TRANSPORTATION RESEARCH RECORD

495

Formerly issued as Highway Research Record

Incidents and Freeway Operations

8 reports prepared for the 53rd Annual Meeting of the Highway Research Board

subject areas

51 highway safety
53 traffic control and operations
54 traffic flow
55 traffic measurements

TRB

TRANSPORTATION RESEARCH BOARD

NATIONAL RESEARCH COUNCIL

Washington, D. C., 1974

NOTICE

These papers report research work of the authors that was done at institutions named by the authors. The papers were offered to the Transportation 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 Transportation Research Board.

Before publication, each paper was reviewed by members of the TRB 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 Transportation 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.

The National Research Council is the principal operating agency of the National Academy of Sciences and the National Academy of Engineering, serving government and other organizations. The Transportation Research Board evolved from the 54-year-old Highway Research Board. The TRB incorporates all former HRB activities but also performs additional functions under a broader scope involving all modes of transportation and the interactions of transportation with society.

TRR 495
ISBN 0-309-02283-5
LC Cat. Card No. 74-12522
Price: $4.00

FOREWORD

Disabled vehicles and other incidents that temporarily reduce the capacity of freeways have a major impact on traffic, and research on them is being given top priority by the Transportation Research Board Committee on Freeway Operations. The papers in this RECORD are concerned with detecting incidents and with related subjects that should be of special interest to all concerned with traffic operations in arterial corridors.

Cook and Cleveland investigate the feasibility of using the traffic flow information from on-line, computer-based surveillance systems to detect incidents. They studied 50 incidents that occurred on that part of the John Lodge Freeway in Detroit for which both television and detector surveillance were available. Ranking 19 computer algorithms for detecting stoppages, they give top place to a model based on occupancy. Changes in occupancy, compared with a 5-minute moving average, are used to detect incidents. Exponential smoothing of traffic data is found to be most effective.

A similar approach is proposed by Dudek, Messer, and Nuckles. Their analysis was tested using 35 incidents on the Gulf Freeway in Houston. Their concept is to evaluate the trends in a control variable such as occupancy and to recognize when the variable changes more rapidly than would be expected in normal traffic flow. They conclude that a strategy with lane occupancy as the control variable, using a 5-minute time base to compute occupancy, gave the best results. It is striking that these authors as well as Cook and Cleveland both find that their best algorithms detect 92 percent of the stoppages. The Detroit-based analysis resulted in a false-alarm rate of 1.87 percent and a mean detection time of 0.74 minute. The Houston-based research had a lower false-alarm level of 1.3 percent but a longer mean detection time of 1.1 minutes.

An altogether different approach is implicit in the work reported by Trabold and Reese, which gives the final report on the use of Citizens Band radio to provide a highway communications service in Ohio. Analyzing a sample of 14,750 calls, the authors report more than 60 percent of them relate to incidents of a type that would reduce capacity. Nearly 25 percent are related to accidents and 22 percent to stalled vehicles. The authors describe other types of calls and responses received, stressing the extensive voluntary participation obtained in Ohio.

Detector reliability is discussed by Dudek, Messer, and Dutt with reference to a motorist information system. Analyzing the 96 traffic detectors on the Gulf Freeway in Houston, the authors found 47 detector failures in 5 months, which they estimated would cause 87 percent of the hardware problems associated with a real-time freeway warning system. In view of failure and repair rates experienced in Houston, the authors calculated that the overall system would be operational from 95 to 99.9 percent of the time, depending on the degree of redundancy provided. The approach they offer should be useful to others, and reports of experience by others would be valuable.

The last four papers are more theoretical, although real data figure in each of them. Gazis and Szeto address the problem of measuring traffic density, which has proved useful in predicting the need for restraining traffic demands to maintain fluid high-flow movement in tunnels. One difficulty in applying the tunnel findings to freeways arises in measuring density on-line, because the methodology used in tunnels requires that lane-changing be at a minimum. The technique evaluated by Gazis and Szeto uses Kalman filtering with promising results. The authors offer a procedure for applying the technique on freeways.

The discontinuous drop in flow that occurs as traffic density increases has been reported earlier by several authors, but Hillegas, Houghton, and Athol examine the phenomenon over a longer space and time interval than previously reported. They use a 7-mile, two-lane expressway with only one intermediate ramp and examine it for a period of 7 hours during which flow passes from fluid to congested and back to fluid.

A method for smoothing flow during congestion is discussed by Kleinman and Wiener. Although the use of advisory speed signs has not in the past proved effective for this purpose, the analytical approach derived by the authors for the control algorithm is of interest.

The final paper is in a continuing series on the use of computer simulation to optimize the design and operation of freeway corridor systems. May reports the development and performance of a model that predicts traffic performance as a function of capacity and demand, that evaluates alternate capacity and ramp control strategies, and that expands to consider arterial street capacity in the corridor as well.

—Robert S. Foote

CONTENTS

DETECTION OF FREEWAY CAPACITY-REDUCING INCIDENTS BY TRAFFIC-STREAM MEASUREMENTS

Allen R. Cook, Department of Civil Engineering, McGill University; and
Donald E. Cleveland, Department of Civil Engineering, University of Michigan

This paper investigates the feasibility of using freeway traffic-flow data compiled by electronic surveillance and control systems for the detection of accidents and other lane blockage incidents that temporarily disrupt traffic flow. Research was conducted on the John C. Lodge Freeway in Detroit where traffic data consisted of 1-minute compilations of volume and occupancy recorded by ultrasonic presence detectors. Nineteen detection algorithms, including one being used in Los Angeles, were evaluated with a sample of 50 representative afternoon peak-period incidents. The technique of exponential smoothing of occupancy or volume data to detect incident-generated shock waves was found to be the most effective. This algorithm detected 42 percent of the incidents with virtually no false alarms and every incident with an 8 percent false-alarm rate. Most of the incidents were detected within 1 minute of the onset of congestion at a detector station. Algorithm effectiveness was not affected by detector spacings ranging from 1,460 to 4,815 ft (445 to 1468 m), volumes from 1,200 to 2,000 vph per lane, occupancies from 9 percent to 45 percent, precipitation, or the particular lane blocked. The algorithms could not distinguish accidents from less serious incidents, but because they directly related each incident to its impact on traffic operations their incorporation in control systems could improve system response to incidents.

•THIS PAPER investigates the feasibility of using traffic-flow information compiled in real-time by freeway traffic surveillance, communication, and control systems for the detection of accidents, vehicular breakdowns, and other incidents that temporarily block traffic lanes and disrupt flow. Freeways of large cities regularly operate at demand levels that approach or exceed capacity, and they are particularly vulnerable to incident disruptions. Although incidents are statistically rare events, occurring once every 20,000 to 30,000 vehicle-miles (30 000 to 50 000 vehicle-km) on high-volume urban freeways, they occur with such frequency that they must be taken into account by surveillance and control systems.

MEASURES OF EFFECTIVENESS FOR DETECTION ALGORITHMS

The role of computer-supervised incident detection in a surveillance system is shown in Figure 1. Implicit in this role is the capability of implementing control measures without the inevitable time lag of detection systems involving human intervention and of relating this response directly to the magnitude of the capacity reduction. To evaluate an incident detection capability, the following measures of effectiveness were investigated: the probability of detection, the probability of false alarms, the time lag to first detection, the impact of detector spacing on algorithm effectiveness, and the probability of detecting the end of the incident. A false alarm is defined as a detection signal when no incident has occurred. Information on these measures of effectiveness was obtained

Publication of this paper sponsored by Committee on Freeway Operations.

by applying detection algorithms to traffic data compiled before, during, and after actual lane blockage incidents.

DATA COLLECTION

The John C. Lodge Freeway Corridor Control and Data System

Research on detection algorithms was conducted at the National Proving Ground for Freeway Surveillance, Control, and Electronic Traffic Aids in the John C. Lodge Freeway corridor in Detroit. The research section of the Lodge Freeway contained a closed network of ultrasonic vehicle presence detectors linked to an IBM 1800 digital computer. The system was capable of sensing fluctuations in traffic conditions that were brought about by congestion and capacity reductions and of responding to them through restrictions of ramp inputs and route diversion.

Incident data were collected and analyzed from a 2-mile (3.2-km) portion of the instrumented freeway containing four detector stations, as shown in Figure 2. Freeway sections between these stations were identified as subsystems; these varied in length from 1,460 ft (445 m) to 4,815 ft (1468 m). Data were collected during a 13-month period from December 1968 to December 1969 when this section was under continuous television surveillance. Descriptions of lane blockage incidents recorded by television observers were correlated with data from the detectors during the afternoon peak period (outbound from Detroit CBD) extending from 2:30 to 6:30 p.m. Traffic data consisted of 1-minute compilations of volume and occupancy recorded by each detector and aggregated for each station. Occupancy, a surrogate for density, is the percentage duration of activation of a presence detector.

Incident Characteristics

A total of 50 lane blockage incidents—18 accidents, 28 stalls and breakdowns, 2 instances of debris, and 2 short maintenance operations—were selected for analysis, with virtually every such incident occurring in the subsystems for which traffic data were available. It is believed that this sample is representative of the lane blockage incidents that take place during the peak period on the Lodge Freeway. In only one case was more than one lane blocked, and the freeway was never completely blocked. In two-thirds of the cases, the lane adjacent to the median, where there was no shoulder refuge for vehicles, was blocked. The average duration of blockage was 6.1 minutes, with times ranging from 1 to 19 minutes. None of the 18 accidents appeared serious or affected traffic flow differently from the other incidents. Therefore, it was not possible to investigate the feasibility of determining incident type by analysis of traffic data alone.

The average volumes observed before each incident took place ranged from 1,200 to 2,000 vph per lane. Prevailing occupancies ranged from 9 to 45 percent, with the great majority extending from 10 to 30 percent. The average prevailing downstream volume before the incidents, 81.4 vehicles per minute (σ = 10.3), was reduced 21 percent to an average of 64.4 vehicles per minute (σ = 11.3) during the incidents. There was considerable variability in the flow reductions, as shown in Figure 3. Similar variability was observed by Goolsby (1) with data measured at the incident site. Although downstream volumes generally decreased, there were seven exceptions where volume increased. These were attributed to congestion prevailing before the incident, where the capacity reduction appeared to have a beneficial effect on operations downstream. Six of these incidents were located upstream of the Seward Avenue entrance ramp, which was capable of admitting up to 25 vehicles per minute depending on the state of flow downstream.

DEVELOPMENT OF COMPUTER ALGORITHMS

Propagation of Disturbances Caused by Incidents

The theoretical impact of an incident on traffic operations when demand exceeds capacity is shown in the hypothetical flow-occupancy diagram of Figure 4, where point 2 is the prevailing flow state prior to the incident. After the incident, traffic upstream

operates in the congested regime of point 4 and at point 5 downstream. The flow state changes are propagated through the traffic stream at a velocity equal to the slope of the chord connecting the two flow states, as hypothesized by Lighthill and Whitham (2) in their continuum theory of traffic flow. A shock wave with velocity c_{52} proceeds downstream from the incident, and a shock wave of congestion growth proceeds upstream at velocity c_{24}. Detection algorithms based on traffic-stream characteristics should signal after these waves have passed upstream and downstream detector stations and either detect their passage or recognize the joint occurrence of states 4 and 5 as being indicative of incident presence. These considerations are reflected in the various algorithms presented in the following section. The following notation applies throughout:

$\Theta(x, t)$ = occupancy in percent at detector station x and minute t,
$q(x, t)$ = volume (vehicles per minute) at detector station x and minute t,
$u(x, t) = q(x, t)/\Theta(x, t)$, surrogate for speed,
$E(x, t) = [q(x, t)]^2/\Theta(x, t)$, surrogate for kinetic energy,
x_i = upstream detector station,
x_{i+1} = downstream detector station, and
t = time in 1-minute increments.

LAAFSCP Algorithm

The algorithm used by the Los Angeles area freeway surveillance and control project (LAAFSCP) consists of three sequential tests all based on occupancy changes over short periods of time at the upstream and downstream detector stations of a subsystem (3). An incident is signaled only when the predetermined threshold values for all three variables are exceeded, indicating that the sequence of events associated with a typical capacity-reducing incident has occurred. The algorithm is applied to occupancy data compiled for the most recent minute and updated every 20 seconds.

The LAAFSCP algorithm was adapted as follows for application to Lodge Freeway data. The initial incident test, which consisted of determining whether or not the numerical difference in occupancies at two adjacent detector stations exceeded a threshold value, was dispensed with because in this study it would not affect the algorithm performance. The remaining two tests (Test 2 and Test 3 respectively) are as follows:

$$\frac{\Theta(x_i, t) - \Theta(x_{i+1}, t)}{\Theta(x_i, t)} \geq \text{Threshold } K_2$$

$$\frac{\Theta(x_{i+1}, t - 1) - \Theta(x_{i+1}, t)}{\Theta(x_{i+1}, t - 1)} \geq \text{Threshold } K_3$$

An incident is signaled if upstream occupancy becomes significantly greater than downstream occupancy (Test 2) and at the same time downstream occupancy has significantly decreased in the past minute (Test 3). Test 3 distinguishes an incident from a bottleneck situation by indicating the reduction in flow past the incident over a short period of time. Detection thresholds were determined by applying the model to 4,860 minutes of occupancy observations in each subsystem (18 peak periods containing at least 8 incidents in each subsystem). Threshold values varied from subsystem to subsystem; K_2 ranged from 0.53 to 0.61 in both dry- and wet-weather conditions, and K_3 ranged from 0.11 to 0.26 in dry weather and from 0.24 to 0.30 in wet weather. Values were selected that produced the most detections with fewest false alarms (in the authors' judgment). With these thresholds there were 81 false detections out of 14,580 applications of the algorithm, a false-alarm rate of 0.56 percent. The performance of the LAAFSCP algorithm during a typical incident is shown in Figure 5.

TTI Lodge Freeway Algorithms

Texas Transportation Institute (TTI) began incident detection research in 1968 as part of the Lodge Freeway corridor study program. Courage and Levin (4) developed five algorithms that used various combinations of volume and occupancy data to form

Figure 1. Role of incident detection in freeway surveillance and control.

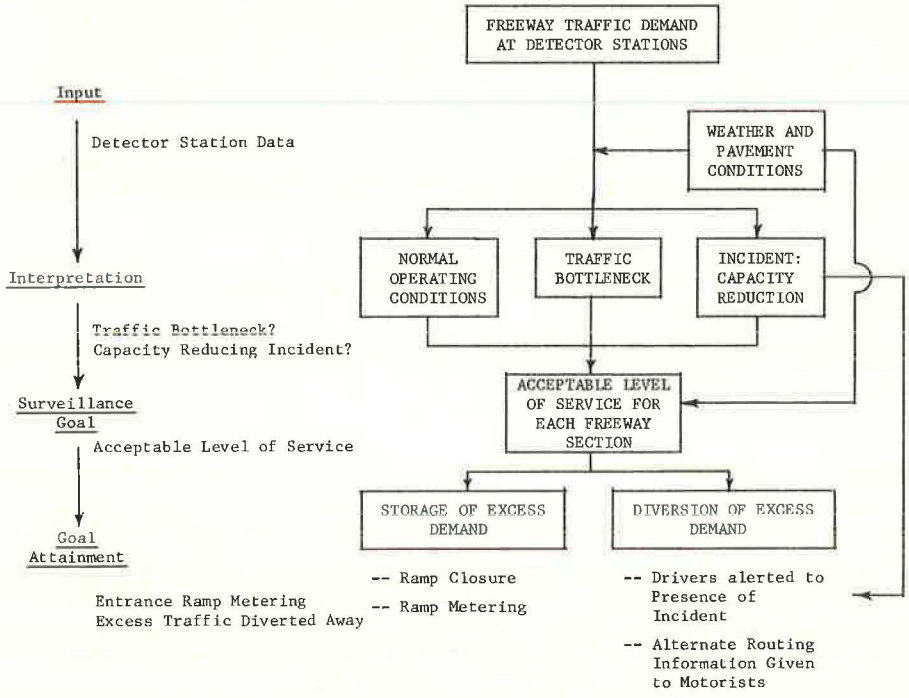


Figure 2. Lodge Freeway detection subsystems and sensing equipment.

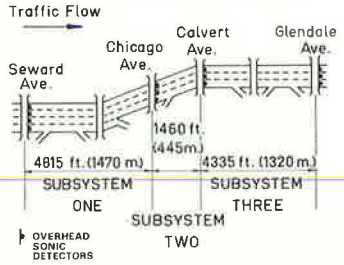


Figure 3. Effect of incidents on downstream traffic volume.

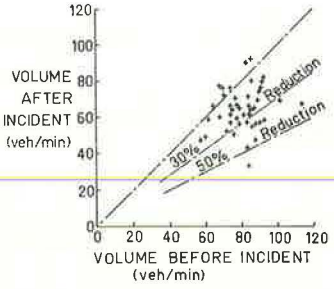


Figure 4. The fundamental diagram and kinematic wave theory.

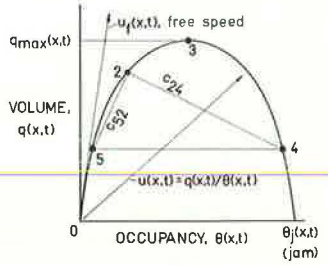


Figure 5. Performance of the LAAFSCP algorithm during an incident.

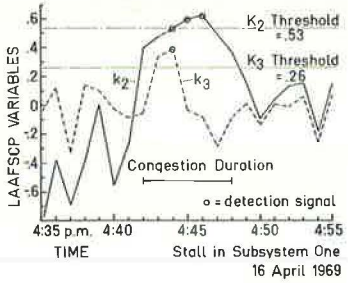


Figure 6. TTI station discontinuity performance during an incident.

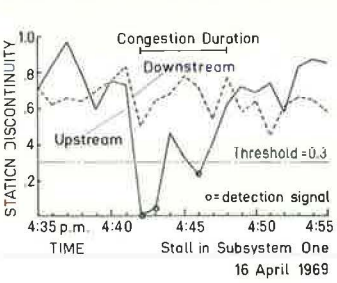
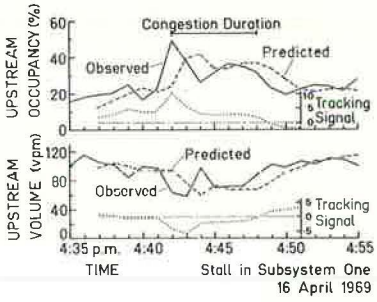


Figure 7. Exponential algorithm performance during an incident.



variables sensitive to incident situations such that each variable would exceed a predetermined threshold of detection only 1 percent of the time during incident-free operations. He concluded from an analysis of 25 peak periods containing 29 incidents that

> all models demonstrated some ability to detect incidents and may therefore merit further consideration. They did, however, exhibit a high false alarm rate, and it is felt that considerable refinement would be required to produce an operational incident detection scheme.

The variables for five alternative TTI algorithms used in this paper are described in the following paragraphs.

Station Kinetic Energy—Traffic kinetic energy $[q(x,t)]^2/\Theta(x,t)$ comes from an energy-momentum analogy between traffic flow and the flow of a compressible fluid. Significantly small variable values may indicate an incident downstream when congestion is generated upstream:

$$\text{Variable } E(x,t)/E_m(x,t)$$

This variable is normalized with the maximum kinetic energy $E_m(x,t)$ determined by fitting representative peak-period data to a linear speed-occupancy relationship (4).

Subsystem Energy—Significantly large numerical differences in normalized kinetic energy variables recorded at adjacent stations may indicate an incident presence in that subsystem:

$$\text{Variable } [E(x_i,t)/E_m(x_i,t)] - [E(x_{i+1},t)/E_m(x_{i+1},t)]$$

Subsystem Shock Wave—Significantly large numerical differences in the volume recorded at adjacent stations may indicate an incident presence in that subsystem:

$$\text{Variable } q(x_i,t) - q(x_{i+1},t)$$

Subsystem Discontinuity—The difference in flow states 2 and 4 shown in Figure 4, when transformed to the approximately linear speed-occupancy relation, may be indicative of an incident situation if it is significantly large:

$$\text{Variable}\left\{\left[\frac{u_f(x_i,t) - u(x_i,t)}{u_f(x_i,t)}\right]^2 + \left[\frac{\Theta(x_i,t)}{\Theta_j(x_i,t)}\right]^2\right\}^{1/2} - \left\{\left[\frac{u_f(x_{i+1},t) - u(x_{i+1},t)}{u_f(x_{i+1},t)}\right]^2 + \left[\frac{\Theta(x_{i+1},t)}{\Theta_j(x_{i+1},t)}\right]^2\right\}^{1/2}$$

Speed values are normalized by the theoretical free speed, $u_f(x,t)$ and occupancy is normalized by the theoretical jam occupancy, $\Theta_j(x,t)$, both obtained from an assumed linear speed-occupancy fit of data (4). This is done to make the data at adjacent stations comparable.

Station Discontinuity—This variable is based on a comparison of the kinetic energies, $E'(x_{i,j},t)$ of individual lanes, j, at a station where there are n lanes:

$$\text{Variable } \frac{(n-1)\{\text{Min}[E'(x_{i,j},t)]_{j=1}^{n}\}}{\sum_{j=1}^{n} E'(x_{i,j},t) - \text{Min}[E'(x_{i,j},t)]_{j=1}^{n}}$$

This variable can range from 0 to 1. A value of 1 would indicate an equal distribution of kinetic energy across the lanes. A value of 0 would indicate no energy in one lane, a result of either no traffic at all in the lane because it is blocked upstream or of traffic stopped altogether in the lane because of a blockage downstream. Both of these may occur while traffic in other lanes is moving. Significantly low values of the variable thus constitute an incident detection signal.

Threshold values for each variable were determined for each station or subsystem and dry- and wet-weather conditions such that false alarms were generated no more than 1 percent of the time during 8 incident-free dry-weather 4-hour peak periods and 4 incident-free wet-weather peak periods. The full set of numerical thresholds thus developed elsewhere (5). The performance of the TTI station discontinuity algorithm during a typical incident is shown in Figure 6.

