes and Their Applications. McGraw-Hill, 1960.
3. Breiman, L. Probability and Stochastic Processes, With a View Towards Applications. Houghton Mifflin Co., 1969.
4. Clarke, B. A., and Disney, R. L. Probability and Random Processes for Engineers and Scientists. John Wiley and Sons, 1970.
5. Davies, E., ed. Traffic Engineering Practice. E. F. N. Spon, Ltd., 1968.
6. Drew, D. R. Traffic Flow Theory and Control. McGraw-Hill, 1968.
7. Edie, L. C., Herman, R., and Rothery, R., eds. Vehicular Traffic Science. Elsevier Publ. Co., 1967.
8. Haight, F. A. Mathematical Theories of Traffic Flow. Academic Press, 1963.
9. Morse, P. M., and Yaffe, H. J. A Queuing Model for Car Passing. Transportation Science, Vol. 5, No. 1, 1971.
10. Parzen, E. Stochastic Processes. Holden-Day, Inc., 1962.
11. Ross, S. M. Introduction to Probability Models. Academic Press, 1972.
12. Solomon, D. Accidents on Main Rural Highways Related to Speed, Driver and Vehicle. Bureau of Public Roads, 1964.
13. Soyster, A. L. A Multiple Birth and Death Process Applied to Highway Traffic on Hills. Pennsylvania Transportation and Traffic Safety Center, Tech. Note 54, April 1972.
14. Taha, H. A. Operations Research: An Introduction. The Macmillan Co., 1971.
15. Wohl, M., and Martin, B. V. Traffic System Analysis for Engineers and Planners. McGraw-Hill, 1967.

INTERNAL ENERGY OF TRAFFIC FLOWS

Joe Lee, University of Kansas; and
Jason C. Yu, Virginia Polytechnic Institute and State University

The objective of this paper is to establish an acceptable parameter for the internal energy of traffic flow so that further exploration of traffic dynamics can be pursued. Through a boundary condition analysis of traffic flows, it has been found that the currently suggested "acceleration noise" is not a good measure of the internal energy. Results of a theoretical analysis of analogous compressible fluid conditions indicate that, if the kinetic energy of a traffic stream is defined as αku^2, the principle of conservation of energy will not apply. This is because, when density is used instead of total number of vehicles, the system is not confined; thus energy will not be conserved. The compressible fluid analogy further suggests that a term $P_1[(k_1/k_o) - 1]$ may be used to represent the internal energy of traffic flows. Because the accuracy of this compressible fluid analogy is questionable, the term $P_1[(k_1/k_o) - 1]$ is not directly applicable to traffic flows. Instead, an empirical approach is used in the search for a suitable internal energy parameter. Aerial photographic traffic data were used in this effort. Four vehicle-interaction-related parameters were analyzed. One of the parameters tested, the coefficient of variation of speed, not only has exhibited a variational pattern that agrees with that of the $P_1[(k_1/k_o) - 1]$ but also satisfies the boundary condition requirements. It is, therefore, proposed as a suitable measure of the internal energy of traffic flow.

•AN UNDERSTANDING of the dynamics involved in traffic movement is no doubt a basis for design of an efficient and safe highway system. However, fundamentals of traffic dynamics have not been so fully developed as have other physical phenomena such as movement of discrete or continuous masses. One difficulty has been defining energy parameters in the involved macroscopic traffic dynamic system. In an attempt to provide a solution to this problem, a traffic parameter is discussed here that can be used to measure the internal energy of a traffic flow.

Drew (1) introduced the energy concept into traffic flow analysis by considering the traffic stream to be analogous to the flow of a compressible fluid in a constant-area duct. He suggested that a kinetic energy term on the order αku^2 might be used to describe certain properties of a traffic stream inasmuch as a similar term, $\frac{1}{2}\rho V^2$, is defined in fluid mechanics as the kinetic energy of a compressible fluid. In the traffic case, α is a dimensionless constant, k is the density of the traffic stream, and u is the average speed of the stream. Then, by applying the well-known principle of conservation of energy, Drew further suggested that an internal energy term be added to the system to yield an expression for total traffic energy. The proposed relationship may be written as

$$T = E + I \tag{1}$$

where

T = total energy of a traffic stream (constant),
E = kinetic energy of the traffic stream (αku^2), and
I = internal energy of the traffic stream.

In most cases, the kinetic energy of a traffic stream can be easily obtained by measuring the density and average velocity of the stream. The internal energy, however, is thought to be related to the interactions among vehicles in the stream, and it is very difficult to define. Drew has proposed that the parameter "acceleration noise" (3) be used as a measure of internal energy. His proposal was based on two observations. First, the acceleration noise obtained by finding the standard deviation of the acceleration distribution of one vehicle traveling along a stretch of roadway has the same dimensions as kinetic energy. Second, a plot of acceleration noise and αku^2 versus density revealed that the acceleration noise values are generally low when the kinetic energy values are high, thus yielding a near constant value for total energy.

Using acceleration noise as a measure of internal energy, we can rewrite Eq. 1

$$T = \alpha ku^2 + \sigma_t = \text{constant} \tag{2}$$

where σ_t is the derived acceleration noise parameter.

Although this expression represents a significant concept for studying traffic characteristics, it appears to have certain shortcomings.

If the expression $\alpha ku^2 + \sigma_t = \text{constant}$ is applied at the boundary conditions of a traffic stream, certain discrepancies become apparent. Consider first the internal energy term σ_t. According to Drew, σ_t is derived from $\sigma - \sigma_n$ where σ is the measured acceleration noise of a vehicle and σ_n is the natural acceleration noise displayed by the same vehicle subjected to no traffic interference.