Exponential Algorithms

The sudden flow-state changes observed during incidents suggest the application of short-term forecasting techniques for detecting irregularities in a time series of traffic data $z(x,t)$. Whitson et al. (6) proposed the use of a moving average of the most recent 5 minutes of volume data as the forecast variable with confidence limits determined from the variance of the data. In this research, the double exponential smoothing algorithm described by Brown (7) was investigated as a means for incident detection. With this technique, the forecast traffic variable $\hat{z}(x,t)$ is a function of the past data observations geometrically discounted back in time. Incident detection is accomplished through use of a tracking signal, which is the algebraic sum to the present minute of all the previous estimate errors divided by the current estimate of the standard deviation. The tracking signal should dwell around zero because the predictions either match the data or compensate for errors in succeeding time periods. A detection is indicated by a significant deviation of the signal from zero, the threshold deviation being either a function of the variability of the data or the likelihood of false alarms.

For computational efficiency the standard deviation of the traffic data was estimated by the mean absolute deviation $m(x,t)$, which Brown (7) has shown to be proportional to the standard deviation. The mean absolute deviation is obtained by single exponential smoothing of the absolute values of the prediction errors, using a smoothing constant of 0.1:

$$m(x,t) = \alpha|e(x,t)| + (1 - \alpha)\, m(x,t-1) = \text{mean absolute deviation}$$

where

$e(x,t)$ = error of prediction, $z(x,t) - \hat{z}(x,t)$ and
α = smoothing constant.

The variable forecast $\hat{z}(x,t)$ is computed by double exponential smoothing with a smoothing constant of 0.3, and the tracking signal is found as follows:

$$TS(x,t) = y(x,t)/m(x,t-1)$$

where $y(x,t) = y(x,t-1) + e(x,t)$ = cumulative error.

Several values of the smoothing constant other than 0.3 were tried on data for several peak periods, but higher values were insensitive to incidents, and lower values greatly increased false alarms. A smoothing constant of 0.1 was used for $m(x,t)$ to make it less sensitive to recent changes in data variability.

Initial estimates for the algorithm parameters were based on the mean and standard deviation of the first 6 minutes of variable observations in each incident data stream. An initial $m(x,t-1)$ was found from the sample standard deviation, σ_z, as follows:

$$m(x,t-1) = [\sqrt{2/\pi}\sqrt{2/(2-\alpha)}]^{\sigma_z}$$

A total of 13 traffic variables were investigated with the exponential smoothing algorithm, including all five variables used with the TTI algorithm; they are as follows:

1. Station volume, $z_1(x,t) = q(x,t)$.
2. Station occupancy, $z_2(x,t) = \Theta(x,t)$.
3. Station speed, $z_3(x,t) = q(x,t)/\Theta(x,t) = u(x,t)$.

4. Station volume-occupancy, where the chord connecting flow states 2 and 4 in Figure 4 is treated as the forecast error and where the forecast variable is the joint volume-occupancy flow state,

$$z_4(x,t) = \{q(x,t),\ \Theta(x,t)\}$$

$$e_4(x,t) = \sqrt{[q(x,t) - \hat{q}(x,t)]^2 + [\Theta(x,t) - \hat{\Theta}(x,t)]^2}$$

5. Station speed-occupancy, which is analogous to volume-occupancy,

$$z_5(x,t) = \{u(x,t),\ \Theta(x,t)\}$$

$$e_5(x,t) = \sqrt{[u(x,t) - \hat{u}(x,t)]^2 + [\Theta(x,t) - \hat{\Theta}(x,t)]^2}$$

6. Station kinetic energy, $z_6(x,t) = E(x,t)$.
7. Station discontinuity, which is the same variable used in the TTI station discontinuity algorithm, $z_7(x,t) = SD(x,t)$, where $SD(x,t)$ is the station discontinuity variable.
8. Subsystem volume, $z_8(x,t) = q(x_i,t) - q(x_{i+1},t)$.
9. Subsystem occupancy, $z_9(x,t) = \Theta(x_i,t) - \Theta(x_{i+1},t)$.
10. Subsystem speed, $z_{10}(x,t) = u(x_i,t) - u(x_{i+1},t)$.
11. Subsystem kinetic energy, $z_{11}(x,t) = E(x_i,t) - E(x_{i+1},t)$.
12. Subsystem volume-occupancy discontinuity,

$$z_{12}(x,t) = \sqrt{[q(x_i,t) - q(x_{i+1},t)]^2 + [\Theta(x_i,t) - \Theta(x_{i+1},t)]^2}$$

13. Subsystem speed-occupancy discontinuity,

$$z_{13}(x,t) = \sqrt{[u(x_i,t) - u(x_{i+1},t)]^2 + [\Theta(x_i,t) - \Theta(x_{i+1},t)]^2}$$

Subsystem variables 8 through 13 are compiled from adjacent detector stations upstream i and downstream i + 1.

The performance of the algorithm with volume and occupancy as the variables during the incident of Figure 5 is shown in Figure 7. Forecasts are observed to correct for trends in the data within several minutes, and it is evident that the most effective variables for detection will be those that respond sharply and promptly to the passage of an incident shock wave. The sensitivity of the algorithm to incident situations was explored for each variable by varying the detection thresholds for the tracking signal from ±1.5 to ±10.0 (-1.0 to -8.0 for variable 7) in increments of 0.5. The resulting operating characteristic curves relating detection effectiveness to false alarms are presented in the following section.

RESULTS

Detections and False Alarms

Figure 8 shows the operating characteristic curves for the three best exponential station algorithms: station occupancy, station volume, and station discontinuity. More detections were achieved at the expense of more false alarms. The exponential station occupancy variable dominated the others in performance and detected all 50 incidents at a 6 percent false-alarm rate. The operating points for the five TTI algorithms and the LAAFSCP algorithm are included in the figure, and none was observed to exceed the performances of the best exponential algorithms. Moreover, they tended to generate more false alarms than anticipated. This may indicate that the incident data contained more false-alarm situations than would be found in typical traffic operations, but it also indicates that the false-alarm rate for these algorithms may vary according to the state of the traffic (an undesirable characteristic).

The differences in total detections achieved by individual algorithms in a similar false-alarm range, about 1 percent to 2 percent, were investigated by means of chi-

Figure 8. Operating characteristic curves for exponential station incident detection algorithms.

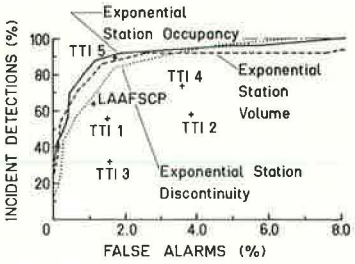


Table 1. Detection performance of the most effective algorithms at three false-alarm levels.

	False-Alarm Level					
	Low		Medium		High	
Incident Detection Algorithm	Incident Detections[a]	False Alarms[b]	Incident Detections[a]	False Alarms[b]	Incident Detections[a]	False Alarms[b]
Exponential station occupancy	48	0.26	92	1.87	96	5.73
Exponential station volume	54	0.19	90	1.90	92	6.37
Exponential station discontinuity	44	0.26	84	1.67	100	6.5
TTI station discontinuity	30	0.19	90	1.74	94	6.18

[a]Percentage of incidents detected, sample size of 50 lane-blockage incidents.
[b]Percentage of false alarms, sample size of 1,554 minutes of incident-free operations.

Table 2. Tracking signal thresholds for the most effective exponential algorithms at three false-alarm levels.

	False-Alarm Level		
Incident Detection Algorithm	Low	Medium	High
Exponential station occupancy	±8.0	±4.0	±2.75
Exponential station volume	±6.0	±3.75	±2.75
Exponential station discontinuity	-5.0	-3.0	-1.5

Table 3. Discontinuity thresholds for TTI station discontinuity algorithm at three false-alarm levels.

	False-Alarm Level					
	Low		Medium		High	
Station	Dry	Rain	Dry	Rain	Dry	Rain
Seward Avenue	0.05	0.05	0.30	0.16	0.44	0.36
Chicago Avenue	0.05	0.05	0.33	0.25	0.41	0.35
Calvert Avenue	0.05	0.10	0.31	0.30	0.43	0.45
Glendale Avenue	0.05	0.15	0.23	0.28	0.38	0.42

Table 4. Mean time lags (minutes) to detection for the most effective algorithms at three false-alarm levels.

	False-Alarm Level		
Incident Detection Algorithm	Low	Medium	High
Exponential station occupancy	1.46 (2.78)	0.74 (1.40)	0.35 (0.81)
Exponential station volume	2.18 (3.09)	0.82 (1.58)	0.30 (0.66)
Exponential station discontinuity	0.82 (1.19)	1.14 (1.78)	0.64 (1.67)
TTI station discontinuity	3.13 (5.62)	2.07 (4.05)	0.83 (1.11)

Note: Standard deviations are in parentheses.

square tests using the null hypothesis that the expected detection performance was the average performance of each pair. The exponential station occupancy and volume algorithms detected significantly more incidents than all but five other algorithms including the LAAFSCP algorithm and all the TTI algorithms except station discontinuity. The TTI station discontinuity algorithm, in turn, detected significantly more incidents than the other TTI algorithms and the LAAFSCP algorithm. All nine station algorithms together detected 74.5 percent of the incidents, a significantly better performance than the 10 subsystem algorithms (63.2 percent, $\chi^2 = 23.2$). Similarly, the combined performance of the 13 exponential algorithms (73.4 percent) was significantly better than the TTI and LAAFSCP algorithms together (62.4 percent, $\chi^2 = 12.3$).

Joint application of the exponential volume and occupancy algorithms yielded the detection of 49 of the 50 incidents, while the exponential occupancy and station discontinuity variables together detected all 50 incidents. The disadvantage of the joint application of independent models is that the false-alarm rate will also double with the use of two independent models. Alternatively, it is seen in Figure 8 that joint algorithm performances can be matched by permitting more false alarms in individual algorithms.

In summary, the most effective detection algorithms were the exponential algorithms, using either occupancy, volume, or station discontinuity as the variables to be smoothed. Information on the performances of these algorithms and the best TTI algorithm, station discontinuity, at three arbitrary false-alarm levels is given in Table 1. The corresponding detection thresholds are given in Tables 2 and 3. The low false-alarm range allows a few false alarms rather than none, and these would be tolerable in many operational detection systems. There was a significant improvement in the number of incidents detected by each algorithm from the low to the medium false-alarm level but no significant improvement in detection from the medium to the high level.

Time Lags to Detection

The most effective algorithms for detection tended to have the shortest mean time lags, eliminating the need to consider a trade-off between total detections and time lag. The relationship between the false-alarm rate and mean time lag is shown in Table 4 for the four algorithms of Table 1. Time lags generally decreased as false alarms increased, as expected, the difference being statistically significant for the combined algorithms between the medium and high false-alarm levels.

Effect of Detector Station Spacing

No significant difference was found between the station and subsystem algorithms in subsystems 2 and 3, but the station algorithms were significantly better in subsystem 1 ($\chi^2 = 22.7$). The LAAFSCP algorithm had particular difficulties in subsystem 1. Inasmuch as subsystems 1 and 3 were similar in length, length was apparently not the explanation for the differences in performance. A possibly significant factor may have been the effect of a geometric discontinuity (a reduction from 4 lanes to 3 in subsystem 1) on upstream and downstream flow states in the subsystem algorithms. In subsystem 1, only 15 of the 26 incidents displayed a flow response similar to the LAAFSCP assumptions, i.e., increases in occupancy upstream and occupancy decreases downstream in a short period of time. This pattern was observed in 22 of the 24 incidents in the other subsystems.

In subsystem 3 those incidents taking place within 1,000 to 1,500 ft (300 to 450 m) of a detector station were no easier to detect than those located midway in the subsystem. This was not the case in subsystem 1, where the subsystem algorithms experienced more difficulties with incidents located near either detector station than the more centrally located incidents. In these long subsystems, the shock waves moving upstream and downstream of an incident located near one end would be less likely to reach both stations in the same minute, but this apparently is not the explanation for the subsystem algorithm difficulties. Difficulties in subsystem 1 were not paralleled in subsystem 3, and simultaneous shock-wave impact at both stations would not appear to be essential for a subsystem algorithm. The lane-drop factor remains, and it was concluded that the range of detector spacings in this study was not a factor in algorithm detection per-

formance. In addition, the lane of occurrence was not a factor in algorithm performance in any subsystem.

An important consideration for any station algorithm is whether the detection signal consistently indicates an incident upstream or downstream of the station. Four incidents resulted in increased occupancy levels downstream that were detected by the exponential station occupancy algorithm. In every case congested operations prevailed before the incidents. Fortunately, the effect of this location error on a ramp-metering strategy response to the incident would be to divert vehicles farther downstream than necessary, a less costly outcome than allowing vehicles to enter upstream of the incident if the location error had been in the other direction. Ambiguous detections were also observed with the station discontinuity variable, where three incidents were detected downstream only because of disproportionately few vehicles in the blocked lane at the downstream station. Many incidents were detected both upstream and downstream with this variable.

Detection of Incident Termination

An algorithm was considered successful in detecting the end of an incident if it either ceased generating detection signals or produced a distinct termination signal within ±3 minutes of the end of noticeable congestion at nearby detector stations. The LAAFSCP termination logic, the restoration of occupancy to the preincident level downstream, was successful for every incident, which suggests that it should be employed with the exponential and TTI algorithms as well.

SUMMARY AND CONCLUSIONS

It has been determined that accidents and other capacity-reducing incidents that are typical occurrences on urban freeways can be detected by the flow perturbations they generate in traffic data at nearby detector stations. The following conclusions on detection algorithm effectiveness apply to prevailing volumes of 1,200 to 2,000 vph per lane and occupancies of 9 to 45 percent. Algorithms were evaluated in terms of probability of detection, probability of false alarm, time lag to detection, impact of detector station spacing on algorithm effectiveness, and detection of the end of the incident.

The technique of exponential smoothing of traffic data to better detect flow disruptions was found to be most effective, and superior to algorithms developed by TTI and LAAFSCP. The most effective traffic variables were volume (vehicles per minute), occupancy (percent, compiled over 1-minute intervals), and station discontinuity (the variable that measures the uniformity of lane kinetic energies at a detector station), for data recorded at individual detector stations. The exponential station occupancy algorithm detected 46 of 50 representative incidents at a false-alarm rate of 1.87 percent and a mean time lag of 0.74 minute after the onset of congestion. Of these incidents, 42 percent would be detected if thresholds for detection were made sufficiently stringent to eliminate most false alarms.

Algorithm effectiveness was not adversely affected by weather conditions, the given range of volumes and occupancies, the particular lane in which the incident took place, or detector spacings up to 4,815 ft (1468 m). Algorithms that used traffic data at two adjacent detector stations were less effective than those that used data from just one station and were adversely affected by a geometric discontinuity in the form of a lane drop between two of the stations. Termination of the incident was most effectively determined using the LAAFSCP termination logic, the restoration of occupancy to the preincident level downstream.

Incorporation of a detection logic in a freeway surveillance system should foster a more flexible and effective response in diverting vehicles away from congested areas. Inasmuch as the magnitude of the capacity reduction is recorded directly, this can serve as a means for determining priorities in the dispatch of aid units. The incidents studied did not permit the evaluation of the feasibility of determining type of incident from traffic-stream characteristics alone. False alarms will likely be an important operational problem, and the research demonstrated that neither the variables used nor the sudden change in variable values over time were exclusive to incident situations as op-

posed to incident-free operations. However, the algorithms should prove to be useful in alerting surveillance personnel to potential incident situations and perhaps eliminate the need, for example, for continuous television surveillance of all sections of a freeway.

ACKNOWLEDGMENT

This research was in part sponsored by the American Association of State Highway Officials in conjunction with the National Cooperative Highway Research Program. The authors express their appreciation for the cooperation and aid of various officials of the Detroit Department of Streets and Traffic, the Michigan Department of State Highways, and the University of Michigan Highway Safety Research Institute.

REFERENCES

1. Goolsby, M. E. Influence of Incidents on Freeway Quality of Service. Highway Research Record 349, 1971, pp. 41-46.
2. Lighthill, M. J., and Whitham, G. B. Hydrodynamic Approaches, Part II: On Kinematic Waves—II. A Theory of Traffic Flow on Long Crowded Roads. Proceedings of the Royal Society, Vol. 229, Series A, May 10, 1955, pp. 317-345.
3. West, J. T. California Makes Its Move. Traffic Engineering, Vol. 41, No. 4, Jan. 1971, pp. 12-18.
4. Courage, K. G., and Levin, M. A Freeway Corridor Surveillance, Information, and Control System. Texas Transportation Institute, Texas A&M Univ., College Station, Res. Rept. 488-8, Dec. 1968, 349 pp.
5. Cook, A. R., and Cleveland, D. E. The Detection of Freeway Capacity Reducing Incidents by Traffic Stream Measurements. HSRI Rept. TrS-1, Univ. of Michigan, Ann Arbor, 1970, 300 pp.
6. Whitson, R. H., Burr, J. H., Drew, D. R., and McCasland, W. R. Real-Time Evaluation of Freeway Quality of Traffic Service. Highway Research Record 289, 1969, pp. 38-50.
7. Brown, R. G. Smoothing, Forecasting and Prediction of Discrete Time Series. Prentice-Hall, Englewood Cliffs, N. J., 1963.

INCIDENT DETECTION ON URBAN FREEWAYS

Conrad L. Dudek and Carroll J. Messer,
Texas Transportation Institute, Texas A&M University; and
Nelson B. Nuckles, City of Dallas

An automatic incident-detection model using the standard normal deviate (SND) of the control variable (energy or lane occupancy) was proposed, developed, and evaluated. Two strategies were tested using a 3- and 5-minute data base for each control variable. Strategy A required one SND value to be critical; strategy B required two successive SND values to be critical. Strategy B, using lane occupancy with a 5-minute time base, produced the best results. It detected 92 percent of the 35 incidents studied during moderate and heavy flow, with a computer response time of 1.1 minutes and a 1 percent false-alarm rate during the peak period. Based on a limited sample size, the study indicated that the SND model was as effective as the composite model, which was considered to be the best existing model. Because the SND model does not require separate distribution curves for various traffic conditions, it may be a more attractive model for an operational system. Relationships were developed and presented that identify sensor spacing requirements for an incident-detection system using a station model.

•OCCURRENCE of an accident or other lane blockage incident on a freeway reduces the capacity of that section of freeway significantly below normal. Freeway incidents occur randomly, are unpredictable, and result in what is termed nonrecurrent congestion. Moskowitz (1) believes that the single most important problem in urban freeway traffic operations is the determination of a methodology to detect stopped vehicles and the necessary steps to remove the stoppage. West (2) indicates that the nonrecurring freeway congestion due to incidents is responsible for as much motorist delay in the urban area as is the recurring congestion due to geometric bottlenecks. Thus, automatic detection of incidents is a very important function of a corridor surveillance, control, and information system.

This paper is concerned with the development, testing, and evaluation of automatic incident-detection algorithms for urban freeways. The emphasis is directed toward incident detection during medium- and heavy-flow traffic conditions.

PREVIOUS WORK ON INCIDENT-DETECTION MODELS

Six approaches to the automatic detection and location of incidents on the John C. Lodge Freeway in Detroit during the peak period were explored by Courage and Levin (3). These approaches were based on vehicle storage, kinetic energy, energy differential (longitudinal), energy distribution (transverse), speed-density characteristics, and metering rates. The researchers conducted limited studies to test the feasibility of the six incident-detection approaches. It was concluded by Courage and Levin that all models demonstrated some ability to detect incidents and may, therefore, merit further consideration. The models exhibited a high false-alarm rate, and the researchers felt that considerable refinement would be required to produce an operational incident-detection scheme.

Publication of this paper sponsored by Committee on Freeway Operations.

California developed an incident-detection model for use on the Los Angeles freeway system (4). The California model consists of three sequential tests all based on occupancy changes at the upstream and downstream stations of a subsystem. An incident is signaled only when the threshold values for all three variables are exceeded, indicating that the sequence of events associated with a typical capacity-reducing incident has occurred. Like most of the models developed by Courage and Levin, the California model requires that cumulative distribution curves be drawn for each location.

Whitson et al. (5) suggested a detection model using volume as the controlling parameter. A running 5-minute mean of the flow rate was plotted with corresponding upper and lower limits that were a constant 2 standard deviations away from the mean. An incident was detected when the 1-minute flow rate fell below the lower limit for 30 seconds. This model was not fully evaluated with field data.

Cook and Cleveland (6) conducted a study on the John C. Lodge Freeway in which the California model and the first five models developed by Courage and Levin were analyzed. These latter models were referred to as the subsystem shock wave model, station energy model, subsystem energy model, station discontinuity model, and subsystem discontinuity model respectively. Cook and Cleveland also combined Courage and Levin's energy distribution model with the speed-density model into a composite model as a means of improving the reliability of detection techniques. In addition, exponential smoothing of traffic stream variables was investigated as an approach to incident detection. Because of the limited work with this latter technique, definite conclusions as to its applicability could not be reached. The effectiveness of the former seven models analyzed was determined, based on a set of 50 incidents. The results of the analysis are given in Table 1.

Cook and Cleveland observed specific weaknesses in the models studied. Particularly, inasmuch as the threshold values for Courage and Levin's models were set at a 1 percent level, the level of false alarms would also be 1 percent. The threshold levels could not be made more stringent without reducing model effectiveness. Cook and Cleveland observed that in most cases the threshold values could be estimated as 1 standard deviation away from the mean value. They suggested that a more effective method for determining the threshold may be the use of real-time estimates of the standard deviation of the parameter values. Thus, the false-alarm rate possibly could be reduced, and the thresholds would be responsive to such factors as time of day, day of week, and environmental conditions. This approach might also eliminate the need for separate frequency distribution curves for each freeway station and for different periods of the day or weather conditions.

In related work, Dudek, Messer, and Friebele (7, 8) developed control logic for automatic operation of safety warning signs at three locations on the Gulf Freeway in Houston. The control logic is not an incident-detection algorithm, but it is responsive to stoppage waves and activates the warning system when a stoppage wave is predicted or sensed downstream of one of the critical overpasses. The logic also turns the system off when the conditions on the freeway no longer warrant the alert provided by the warning system. Three control algorithms were developed using one of the following control variables: energy, speed, or lane occupancy. Each program has been successfully implemented on an operating freeway. The logic using the energy variable has been operating the safety warning system on the Gulf Freeway since April 3, 1972, and the system has responded very satisfactorily to the shock waves on the freeway. Of significance is the fact that, barring the occurrence of hardware failures, the warning system responds to shock waves with a high degree of reliability and is operating such that no false alarms are generated. These are two important criteria for establishing incident-detection algorithms as well.

FACTORS AFFECTING INCIDENT DETECTION

Incident-Detection Time

Incident-detection time can be defined as the elapsed time from the moment the incident occurs until it is detected. In an automatic incident-detection system using a station model, the detection time will be influenced by the following two components:

(a) the shock-wave travel time—the elapsed time between the occurrence of the incident and the time the shock wave crosses the upstream detector station, and (b) the computer algorithm response time—the time required for the computer algorithm to recognize an incident after the shock wave crosses the sensor station. The shock-wave travel time depends on the level of freeway operation and the spacing between the incident and the upstream detector station. The computer algorithm response time depends on the strategy used to detect the incident.

Theoretical Analysis

Messer, Dudek, and Friebele (9) have developed relationships that express the movement of shock waves as a function of speed when a freeway incident occurs and immediately following the removal of the incident. These relationships are shown in Figure 1.

The representation illustrates that when an incident occurs on a freeway, the first shock wave that generates upstream will travel at a speed W_{u1}. The speed of the metered wave that travels downstream is W_{d1}. The shock wave and the metered wave are the boundary vectors emanating upstream and downstream from point A, which defines the beginning of the incident.

After a time T has elapsed, the incident is assumed to be completely removed from the freeway. When the incident is removed (point B), the capacity of the freeway is increased to normal, and the vehicles stored upstream of the incident site then begin to travel downstream. The flow of these vehicles out of the downstream end of the congested queue also begins to shorten or clear up the queue upstream of the incident site. Associated with removal of the incident is the movement upstream of the capacity flow wave at a speed W_{u2}. Likewise, a wave that defines the boundary between the capacity flow and the metered regions moves downstream from the site of the incident (when it is removed) at a speed W_{d2}.

As shown in Figure 1, one remaining wave occurs before the freeway traffic conditions return to normal. Sometime after the incident is removed, the capacity flow wave W_{u2} will catch the shock wave W_{u1}, and the congested queue will have been dissipated. At this point, the final clearing wave forms and begins to move downstream at a speed W_{d3}. This wave defines the boundary between the high-density capacity flow region and normal traffic flow. It is important to note that the location where W_{u1} and W_{u2} intersect defines the maximum distance from the incident that the sensors can be positioned to detect the shock wave for the given duration of incident. Thus, if the first sensor station were farther upstream, the incident would probably not be detected regardless of the efficiency of the computer algorithm.

Therefore, the two important variables, as far as sensor spacing is concerned, are the speed of the shock wave moving upstream W_{u1} when an incident occurs and the speed of the capacity flow wave W_{u2}. The shock-wave speed can be determined from the following relationship (9):

$$W_{u1} = -u_f + u_n + u_q \tag{1}$$

where

u_f = mean free speed,
u_n = normal speed before the incident, and
u_q = average speed within the congested queue.

The capacity flow wave can be expressed as follows (9):

$$W_{u2} = -\frac{u_f}{2} + u_q \tag{2}$$

Negative values in the expression for W_{u1} and W_{u2} indicate that the waves are moving upstream.

Table 1. Review of individual detection model performance (6).

Model	Specific to Incidents	Average Time Lag for First Detection (minutes)	Incidents Detected (percent)	False-Alarm Rate (percent)
Composite	Yes	0.81	96	2
Station discontinuity	Yes	2.07	90	1
Subsystem discontinuity	No (bottlenecks, also)	2.14	74	1
Subsystem energy	No (bottlenecks, also)	5.83	58	1
Station energy	No	2.58	56	1
California	Yes	0.96	52	~0.1
Subsystem shock wave	No	3.06	32	1

Figure 1. Model of freeway traffic conditions due to an accident (9).

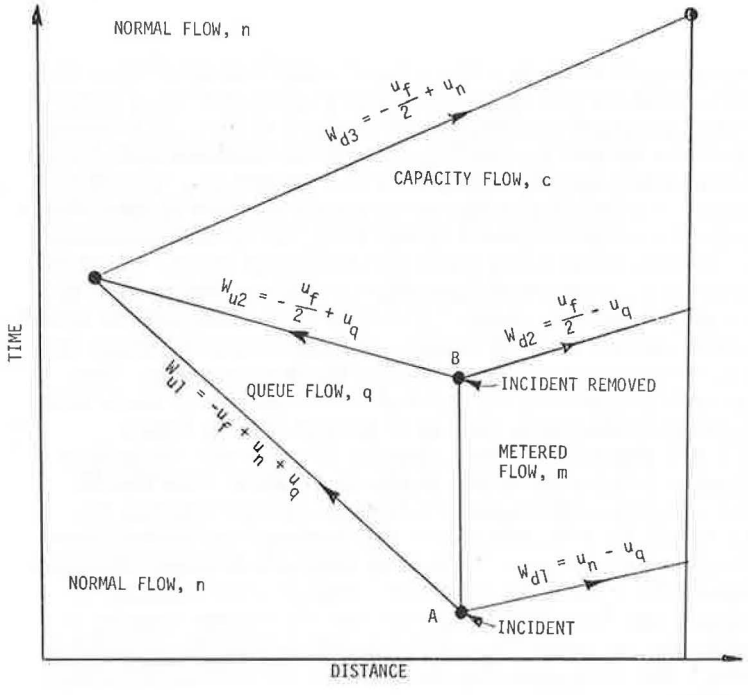
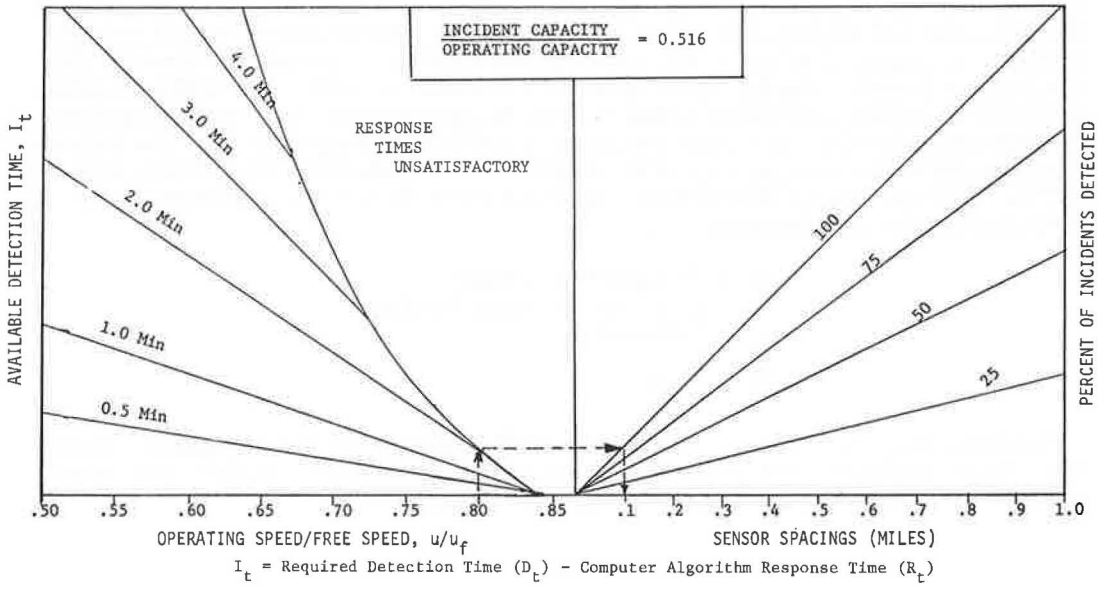


Figure 2. Sensor spacings for incident duration of 2 minutes or more.



It has also been shown that the average speed within the congested queue u_q can be expressed as a function of the flow during incident conditions and the available capacity under normal conditions as follows (9):

$$u_q = \frac{u_f}{2}\left[1 - \sqrt{1 - \frac{q}{q_m}}\right] \quad (3)$$

where

q = the flow under incident conditions and
q_m = the available capacity under normal conditions.

Sensor Spacing

The maximum sensor spacing required to detect the shock wave resulting from an incident is the intersect distance between the shock wave traveling upstream at a speed of W_{u1} and the capacity flow wave propagating upstream at a speed of W_{u2}. It is noted, from Eqs. 1, 2, and 3, that the wave speeds W_{u1} and W_{u2} are functions of the flow during incident conditions q and the available capacity under normal condition q_m. The intersection distance of the two waves is a function of the wave speeds and the incident duration. Thus, the lower the level of service is before an incident, the farther upstream the intersection of the two waves will occur for a given duration of incident. Likewise, the longer the incident duration for a given set of operating conditions, the farther upstream the intersection of the two waves will occur. It is therefore necessary to select the minimum duration of incident and the level of freeway operations during which the incident must be detected in order to determine sensor spacing requirements. For example, the operating agency may decide that all incidents of 2 minutes or more must be detected while the freeway is operating at speeds up to 50 mph (80 km/hour).

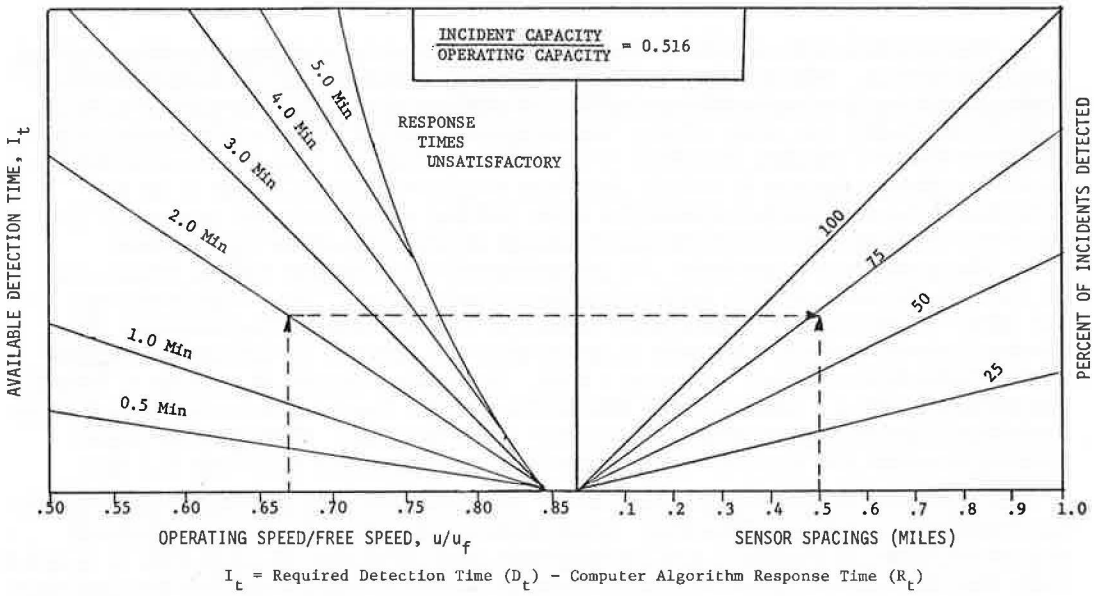
Another criterion that must be established is the required incident-detection time. Because incident-detection time is influenced by the shock-wave travel time and the computer algorithm response time, these two factors will affect sensor spacing requirements. The shock-wave travel time is affected by the freeway operational characteristics previously described. The computer algorithm response time is a function of the particular algorithm selected for incident detection. For a given duration of incident, freeway level of service, and incident-detection time, the sensor spacing requirements will be influenced by the response time of the incident-detection algorithm. Thus, if the algorithm provides a fast response, the sensors can be positioned farther apart than for a sluggish algorithm.

From Eqs. 1, 2, and 3, two graphs were developed that relate maximum sensor spacing to normal operating speed, required incident-detection time, and percentage of incidents that will be detected, based on characteristics observed on the Gulf Freeway in Houston. Figure 2 shows sensor spacings for incidents of 2 minutes or more; Figure 3 shows sensor spacings for incidents of 4 minutes or more. The available detection time (available shock-wave travel time) I_t is the difference between the required incident-detection time D_t and the computer algorithm response time R_t. The assumption inherent in the two figures is that incidents occur randomly and uniformly over the freeway section. These figures were developed using the following characteristics measured on the Gulf Freeway:

u_f = 60 mph (97 km/hour)
q = 2,880 vph (one-lane blockage)
q_m = 5,560 vph
$\frac{q}{q_m}$ = 0.516

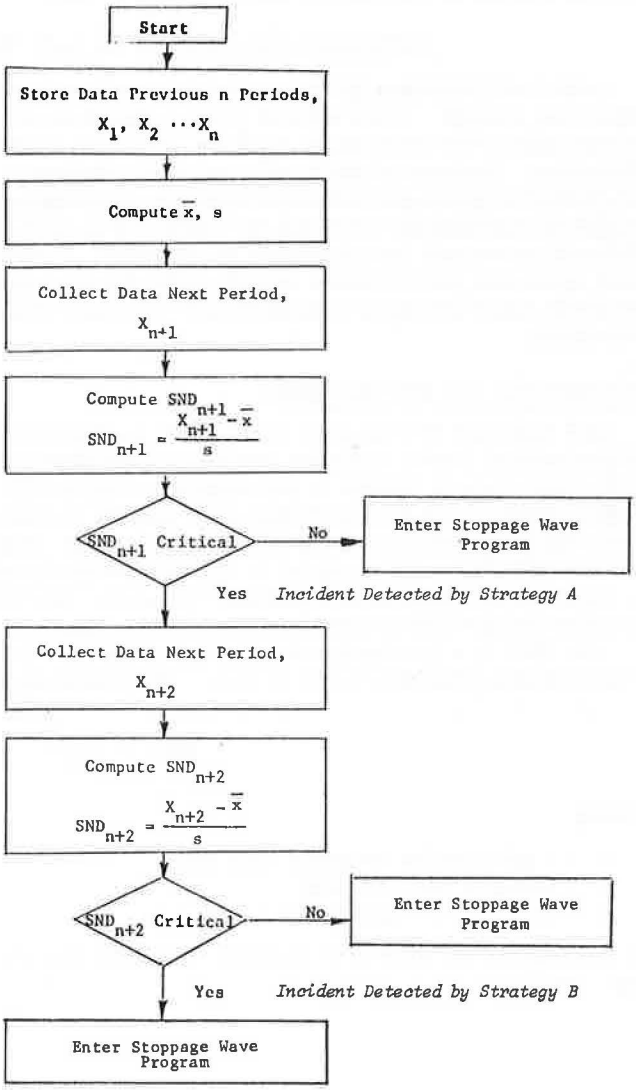
Two hypothetical examples are presented that illustrate the use of Figures 2 and 3. Both examples assume that the computer algorithm can detect an incident as soon as the shock wave reaches the upstream sensors, that is, R_t = 0:

Figure 3. Sensor spacings for incident duration of 4 minutes or more.



I_t = Required Detection Time (D_t) - Computer Algorithm Response Time (R_t)

Figure 4. Schematic of incident detection strategies A and B.



1. The problem is to determine the maximum sensor spacing for a freeway incident detection system. The system should be capable of detecting all incidents blocking a freeway lane for 2 or more minutes while the freeway is operating at speeds up to 48 mph (77 km/hour), i.e., $u/u_f = 0.80$. The system should be capable of detecting the incidents within 2 minutes after they occur. From Figure 2, it is determined that the maximum sensor spacing to satisfy the above requirements is 0.1 mile (0.16 km). Note that if the sensors are spaced 0.4 mile (0.6 km) apart, only 25 percent of the incidents would be detected within the 2-minute incident-detection requirement.
2. The problem is to estimate the percentage of incidents that will be detected by an incident-detection system having a given sensor spacing. The sensors are spaced 0.5 mile (0.8 km) apart. The system should be capable of detecting incidents that block a freeway lane for 4 minutes or more while the freeway is operating at speeds up to 40 mph (64 km/hour), i.e., $u/u_f = 0.67$. The system should be capable of detecting the incidents within 2 minutes after they occur. From Figure 3, it is observed that the system will detect approximately 70 percent of the incidents. Note that the sensor spacing to detect 100 percent of the incidents is approximately 0.37 mile (0.6 km).

The foregoing results apply to only those freeways that have the same traffic operating conditions as the Gulf Freeway. When these conditions are different, separate curves as shown in Figures 2 and 3 would need to be developed by using Eqs. 1, 2, and 3. Also, this discussion applies to station incident-detection models that use the upstream sensor station to analyze the discontinuity in flow.

INCIDENT-DETECTION MODEL DEVELOPMENT

Cook and Cleveland (6) suggested inherent weaknesses in the existing incident-detection models. Most models require development of frequency distributions for the measured traffic variable as the procedure for identifying threshold values for incident detection. Because of the hourly and daily variations in traffic flow and the effects attributed to pavement and environmental conditions, several frequency distributions would be required for each set of conditions at each freeway sensor station. It may be difficult to account for all variables involved. If threshold values are set at the 1 percent level, the level of false alarms will also be 1 percent. The threshold values cannot be made more stringent without reducing model effectiveness in terms of detecting incidents.

Standard Normal Deviate Model

One approach to circumvent the given weaknesses is to consider the rate of change of the control traffic variable rather than an absolute threshold value. It is hypothesized that a high rate of change in the control variable will reflect an incident situation as distinguished from a normal demand-capacity problem caused by geometrics. The statistic proposed for the control function is the standard normal deviate (SND) of the control variable. The concept is to evaluate the trends in the control variable (e.g., occupancy, energy) and to recognize when the variable changes rapidly in relationship to expected changes caused by normal fluctuations in traffic flow.

The SND is a standardized measure of the deviation from the mean in units of the standard deviation and is expressed by the following relationship:

$$SND = \frac{x - \bar{x}}{s} \tag{4}$$

where

x = a given value from the data set,
$\bar{x}$ = mean of data set, and
s = standard deviation of data set.

In the application of SND to incident detection, the above variables take on the following meaning:

x = value of control variable at time t,
$\bar{x}$ = mean of control variable over previous n sampling periods, and
s = standard deviation of control variable over previous n sampling periods.

The value of SND will thus reflect the degree to which the control variable has changed in relationship to the average trends measured during previous intervals. A large SND value would reflect a major change in operating conditions on the freeway.

The overall incident-detection concept suggested in this paper is to incorporate an incident-detection algorithm with a stoppage-wave-detection algorithm previously developed for operation of the safety warning devices on the Gulf Freeway (7, 8). When stoppage waves are detected by the latter algorithm, each wave will be analyzed by the incident-detection algorithm to ascertain whether the waves resulted from a freeway incident (e.g., accident, stalled vehicle) in contrast with a geometric bottleneck. The model proposed is a station model that detects incidents that occur downstream of the sensor station.

Operational Approaches

Two operational approaches or strategies are identified and evaluated. Strategy A requires one SND value to be critical; strategy B requires two successive SND values to be critical. These strategies are consistent with techniques developed in earlier research on incident detection. Schematics of strategies A and B are shown in Figure 4.

STUDY METHOD

The SND incident-detection model was tested and evaluated on the Gulf Freeway in Houston. The facility is a 6-lane freeway with surveillance and control implemented in the inbound direction only. For the purposes of this study, five inbound freeway locations having double-loop sensors on each lane, as shown in Figure 5, were used to evaluate the model.

Lane occupancy and energy were evaluated as control variables. Energy was computed from volume and speed measurements.

The SND models were tested using 3- and 5-minute time bases. The first method used data from the previous 3-minute sampling periods to compute the mean $\bar{x}$ and standard deviation s. The second method considered parameters from the previous 5 minutes. With the two strategies A and B, two variables of energy and occupancy, and two time bases, a total of eight combinations were tested.

Data were collected during 35 incidents that occurred on the inbound section of the Gulf Freeway. Three peak-hour periods (7 to 8 a.m.) in which no incidents were observed were used for the investigation of false alarms. These peak hours had many slowdowns and stoppage waves that provided a good test for the SND model. The percentage of false alarms in this paper refers to the percent of analysis time periods during operation that a false alarm is indicated. For example, because the computer evaluated the system once each minute, two false alarms during a 2-hour operational period would result in a false-alarm percentage of $(2/120)(100) = 1.7$ percent.

RESULTS OF SND MODEL ANALYSIS

Characteristics of Incidents Evaluated

Because the duration of an incident and the operating conditions on the freeway both influence the capabilities for incident detection, it was of particular interest to evaluate the characteristics of the 35 incidents. Cumulative distributions of the duration of incidents and the operating speed/free speed ratio (u/u_f) are shown in Figures 6 and 7 respectively. The results show that 11 percent of the incidents studied blocked a freeway lane for a duration of 2 minutes or less. Approximately 90 percent of the incidents occurred when the freeway was operating at or below 50 mph (80 km/hour), i.e., $u/u_f = 0.83$.

Figure 5. Schematic of Gulf Freeway.

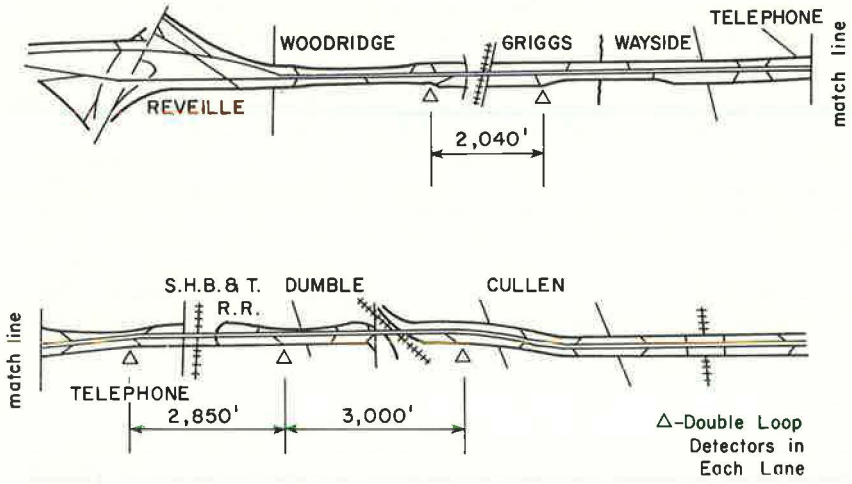


Figure 6. Duration of incidents investigated.