For the boundary condition where the density is zero ($k = 0$), the acceleration noise value σ has to equal σ_n by definition. Therefore, $\sigma_t = \sigma - \sigma_n$ would reduce to $\sigma_n - \sigma_n$ or zero. Because there are no vehicles on the road at zero density, the kinetic energy at this point would also be equal to zero. Consequently, the total energy of the traffic stream when $k = 0$ would be $T = E + I = 0$. At the other end of the density domain, jam density ($k = k_j$), all vehicles on the roadway are stopped. Because there is no movement, σ would necessarily be zero. Also, because the idea of a natural acceleration noise makes no sense for such extremely high-density conditions, σ_n is undefined at k_j. Thus, no meaningful value for total energy can be found for the jammed condition by using the proposed definition of internal energy. If the principle of conservation of energy holds true for a traffic stream using the parameters suggested by Drew, σ_t does not seem to represent a good measure of internal energy.

Intuitively, the internal energy of a traffic stream should express the degree to which vehicle interactions exist in the stream. From this point of view, the internal energy should be equal to zero when there are no vehicles on the road and should reach its maximum value when the density is maximum inasmuch as the greatest amount of vehicle interaction can be expected to occur at this point. A parameter that fulfills these boundary conditions is required. If it is assumed that such a parameter, call it I, exists, then the condition for the conservation of energy would be written as

$$T = \alpha ku^2 + I = \text{constant} \tag{3}$$

where $I = 0$ at $k = 0$ and $I = I_{max}$ at $k = k_j$.

To this point in the analysis, it has been assumed that the principle of conservation of energy can be applied to a traffic stream with the suggested parameters. If Eq. 3 is evaluated at the appropriate boundary conditions, however, the following results are obtained:

$$k = 0 \rightarrow E = \alpha ku^2 = 0 \text{ and } I = 0$$

$$\therefore T = E + I = 0$$

$$k = k_j \rightarrow E = \alpha ku^2 = 0 \text{ since } u = 0 \text{ and } I = I_{max}$$

$$\therefore T = E + I = I_{max}$$

Because I_{max} must be greater than 0, the conservation of energy (Eq. 3) does not hold.

From the analysis documented above, two general conclusions can be drawn. First, acceleration noise is not an adequate parameter for representing the internal energy of a traffic stream if the internal energy is defined in terms of vehicular interaction; and, second, if the kinetic energy of the traffic stream is defined as αku^2 and the internal energy is defined in terms of vehicular interaction, which is zero at zero density and a maximum at jam density, the principle of conservation of energy does not apply.

From these conclusions, it is apparent that some modifications must be made in the energy concept if it is to be used in traffic flow conditions.

THEORETICAL INVESTIGATIONS

Energy System of a Traffic Stream

Consider a platoon of n vehicles. At time t_o, assume that these vehicles are spread along a section of roadway at a low density k_o and are moving at an average speed u_o. Due to a disturbance of some sort, the first vehicle slows down and the vehicles start backing up. At time t_i the average speed has dropped to u_i and the density has increased to k_i. If the cause of the disturbance continues to prevail, a complete stoppage of the platoon will eventually occur. At this time a bumper-to-bumper situation will exist, and the density will have reached its maximum value of k_j. This sequence of occurrences is shown in Figure 1.

These conditions can be considered analogous to the system shown in Figure 2. In this system, a bulk of compressible fluid with mass m is moving through a frictionless pipe with unit cross section. The initial conditions are that at time t_o this bulk of fluid is moving at a velocity v_o with density ρ_o and has length l_o. A varying resistant force is introduced into the system. Because of the resistance, the movement of the fluid mass is retarded and the fluid slows, eventually coming to a stop. At the same time, because of the compressive action of the variable force and the inertia of the fluid, the density of the fluid increases and reaches a maximum density ρ_j when the stoppage occurs.

Now suppose that the fluid mass was completely stopped at time t_j and that the average resistant force from time t_o to t_j was measured as P_j. Also assume that the length of the mass at t_j was l_j. In the intermediate condition at time t_i (Fig. 2b), the mass is moving at a velocity v_i, the density is ρ_i, the length of the mass has been reduced from l_o by an amount Δl_i to l_i, and the average resistant force from time t_o to t_i is represented as P_i.

Consider the condition shown in Figure 2a. There is no external force in the system, and the total energy involved is simply equal to the kinetic energy of the moving mass, $\frac{1}{2}mv_o^2$. After the resistance is applied to the system, the speed of the mass is reduced, and part of the kinetic energy is lost and is transferred to another form of energy. In this confined system the only other form of energy possible is that stored in the fluid itself due to the work done by the compressive action of the resistant force and the inertia of the fluid itself. At time t_i the kinetic energy of the fluid has been reduced to $\frac{1}{2}mv_i^2$. The work done to this time by the resistance is equal to the average compressive force P_i times the distance by which the fluid was compressed Δl_i. The total energy at time t_i is then

$$\frac{1}{2}mv_i^2 + P_i \Delta l_i = T \tag{4}$$

When the fluid mass is stopped at time t_j, there is no kinetic energy in the system. The stored energy at this time is equal to $P_j \Delta l_j$ where $\Delta l_j = l_o - l_j$. Hence, the total energy would be

$$P_j \Delta l_j = T \tag{5}$$

With the system confined and no other forces or energy involved, the principle of conservation of energy states that the total energy of the fluid for all three points in time must be equal.

$$\tfrac{1}{2}mv_o^2 = \tfrac{1}{2}mv_1^2 + P_1\Delta l_1 = P_j\Delta l_j \quad (6)$$

Now, if the intermediate condition is taken as a reference, the following general expression can be written:

$$\tfrac{1}{2}mv_1^2 + P_1\Delta l_1 = C \text{ (constant)} \quad (7)$$

Dividing both sides of Eq. 7 by l_1 gives Eq. 8.

$$\tfrac{1}{2}\frac{m}{l_1}v_1^2 + \frac{P_1\Delta l_1}{l_1} = C/l_1 \quad (8)$$

Now, m/l_1 is the density of the fluid mass at time $t_1(\rho_1)$ and the term $(P_1\Delta l_1)/l_1$ is simply the energy stored in a unit section (I_1). Thus another form of Eq. 8 is

$$\tfrac{1}{2}\rho_1 v_1^2 + I_1 = C/l_1 \quad (9)$$

Inasmuch as C/l_1 is not a constant but is a function of l_1, the conclusion extracted from this analysis is that, if the kinetic energy of a compressible fluid is expressed as $\tfrac{1}{2}\rho_1 v_1^2$ and the internal energy is expressed as the energy stored in a unit section of the fluid, then the principle of conservation of energy does not hold because the system is no longer confined; we are not dealing with a certain amount of mass but, instead, the variable mass in a unit volume. From the analogous point of view, if the kinetic energy of a traffic stream is expressed as αku^2, then the energy of the stream will not be conserved, no matter how the internal energy is defined, because the system is no longer confined. This conclusion agrees with the observation made in the previous section from examination of the traffic stream boundary conditions.