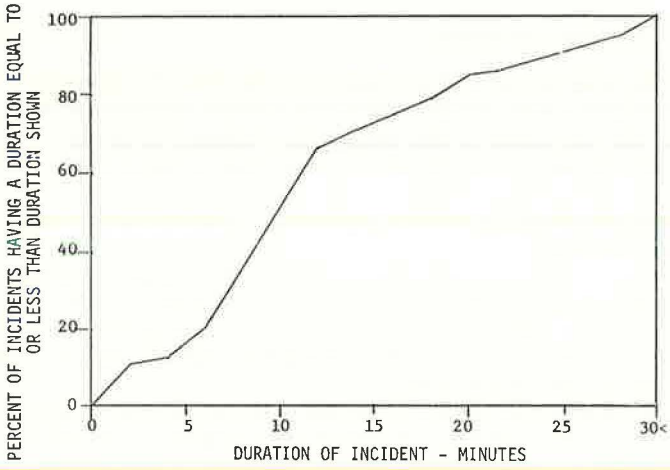
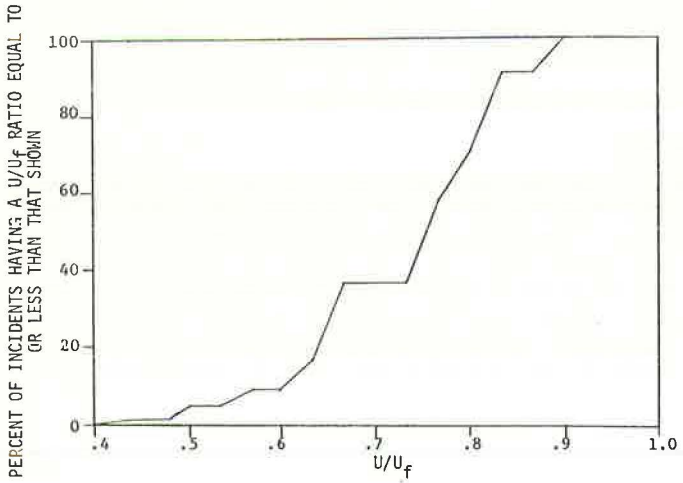


Figure 7. Operating speed/free speed ratio of incidents investigated.



Effectiveness of SND Models

The effectiveness of the incident-detection models can be evaluated in part by the percent of incidents detected and the frequency of false alarms. Cumulative distributions of the percent of incidents detected and the percent of false alarms for each strategy, variable, and time base tested were developed. Distributions using strategy B with lane occupancy as the control variable with a 5-minute time base are shown in Figure 8.

A study of the distributions indicated that there is probably an optimum SND value that can be used for each strategy. One would need to trade off incident-detection capabilities with false alarms. The authors decided that SND values producing results approaching 90 percent incident detection and 1 percent false alarms would be the critical SND values. Based on this selection, the performance of the strategies was evaluated and is given in Table 2.

With strategy A, which required only one SND value to be critical, the occupancy and energy variables both detected 86 percent of the incidents studied. The performance of the occupancy variable was considered somewhat better, however, because of the lower frequency of false alarms. Changing the time base seemed to have little effect on the performance of the variables.

Strategy B, which required two successive SND values to be critical, resulted in a higher percentage of incidents detected using occupancy and a lower percentage using energy in comparison to strategy A. Both variables resulted in a lower frequency of false alarms. Changing the time base did not affect the performance of the energy variable. However, a larger time base using the occupancy variable resulted in a higher percentage of incidents detected.

A review of Table 2 shows that strategy B, using a 5-minute time base with lane occupancy as the control variable, resulted in the best performance. This approach detected 92 percent of the 35 incidents with an average computer response time of 1.1 minutes. The false-alarm level during the peak period was 1.3 percent.

Comparison of SND Model to Other Detection Models

It was interesting to compare the SND model to existing incident-detection models to evaluate their relative performances. It is not appropriate to use the results reported in the literature directly because the conditions are different from those on the Gulf Freeway. In particular, the sensor spacings and the relative location of the incidents to the sensors do not compare with the Gulf Freeway data. The conditions must be the same to permit a fair comparison. Unfortunately, during the early stages of the research reported here, sufficient data were not collected for the other models. As of this writing, data for only 26 incidents are available for proper comparison. Therefore, these results should be considered as provisional and not necessarily conclusive.

So that a good comparison could be realized, distribution curves similar to those discussed by Courage and Levin (3) and Cook and Cleveland (6) were developed at each sensor at the five sensor stations on the Gulf Freeway for the most effective models reported in the literature. These distribution curves permitted the authors to select critical values that would result in approximately the same percentage of false alarms for each model. The percent of incidents detected was then determined using these critical values.

A comparison of the SND model to the existing incident-detection models is presented in Table 3. The results indicate that the SND model is as effective as the composite model. The SND and composite model detected 26 and 25 incidents while operating at a 1 percent level of false alarms during the peak period. The station discontinuity and subsystem discontinuity models detected 22 and 23 incidents while operating at a 1 percent level of false alarms during the peak period. Inasmuch as the SND model does not require separate distribution curves for various traffic conditions, it appears that it may be a more attractive model for an operational system.

Figure 8. Performance curves using lane occupancy, strategy B, and 5-minute time base.

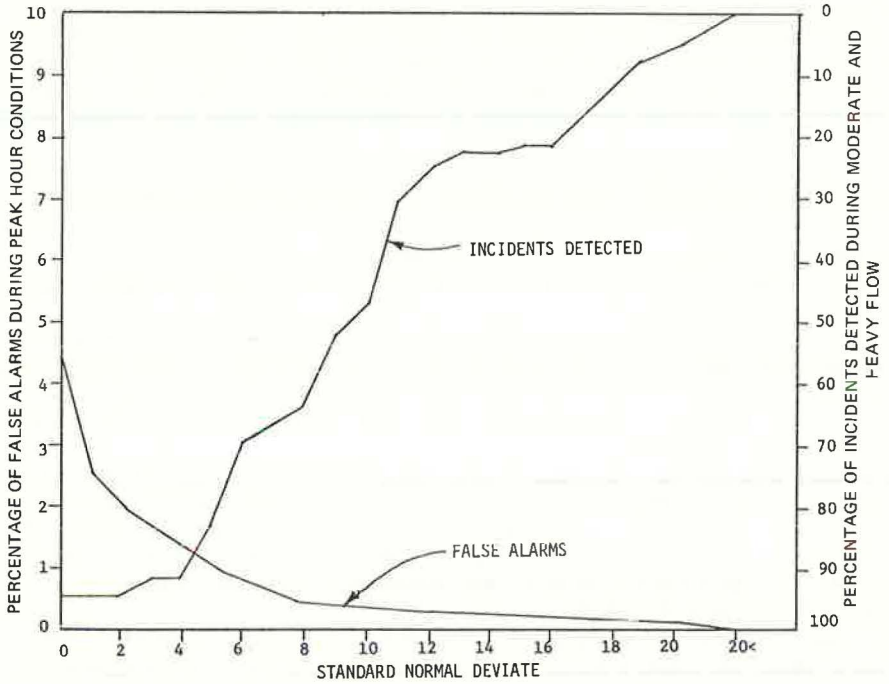


Table 2. Effectiveness of SND detection strategies.

Strategy	Variable	Time Base (minutes)	Critical SND Values	Average Computer Response Time (minutes)	Standard Deviation Computer Response Time	Incidents Detected[a] (percent)	False Alarms[b] (percent)
A	Occupancy	5	6	0.5	1.1	86	1.7
	Occupancy	3	6	0.7	1.9	86	2.0
	Energy	5	-4	0.8	2.6	86	2.4
	Energy	3	-3	0.3	0.7	86	2.5
B	Occupancy	5	4	1.1	0.6	92	1.3
	Occupancy	3	4	1.1	1.5	89	1.4
	Energy	5	-3	1.1	0.5	83	1.4
	Energy	3	-3	1.1	0.5	83	1.4

[a]During moderate and heavy flow conditions. [b]During peak periods.

Table 3. Comparison of incident-detection models.

Model	Control Variable	False Alarms per Station (percent)	False Alarms per Subsystem (percent)	Number of Incidents Detected[a]
Station discontinuity	Energy	1	-	22
Subsystem discontinuity	Energy	-	1	23
Composite	Energy	½	½	25
SND, strategy B, 5-minute time base	Occupancy	1	-	26

[a]Based on 26 observed incidents during peak and off-peak periods.

SUMMARY AND DISCUSSION OF RESULTS

The results of the study indicated that the 5-minute occupancy SND model using strategy B produced the best results of the strategies and variables tested. The model detected 92 percent of the 35 incidents that occurred during moderate and heavy flow and operated with 1.3 percent false alarms during the peak period on the Gulf Freeway system. Although the control parameter could be changed to detect a higher percentage of incidents on the existing system, the desire for this capability would be at the expense of a higher frequency of false alarms.

The failure by the model to detect all the incidents could conceivably be caused from one or a combination of the following: (a) failure in the model logic, (b) very short duration of an incident, (c) sensor spacing, and (d) a high operating speed before an incident. The SND model depends on the passage of a shock wave over a set of sensors. If the shock wave passes over the sensors, its effect must be noticeable. As discussed earlier, the duration of the incident, the sensor spacing, and the normal operating speed of the freeway before occurrence of the incident affect the passage of the shock wave over the sensors.

An analysis of the data revealed that, of the three incidents missed by the 5-minute occupancy SND model, one incident blocked the freeway for only 2 minutes. The relative location of the incident to the upstream detectors was a factor in the lack of detection. The other two incidents that were not detected occurred when the operating speeds were 48 and 53 mph (77 and 85 km/hour) respectively. The relative degree of queuing was lessened and contributed to the model's failure.

It is important to emphasize that the efficiency of the strategies using the SND model evaluated in this paper apply to the Gulf Freeway with the given sensor spacings shown in Figure 5. It is the opinion of the authors that the 5-minute occupancy model is capable of detecting close to 100 percent of the incidents of 2-minute duration or more during moderate and heavy flow if the sensor spacing were adequate.

The authors wish to reemphasize that the inability of the SND and other incident-detection models to detect all the incidents is not necessarily a reflection of the individual model inadequacies. The duration of an incident, the sensor spacing, the relative location of the incident to the sensors, and the operating conditions immediately before the incident are all important factors that affect the capabilities of any incident-detection model. Therefore, the results of incident-detection model capabilities reported in the literature must be placed in proper perspective.

It is the belief of the authors that, with proper sensor spacing, it is possible to approach 100 percent detection of incidents blocking a freeway lane for a duration equal to or greater than a preselected time during moderate or heavy flow. One of the most perplexing problems is the frequency of false alarms during the peak periods. A false-alarm rate of 1 percent outwardly appears insignificant. However, the false-alarm rate applies to each sensor station or each subsystem (depending on the model). Therefore, the number of false alarms can become very significant in an operational system. Greater effort should be directed toward minimizing and ideally eliminating the false-alarm problem.

ACKNOWLEDGMENT

This paper discusses one phase of a research project on the development of urban traffic management and control systems conducted by the Texas Transportation Institute and the Texas Highway Department in cooperation with the Federal Highway Administration. This paper reflects the views of the authors, who are responsible for the facts and the accuracy of the data presented. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration. This paper does not constitute a standard, specification, or regulation.

REFERENCES

1. Moskowitz, K. Analysis and Projection of Research on Traffic Surveillance, Communication, and Control. NCHRP Report 84, 1970.

2. West, J. Proposed Real-Time Surveillance and Control System for Los Angeles. Paper presented to HRB Committee on Freeway Operations, Aug. 1969.
3. Courage, K. G., and Levin, M. A Freeway Corridor Surveillance Information and Control System. Texas Transportation Institute, Texas A&M Univ., College Station, Res. Rept. 488-8, Dec. 1968, 349 pp.
4. Schaefer, W. E. California Freeway Surveillance System. California Division of Highways, Department of Public Works, Nov. 1969.
5. Whitson, R. H., Buhr, J. H., Drew, D. R., and McCasland, W. R. Real-Time Evaluation of Freeway Quality of Traffic Service. Highway Research Record 289, 1969, pp. 38-50.
6. Cook, A. R., and Cleveland, D. E. The Detection of Freeway Capacity Reducing Incidents by Traffic Stream Measurements. HSRI Preliminary Rept. TrS-1, 1970. See also their paper in this Record.
7. Dudek, C. L., Messer, C. J., and Friebele, J. D. Investigation of Lane Occupancy as a Control Variable for a Safety Warning System for Urban Freeways. Texas Transportation Institute, Res. Rept. 165-6, March 1973.
8. Dudek, C. L., and Messer, C. J. Detecting Stoppage Waves for Freeway Control. Highway Research Record 469, 1973, pp. 1-15.
9. Messer, C. J., Dudek, C. L., and Friebele, J. D. Method for Predicting Travel Time and Other Operational Measures in Real-Time During Freeway Incident Conditions. Highway Research Record 461, 1973, pp. 1-16.

PERFORMANCE OF VOLUNTEER MONITORS USING CITIZENS BAND RADIO FOR A HIGHWAY COMMUNICATIONS SERVICE

William G. Trabold, General Motors Research Laboratories, Warren, Michigan; and Gerald H. Reese, REACT National Headquarters, Chicago

A 2-year study of the performance of Ohio REACT volunteer monitors using Citizens Band (CB) radio to provide a highway and emergency communication service has been completed. The report describes how CB radio is used for aid and information purposes. Measured performance data are used to analyze monitoring coverage in the state. It is shown that in Ohio volunteer CB monitors annually contribute a public service having an economic value of approximately $10.2 million.

•REACT (radio emergency associated citizens teams) is a nationwide organization of volunteer monitors who use the official U.S. emergency channel, No. 9, in the Citizens Radio Service band to provide assistance to motorists and others in need.

The Ohio REACT program was initiated in June 1970, in cooperation with the Ohio State Highway Patrol, to measure the performance and evaluate the potential of a volunteer-operated radio link as a highway communication service. The program proved conclusively that thousands of volunteers, donating time worth an estimated $10 million while providing emergency highway communications, are a major public service resource. The Ohio study represents only the tip of the iceberg, because some 40,000 REACT volunteers around the country use Citizens Band (CB) two-way radios to communicate pleas for assistance to monitoring REACT teams. The program is significant in that Class D Citizens Radio Service (CB) provides the largest possible user population with which to study a highway communications system using two-way radio from the motorist to an emergency monitor. The Citizens Radio Section of the Electronic Industries Association (EIA) claims that one out of every 60 automobiles and one out of every 10 recreational vehicles are equipped with CB radio. The Federal Communications Commission (FCC) states that approximately 800,000 licenses are active. Overall, CB radio provides a low-cost, widely deployed means of providing the motorist with a two-way communications system.

Our objective here is to report the phase 2 performance of the Ohio REACT program (2). Phase 2 extended from September 1, 1971, through August 31, 1972, second and final year of the program. Phase 2 differed from phase 1 (1, 2, 3) in two respects. The call log form used by REACT monitors was expanded to provide documentation of more categories. This step was undertaken to better define the "other incident" category that accounted for 22.3 percent of all calls reported during phase 1. The "action taken" category was expanded for the same reason.

In addition to changes in the call log form, a monitoring hours form was introduced to document how comprehensively REACT teams were achieving their primary goal of monitoring the national emergency channel 24 hours per day. This is the first attempt that has been made to directly document monitoring hours and is the most interesting aspect of this report.

The role of the state highway patrol has been vital to the success of the program. When the program started in 1970, 23 of the patrol's 57 posts were equipped with CB radio. Today, 40 posts are so equipped. The REACT teams have been primarily responsible for this growth, because they donate and install the equipment. State highway

Publication of this paper sponsored by Committee on Communications.

patrol dispatchers monitor these radios on a volunteer basis. Calls received by the state highway patrol directly are not included in this report. More important than any monitoring conducted by the patrol has been their enthusiastic support of the program, which continues to be a source of encouragement to REACT volunteers in Ohio.

TEAM PARTICIPATION

There are 80 REACT teams located in 75 cities in Ohio. These teams are distributed according to city population as shown in Figure 1. The distribution has a pronounced mode with 40 percent of the teams located in cities with populations of 10,000 to 50,000. Team participation in the program is given in Table 1. Of the 80 teams, 52, or 65 percent, consistently participated in the reporting process. In view of the fact that this is a completely voluntary operation with no monetary compensation, we judge this to be good participation. The other 28 teams accounted for less than 4 percent of the call reports and less than 2 percent of the reported gross monitored hours. From the monitoring hours reports, we were able to document the participation of 889 individuals in the program. This number is about 10 percent below that expected because husband and wife pairs reported their monitoring hours on a common form.

CALL REPORTS

The purpose of the call reports was to document how motorists make use of CB radio to satisfy their needs for aid and information. A completed report form is shown in Figure 2. The forms, as received from the monitors, were used by keypunch operators to prepare record cards. One record was prepared for each column of information with the boxed header information common to each record. The records were then machine-tabulated by entry code (an accident is code 20) and by month. The total calls reported were also tabulated according to team number and month.

Type of Incident

A total of 14,750 calls were documented. The callers used CB radio for the purposes given in Tables 2 and 3. About 82.8 percent of the calls were road-related, and 21.96 percent were non-road-related. These data are approximate because 4.78 percent of the calls involved more than one incident category. The statistics on road-related uses are consistent with previously published data (1, 3). The significant categories of accidents, occupied stalled vehicles, and requests for road information account for 58.2 percent of all calls or 71 percent of all road-related calls.

Documentation of non-road-related calls has not been done before by us. There is a residual unidentified call category that accounts for 7.39 percent of all calls, and we believe that most of these calls are time checks and information exchanges between monitors. The net non-road, non-trivial use of CB radio by the callers accounted for 14.57 percent of all calls.

Sources of Calls

The sources of calls are given in Table 4. Note that 76.14 percent of all calls (85.62 percent of the calls that specified source) originated from a passerby, mobile REACTer, or trucker. These latter two categories might also be classified as passersby. In only 9.18 percent (10.32 percent of calls that identified the source) was the caller involved in the incident.

Time to Complete Calls

Figure 3 shows the distribution of time in minutes to complete call transactions. This is the time from which the call was received by the monitor until appropriate action was initiated and the caller was advised of the action taken. The most frequently reported time to complete a transaction is 1 minute or less. We calculate that 84.9 percent of the transactions are reported as complete in 5 minutes or less. Other researchers (4) have estimated that the use of CB radio on the Detroit expressway system saves an average of 17 minutes in detection time.

Figure 1. Distribution of Ohio REACT teams according to city population.

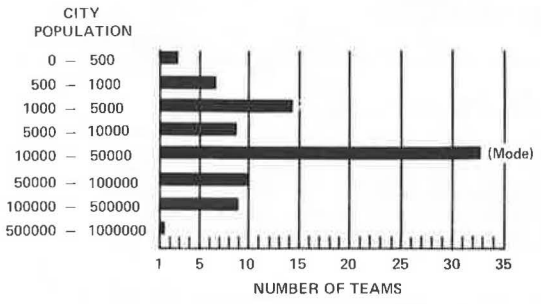


Table 1. Participation of REACT in phase 2.

Participation	Number of Teams	Percent of Teams
None (neither call nor hour reports)	7	0.88
Call but no hour reports	9	11.25
Trivial hour reporting[a]	12	15.00
Consistent participation	52	65.00
Total	80	
Teams reporting any monitored hours	64	

[a]Monitoring of less than 10 percent of the available time (8,784 hours) was reported.

Table 2. Call reports by type of road-related incident.

Incident	Total Calls	Percent
Accident count	3,476	23.57
Vehicle count	6,618	
With injuries	579	
With fatalities	56	
Stalled vehicle, occupied	2,497	16.93
Stalled vehicle, unoccupied	817	5.54
Abandoned vehicle, no plates	94	0.64
Road obstruction or traffic hazard	1,216	8.24
Major traffic jam	310	2.10
Traffic control equipment malfunction	707	4.79
Reckless or drunk driver	351	2.38
Request for road information	2,611	17.70
Vehicle fire	134	0.91
Total		82.82

Table 3. Call reports by type of non-road-related incident.

Incident	Total Calls	Percent	Incidents	Total Calls	Percent
Aircraft accident	9	0.06	Natural disaster	4	0.03
Alarm ringing	35	0.24	Non-vehicle fire	87	0.59
Animal on road	91	0.62	Persons fighting	52	0.35
Boating emergency	14	0.09	Railroad accident	9	0.06
Civil disturbance	43	0.29	Red Cross business	22	0.15
Crime report	175	1.19	Relay personal call	568	3.85
Dead animal	73	0.49	Severe weather	116	0.79
Explosion	17	0.12	Street lights out	59	0.40
Family emergency	86	0.58	Telephones out	59	0.40
Flood	39	0.26	Unconscious person	44	0.30
Gas leak	10	0.07	Vandalism	137	0.93
Industrial accident	5	0.03	Water leak	32	0.22
Medical emergency	103	0.70	Wires down	80	0.54
Missing child	116	0.79	None of these	1,090	7.39
Missing person	64	0.43	Total		21.96

Table 4. Sources of call reports.

Source	Total Calls	Percent
Caller involved in incident	1,354	9.18
Passerby	7,212	48.89
REACTer	3,751	25.43
Police	259	1.76
Base station	273	1.85
Trucker or commercial vehicle	269	1.82
Total		88.93

Figure 2. Call report form.