Internal Energy of a Traffic Stream

Although it has been demonstrated that the principle of conservation of energy does not hold for a traffic stream when kinetic energy is defined as αku^2, it is thought nevertheless that the concepts of kinetic and internal traffic stream energy are valuable contributions to the understanding of the dynamics of traffic flow. To apply these concepts, however, we must find a parameter that accurately reflects internal energy. This parameter must satisfy the boundary conditions for internal energy, which were discussed previously, and should in general exhibit a compensatory pattern with corresponding kinetic energy.

Consider the compressible fluid discussed previously. The internal energy in general can be expressed as $(P_1\Delta l_1)/l_1$. Because $\Delta l_1 = l_o - l_1$ and $m = \rho_o l_o = \rho_1 l_1$, then Δl_1 can be written as $m/\rho_o - m/\rho_1 = m(1/\rho_o - 1/\rho_1)$. Thus, the internal energy term becomes $P_1\rho_1(1/\rho_o - 1\rho_1)$. The traffic stream analogy of this term would be $P_1k_1(1/k_o - 1/k_1)$ where P_1 is the average of an imaginary resistant force acting on the traffic stream from time t_o to time t_1.

If this resistant force were constant (call it P_o), then the term $P_1k_1(1/k_o - 1/k_1)$ could be written as a linear function of k_1, that is, as $P_o(k_1/k_o - 1)$. A graphical presentation of this force is shown in Figure 3. It can be seen that the greater the density k_1 becomes, the greater the internal energy becomes, and when $k_1 \to k_o \to 0$ the internal energy also approaches zero. This behavior satisfies the boundary conditions previously postulated for the internal energy of a traffic stream.

The relationship shown in Figure 3 was based on the assumption that the resistant force was constant. When a traffic stream is considered, however, this force is invisible and might be imagined to be a function of the internal friction inherent in traffic flow. From our general knowledge of traffic behavior, it seems more logical to assume a variable force in these circumstances than a constant force. In mechanics a force F is related to the mass of an object m and its acceleration a by Newton's second law of motion, $F = ma$. Because m is a constant, the force can be written simply as a function of acceleration: $F = f(a)$. This argument suggests that the imaginary force that acts

Figure 1. Traffic queue-forming condition.

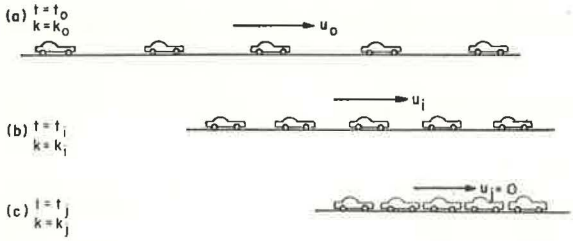


Figure 2. Compressible fluid condition analogous to traffic flow.

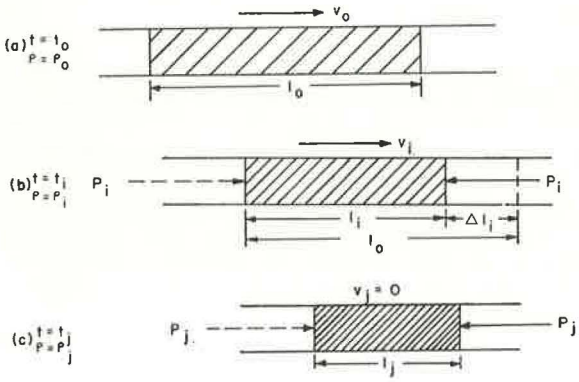
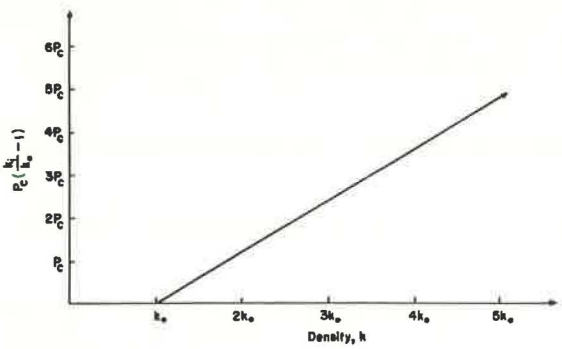


Figure 3. Density versus $P_c(k_j/k_o - 1)$.



on a vehicular platoon is a function of the acceleration distribution of the stream with mean value $\dot{u}$. It is generally accepted that the velocity of traffic flow is a function of traffic density: $u = f(k)$. If we differentiate this expression with respect to time, the following relationship is obtained:

$$\dot{u} = f'(k)\frac{dk}{dt} \tag{10}$$

where $f'(k) = df/dk$. This implies that the imaginary resistant force P is a function of $f'(k)(dk/dt)$. In this expression, $f'(k)$ would be a known function if the relationship between speed and density were defined. The term dk/dt, which is the time rate of change of density, however, does not present a functional pattern according to existing knowledge. For this reason, no exact expression for the variation of internal energy as a result of the analysis of this section provides a valuable guide in the search for a suitable internal energy parameter.

EXPERIMENTAL INVESTIGATION

Methodology

From the previous analysis, it seems that, if a traffic flow could closely resemble the properties of a compressible fluid, the term $P_1(k_1/k_o - 1)$ would be used to indicate its internal energy. Because the fluid analogy is not strictly applicable throughout the density domain and because it has been shown that the imaginary resistance for a traffic flow does not take a specific functional pattern, the term $P_1(k_1/k_o - 1)$ cannot be used as a direct measure of internal energy in itself. It does seem to provide, however, a good approximation of the true internal energy pattern.

A good internal energy parameter should satisfy the following requirements:

1. It should be a measure of vehicular interaction,
2. It should satisfy the boundary requirements of traffic conditions, and
3. It should have a variational pattern that approximates the variational pattern of the fluid analogous term $P_1(k_1/k_o - 1)$.

With these criteria in mind, an empirical approach is used in the search for an internal energy parameter. This approach is dictated by our inability to establish a theoretical expression for the internal energy of a traffic system. The following activities direct the empirical analysis:

1. Establish the variational pattern of the term $P_1(k_1/k_o - 1)$ versus density from appropriate data;
2. Choose vehicular interaction related parameters, and plot their variational pattern against density;
3. Compare the plots obtained from procedures 1 and 2 to see whether they agree; and
4. If they do, check the boundary requirements.

The data used for this investigation were collected by an aerial photogrammetry technique (2). The selected platoon is displayed on the vehicle trajectories shown in Figure 4.

Variations of the Imaginary Resistant Force for the Platoon Studied

To calculate the imaginary resistant force, we treat the platoons as confined masses of compressible fluid. The arithmetic mean speed of the platoon is taken as the speed of the fluid mass, and the force is considered to be a function of the time rate of change of the average speed (acceleration). Figure 5 shows the relationships between average acceleration and density for the platoon. Disregarding scale differences, the average imaginary force would have the same variational pattern with density as does average acceleration.

Figure 4. Identification of platoon studied.

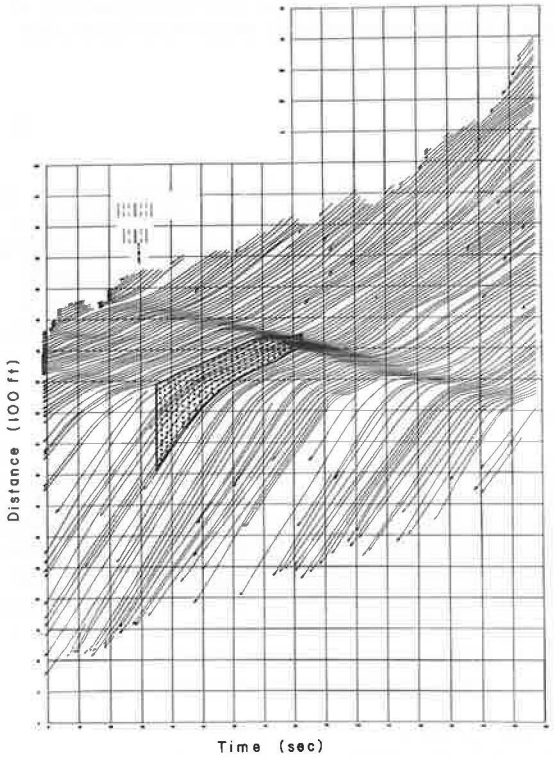
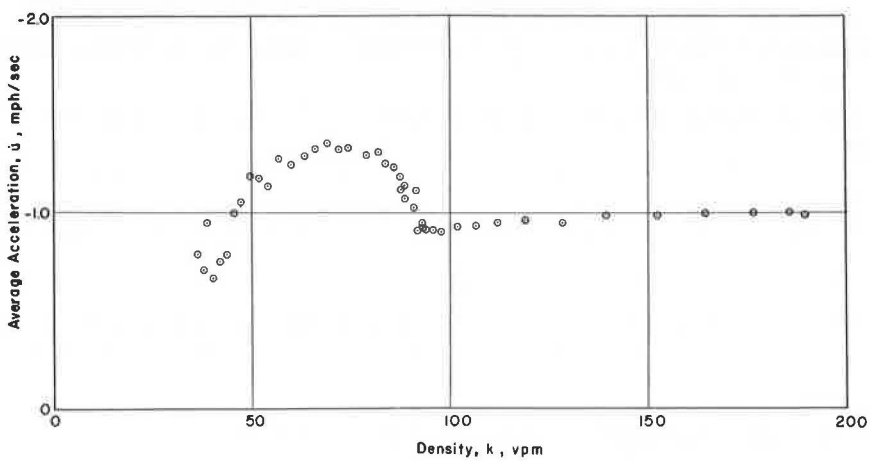


Figure 5. Average acceleration versus density for platoon studied.



If we recall that the expression for internal energy of the traffic stream is $P_1(k_1/k_o - 1)$ and that $(k_1/k_o - 1)$ is an increasing linear function of density, then the variation of internal energy with density can be specified. The internal energy will be a generally increasing function of density with a hump at that value of density where the average resistance is at a maximum (Fig. 5).

Alternative Internal Energy Parameters

With the theoretical pattern for the variation of internal energy with density determined for the selected platoon, it is now possible to investigate the applicability of several possible internal energy parameters. Four different parameters that are considered to be vehicular interaction related have been analyzed. These are

1. Standard deviation of the acceleration distribution of a platoon σ_a (this is in contrast to the acceleration noise value, discussed earlier, that considers only one vehicle),
2. Average of the absolute value of acceleration of the vehicles in a platoon $|\bar{a}|$,
3. Standard deviation of the platoon speed distribution σ_s, and
4. Coefficient of variation of the platoon speed distribution defined as the standard deviation of speed divided by the arithmetic mean speed (CV_u).

Standard Deviation of Acceleration—The relationship between the standard deviation of acceleration and density for the selected platoon is shown in Figure 6. No recognizable pattern similar to the one desired for internal energy can be identified. In addition, this parameter does not satisfy the boundary condition that requires that it be a maximum at maximum density.