TEAM NAME: Summit County REACT TEAM # C450

TEAM ADDRESS: 865 Douglas Rd OPERATOR'S NAME: Lois Prokey UNIT # 3

COUNTY: Summit STATE: Ohio ZIP 44306 BASE STA. [X] OR MOBILE [ ]

		CALL	CALL	CALL	CALL	CALL
ENTER DATE OF CALL	(MO/DA/YR)	4/21/72	4/21/72	4/21/72	4/21/72	4/22/72
ENTER TIME (24 Hr Clock) OF FIRST CONTACT	(XXXX)	0935	1035	1605	2345	1130
TYPE OF INCIDENT:						
ACCIDENT	(20)		20			
THE ACCIDENT INVOLVED INJURY/S	(21)					
THE ACCIDENT INVOLVED FATALITY/S	(22)					
ENTER THE NUMBER OF VEHICLES INVOLVED	(XX)		3			
STALLED VEHICLE - OCCUPIED	(23)	23		23		23
STALLED VEHICLE - UNOCCUPIED	(24)					
ABANDONED VEHICLE - PLATES REMOVED	(25)					
ROAD OBSTRUCTION AND/OR TRAFFIC HAZZARD	(26)					
MAJOR TRAFFIC JAM	(27)		27			
TRAFFIC CONTROL EQUIPMENT MALFUNCTION	(28)					
RECKLESS AND/OR DRUNK DRIVER	(29)					
REQUEST FOR ROAD INFORMATION	(30)				70	
VEHICLE ON FIRE	(31)					
OTHER INCIDENT (Enter Code #, See List)	(XX)					
TYPE OF ROAD:						
LIMITED ACCESS (Expressway)	(32)		32	32	32	32
TOLL ROAD (Turnpike)	(33)					
URBAN (City Street)	(34)	34				
RURAL (Out-City Street/Road)	(35)					
SOURCE OF CALL:		KDY	KDY	KAK	KDZ	KES
CALLER'S SIGN	(XXXXXXX)	6054	7040	2975	1538	3559
CALLER INVOLVED IN THE INCIDENT	(36)					36
PASSERBY	(37)		37	37	37	37
REACTer	(38)	38				
POLICE	(39)					
BASE STATION	(40)					
TRUCKER AND/OR COMMERCIAL COMPANY	(41)					
ACTION TAKEN BY YOU:						
NOTIFIED STATE HIGHWAY PATROL (State Police)	(42)					
NOTIFIED CITY POLICE	(43)	43	43	43		
~~NOTIFIED SHERIFF~~	(44)					
NOTIFIED FIRE DEPARTMENT	(45)					
NOTIFIED HIGHWAY DEPARTMENT	(46)					
NOTIFIED HOSPITAL AND/OR AMBULANCE	(47)					
CALLED SERVICE STATION AND/OR WRECKER	(48)					48
CALLED UTILITY COMPANY	(49)					
CONTACTED OTHER REACT TEAM	(50)					
COMPLETED PERSONAL CALL FOR CALLER	(51)					
GAVE PERSONAL SERVICE TO MOTORIST	(52)	52				
GAVE CALLER INFORMATION	(53)					
NONE	(54)					
ACTION COMPLETED VIA LANDLINE (TELEPHONE)	(55)		55	55		55
ACTION COMPLETED VIA CB RADIO	(56)	52			52	
TIME (24 Hr. Clock) MOTORIST NOTIFIED	(XXXX)	0940	1035	1610	2350	1135

"OTHER INCIDENTS" CODE LIST:

AIRCRAFT ACCIDENT	(60)	EXPLOSION	(67)	MISSING PERSON/S	(74)	SEVERE WEATHER	(81)
ALARM RINGING	(61)	FAMILY EMERGENCY	(68)	NATURAL DISASTER	(75)	STREET LIGHTS OUT	(82)
ANIMAL ON ROAD	(62)	FLOOD	(69)	NON-VEHICLE FIRE	(76)	TELEPHONES OUT	(83)
BOATING EMERGENCY	(63)	GAS LEAK	(70)	PERSONS FIGHTING	(77)	UNCONSCIOUS PERSON	(84)
CIVIL DISTURBANCE	(64)	INDUSTRIAL ACCIDENT	(71)	RAILROAD ACCIDENT	(78)	VANDALISM	(85)
CRIME REPORT	(65)	MEDICAL EMERGENCY	(72)	RED CROSS BUSINESS	(79)	WATER LEAK	(86)
DEAD ANIMAL	(66)	MISSING CHILD	(73)	RELAY PERSONAL CALL	(80)	WIRES DOWN	(87)
						NONE OF THESE	(88)

Distribution of Teams by Number of Calls

Figure 4 shows the distribution of teams according to the number of call reports processed. Twenty teams reported 51 to 100 calls. The next most frequent statistic was 1 to 20 calls reported by 18 teams. Of the 71 teams that submitted any call reports, 69 percent submitted 100 or less during the 12-month period of the program. The average number of calls logged by each REACT monitor was 16.6.

MONITORED HOURS

Our major purpose during phase 2 of the program was to determine the comprehensiveness of monitoring coverage in Ohio. In the past, coverage has been implied from the quantity of call reports, but it has never been directly measured. To obtain a direct measure we distributed the form shown in Figure 5 to each REACT team. The team monitors (unit operators) were requested to use the form to document their monitored hours for each month and to total the hours in the right-hand column and bottom row. Each unit operator was requested to use a separate sheet for each channel monitored and for base-station and mobile hours. The sheets were used to keypunch record cards for date code, channel number, team number, unit number, base station or mobile operation, and the hours data in the right-hand column.

Teams frequently have more than one unit operator monitoring at the same time. Thus, the gross hours submitted by the team do not reflect its net performance factor. We define the team monitoring performance factor as the percent of available hours monitored by at least one unit operator. (Available hours equal 366 days $\times$ 24 hours/day = 8,784 hours for the phase 2 period.) To calculate team performance factors, we prepared forms to show net monitoring hours for each team. A sample of these data is shown in Figure 6.

The data were prepared by overlaying the gross time reports, submitted by each unit operator in a given month, and inking in each hour that was monitored by at least one person. The result is a composite of the net hours monitored for each team for each month. These data were then processed in a manner similar to the gross hours data.

The results of the tabulations are summarized in Table 5. A total of 388,976 gross monitored hours were reported. Of these hours, 89.9 percent were spent in monitoring channel 9 and 96.6 percent of the gross hours were reported as monitoring Channel 9 or 11.

The net monitored hours tabulation yielded 220,628 hours. This is 56.7 percent of the gross hours and indicates the degree of redundance (43.3 percent) in team monitoring activities. The net average performance factor for participating teams was 39.24 percent of available hours. Redundancy in monitoring coverage may not be wasteful; it is often necessary to achieve full coverage of a team's geographical area.

Team Reporting Performance

The conscientiousness with which teams participated in the program by submitting their monitored hours reports is shown in Figure 7. Of the 80 teams, 64, or 80 percent, participated in the program to some extent. Of those teams, 28.75 percent reported for all 12 months of the program and 79.64 percent reported for 6 or more months.

Monitoring Performance Factor

The distribution of teams according to their monitoring performance factor is shown in Figure 8. Two (3.12 percent) of the 64 participating teams achieved more than a 90 percent performance factor. The average monitoring performance factor was 39.1 percent.

Number of Monitors

During the 12-month period we received monitored hours reports from 889 unit operators. The distribution of teams according to their number of unit operators is shown

Figure 3. Relative density of time required for monitors to complete call transactions.

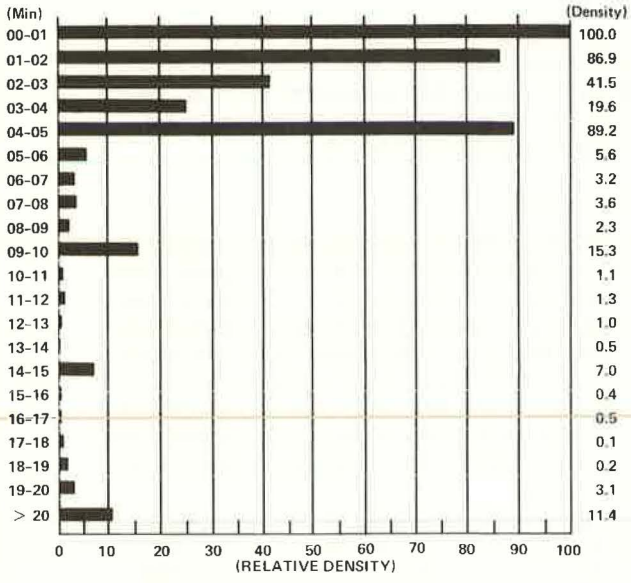


Figure 4. Distribution of teams by number of call reports.

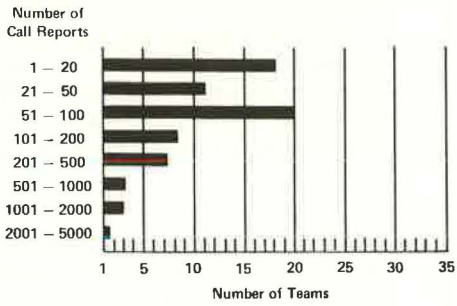


Figure 5. Monitored hours reporting form.

STATE OHIO
MONTH SEPT. TEAM NAME SUMMIT CO REACT, TEAM# C480 BASE STATION
CB CHANNEL # 9 OPERATOR'S NAME KENNY COX, UNIT# 43 MOBILE

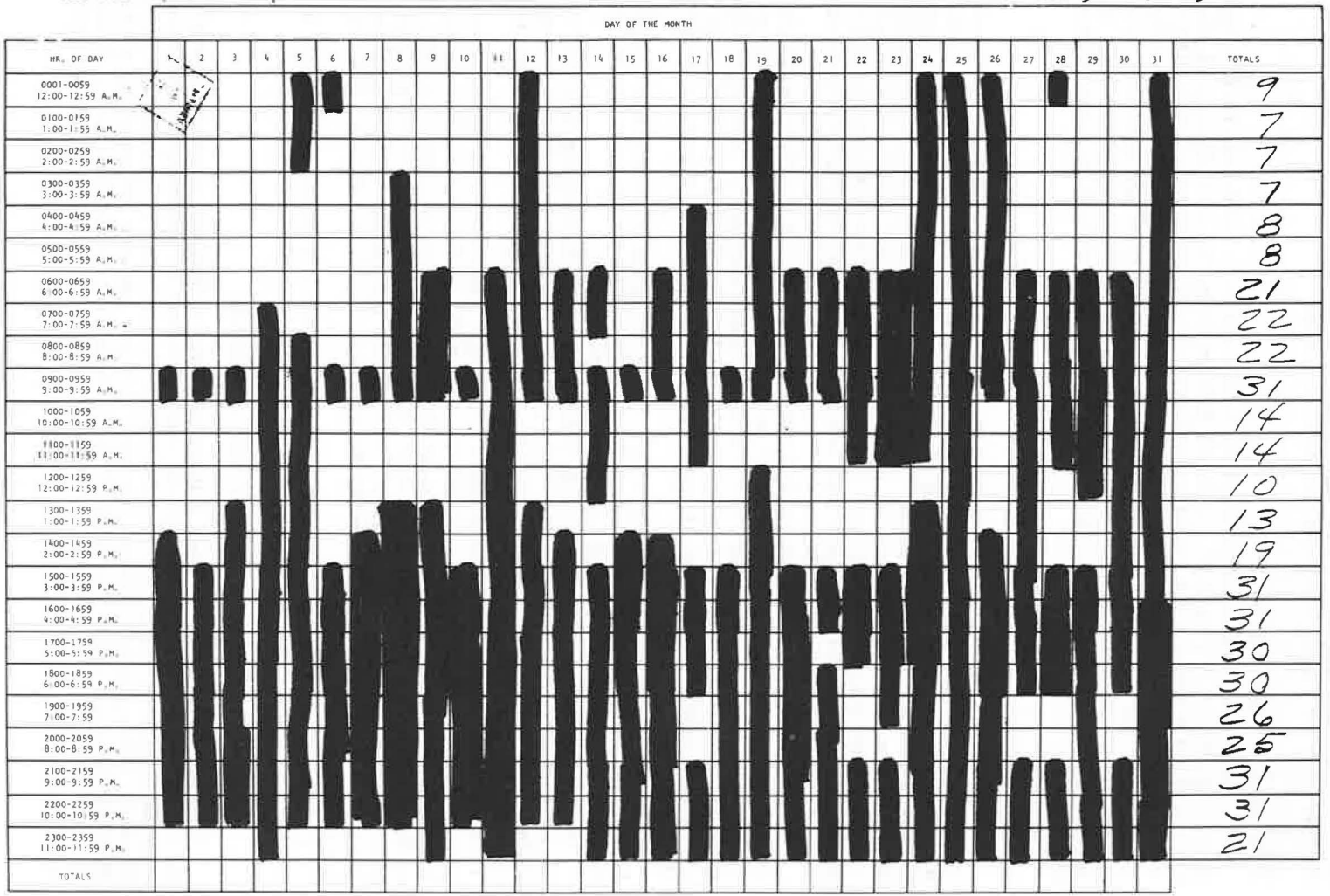
DAY OF THE MONTH

HR. OF DAY	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	TOTALS
0001-0059 12:00-12:59 A.M.														X	X	X								X					X			5
0100-0159 1:00-1:59 A.M.														X	X	X								X					X			5
0200-0259 2:00-2:59 A.M.														X	X									X					X			4
0300-0359 3:00-3:59 A.M.														X																		1
0400-0459 4:00-4:59 A.M.																																
0500-0559 5:00-5:59 A.M.																																
0600-0659 6:00-6:59 A.M.																																
0700-0759 7:00-7:59 A.M.																																
0800-0859 8:00-8:59 A.M.												X	X	X	X	X	X	X				X			X				X			10
0900-0959 9:00-9:59 A.M.												X	X	X	X	X	X	X				X			X				X			10
1000-1059 10:00-10:59 A.M.				X				X				X	X	X	X	X	X	X				X			X				X			13
1100-1159 11:00-11:59 A.M.				X				X				X	X	X	X	X	X	X				X			X				X			13
1200-1259 12:00-12:59 P.M.				X				X				X	X	X	X	X	X	X				X			X				X			13
1300-1359 1:00-1:59 P.M.				X				X				X	X	X	X	X	X	X				X			X				X			12
1400-1459 2:00-2:59 P.M.				X				X				X	X	X	X	X	X	X				X			X				X			12
1500-1559 3:00-3:59 P.M.				X				X				X	X	X	X	X						X		X					X			11
1600-1659 4:00-4:59 P.M.				X				X				X	X	X	X	X						X		X					X			11
1700-1759 5:00-5:59 P.M.				X				X				X	X	X	X	X						X		X					X			11
1800-1859 6:00-6:59 P.M.				X				X				X	X	X	X	X						X		X					X			11
1900-1959 7:00-7:59				X				X				X	X	X	X						X	X		X					X			12
2000-2059 8:00-8:59 P.M.				X				X				X	X	X	X						X	X		X					X			12
2100-2159 9:00-9:59 P.M.				X				X				X	X	X	X							X							X			11
2200-2259 10:00-10:59 P.M.				X				X				X	X	X	X							X							X			11
2300-2359 11:00-11:59 P.M.				X				X				X	X	X	X							X							X			11
TOTALS				14				14				16	16	20	19	18	16	7			5	16		9	7			14	8			11 9 HRS

Figure 6. Net monitored hours form.

MONTH 121 DEC. TEAM NAME _____, TEAM# 480

5, 9, 13, 68, 69, 72



(Use separate sheets for each channel monitored and for base station and mobile operations. Submit reports monthly.)

in Figure 9. The mode of the distribution is 6 to 10 monitors, with a population of 16 teams. The average is 13.89 unit operators per team.

Geographic Considerations

The geographic location of participating teams and their monitoring performance factors are shown in Figure 10. Of the 88 Ohio counties, 42, or 47.72 percent, had a team that participated in the reporting program. The distribution of Ohio cities and REACT teams, by city population, is given in Table 6. It is interesting to note that, as city size increases, the probability of a REACT team being formed there also increases. Although there is very good probability that cities with populations greater than 50,000 people will have a REACT team, a team performance factor does not correlate with city size.

CB RADIO AS A RESOURCE

Several projections based on the data presented here and in the 1969 FCC survey (5) were made. We have documented the participation of 889 individuals who contributed a total of 388,976 hours of monitoring time. This averages to 437.5 hours per man per year or 1.19 hours per day. The economic value of this contribution is estimated at $661,259 when a minimum labor rate of $1.70 is used. Furthermore, the average capital investment of a REACT team member is estimated by REACT National Headquarters at $600. The equipment used in this study then represents a capital investment of approximately $533,400.

The FCC survey reports that in 1969 there were 39,587 CB licensees in Ohio and that 74 percent of all licensees, or 29,294, were actively using CB radio at the time the survey was taken. Furthermore, the FCC report estimates that 47 percent of active licensees use CB radio for emergency communications. These data indicate that there are approximately 13,768 active licensees using CB radio for emergency communications in Ohio.

The FCC report further estimates the average operating licensee's investment at between $311 and $793. By projecting these data for Ohio, we estimate the capital investment at between $4.3 million and $10.9 million.

We note that the FCC report indicates that CB radio is used an average of 17 times a year for emergency communications per qualifying licensee. Our data agree very closely with this figure. The 889 REACT participants documented 14,750 calls in phase 2 of the program, or 16.6 calls per participant.

We observed that these calls required an average monitoring time of 437.5 hours per participant. Extending this monitoring time to the 13,768 Ohio licensees, we calculated 6.0 million man-hours. This volunteer labor has an estimated economic value of $10.2 million even at a minimum wage rate of $1.70 per hour. The investment in time and capital is continuing to grow. EIA reports 180,000 new transmitters are being put into service annually in the United States. Ohio's share is estimated at 7,200 new installations per year.

The use of CB radio for emergency communications is a substantial existing and constantly growing resource that makes no demands on public money.

ACKNOWLEDGMENTS

We wish to thank the 889 REACT team operators in Ohio who devoted personal time to seeing that we received their documentation: Robert M. Chiaramonte, Superintendant, Ohio State Highway Patrol, and his officers and men for their splendid cooperation; and Henry B. Kreer, REACT National Director, for his assistance and encouragement. Special thanks are due to Vincent W. Coppola of General Motors Research Laboratories, who reviewed many of the call and hour reports and who compiled the net monitored hours data.

REFERENCES

1. Summary of REACT Log Reports, Ohio Monitoring Project: Phase 1 Final Report. General Motors Research Laboratories, Warren, Michigan, Nov. 1971.

Table 5. Classification of monitored hours.

Monitoring Subject	Gross Hours	Percent of Gross
Gross monitored hours, total	388,976	100.0
Channel 9	349,885	89.9
Channel 11	25,996	6.7
Other channels	13,095	3.4
Base station operators	378,377	97.3
Mobile operators	10,599	2.7
Net monitored hours[a]	220,628	56.7

[a]Net hours = number of available hours that at least one team member reported monitoring.

Figure 7. Distribution of teams making monthly monitored hours reports.

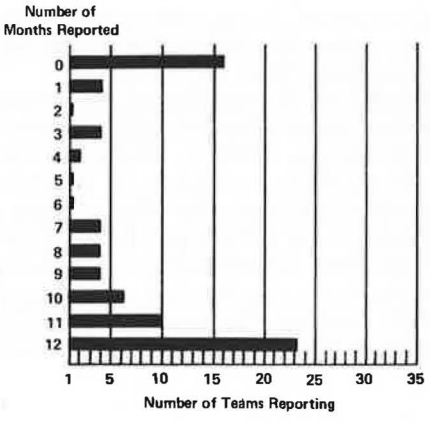


Figure 8. Distribution of teams by net monitored hours performance factor.

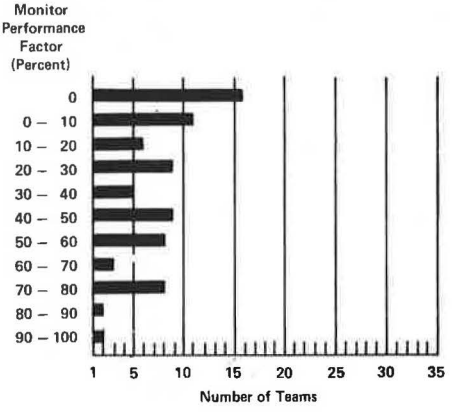


Figure 9. Distribution of teams by total number of participating unit operators (monitors) on team.

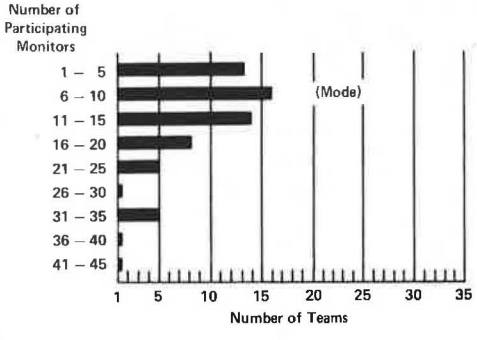


Figure 10. Geographic distribution of Ohio REACT teams and their monitoring performance factors.

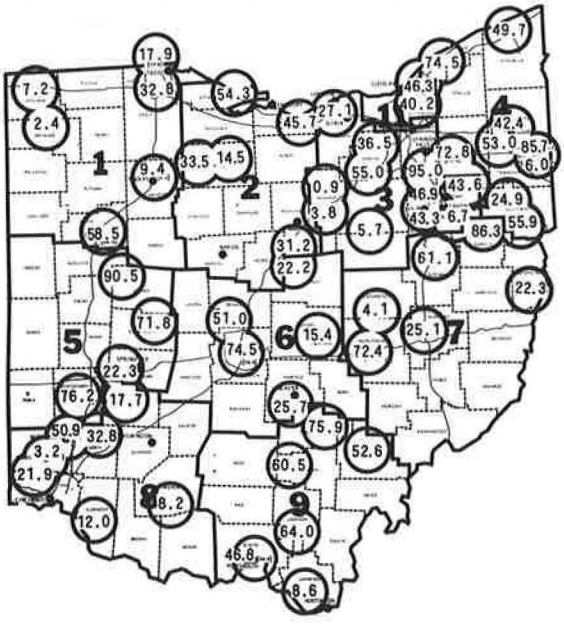


Table 6. Distribution of Ohio cities and REACT by city population.

City Population	Number of Cities	Number of Teams	Percent of Cities With Teams
<1,000	No data	4	
1,000-5,000	106	15	14.15
5,000-10,000	94	5	5.31
10,000-50,000	133	33	24.81
50,000-100,000	11	7	63.63
100,000-500,000	7	6	85.70
500,000-1,000,000	2	2	100.00

2. Chiaramonte, R. M., and Kreer, H. B. Measuring the Effectiveness of a Volunteer Emergency-Monitoring System in the Citizens Radio Service. Highway Research Record 402, 1972, pp. 16-27.
3. Trabold, W. G. Measuring the Effectiveness of a Volunteer Emergency-Monitoring System in the Citizens Radio Service: Discussion Paper. General Motors Research Laboratories, Warren, Michigan, GMR-1164, April 14, 1972.
4. Satterly, G. T., and Basu, S. Analysis of Citizens Band Radio Communication System as a Motorist Aid. Wayne State Univ., 1970.
5. Renner, J. A Survey and Analysis of Citizens Radio Service. Prepared for the Federal Communications Commission by Advanced Technology Systems, April 1971.
6. Bauer, H. J., Quinn, C. E., and Malo, A. F. Response to a CB Radio Driver Aid Network. Highway Research Record 279, 1969, pp. 24-39.
7. Citizens Two-Way Radio. Electronic Industries Association, 1973.

STUDY OF DETECTOR RELIABILITY FOR A MOTORIST INFORMATION SYSTEM ON THE GULF FREEWAY

Conrad L. Dudek and Carroll J. Messer, Texas Transportation Institute, Texas A&M University; and
A. K. Dutt, University of North Carolina at Chapel Hill

An experimental warning system has been installed on the Gulf Freeway in Houston as a means of alerting drivers approaching crest type of vertical curves to stoppages downstream of the crest. Successful automatic operation of the warning system depends on the reliability of system components. Earlier studies showed that developed control logic is responsive to stoppage waves, provided that the hardware functions properly. A one-lane control criterion resulted in 100 percent detection whereas 96 percent of the waves were detected using a two-lane control criterion. The studies also indicated a relatively high frequency of detector failures. The frequency of detector failures prompted a study to evaluate reliability of the warning system based on detector failure and repair rates experienced on the Gulf Freeway surveillance and control system and to ascertain whether detector redundancy or improved maintenance would be necessary. Malfunctions and repairs of all the Gulf Freeway surveillance and control hardware, including the 96-detector subsystem, were recorded for a 5-month period. The data revealed that the detectors on the Gulf Freeway failed at a rate of 3.78×10^{-4} failures per detector hour. Detectors were repaired at a rate of 0.23 repairs per hour. The reliability in terms of availability of the safety warning system was analyzed using these data and classical models for maintained systems. Availability of the system was 0.95 and 0.995 for the one- and two-lane criteria respectively. The results indicated that the current detector configuration and maintenance practices were adequate.

•RAMP CONTROL has resulted in significant improvements in peak-period freeway operation and reduction of accidents. Certain safety and operational problems continue to exist because of geometric features of the freeway and environmental phenomena that restrict driver sight distances. For example, the grade line and alignment of several freeways are such that sufficient sight distance is not always available for the motorist to confirm his expectations of traffic flow downstream. Problems arise because of unexpected traffic stoppages resulting from accidents and stalled vehicles or from stoppage waves generated during peak-period flow.

An experimental warning system has been installed on the inbound control section of the Gulf Freeway in Houston to reduce the effects of traffic incidents and congestion (1). The purpose of the system is to assist the freeway driver approaching crest type of vertical curves in formulating his expectations of the actual downstream traffic flow by alerting him to stoppage waves downstream of the crest.

Three overpasses were selected as the sites for pilot installations to study the effectiveness of the warning system, to develop automatic control algorithms, and to further evaluate the design concepts. The system currently consists of a static sign with

Publication of this paper sponsored by Committee on Freeway Operations.

attached flashing beacons (Fig. 1) located upstream of each overpass crest and a flashing beacon mounted on the bridge rail on the top of each crest (Fig. 2). The warning signs are controlled automatically by a digital computer. Double loop detectors are installed on each lane and located on both sides of the three overpasses (Fig. 3). Inasmuch as one detector station serves as the downstream station for one subsystem and as the upstream station for the next subsystem, only 30 detectors are included in the installation. The primary function of the detectors downstream of the overpass is to sense stoppage waves in order to activate the warning sign. The upstream detectors would indicate the time that the sign should be turned off.

PROBLEM STATEMENT

Successful automatic operation of the warning system depends on the reliability of the software and hardware components. Earlier studies (2, 3) showed that the developed control logic is responsive to all stoppage waves, providing that the detectors function properly. During the development of the computer control logic, a relatively high frequency of detector failures was noted. Because of the function of the system, it is important that it respond to all stoppage waves and maintain an extremely low level of false activations. Detector failures, of course, would have adverse consequences on the system. Because of the relatively high frequency of detector failures while the system was being developed, a study was conducted to evaluate the reliability of the warning system based on the detector failure and repair rates experienced and to ascertain whether detector redundancy or improved maintenance practices were necessary to increase the reliability of the warning system. The study also provided some insight about hardware failures and the suitability of maintenance activities for the entire surveillance and control system.

CONTROL PARAMETERS AND CRITERIA

Computer algorithms that have been successfully developed and implemented for the Gulf Freeway warning system use either traffic energy, speed, or occupancy as control variables (2, 3). Stoppage waves are predicted at the downstream detector station whenever the control variable reaches a predetermined critical value. Likewise, the stoppage waves are sensed as passing over the crest of the overpass when the variable at the upstream detector station reaches a critical value. Although the performance of each control variable is about the same on the average, traffic energy was selected for the system in Houston because of certain desirable features. The energy variable is more responsive to slow-moving trucks during the off-peak period and in many cases sounds an alarm when particular hardware problems arise.

Two approaches to control have been previously tested. In one approach, referred to as the one-lane criterion, a warning device is activated whenever any one of the three lanes indicates the presence of a stoppage wave. The second approach was developed in an attempt to compensate for the detector failures experienced at the time of system development. This approach is referred to as a two-lane control criterion, and it relies on information from a second lane to verify conditions on the first. In other words, the warning device is not activated until detectors on two lanes sense the presence of stoppage waves. Tests have shown that the one-lane criterion logic was responsive to all stoppage waves studied in relation to the existing detector locations and was subject to the proper functioning of the detectors. The two-lane control logic was responsive to 96 percent of the cases studied. The relative responsiveness of the system for each of the criteria using energy as the control variable is shown in Figure 4.

METHOD OF STUDY

Malfunctions and repairs of all the Gulf Freeway control and communications hardware subsystems, including the 96-detector subsystem, were documented for a 5-month period. The data were collected to establish the relative degree of subsystem failures and specifically to determine the failure rates and repair rates for the detector subsystem. Classical models relating to reliability of maintained systems were employed

to ascertain the reliability of the detector subsystem to establish whether detector redundancy or changes in maintenance practices would be required if the automatic warning system were to operate with a high degree of confidence.

HARDWARE SUBSYSTEMS

Table 1 gives the outages experienced on the Gulf Freeway during the 5-month analysis period (December 1971 through April 1972) for the primary subsystems that are related to the operation of the safety warning system. The data do not include the outages experienced with the closed-circuit television subsystem or with the ramp control signals.

The results reveal that 47 detector failures were experienced during the analysis period. In addition, 19 outages relating to the computer hardware, 1 cable outage, and 3 wiring outages were experienced. These data include both failures and outages due to maintenance, installation of additional equipment or control subsystems, and the like. Removing the latter from the totals, in addition to failures of those components of the computer peripheral equipment that would not have a bearing on the operation of the warning system, gives the total number of subsystem failures (Table 1).

Eleven percent of the subsystem failures related to the computer hardware. Generally, these failures were attributed to electrical power failures. In addition, one incident of a cable problem occurred when the main cable was accidentally cut by a construction crew. In general, this type of problem is rare and in the long run would constitute an insignificant percentage of subsystem failures.

The data also reveal that detector failures represented 87 percent of the problems experienced with the hardware that would be associated with a real-time freeway warning system. These data, from the individual detector failures of 96 detectors within the Gulf Freeway system, illustrate the relative frequency of failures that have been experienced in an operational control system.

When the computer fails, the entire system is inoperative. When a detector fails, a portion of the control and communication system becomes inoperative. A computer failure is easily recognized, but detector failures are more difficult to detect during control operations, and thus the control strategies can easily become ineffective.

Table 2 gives the types of detector problems experienced on the Gulf Freeway during the 5-month analysis period. Relay burns and internal circuitry problems accounted for 81 percent of the 47 failures (40.5 percent relay contact burns, 40.5 percent circuitry failures). There was only one case of failure of the loop itself. In addition, 17 percent of the failures were attributed to other problems such as blown fuses and defective wiring from the freeway lanes to the control box.

The relatively high frequency of detector failures, particularly the relay contact burns, was due in part to the equipment configuration on the Gulf Freeway. During the development of the safety warning system, the surveillance subsystem was operating between 48 and 72 volts DC, whereas the relay contacts were rated for 24-volt DC operation. The increased voltage was necessary because of the extensive length of the communications cable and associated interconnections. The communications subsystem was modified after this study was conducted.

ANALYSIS OF DETECTOR SUBSYSTEM RELIABILITY EFFECTIVENESS

There is a similarity between reliability problems of maintained systems and problems of queuing theory. For example, in the general queuing problem, one is concerned with serving arrivals with the objective of minimizing the length of the waiting line. The analogy here is that the repairman is the server of equipment failures, and the objective is to minimize system downtime.

A full description of the reliability of a given system that can be maintained requires specification of the equipment failure process, the system configuration, the repair process, and the state in which the system is to be defined as failed. If the times between individual equipment failures follow the negative exponential distribution and the times-to-repair are also exponentially distributed, then a Markovian representation can be used.

Figure 1. Warning sign with flashers.

Figure 2. Flasher unit at crest of overpass.

Figure 3. Locations of detectors for warning signs.

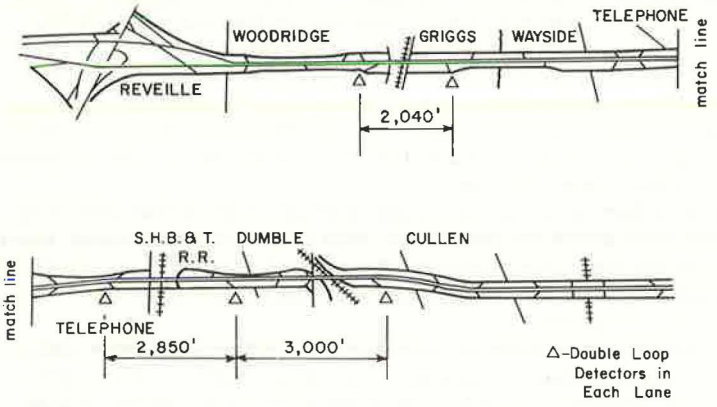


Table 1. Subsystem outages and failures.

Type	Outages		Failures	
	Number	Percent	Number	Percent
Detector	47	67	47	87
Computer hardware	19	27	6	11
Wiring (office)	3	4	0	0
Cable	1	2	1	2
Total	70	100	54	100

Table 2. Detector failures.

Type	Failures	Percentage of Failures
Relay contact burn	19	40.5
Internal circuitry	19	40.5
Other	8	17.0
Loop	1	2.0
Total	47	100.0

Figure 4. Stoppage wave detection performance curves using energy as control variable (2).

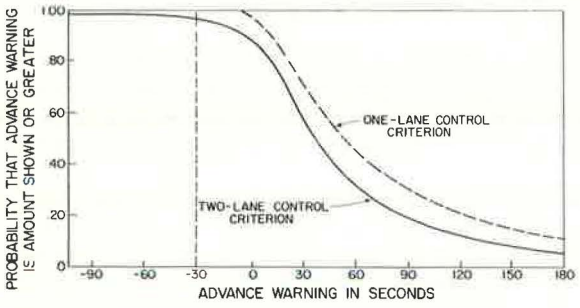


Table 3. Detector failures by month.

Month	Failures	Total Time of Failures (hours)
December	6	32.00
January	19	99.17
February	10	20.16
March	8	37.67
April	4	16.50
Total	47	205.50

Note:

Failure rate,

$$\lambda = \frac{47 \text{ failures}}{96 \text{ detectors} \times 108 \text{ days} \times 12 \text{ hours/day}} = 3.78 \times 10^{-4} \text{ failures/detector-hour.}$$

Repair rate,

$$\mu = \frac{47 \text{ repairs}}{205.5 \text{ hours}} = 0.23 \text{ repairs/hour.}$$

Measure of System Reliability Effectiveness

Several measures of system reliability effectiveness are available for consideration (4). The selection of an appropriate measure of effectiveness is determined primarily by the mission of the system. Availability is one measure of system reliability effectiveness that is applicable to maintained systems. It is defined as the proportion of time that the system will spend in acceptable states. Because of the particular mission of the safety warning system, it was of particular concern to establish the system's availability. Consequently, detector availability was selected as the measure of system reliability effectiveness.

Assumptions

The control and communication system on the Gulf Freeway is operated each weekday from 6 a.m. to 6 p.m. Consequently, any malfunction that developed after 6 p.m. was noted the following morning. In addition, all repairs were made during the 12-hour operational period. For the purposes of this study, the assumption was made that one repairman would be used to service the 96-detector subsystem. Inasmuch as the 30 detectors used for the warning system are more critical than the remaining detectors because of the function they serve, these 30 detectors would receive priority by the repairman on a first-come-first-served basis.

Because the detectors fail randomly, the detector failures can be assumed to be Poisson distributed; thus the time between failures will be negatively exponentially distributed. Likewise, the times-to-repair were assumed to follow a negative exponential distribution. A chi-square goodness of fit test was applied to the repair data, and the results indicated that the data did not quite fit a negative exponential distribution. However, the fit was relatively close, and it was felt that the small sample size may have influenced the fit. To make the analysis tractable, the negative exponential distribution was assumed.

The following implications relate to the assumptions that the individual detectors fail in accordance with the negative exponential distribution and that the times-to-repair are also exponentially distributed:

1. The conditional probabilities of failure and of repairing a detector are constant.
2. The probability of a single detector failure in the time interval t to t+dt, given that it was working at time t, is λdt where λ is the failure rate.
3. The probability of repairing a detector in the time interval dt, given that it was not working at time t, is μdt where μ is the repair rate.
4. The major portion of failures can be repaired in a short time, and those that take a long time to repair occur infrequently.
5. Only one detector will fail during time interval dt; similarly, only one detector can be repaired at a time.

Failure and Repair Rates

The frequency of detector failures during the 5-month analysis period is shown in Table 3. Also presented is the total time of failure, which in effect constituted the repair time for the detectors. From these data, a failure rate λ of 3.78×10^{-4} failures per detector-hour and a repair rate μ equal to 0.23 repairs per hour are computed.

One-Lane Criterion

For the purposes of this analysis, the three warning devices are considered as one complete system. All the detectors must function to have an operating system. As mentioned earlier, 30 basic detectors are used to operate the three warning devices on the Gulf Freeway. Thus, the system is considered to be in a failed state when any one detector is defective. The reliability analyzed in the following paragraphs refers to the availability of all three warning devices operating simultaneously.

It has been shown that the following steady-state probabilities apply for the general case of n detectors and r repairmen (4):

$$P_k = \frac{n!}{(n-k)!k!} \rho^k P_o \qquad \text{for } k < r \tag{1}$$

$$P_k = \frac{n!}{(n-k)!r!} \rho^r \left(\frac{\rho}{r}\right)^{k-r} P_o \qquad \text{for } k \geq r \tag{2}$$

where

P_k = probability of being in state P_k,
k = number of detectors down,
n = number of detectors in the system, and
$\rho = \frac{\lambda}{\mu}$

Availability is the measure of system reliability effectiveness selected for the analysis of the safety warning system. Availability is the proportion of time that the system will spend in acceptable states. The steady-state availability A of a system can be computed from the following relationship (4).

$$A = P_o = \left[\sum_{l=0}^{r-1} \frac{n!}{(n-k)!k!} \rho^k + \sum_{k=r}^{n} \frac{n!}{(n-k)!r!} \rho^r \left(\frac{\rho}{r}\right)^{k-1} \right]^{-1} \tag{3}$$

For the case of one repairman, Eq. 3 reduces to the following:

$$A = P_o = \left[\sum_{k=0}^{n} \frac{n!}{(n-k)!} \rho^k \right]^{-1} \tag{4}$$

For the special case of a 30-detector system, Eq. 4 can be written as follows:

$$A = P_o = \left[\sum_{k=0}^{30} \frac{30!}{(30-k)!} \rho^k \right]^{-1} \tag{5}$$

Substituting the computed values for the failure rate $\lambda = 3.78 \times 10^{-4}$ and the repair rate $\mu = 0.23$, the steady-state availability of the warning system becomes

$$A = P_o = 0.95 \tag{6}$$

Under steady-state conditions, there is a 95 percent chance that all 30 detectors will be functioning. Thus, there is a 95 percent chance that all three safety warning subsystems would be available, assuming that all other hardware components are functioning.

The probability that one, two, or three detectors will be out of operation can likewise be computed:

$$P_1 = 30\rho \, P_o = 0.045 \tag{7}$$

$$P_2 = \frac{(30)(29)}{2!} \rho^2 \, P_o = 0.001 \tag{8}$$

$$P_3 = \frac{(30)(29)(28)}{3!} \rho^3 \, P_o = 0.00002 \tag{9}$$

Two-Lane Criterion

For a two-lane criterion, the availability function is slightly more complex. This criterion requires that the energy variable can be measured in at least two lanes. Because the system consists of 3 warning devices having a total of 30 detectors and is in a failure state when any one of the devices is inoperative, then, at most, two detectors in the same lane can fail at each detector station without the system reaching a failure state. Thus, if a total of 10 detectors failed in one lane, the system would still be operational. However, if detectors fail in two lanes at a particular station, the system is considered unavailable.

The availability function under steady-state conditions for this case is given by the following relationship:

$$A = P_o + \sum_{k=1}^{10} C_k P_k \tag{10}$$

where C_k is a coefficient that is equal to the ratio of the number of ways in which k detectors can fail and yet the system be operable to the total number of ways in which k detectors can fail. Thus, the availability of the 30-detector system is

$$A = 0.95 + C_1(0.045) + C_2(0.001) + C_3(0.00002) + \ldots \tag{11}$$

The coefficient C_1 is computed as follows:

$$C_1 = \frac{\text{Number of ways in which one detector can fail and yet the system be available}}{\text{Number of ways in which one detector can fail}} \tag{12}$$

If one detector fails, the system would still be available; therefore

$$C_1 \frac{\binom{30}{1}}{\binom{30}{1}} = 1 \tag{13}$$

Likewise for C_2 and C_3,

$$C_2 = \frac{\binom{30}{2} - 5\left[\binom{6}{2} - 3\right]}{\binom{30}{2}} = 0.862 \tag{14}$$

$$C_3 = \frac{\binom{30}{3} - 5\left[\binom{6}{3}\right] - 120\binom{6}{2} - 3}{\binom{30}{3}} = 0.621 \tag{15}$$

The availability of the 30-detector system using a two-lane control criterion, therefore, is

$$\begin{aligned} A &= 0.95 + 1(0.045) + 0.862(0.001) + 0.621(0.00002) \\ A &= 0.995 \end{aligned} \tag{16}$$

The results indicate that the system availability using a two-lane control criterion is quite acceptable. However, it must be emphasized that, based on the results of previous studies (2), it would be expected that the warning system would be late in respond-

ing to 4 percent of the stoppage waves. Although the availability for the one-lane criterion is lower (i.e., 0.95), it is expected that this control approach would be responsive to all stoppage waves. Based on these results, it does not seem imperative to add redundant detectors to the system. This does not rule out the desirability of adding an additional detector station farther downstream to provide earlier warnings of stoppage waves.

Analysis of a Single Warning Device

The foregoing analysis is specific to the 3 warning devices on the Gulf Freeway having a total of 30 detectors. It was of interest to determine the reliability of a single warning device that may operate in isolation. The single warning device would require 12 detectors if the energy control variable is used. If one assumes that the same failure and repair rates apply as experienced on the Gulf Freeway and uses the same repair policy with a single repairman, then from Eq. 4 the system availability using a one-lane control criterion becomes

$$A = P_o = \left[\sum_{k=0}^{12} \frac{12!}{(12-k)!}\rho^k\right]^{-1} \tag{17}$$

$$A = P_o = 0.98 \tag{18}$$

Under steady-state conditions, there is a 98 percent chance that all 12 detectors will be functioning, assuming that all other hardware components are operative.

The probability that one, two, or three detectors will be out of operation can likewise be computed:

$$P_1 = 12\rho\ P_o = 0.019 \tag{19}$$

$$P_2 = \frac{(12)(11)}{2!}\rho^2\ P_o = 0.0002 \tag{20}$$

$$P_3 = \frac{(12)(11)(10)}{3!}\rho^3\ P_o = 0.0000 \tag{21}$$

For a two-lane criterion, the coefficients C_1, C_2, and C_3 are computed as follows:

$$C_1 = \frac{\binom{12}{1}}{\binom{12}{1}} = 1 \tag{22}$$

$$C_2 = \frac{\binom{12}{2} - 2\left[\binom{6}{2} - 3\right]}{\binom{12}{2}} = 0.636 \tag{23}$$

$$C_3 = \frac{\binom{12}{3} - 2\left[\binom{6}{3}\right] - 12\left[\binom{6}{2}\right] - 3}{\binom{12}{3}} = 0.164 \tag{24}$$

The availability of the 12-detector warning system using a two-lane control criterion is

$$A = 0.98 + 1(0.019) + 0.636(0.0002) + 0.164(0.0000)$$
$$A = 0.999 \tag{25}$$

The results indicate that the one-lane criterion and the two-lane criterion are both acceptable based on the maintenance practices on the Gulf Freeway.

DISCUSSION OF FINDINGS

The analyses presented in this paper relate to the detector failures and maintenance practices experienced on the Gulf Freeway surveillance and control system. The results may not be directly translatable to other systems because the hardware failures and maintenance practices may differ. We believe that comparable data from other systems will help shed some light on hardware problems so that greater effort can be made to solve common problems.

It is our hope that the results have focused attention on the degree of maintenance necessary for reliable systems, especially with respect to detectors, and the types of hardware problems that have been experienced on one operational system.

ACKNOWLEDGMENT

This paper discussed one phase of a research project on development of urban traffic management and control systems, conducted by the Texas Transportation Institute and the Texas Highway Department in cooperation with the Federal Highway Administration. The authors wish to thank W. R. McCasland, J. Gibbs, F. Guardiano, and G. Ritch of TTI for their efforts in collecting the data and their review of this paper. Special thanks are due L. Ringer of TTI for his review and comments.

The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration. This paper does not constitute a standard, specification, or regulation.

REFERENCES

1. Dudek, C. L., and Biggs, R. G. Design of a Safety Warning System Prototype for the Gulf Freeway. Texas Transportation Institute, Res. Rept. 165-4, May 1972.
2. Dudek, C. L. Development of a Technique for Digital Computer Control of a Safety Warning System for Urban Freeways. Texas Transportation Institute, Res. Rept. 165-5, May 1972.
3. Dudek, C. L., and Messer, C. J. Detecting Stoppage Waves for Freeway Control. Highway Research Record 469, Jan. 1973, pp. 1-15.
4. Sandler, G. H. System Reliability Engineering. Prentice-Hall, 1963.

DESIGN OF DENSITY-MEASURING SYSTEMS FOR ROADWAYS

Denos C. Gazis and Michael W. Szeto, IBM Research Center,
Yorktown Heights, New York

A Kalman filtering methodology for the estimation of traffic densities on multilane roadways is tested by using aerial photography data. The method gives very satisfactory estimates even when the sensor separation is as great as 3,000 ft. A systematic procedure is given for designing and calibrating a density-measuring system for a roadway.