Average Absolute Acceleration—Figure 7 shows a plot of the average absolute acceleration versus density for the selected platoon. The pattern is similar to that obtained for the standard deviation of acceleration and has no value as a representative of internal energy.

Standard Deviation of Speed—Investigation of the standard deviation of the platoon speed distributions yielded much more encouraging results than the acceleration-oriented studies. Figure 8 shows the variation of the standard deviation of speed with density for the selected platoon. A functional variational pattern is presented: The dispersion of speed decreases as density increases until a region is reached where almost all the vehicles in the platoon are moving at about the same speed. As density continues to increase, the dispersion of speeds begins to increase as well. This phenomenon can be explained by the fact that traffic flow at high densities tends to be unstable, and there can exist a large variance among the speeds of the individual vehicles in such a disturbed flow situation. With still further increases in density, the dispersion of speed once again drops because the space available to each vehicle for maneuvering has become severely limited. Finally, when jam density is reached, σ_s falls to zero, for all movement on the roadway has ceased.

This parameter appears to be a good indicator of internal energy in that it is representative of prevailing vehicle interactions. It presents a consistent and recognizable pattern with density and is simple to calculate. It does not, however, satisfy the boundary condition that internal energy be a maximum at jam density.

Coefficient of Variation of Speed—To correct the boundary condition shortcoming displayed by σ_s requires that a modified parameter be formed by dividing the standard deviation of the speed distribution by the arithmetic mean speed at each density level. This parameter, CV_u, is referred to in statistical terms as the coefficient of variation of speed and provides a measure of the relative dispersion of the speed values as a percentage of the mean speed. A plot of CV_u versus density is shown in Figure 9 for the selected platoon. Comparing the exhibited patterns with Figure 5, which were derived earlier, evidences a superb agreement.

Now we check the boundary conditions. According to the definition of the coefficient of variation of speed, we have

$$CV_u = \frac{\left[\frac{1}{n-1}\sum_{i=1}^{n}(\mu_i - \mu)^2\right]^{1/2}}{\bar{\mu}}$$

Figure 6. Standard deviation of acceleration versus density for platoon studied.

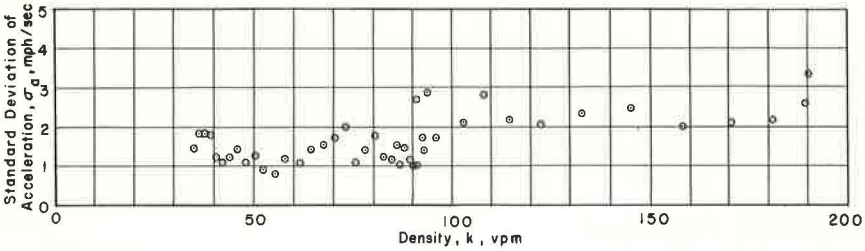


Figure 7. Average absolute acceleration versus density for platoon studied.

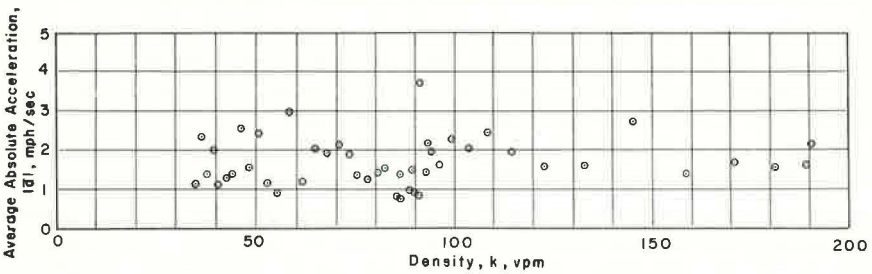


Figure 8. Standard deviation of velocity versus density for platoon studied.

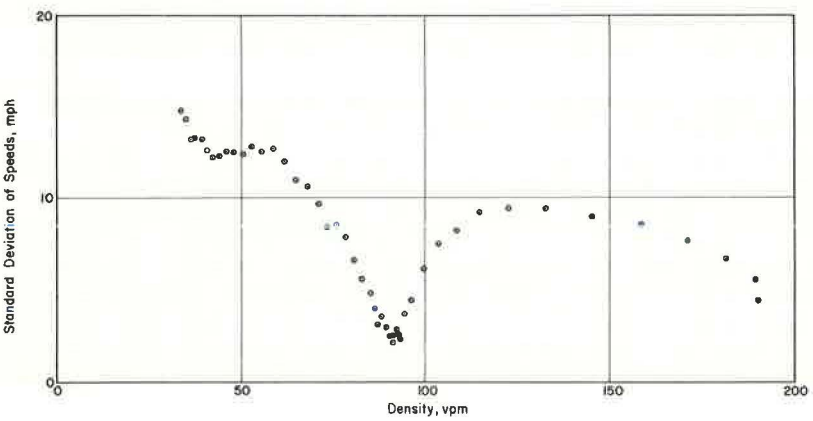
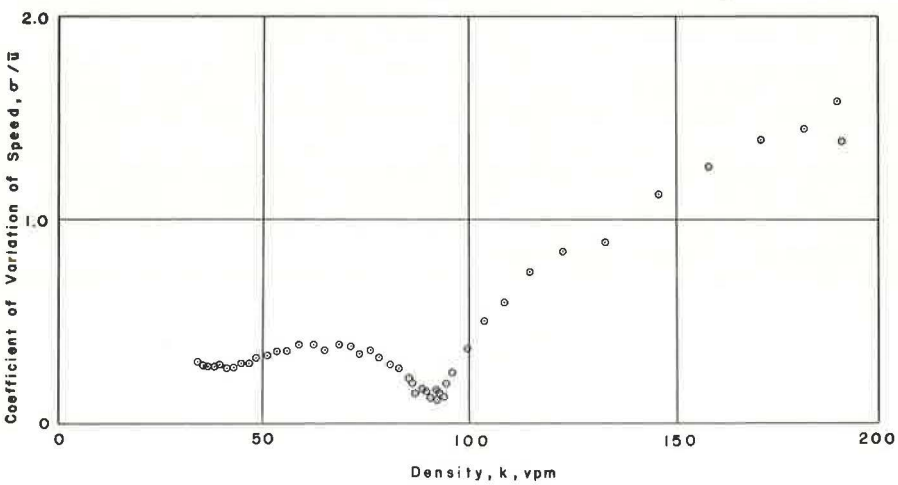


Figure 9. Coefficient of variation of speed versus density for platoon studied.



where

$\bar{\mu} = \sum_{i=1}^{n} \mu_i/n$ = average speed,
μ_i = speed of ith vehicle in a traffic stream, and
n = number of vehicles.