A SERIES OF PAPERS (1, 2, 3) demonstrated that it is possible to obtain accurate estimates of the number of vehicles on a section of a roadway by using the technique of Kalman filtering on noisy measurements of flow and velocity taken by sensors at the entrance and exit of the section. The technique was used with sensor data obtained at the Lincoln Tunnel in New York City during the tunnel control experiment (4) carried out from 1966 to 1969. The data corresponded to the particular configuration and conditions of the experiment; namely, the sensors were placed about ½ mile (0.8 km) apart, and lane-changing was illegal and therefore infrequent (but not altogether absent).

Many questions arose as a result of the investigation discussed in the series of papers. Would the method work equally well in a freeway environment where lane-changing is much more frequent? How close must the sensors be placed in order to yield accurate density estimates? What are the trade-offs between cost and accuracy? The purpose of this paper is to provide some answers to these questions. A partial answer to the question of the effectiveness of the Kalman filtering approach in measuring freeway densities has been provided by the successful use of a somewhat different Kalman filtering procedure by Nahi and Trivedi (5).

The data used in this paper were obtained by the System Development Corporation through an analysis of aerial photographs of sections of the Long Island Expressway taken at the rate of one frame every 2 seconds. Using these data, we could assume the placement of imaginary sensors at any location of the observed section of the expressway, which was about ¾ mile (1.2 km) long. The aerial data also provided exact counts of vehicles between the assumed location of sensors, and these counts were used as a benchmark for estimating the effectiveness of the density estimation algorithm. In view of the nature of the aerial data, the assumed sensors gave a very accurate measurement of the flow and a reasonably accurate measurement of the speed of each vehicle past the sensor point. Noisy flow measurements could be simulated by the addition of noise to the aerial data for flow. The density estimation algorithm used (1) in this investigation was chosen because of its advantage of linearity in both the state equations and the observation equations entering in the Kalman filtering procedure and because of the satisfactory quality of what is taken as observation.

PROBLEM STATEMENT

Let us briefly review the essential features of the density estimation algorithm. Suppose there are N + 1 sensors placed at discrete distances on a roadway for the measurement of count and velocity and that these sensors divide up the roadway into N sections identified by the index of the upstream sensor. Let Y_k^i be the number of vehicles in section i at time k, where i = 1, 2, . . ., N and k = 1, 2, 3, . . ., and let n_k^i be the number of vehicles that pass over sensor i between times k and k + 1, where i = 1, 2, . . .,

Publication of this paper sponsored by Committee on Freeway Operations.

N + 1 and k = 1, 2, 3,.... Then the following equation, which merely states the conservation of cars, should be true:

$$y_{k+1}^i = y_k^i + n_k^i - n_k^{i+1} \tag{1}$$

If there are no lane-changing effects or errors on the flow measurements, Eq. 1 is an exact description of the transition from time k to time k+1. However, in general it is necessary to add a "noise" term in Eq. 1 to account for occasional deviations, yielding the state equation

$$y_{k+1}^i = y_k^i + n_k^i - n_k^{i+1} + w_k \tag{2}$$

where w_k is the noise term, which is assumed to have the following statistical properties:

$$\left.\begin{aligned} E\{w_k\} &= 0 \\ E\{w_j w_k\} &= \begin{cases} 0, & j \neq k \\ Q, & j = k \end{cases} \end{aligned}\right\} \tag{3}$$

From the measurements of velocities and counts at the entrance and exit of section i, one can generate a rough measurement of y_k^i, which we will call z_k^i, using the travel-time algorithm (1). The travel-time algorithm involves smoothness arguments on the distribution of velocities inside the section, and these arguments allow the computation of some average travel time for the cars that enter section i between times k and k + 1, and hence also for the last car to exit during this period. All the cars that entered the section while the last vehicle was traversing it are assumed to be still within the section. As might be expected, the rough count z_k^i deviates from y_k^i, and if we denote the difference between y_k^i and z_k^i by v_k, we have the observation equation,

$$z_k^i = y_k^i + v_k \tag{4}$$

Experience has shown that it is not unreasonable to assume that the noise term v_k has the following statistical properties:

$$\left.\begin{aligned} E\{v_k\} &= 0 \\ E\{v_j v_k\} &= \begin{cases} 0, & j \neq k \\ R, & j = k \end{cases} \end{aligned}\right\} \tag{5}$$

From now on, w_k will be referred to as system noise and v_k will be referred to as observation noise. The quantity Q from Eq. 3 is then the variance of the system noise, and R from Eq. 5 is the variance of the observation noise. With the state equation (Eq. 2) and the observation equation (Eq. 4), one can use Kalman filtering techniques to generate optimal estimates, $\hat{y}_k^i$, of the y_k^i. The equations for the Kalman filter are given in the Appendix. Here, it is sufficient to recall that the Kalman filter produces, recursively, best estimates of y_k^i as a weighted average of a value obtained by using the previous best estimate and the flow data, as shown in Eq. 1, and the rough count z_k^i. The weights of the averaging process are the Kalman gain, G_k, defined in the Appendix, and its unit complement.

Before the Kalman filter can be used, four parameters must be set: the initial mean of the state vector, the covariance matrix Σ_0 of the initial state, Q, and R. For any finite run, the accuracy of the filter estimates is dependent on the choice of these parameter values of which Σ_0, Q, and R are particularly important. In practice, it is difficult to guess the best values for Σ_0, Q, and R, especially when one is designing a surveillance system for an unknown environment. A systematic methodology should therefore include a procedure for choosing appropriate values of Σ_0, Q, and R.

In view of these observations and those made in the introduction, we pose the following questions:

1. Is there a systematic way of choosing the initial values for the Kalman filter in this particular application?
2. How sensitive is the algorithm to the distance between sensors?
3. Is there a systematic methodology for designing a surveillance system for roadways?

CHOICE OF INITIAL VALUES OF THE KALMAN FILTER

Because we are working with linear state equations and linear observation equations, the effect of the choice of values for the initial mean of the state vector and the initial covariance matrix Σ_0 on the accuracy of the filter estimates is minimal (6). However, the choice of Σ_0 affects the rate of convergence of the estimates. In practical terms, we are faced with the problem of choosing Σ_0, Q, and R so that the filter will yield good estimates of the state (the number of vehicles between sensors) for a finite run of the algorithm.

At this point we must define how we measure the goodness of the estimates. The Kalman filter is supposed to be an unbiased estimator, but this is never the case in practice. For each section i between sensors, let us define the following measures:

$$m_i = \sum_{k=1}^{T} (y_k^i - \hat{y}_k^i) \tag{6}$$

$$s_i = \sum_{k=1}^{T} (y_k^i - \hat{y}_k^i)^2 - m_i^2 \tag{7}$$

where T is a finite time horizon and y_k^i and $\hat{y}_k^i$ are respectively the number of cars in section i at time k and its corresponding estimate. Over a certain finite time horizon, one always finds a certain amount of bias in the filter estimates. The quantity m_i is this bias and is never zero in practice. It is one measure of the goodness of the estimates. The quantity s_i measures the amount of dispersion of the estimates about the actual values of the state and as such is also a measure of the goodness of the estimates. Experience has shown that by appropriately choosing Σ_0, Q, and R it is possible to bring both $|m_i|$ and s_i down to a certain point, beyond which varying the values for Σ_0, Q, and R will only decrease $|m_i|$ at the expense of increasing s_i, and vice versa. It is therefore necessary to take both m_i and s_i into account when considering the performance of the filter. We therefore chose a measure of error ϵ_i for each section i, defined by

$$\epsilon_i = \tfrac{1}{2} (m_i^2 + s_i)^{1/2} \Big/ \sum_{k=1}^{T} y_k^i \tag{8}$$

The values of the parameters Σ_0, Q, and R are then to be chosen so that ϵ_i is minimized for each section of the roadway.

Experimentation with the choice of these initial values leads to the observation that the accuracy of the filter estimates does not depend on the individual values of Σ_0, Q, or R; instead, it depends crucially only on the choice of the ratio ρ defined by

$$\rho = Q/R \tag{9}$$

By keeping ρ constant, it is possible to vary Σ_0, Q, and R individually by one or two orders of magnitude without affecting the accuracy of the filter estimates. This is an important observation, because it means that in designing a surveillance system for a roadway the engineer has to worry about the choice of the proper value for only one unknown parameter, instead of three, as the equations of the Kalman filter would seem

to indicate. This result is also justifiable theoretically. It can be shown that for our particular application of the Kalman filter the steady-state feedback gain G_∞ is of the form

$$G_\infty = \frac{1}{1 + \frac{R}{\Sigma_\infty + Q}} \tag{10}$$

where Σ_∞ is the steady-state variance of the estimator. But $\Sigma_\infty \to 0$ and $G_\infty \to (1 + 1/\rho)^{-1}$, where ρ is given in Eq. 9. In other words, the steady-state gain approaches a function of ρ alone, and, because the steady-state gain is the most crucial factor in determining the accuracy of the filter estimates in the long run, it is not surprising to find the filter performance dependent on ρ alone.

The next logical step at this point is to investigate how the optimal values for ρ relate to the operating conditions of the corresponding sections. In our experiments with the choice of initial values, it was found that the optimal values for ρ (i.e., those values that yield minimum ϵ_i, where $i=1, \ldots, N$ for the various sections) do not have any systematic relation to the separation between sensors, the mean speed of vehicles in the section, or the mean density of vehicles in the section. They are, however, related to the average frequency of lane-changing in the section. In general, a high frequency of lane-changing inside the section implies a high optimal value for ρ, and vice versa. The results are shown in Figure 1, where the solid line indicates the mean and the dotted lines indicate the spread of the optimal values for ρ versus the level of lane-changing frequency. This dependence of ρ on the frequency of lane-changing is understood if one recalls that ρ is the ratio of the system noise variance to the observation noise variance. System noise is largely caused by lane-changing, and observation noise accounts for the crudeness of the travel-time estimates (used as noisy observations of the state). If the quality of the travel-time estimates is uniform, the observation noise variance is more or less constant, and ρ is a monotonically increasing function of the system noise variance. Because higher frequencies of lane-changing imply higher values for the system noise variance, they also imply higher values for the optimum ρ.

This discussion pertains to the density estimate of a single lane that is affected by lane-changing. If we apply the Kalman filter combining all lanes of a roadway, then the system noise may reasonably be expected to be close to zero, and hence the optimum ρ is also very close to zero.

OPTIMAL SENSOR PLACEMENT

As mentioned, we set up imaginary sensors along the roadway to measure count and velocity and used these measurements to generate estimates of the numbers of vehicles between sensors. We first created several sections of roadway by placing sensors 500 ft (152 m) apart in each of the three lanes of the expressway, chose optimal values for the initial parameters for each section between sensors, and ran the estimation algorithm for all the sections. We then increased the distance between sensors from 500 ft to 4,000 ft (1219 m) in 500-ft steps, repeating the estimation experiment at each step. In the cases of wide sensor separation, overlapping sections were defined to ensure an adequate number of samples. For any given sensor separation, the minimum errors obtained in all the sections were averaged to remove possible effects of geometric peculiarities of the roadway. Our objective was to find out how the accuracy of the estimation algorithm was affected by the separation between the sensors. The results are shown in Figure 2, where the average minimum error is plotted against the sensor separation. It is significant that the error remains almost constant at a low of approximately 12 to 17 percent as the sensor separation is increased from 500 ft to 3,000 ft (914 m), beyond which it starts to rise slowly.

The same experiments were repeated with all three lanes treated together (2). For any given sensor separation, a section was defined across the three lanes, and an estimate was obtained for the total number of vehicles in the three lanes within the section.

Figure 1. Optimum value of ρ versus frequency of lane-changing.

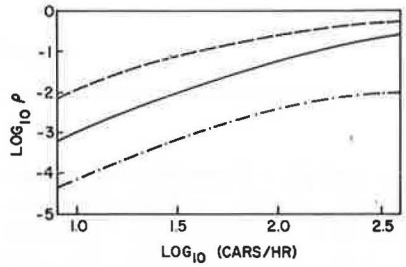


Figure 2. Percent error in density estimation versus sensor separation when all lanes are treated singly.

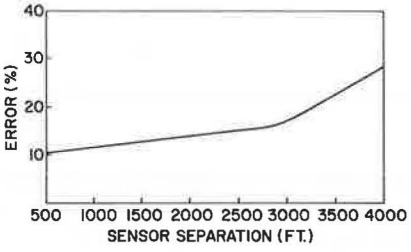


Figure 3. Percent error in density estimation versus sensor separation when all lanes are treated together.

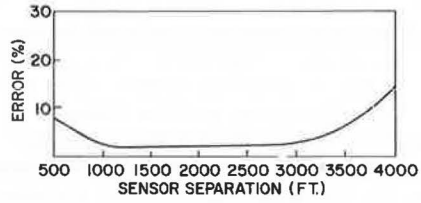


Table 1. Errors in density estimates for different sensor separations.

Sensor Separation (ft)	Errors (percent)						Speed (ft/sec)		
	Separate Lanes			Combined Lanes					
	Max.	Min.	Mean	Max.	Min.	Mean	Max.	Min.	Mean
500	15.0	10.8	11.7	14.3	4.7	8.1	92.2	74.7	84.2
1,800	20.0	7.1	11.8	5.3	0.8	1.8	91.8	75.2	83.1
1,500	21.4	9.7	13.7	7.5	1.9	2.9	91.4	75.2	83.9
2,000	23.8	9.5	14.7	3.5	0.8	1.8	89.2	75.5	82.8
2,500	21.4	10.7	15.4	2.7	1.6	1.8	89.4	76.2	83.2
3,000	22.4	11.1	17.1	—[a]	—	2.1	90.8	76.6	83.6
3,500	32.0	18.0	23.8	—	—	5.7	89.5	76.9	83.4
3,850	45.1	24.8	28.0	—	—	13.1	89.1	77.3	83.0

Note: 1 ft = 0.3048 m; 1 ft/sec = 0.3048 m/s.

[a]The maximum and minimum figures are not given where the sample size was one.

The results, shown in Figure 3, lead to three observations. First, the errors in this case were uniformly smaller than they were when the lanes were treated independently. This is to be expected because by combining the lanes we remove the errors that may result from lane-changing of vehicles. Second, as before, the error curve has a flat region. In this case, it occurs between 1,000 and 3,000 ft (305 to 914 m), where the error stays at a constant low level of about 2 percent. Third, the error curve has a slight dip between 500 and 1,000 ft. This may have been because the true counts have integer values whereas the estimated counts do not, and a count error may be a greater fraction of the true value when the sensor separation, and hence the vehicle count, is small. Such an effect is not observed in the case of lanes treated individually, because lane-changing in that case raises the level of error as the sensor separation is increased.

Clearly the insensitivity of the accuracy of the estimation algorithm to the separation between sensors has significant practical implications for the design of surveillance systems, simply because it costs less to place sensors farther apart.

The numerical results corresponding to Figures 2 and 3 are given in Table 1. Also given in this table are the average speeds corresponding to the various runs of the Kalman filter. It is seen that all the runs correspond to relatively light, free-flowing traffic. Therefore, the tests of the estimation algorithm given here are somewhat incomplete in that they do not show how well the algorithm works during periods of heavy traffic. Undoubtedly, periods of high-density, slow-speed traffic will be handled well if one leads up to them starting with light traffic. It is not clear, however, how well the filter initializes during periods of persistently heavy traffic. A test of the algorithm for such traffic was not possible because of lack of data but would be very desirable.

GENERAL DESIGN METHODOLOGY

We can now suggest a general methodology that uses the estimation algorithm (1) for designing a density measuring system for a roadway:

1. Obtain some aerial data of the roadway—about 1 hour's duration of traffic, with photographs taken at 1 frame per 2 seconds, is more than adequate. Ideally, one should obtain data for different degrees of congestion, because the accuracy of the estimation depends on the degree of congestion (1).
2. Compute the speed and flow past imaginary sensors placed at varying distances on the roadway.
3. Start with reasonable guesses for the initial count of vehicles in a section and the initial variance Σ_0 of this count. Run the estimation algorithm for values of ρ varying from, say, 10^{-4} to 1, keeping the observation noise variance R at some constant level. Find the best value of ρ for each section, corresponding to the minimum error, and compute the average minimum error for each value of the sensor separation, varying the sensor separation in steps of 500 ft between 500 and 4,000 ft.

A plot of the minimum error versus sensor separation may then be used in conjunction with economic considerations of installation and operation costs in order to determine the optimum placement of sensors. A high density of sensors entails a high installation and maintenance cost for the sensors themselves and for communication and possibly a high cost of computing power required for processing the sensor data. Reduction of this cost may be traded-off against some degradation of density estimates.

Frequently, the maximum separation of sensors is dictated by the need to detect incidents such as vehicle stoppages with a high reliability. The problem of reliable incident detection is still not completely understood, and it is not clear how one can best combine incident detection and density estimation.

CONCLUSIONS

We have provided, in this paper, some answers to questions concerning Kalman filter application, algorithm sensitivity, and roadway surveillance system design. We have found a systematic way of choosing proper initial values for the Kalman filter as used in the density estimation algorithm (1), and we have shown that the algorithm gives

satisfactory results for freeways, at least where the densities are relatively low. Contrary to intuition, the minimum achievable error of the density estimation algorithm is not necessarily a strictly increasing function of the distance between sensors, but the variation of this error versus sensor separation may have a flat region, offering an opportunity for substantial savings in sensor cost.

There are a number of possible extensions of the work presented here. Analytical modeling of the flow of traffic from section to section may improve the overall accuracy of a surveillance system or even provide an analytical solution to the problem of optimal sensor placement and a feedback solution for optimizing traffic flow. A simple correlation of the density estimation in two adjacent sections should improve the overall accuracy by use of the argument of conservation of cars over many contiguous sections of a roadway.

ACKNOWLEDGMENT

We gratefully acknowledge the permission of the U.S. Bureau of Public Roads to use the aerial photography foils obtained under their auspices by the System Development Corporation.

REFERENCES

1. Gazis, D. C., and Knapp, C. H. On-Line Estimation of Traffic Densities From Time-Series of Flow and Speed Data. Transportation Science, Vol. 5, 1971, pp. 283-301.
2. Gazis, D. C., and Szeto, M. W. Estimation of Traffic Densities at the Lincoln Tunnel From Time Series of Flow and Speed Data. In Stochastic Point Processes (P., Lewis, ed.), John Wiley, 1972, pp. 151-165.
3. Szeto, M. W., and Gazis, D. C. Application of Kalman Filtering to the Surveillance and Control of Traffic Systems. Transportation Science, Vol. 6, 1972, pp. 419-439.
4. Gazis, D. C., and Foote, R. S. Surveillance and Control of Tunnel Traffic by an On-Line Digital Computer. Transportation Science, Vol. 3, 1969, pp. 255-275.
5. Nahi, N. E., and Trivedi, A. N. Recursive Estimation of Traffic Variables: Section Density and Average Speed. Transportation Science, in press.
6. Jazwinski, A. H. Stochastic Processes and Filtering Theory. Academic Press, 1970.

APPENDIX

THE KALMAN FILTER EQUATIONS

Let $k = 0, 1, 2, \ldots,$ k be an index of time where $k = 0$ corresponds to the initial time. Consider a discrete-time, possibly nonlinear, and time-varying system whose state vector $\underline{x}_k$ is an n-dimensional vector, $\underline{x}_k \epsilon R_n$. Assume that the state propagates according to the stochastic difference equation

$$\underline{x}_{k+1} = \underline{f}_k(\underline{x}_k) + \underline{w}_k \tag{11}$$

Suppose that the measurement vector $\underline{z}_k$, $(\underline{z}_k \epsilon R_r)$, is related, possibly nonlinearly, to the state vector by the observation equation

$$\underline{z}_k = \underline{h}_k(\underline{x}_k) + \underline{v}_k \tag{12}$$

Assume that $\underline{f}_k(\cdot) : R_n \rightarrow R_n$ and $\underline{h}_k(\cdot) : R_n \rightarrow R_r$ are known, continuous, and sufficiently differentiable. Assume zero-mean, uncorrelated "white" noise in both the state and observation equations, i.e.,

$$\left.\begin{aligned} E\{\underline{w}_k\} &= \underline{0} \\ E\{\underline{v}_k\} &= \underline{0} \\ E\{\underline{w}_k \underline{w}_j'\} &= \underline{Q}_k \delta k_j, \ \underline{Q}_k > \underline{0} \\ E\{\underline{v}_k \underline{v}_j\} &= \underline{R}_k \delta_{kj}, \ R_k > \underline{0} \end{aligned}\right\} \quad \text{for all } k \qquad (13)$$

Define

$$\left.\begin{aligned} \underline{x}_0 &= E\{\underline{x}_0\} \\ \Sigma_0 &= E\{(\underline{x}_0 - \overline{\underline{x}}_0)(\underline{x}_0 - \overline{\underline{x}}_0)'\} \end{aligned}\right\} \qquad (14)$$

Assume that $\underline{x}_0$, $\underline{w}_k$, and $\underline{v}_j$ are mutually independent for all k, j.

The following notation will be used:

$\hat{\underline{x}}_{k|k}$ = estimate of state vector $\underline{x}_k$ based on the observation $\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_k$

$\hat{\underline{x}}_{k+1|k}$ = predicted estimate of the vector $\underline{x}_{k+1}$ based only on the measurements $\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_k$ (i.e., before measurement $\underline{z}_{k+1}$ is made)

$\Sigma_{k|k} = E\{(\underline{x}_k - \hat{\underline{x}}_{k|k})(\underline{x}_k - \hat{\underline{x}}_{k|k})'\}$

$\Sigma_{k+1|k} = E\{(\underline{x}_{k+1} - \hat{\underline{x}}_{k+1|k})(\underline{x}_{k+1} - \hat{\underline{x}}_{k+1|k})'\}$

The discrete-time Kalman filter algorithm is best described by decomposing it into three distinct steps: initializing, predicting, and updating. Start the algorithm by setting

$$\left.\begin{aligned} \hat{\underline{x}}_{0|0} &= \overline{\underline{x}}_0 \\ \Sigma_{0|0} &= \Sigma_0 \end{aligned}\right\} \qquad (15)$$

For predicting, generate $\hat{\underline{x}}_{k+1|k}$, $\Sigma_{k+1|k}$ by

$$\left.\begin{aligned} \hat{\underline{x}}_{k+1|k} &= \underline{f}_k(\hat{\underline{x}}_{k|k}) \\ \Sigma_{k+1|k} &= \hat{F}_k \Sigma_{k|k} \hat{F}_k' + Q_k \end{aligned}\right\} \qquad (16)$$

where

$$\hat{F}_k \triangleq \frac{\partial \underline{f}_k(\underline{x}_k)}{\partial \underline{x}_k} \Bigg|_{\underline{x}_k = \hat{\underline{x}}_{k|k}} \qquad (17)$$

For updating, generate $\underline{x}_{k+1|k+1} = \underline{x}_{k+1|k}, \Sigma_{k+1|k+1}$ by

$$\left.\begin{aligned} \underline{x}_{k+1|k+1} &= \underline{x}_{k+1|k} + G_{k+1}[\underline{z}_{k+1} - \underline{h}_{k+1}(\hat{\underline{x}}_{k+1})] \\ \Sigma_{k+1|k+1} &= \Sigma_{k+1|k} - G_{k+1} \hat{H}_{k+1} \Sigma_{k+1|k} \end{aligned}\right\} \qquad (18)$$

where

$$G_{k+1} \triangleq \Sigma_{k+1|k} \hat{H}_{k+1}' [\hat{H}_{k+1} \Sigma_{k+1|k} \hat{H}_{k+1}' + R_{k+1}]^{-1} \qquad (19)$$

$$\hat{H}_{k+1} = \frac{\partial \underline{h}_{k+1}(\underline{x}_{k+1})}{\partial \underline{x}_{k+1}} \Bigg|_{\underline{x}_{k+1} \hat{=} \underline{x}_{k+1|k}} \qquad (20)$$

The algorithm is iterative in nature. Starting with an initial guess, it generates a new state estimate each time an observation vector becomes available.