We can see that, when (a) $k = 0$, $\mu_i \rightarrow \mu_f$ and $\bar{\mu} \rightarrow \mu_f$, and thus CV_u approaches zero; and that when (b) $k = k_j$, $\mu_i \rightarrow 0$, and $\bar{\mu} \rightarrow 0$, and thus CV_u approaches maximum when $k \rightarrow k_j$. From this analysis, it is evident that the boundary conditions have been satisfied. An interesting point to be noted here is that, when every vehicle is moving at the same speed u_i (in this case $\mu_i = \bar{\mu}$), CV_u approaches zero also.

From the evidence presented, it seems that the coefficient of variation of speed is an excellent choice for measuring the internal energy of traffic flows.

CONCLUSIONS

From the analyses relating to traffic energy presented in this paper the following general conclusions may be drawn:

1. If the kinetic energy of a traffic stream is defined as αku^2 and the internal energy is defined in terms of vehicular interactions, the principle of conservation of energy does not hold. In fact, it will not hold regardless of how internal energy is defined so long as kinetic energy is taken to be αku^2 because we are not dealing with a confined system.
2. Acceleration noise does not represent a good indication of internal energy throughout the entire density domain.
3. If traffic flow is taken to be exactly analogous to compressible fluid flow, internal energy can be expressed as $P_i(k_i/k_o - 1)$ for the ith traffic state. If the analogy is only approximately correct, as seems logical, the term $P_i(k_i/k_o - 1)$ serves as an approximation of the true internal energy.
4. Of the four alternative internal energy parameters studied, only the coefficient of variation of speed fulfilled all the requirements postulated for the desired parameter. It is, therefore, proposed as a suitable measure of the internal energy of a traffic stream.

It is thought that the material contained in this paper represents a further step toward the attainment of an understanding of the dynamics involved in traffic movement. Such an understanding is a necessary prerequisite to the establishment of a safe and efficient highway system and forms a basis for determining control strategies for that system.

REFERENCES

1. Drew, D. R. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968.
2. Investigation of Traffic Dynamics by Aerial Photogrammetry Techniques. Transportation Research Center, Ohio State Univ., Interim Rept. EES 278-2, June 1969.
3. Montroll, E. W., and Potts, R. B. Car Following and Acceleration Noise. HRB Spec. Rept. 79, 1964, pp. 37-48.
4. Jones, T. R., and Potts, R. B. The Measurement of Acceleration Noise—A Traffic Parameter. Operations Research, Vol. 10, No. 6, 1962, pp. 745-763.
5. Montroll, E. W. Acceleration Noise and Clustering Tendency of Vehicular Traffic. In Theory of Traffic Flow (Herman, R., ed.), Elsevier Publ. Co., 1961, pp. 147-157.
6. Neville, A. M., and Kennedy, J. B. Basic Statistical Methods for Engineers and Scientists. International Textbook Co., 1964.

MULTIPLE RAMP CONTROL FOR A FREEWAY BOTTLENECK

Patrick J. Athol and A. G. R. Bullen, University of Pittsburgh

Many current approaches to freeway control use deterministic models of traffic flow based on the continuous flow-density curve. This paper proposes a control strategy based on a two-state traffic flow pattern with the primary control parameter being the probability of transition from uncongested flow to congested flow. The objective of the control is to maximize the reward associated with free flow. Trial solutions indicate that feasible numerical values for optimum control can be easily obtained, and these will be dependent on the length of the peak period. The approach should have direct applicability to existing surveillance and control hardware.

•BOTTLENECK, which is the primary cause of congestion on a limited-access highway, is a term defining some operational constriction. It is usually identified with a local area rather than a precise point of the highway. Physical bottlenecks are related to the design features of the highway and are fixed in space, and dynamic bottlenecks are related to traffic incidents and can occur at any location. Regardless of the type, all bottlenecks have a disruptive effect on traffic, which will be some combination of increased accident potential, reduced traffic volume, and detrimental environmental effects. Bottleneck control, simplistically overstated, contends that more traffic can be served at a higher level of service if congestion is eliminated. The control concept is to sustain the best operational level and, by preventing congestion, to yield benefits in increased safety and reduced delay.

Many control methodologies, theoretical and applied, have been based on the traditional flow-density relationship (1), which suggests a point of maximum flow (capacity). Initial controls in the Lincoln Tunnel in New York (2) and the Eisenhower Expressway experiments in Chicago (3) were based on the assumption that traffic could be controlled to this maximum flow condition. This maximum flow point, however, turned out to be very sensitive to breakdown, and it could not be maintained in practice without the rapid onset of congestion. Accordingly, most strategies have backed off from the theoretical ideal of maximum flow, and the emphasis is now on delaying or preventing congestion.

New York used density as the control parameter, whereas Chicago used the directly measurable equivalent, occupancy (4). Experimental work on the Gulf Freeway in Houston (5) combined parameters in various functions of volume, speed, and density. The deterministic approaches to these systems required ongoing empirical refinement of their control functions to balance the risk of congestion against higher allowable flows.

PEAK-PERIOD BOTTLENECK CONTROL

Early literature on freeway characteristics alludes to traffic operation as a two-state process. Mika, Kreer, and Yuan (5) identified two modes of operation corresponding to congested and uncongested flow on a freeway. Refined measurements of flow and density indicate that the q-k curve masks the underlying traffic process. The curve is a regression fitted to historical data and, as such, does not necessarily provide a suitable model for real-time control. In particular, the curve does not model operating differences from day to day caused by weather or short-term variations caused by individual driver characteristics.

A pilot control scheme that has been tested in Chicago (6) modifies some of the previous approaches. The control is aimed at peak-period flow, and its primary objective is to delay the onset of congestion by limiting bottleneck flow. During any time period,

which in this case is 1 min, the probability of congestion setting in, i.e., the probability of breakdown, is assumed to be a function of flow and density in the bottleneck. Given a suitable probability and the bottleneck density, the controller sets the desired bottleneck flow for each time period. The ramps upstream of the bottleneck are then metered to achieve the appropriate bottleneck flow. In the Chicago experiment the probability of breakdown and its functional relationships were heuristically determined and then empirically refined from freeway data.

In the following sections, a more analytical approach is suggested. This approach involves techniques that should be within the capability of current controllers and uses functions that, although not yet empirically validated, will require only currently available freeway data for their estimation.

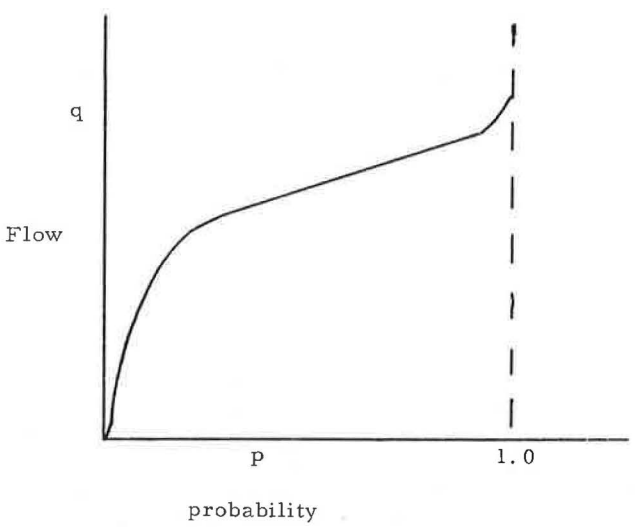
PROBLEM FORMULATION

Bottleneck operation is formulated as a process in which finite probabilities of breakdown are associated with each level of operation. The control stragegy considers the peak period as a series of successive time intervals where the probability of breakdown is set to optimize overall performance. It is assumed that, once congestion has set in, recovery during the peak period cannot be effected, a characteristic common in many practical situations.

The peak-period operation of the bottleneck consists of an uncongested period of some length followed by a congested period. If some reward is associated with the uncongested period, then the control objective would be to maximize the expected reward of the system. The reward, which can be some combination of increased flow, reduced accident risk, reduced emissions, and the like, will be some function of the probability of breakdown and the length of the uncongested period.

The reward function considered here will be in terms of traffic flow, and the objective considered for this problem is that of maximizing the expected value of the uncongested peak-period flow. The exact form of the reward function will require field testing and estimation. Its general characteristics, however, can be deduced from operational bottleneck experience and should follow the approximate form shown in Figure 1. Initially the function will rise rapidly to a substantial traffic flow before the probability of breakdown becomes significant, but then the rate of flow increase will decline. Because only uncongested traffic flow is being considered, the function will be continuous and for this particular problem only the left region ($p < 0.25$) is significant. This will simplify empirical validation.

Figure 1. Flow at a bottleneck as a function of the probability of breakdown.



For a suitable function to fit this general shape we use here the incomplete beta function:

$$R(p) = \frac{\gamma(a + b)}{\gamma(a)\gamma(b)} \int_0^p t^{a-1}(1 - t)^{b-1}dt \tag{1}$$

where

$0 \leq p < 1$,
a and $b > 0$, and
$\gamma(x)$ = the gamma function.

AN INFINITE PEAK PERIOD

Consider a controlled bottleneck where the probability of breakdown is set equal to p for each time period during uncongested flow. Then the probability P(k) of an uncongested period of k time periods will be given by

$$P(k) = (1 - p)^k \times p \qquad k = 0, 1, 2... \tag{2}$$

Because we are dealing with traffic at the macro level (1-min averages) and not the micro level of individual vehicles, the assumption of independence of trials is valid.

The expected return E(R) will be

$$E(R) = \sum_{k=0}^{\infty} (1 - p)^k p \times R(k, p) \tag{3}$$

The optimal control strategy will be to choose p to maximize the return.

$$\frac{dE(R)}{dp} = \frac{d}{dp} \sum_{k=0}^{\infty} (1 - p)^k p \times R(k, p) = 0 \tag{4}$$

for optimum and

$$\frac{d^2E(R)}{dp^2} = \frac{d^2}{dp^2} \sum_{k=0}^{\infty} (1 - p)^k p \times R(k, p) < 0 \tag{5}$$

for maximum.

To check the feasibility of the reward function developed in the previous section we assume a form

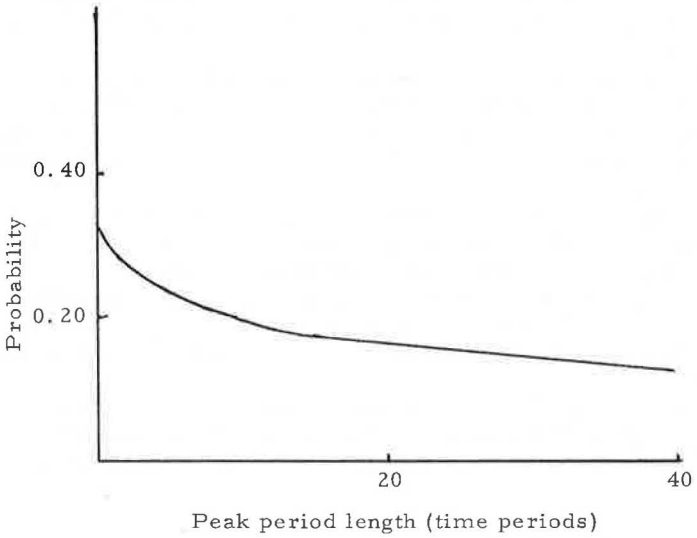
$$R(k, p) = k \frac{\gamma(a + b)}{\gamma(a)\gamma(b)} \int_0^p t^{a-1}(1 - t)^{b-1}dt \tag{6}$$

and the optimal strategy would be given by

$$\frac{d}{dp} \left\{ \sum_{k=0}^{\infty} \left[(1 - p)^k \times p \times k \frac{\gamma(a + b)}{\gamma(a)\gamma(b)} \int_0^p t^{a-1}(1 - t)^{b-1}df \right] \right\} = 0 \tag{7}$$

For values of a = 2 and b = 9, this gave a numerical solution of p = 0.12. Although this is a hypothetical example, the solution is in a practical range although perhaps rather high inasmuch as it gives an expected length of the uncongested period of only 7.5 time periods. Values of p below 0.1 would give somewhat better results.

Figure 2. Optimum probability of breakdown for a finite peak period.



A FINITE PEAK PERIOD

Suppose now that the peak period has a finite length of n time periods. Then the probability of an uncongested period of k time periods is given by

$$P(k) = (1 - p)^k p \qquad k = 0, 1, \ldots (n - 1)$$

$$P(n) = (1 - p)^n$$

The expected reward during the uncongested period will be

$$E(R) = \sum_{k=0}^{n-1} \left[(1 - p)^k \, pR(k, p) \right] + (1 - p)^n \, R(n, p)$$

And, again, setting $[dE(R)]/dp = 0$ will give the optimum control value for the probability of breakdown. In this case, however, this optimum value will be a function of n.

This affects the bottleneck control strategy in at least two ways. First, if uncongested flow has been continuously maintained during the peak period, then the control stragegy at any time period is dependent only on the length of the peak period remaining. As the peak period progresses, therefore, the value of n steadily declines, and, accordingly, the parameter values for the control may change.

The second circumstance is when uncongested flow is recovered from the congested state during the peak period, which can occur, for example, when demand temporarily declines because of an incident upstream. The control strategy for uncongested flow will then depend on the length of the peak period remaining.

To indicate the possible magnitude of this dependence, we used the same reward function used as an example in the previous section as a numerical example of the finite length case. The optimum values for p are shown in Figure 2, which clearly indicates the effect of the peak period length n.

CONCLUSION

This paper offers a different analytical approach to freeway control. Inasmuch as it provides a limited theoretical supplement to the empirical control algorithms already in operation, it should be suitable for practical implementation. The essential traffic

functions required, such as the probability of system breakdown, are not yet generally available; their estimation, however, will require only data that are a normal output of most current freeway surveillance systems.

The methodology can be used on existing freeway control systems by changing the computer programming but without modification to the hardware. New operational parameters can be developed from the system itself based on a new datum of controlled bottleneck operations. Data from an uncontrolled bottleneck serve as the first approximation in developing the control strategy.

To sustain congestion-free bottleneck operations for longer control periods requires that the optimal value of the probability of breakdown be maintained considerably below 0.1 (assuming 1-min time periods). This finding is contrary to the control strategy of operating at maximal flow developed by most theoretical studies but agrees with operational experience where "overcontrol" is necessary to prevent breakdown. The explanation lies with the q-k relationship, its probable discontinuous character, and its nonregular short-term behavior.

For a normal peak period, the length of the peak period has little effect on the selection of the optimum probability. Where the control is operating near the end of the peak period, however, the length of the peak period remaining should be taken into consideration.

REFERENCES

1. Lighthill, M. T., and Whitham, G. B. On Kinematic Waves, A Theory of Traffic Flow on Long Crowded Roads. Proc. Royal Society of London, Series A, Vol. 33, 1955.
2. Edie, L. C., and Foote, R. S. Experiments on Single-Lane Flow in Tunnels. Proc. Symposium on the Theory of Traffic Flow, Warren, Mich., Elsevier Publ. Co., 1961.
3. May, A. D. Experimentation With Manual and Automatic Ramp Control. Highway Research Record 59, 1964.
4. Athol, P. J. Interdependence of Certain Parameters. Presented to Committee on Freeway Operations, Highway Research Board, 1964.
5. Mika, H. S., Kreer, J. B., and Yuan, L. S. Dual Mode Behavior of Freeway Traffic. Highway Research Record 279, 1969, pp. 1-12.
6. Bullen, A. G. R. Strategies for Practical Expressway Control. Jour. Transportation Engineering, Proc. ASCE, Vol. 98, Aug. 1972.
7. Howard, N. A. Dynamic Probabilistic Systems, Vol. 1: Markov Models. John Wiley and Sons, 1971.
8. Parzen, E. Stochastic Processes. Holden Day, 1962.

SPONSORSHIP OF THIS RECORD

GROUP 3—OPERATION AND MAINTENANCE OF TRANSPORTATION FACILITIES
Harold L. Michael, Purdue University, chairman

Committee on Traffic Flow Theory and Characteristics (as of December 31, 1972)
Donald G. Capelle, Alan M. Voorhees and Associates, Inc., chairman
Patrick J. Athol, John L. Barker, Martin J. Beckmann, Martin J. Bouman, Kenneth A. Brewer, Donald E. Cleveland, Kenneth W. Crowley, Lucien Duckstein, Leslie C. Edie, H. M. Edwards, A. V. Gafarian, Denos C. Gazis, Daniel L. Gerlough, John J. Haynes, Edmund A. Hodgkins, James H. Kell, John B. Kreer, Leonard Newman, O. J. Reichelderfer, Richard Rothery, August J. Saccoccio, A. D. St. John, William C. Taylor, Joseph Treiterer, William P. Walker, Sidney Weiner, W. W. Wolman

K. B. Johns, Highway Research Board staff