In a particular case (1), the state of the system is a scalar quantity, x_n, the count of vehicles on a roadway section during the nth time interval. The state propagates according to the linear equation

$$x_{n+1} = x_n + \Delta N_n + w_n \qquad (21)$$

where ΔN_n is the net input into the section (input at the entrance minus output at the exit) and w_n a random error. The measurement is a rough estimate of the count of vehicles z_n obtained directly from the speed and flow measurements. This estimate was obtained by first estimating the travel time of individual vehicles through the section (1). Other methods for obtaining z_n—for example, a phenomenological relationship between speed and density—can also be used satisfactorily (3). The z_n is related to x_n according to

$$z_n = x_n + v_n \tag{22}$$

Thus, both $f(\cdot)$ and $h(\cdot)$ are simple linear functions, and the preceding formulas are reduced to the simpler form

$$\left.\begin{aligned}
\hat{x}_{k+1|k} &= \hat{x}_{k|k} + \Delta N_k \\
\Sigma_{k+1|k} &= \Sigma_{k|k} + Q_k \\
\Sigma_{k+1|k+1} &= \Sigma_{k+1|k}\,(1 - G_{k+1}) \\
G_{k+1} &= \Sigma_{k+1|k}\,(\Sigma_{k+1|k} + R_{k+1})^{-1} \\
\hat{x}_{k+1|k+1} &= \hat{x}_{k+1|k} + G_{k+1}\,(Z_{k+1} - \hat{x}_{k+1|k})
\end{aligned}\right\} \tag{23}$$

where

$$\left.\begin{aligned}
Q_k &= E\{w_k^2\} \\
R_k &= E\{v_k^2\}
\end{aligned}\right\} \tag{24}$$

and the process is initialized by selecting some initial values $\hat{x}_0$ and $\Sigma_{0|0}$.

INVESTIGATION OF FLOW-DENSITY DISCONTINUITY AND DUAL-MODE TRAFFIC BEHAVIOR

Barry D. Hillegas, Donald G. Houghton, and Patrick J. Athol, School of Engineering, University of Pittsburgh

This investigation of traffic behavior was based on an unusually strong data set: The data were taken from a two-lane expressway with only one ramp operation in the 7-mile (11-km) length. Truck and bus traffic was not allowed to operate on the two-lane expressway, and the data set spanned a 7-hour period and reflected all phases of traffic behavior. Autocovariance and cross-covariance time series analyses were applied to traffic-stream occupancy. The autocovariance functions indicated random flow-density behavior for occupancy less than 15 percent (free-flow behavior). The autocovariance functions for higher occupancies indicated varying degrees of Markovian behavior. Cross-covariance analysis indicated that, under free-flow conditions, disturbances in the traffic stream were propagated with the flow of traffic at nearly the free-flow traffic speed. Analysis of flow-density behavior yielded distinct and discontinuous ranges of linear and nonlinear behavior. Further investigation through multivariate discriminant analysis indicated that, although density was the more important parameter, a flow-density criterion function was superior to a simple density criterion function. Furthermore, such a flow-density criterion function would change over time because of differences in the breakdown and recovery processes.

• UNDERSTANDING the behavior and interrelationships of traffic parameters, especially flow and density, is fundamental to techniques of traffic flow control. Various models have been advanced. The early models assume the well-known continuous smooth parabolic relationship between flow and density. However, Edie (1), after observing a number of data sets from the Port Authority of New York and New Jersey files relating to traffic flow in tunnels, noted that the empirical flow-density plots could be represented by two curves, one for the uncongested or free-flow state and one for the congested state. He proposed a distinct discontinuity in the region of maximum flow (Fig. 1) and showed that two curves fit the data better than a single curve. Drake, Schoffer, and May (2), in a statistical comparison of several hypotheses that describe stream flow characteristics, found that the Edie hypothesis yielded a comparatively low value for the standard error of estimate.

Athol (3) showed the flow-density relationship to be discontinuous with a linear trend to maximum volume for free flow and a breakdown to a flow less than maximum volume in congestion (Fig. 1).

According to Athol, congestion results from a driver behavioral response, i.e., drivers have a threshold tolerance of other vehicles. Once this threshold is exceeded, a reaction sets in that results in less effective individual driving and lower speeds. This interpretation of the onset of congestion is compatible with Edie's theory of discontinuity.

In a different approach, Mika, Kreer, and Yuan (4) investigated data from the John C. Lodge Freeway in Detroit. They categorized flow into two distinct modes—a steady

Publication of this paper sponsored by Committee on Freeway Operations and Committee on Traffic Flow Theory and Characteristics.

state with a parabolic relationship between flow and density and an oscillatory mode in which speed and density exhibit out-of-phase periodicities when plotted as a function of time. They found that the transition between these two modes of behavior is near the maximum flow value. If the steady-flow mode can be interpreted as free-flow operation and the oscillatory mode can be construed as congested operation, the work of Mika, Kreer, and Yuan is further evidence that the flow-density curve involves a discontinuity, or at least instability, about the maximum flow value.

This paper describes a further investigation of the flow-density relationship and dual-mode behavior.

DATA, SYSTEM, AND INSTRUMENTATION

Data used in this investigation were collected by the Expressway Surveillance Project in Oak Park, Illinois. The John F. Kennedy Expressway reversible lane section was the monitored segment of expressway. It is used to relieve the traffic load on the adjacent freeway system during peak periods. The freeway system configuration is shown in Figure 2. During the morning peak period, the reversible lanes are available to traffic flowing southeast toward Chicago. The roadway is open to northwest-bound, outward-flowing traffic in the afternoon and early evening. The slip ramp is open only to afternoon traffic.

There are seven monitoring stations along the 7-mile (11-km) length of the highway. As shown in Figure 2, the slip ramp is situated between Stations 3 and 4. Afternoon traffic flows from Station 7 towards Station 1. The monitors are 6-ft (1.8-m) square electromagnetic coils centered in the pavement of each 12-ft (3.6-m) lane.

The equipment detects traffic flow in units of vehicles per 20 seconds and occupancy in units of percentage of the 20-second sampling period (× 100) in which an automobile occupied the coil area. [Volume may be converted to vehicles per hour by multiplying by 180. Occupancy may be converted to appropriate density in vehicles per mile by multiplying by 2.8 (assuming 18.75 ft (5.72 m) = average automobile length) (3).] The data were collected from 1:04 to 8:01 p.m. on a day of "clean" conditions—dry pavement, fair weather, and average weekday volume.

In considering this investigation, it is very important to realize the strength of the data set. First, there was less than 1 percent detectable error (the difference in total occupancy at each station over the 7-hour sampling period) in the data. Second, the monitored highway segment was 7 miles long with only one ramp in operation. No trucks or buses were allowed to operate in the reversible lanes. Thus, the data had a minimum disturbance because of ramp activity and nonautomobile traffic, and this is important in analyzing driver behavior. This minimization of noise would hopefully enable us to see stream flow characteristics much more clearly. Third, the data were collected over a 7-hour time period that included free flow, transitional buildup to congestion, congested behavior, transitional decay to free flow, and free flow again. Undoubtedly, any strong trends in the data would affect the relationship between adjacent data points. However, the measurements are on a sufficiently small time scale compared with the time scale of a trend that such trending effects would be rather small.

TIME SERIES ANALYSIS: A CASE FOR DUAL-MODE BEHAVIOR

Consideration of Entire 7-Mile Section by Autocovariance

Consideration of the wide range of traffic stream behavioral characteristics might lead to classification of two ranges of stream behavior: (a) the state in which there is little or no vehicle interaction affecting stream behavior, and (b) the state in which there is some varying intensity of vehicle interaction that affects stream behavior. It seems intuitive that the state of little vehicle interaction might be characterized as random behavior. With regard to the state of high vehicle interaction, one might propose an autoregressive model in which the behavior of the traffic stream for a given time interval is influenced by the behavior of the previous time intervals (5).

In general, we can allow Z_t (the event or observation at time t) to be influenced by all previous events. For a discreet autoregressive time series,

$$Z_t = \alpha_1 Z_{t-1} + \alpha_2 Z_{t-2} + \ldots + \alpha_j Z_{t-j} + \epsilon_t$$

where ϵ_t is a random variable with zero mean and uncorrelated values for $p \neq q$. Then, a random time series is defined by

$$Z_t = \epsilon_t \quad (\text{i.e., } \alpha_1 = \alpha_2 = \ldots = \alpha_j = 0)$$

For a discreet time series, such as our data represent, one can investigate the dependence of Z_t on previous events by examining the autocovariance function for the series (5, 6).

The covariance of two variables measures the degree to which the two variables vary together. If the two variables are not independent, then their covariance is different from 0 (7). The autocovariance coefficient γ_k at time lag k measures the covariance between two values Z_t and Z_{t-k}, a distance k apart. The plot of γ versus the time lag k is the autocovariance function of the process (5). Box and Jenkins conclude that the most satisfactory Kth lag autocovariance is C_k where

$$C_k = \frac{1}{N}\sum_{t=1}^{N-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z}), \quad k = 0, 1, 2, \ldots, K$$

At this point we can say that the autocavariance function of a random process would appear similar to that shown in Figure 3, whereas a process of some autoregressive nature would appear similar to that shown in Figure 4.

Probably the simplest form of autoregressive behavior is Markovian. A Markovian time series is defined by

$$Z_t = \alpha_1 Z_{t-1} + \epsilon_t$$

According to Kendall and Stuart (8),

$$\alpha_1 = \rho$$

for Markovian behavior, where ρ is the correlation coefficient between Z_t and Z_{t-1}.

One statistical test of the randomness or nonrandomness of a time series can be performed using serial correlation (6). The serial correlation coefficient of order 1, r_1, is defined by

$$r_1 = \frac{\text{Cov}(Z_t, Z_{t+1})}{\sqrt{\text{Var}(Z_t)\,\text{Var}(Z_{t+1})}}$$

so that

$$r_1 = \frac{C_k}{\sigma^2}$$

where σ^2 is the variance. Serial correlation r_1 is the sample estimate of autocorrelation ρ, so that verification of the hypothesis

$$r_1 = 0$$

means the series is random, and verification of the hypothesis

$$r_1 \neq 0, \quad r_2 = r_1^2, \quad r_3 = r_1^3, \ldots$$

Figure 1. Discontinuous flow-density curves.

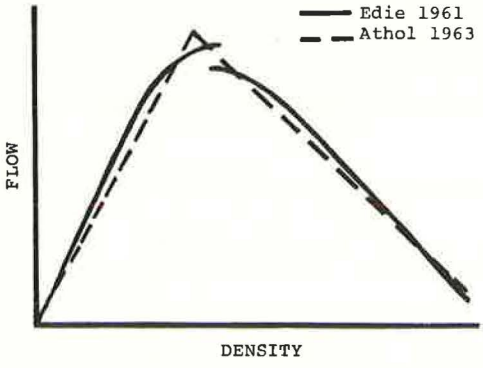


Figure 2. John F. Kennedy Expressway reversible lane section

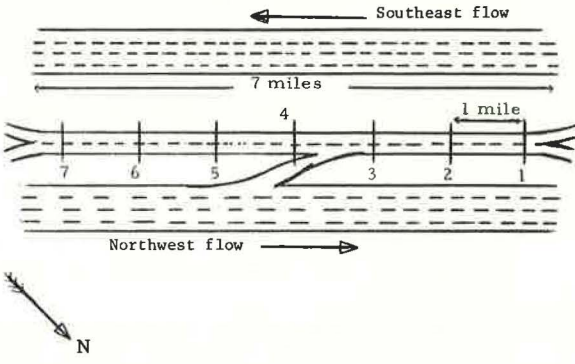


Figure 3. Autocovariance function of random time series.

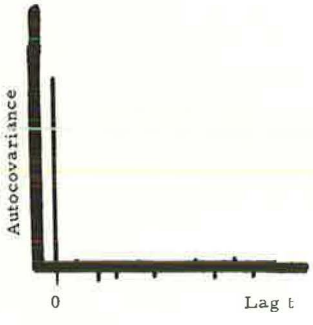


Figure 4. Autocovariance function of autoregressive time series.

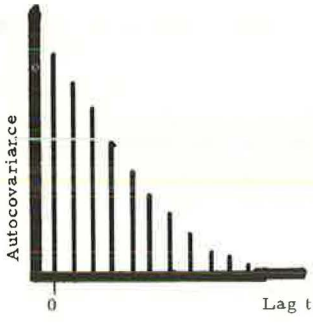


Figure 5. Autocovariance function of occupancy for Station 1, Lane 2, 2:00 to 3:00 p.m.

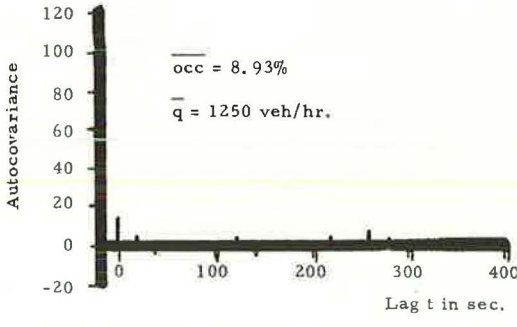


Figure 6. Autocovariance function of occupancy for Station 1, Lane 2, 3:00 to 4:00 p.m.

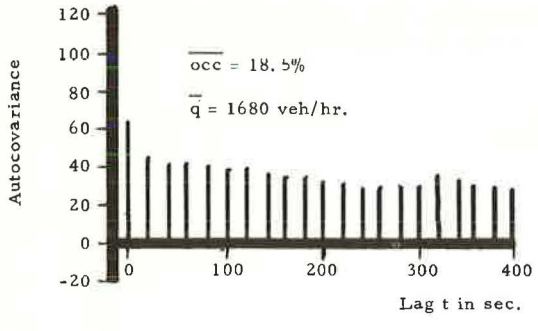
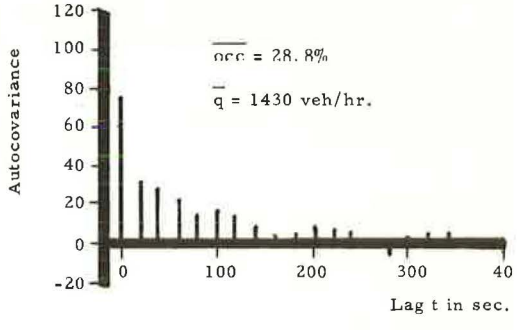


Figure 7. Autocovariance function of occupancy for Station 1, Lane 2, 4:00 to 5:00 p.m.



would mean that the series is Markovian, for example.

The autocovariance functions of occupancy with up to 40 time lags of 20 seconds each were generated for all seven stations for each 1-hour interval from 2:00 to 8:00 p.m. [An autocovariance and power spectral analysis package known as BMDØ2T of the Biomedical Computer Programs (9) was used. The method of calculation agrees with that recommended by Box and Jenkins (5).] Occupancy (or density in vehicles per mile—note the 2.8 scalar multiplier) was used because of its unique nature in defining the traffic stream. Lane 2 data were used to attempt to minimize noise, because Lane 1 was being fed by the slip ramp. Figures 5, 6, and 7 show some of the autocovariance functions of occupancy for Station 1, Lane 2, over 1-hour intervals from 2:00 to 5:00 p.m.

As shown in Figure 5, Station 1 appears to behave randomly from 2:00 to 3:00 p.m. Similar behavior was observed from 7:00 to 8:00 p.m. For the hours of 3:00 to 5:00 p.m. however, Station 1 appears to have varying degrees of autoregressive behavior as shown by Figures 6 and 7. When similar characteristics were observed from 5:00 to 7:00 p.m., an interesting point arose: The apparent autoregressive behavior is not so pronounced during the time of highest occupancy—28.8 percent from 4:00 to 5:00 p.m., and 25.4 percent from 5:00 to 6:00 p.m.—as it is during the transitional times from 3:00 to 4:00 and 6:00 to 7:00 p.m. with respective occupancies of 15.3 percent and 16.3 percent. This might be due to one of two things:

1. Because congestion (the condition of suboptimal flow with high-density, low-speed behavior) may occur downstream of Station 1 and back up through Stations 1, 2, and 3 during peak periods, a well-behaved Markovian autocovariance function may be most apparent only while there is very high vehicle interaction and congestion occurring in the immediate vicinity of the detector. This conjecture is supported in that further analysis showed that Stations 2 and 3 from 4:00 to 5:00 p.m. had a very random autocovariance function, yet their mean occupancies were 31.8 percent and 46.0 percent, and mean flows were 1,410 and 1,470 vph respectively (Table 1). Also from 3:00 to 4:00 p.m. the autocovariance functions for Stations 2 and 3 were similar to that for Station 1.
2. A higher order autoregressive process or integrated autoregressive moving average process may be occurring that cannot be analyzed using only autocovariance techniques.

Although Stations 1, 2, 3, and 4 show varying degrees of autoregressive behavior at different times, Stations 5, 6, and 7 are continuously random in behavior. The autocovariance functions for these three stations from 2:00 through 8:00 p.m. were all quite similar to that shown in Figure 8 for Station 7, Lane 2, 4:00 to 5:00 p.m.

Station 4, a Special Case

The behavior of Station 4 appears to be unique. This is probably due to the effect of the merging slip ramp traffic that enters the roadway between Stations 3 and 4. Lane 2 shows lower maximum volumes and densities for Stations 1, 2, 5, 6, and 7, but higher maximum volumes and densities for Stations 3 and 4. Generally, then, Lane 2 is used less than Lane 1 except in the area where the on-ramp is added to Lane 1 (Stations 3 and 4), where drivers tend to move into Lane 2 to avoid merging disturbances. Once merging is completed, however, these drivers tend to move back into Lane 1.

This driver reaction to anticipated merging disturbance may show itself in two ways. First, it might make drivers more conscious of vehicle interaction in this particular area. This increased awareness seems to be indicated by the autocovariance function. Figure 9 shows the occupancy autocovariance function for Station 4 from 4:00 to 5:00 p.m. Although the autocovariance function for 5:00 to 6:00 p.m. is nearly identical, the function during the other time intervals is random. The driver reaction to anticipated merge behavior may also increase dual-mode behavior.

Cross Covariance, a Measure of Simultaneous Similar Behavior

The cross-covariance function is the covariance between two time series, and it analyzes the behavior of two points of the traffic stream over time. A high cross-covariance value or peak in the function indicates that the two points under analysis

Table 1. Mean operating characteristics.

Station	2:00 to 3:00 p.m.				4:00 to 5:00 p.m.			
	Percent Occupancy	Density (vehicle/mile)	Flow (vph)	Speed (mph)	Percent Occupancy	Density (vehicle/mile)	Flow (vph)	Speed (mph)
1	8.9	25.0	1,250	50.0	28.8	80.6	1,430	17.7
2	7.3	20.4	1,200	58.8	31.8	89.0	1,410	15.8
3	12.9	36.1	1,510	41.8	46.0	128.8	1,470	11.4
4	6.1	17.2	1,100	64.0	11.6	32.5	1,330	40.9
5	4.7	13.2	801	60.6	6.3	17.6	1,020	58.0
6	5.9	16.5	833	50.5	7.4	20.8	1,000	48.1
7	7.6	21.1	1,050	49.8	9.8	27.6	1,320	47.8
Slip ramp			899				603	

Note: Mean speeds were calculated from flow/occupancy using $\bar{q}$ vph/$\bar{k}$ vehicles per mile. This will be accurate only for linear q-k behavior; it ignores nonlinear and discontinuous considerations. However, it should roughly reflect speed ranges in nonlinear behavior.

Figure 8. Autocovariance function of occupancy for Station 7, Lane 2, 4:00 to 5:00 p.m.

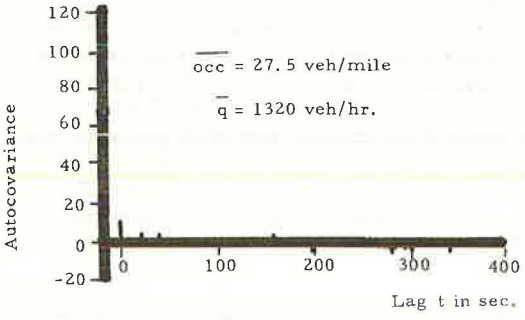


Figure 9. Autocovariance function of occupancy for Station 4, Lane 2, 4:00 to 5:00 p.m.

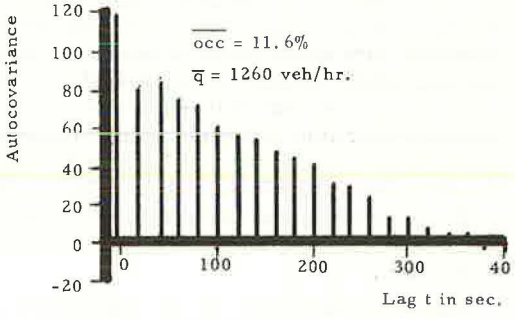


Table 2. Free-flow and congested variances.

Station n	2:00 to 3:00 p.m.				4:00 to 5:00 p.m.			
	Variance at Station n	Maximum Cross Covariance of Station n With Station 1	-Time Lag (second)	Wave Speed (mph)[a]	Variance at Station n	Maximum Cross Covariance of Station n With Station 1	+Time Lag (second)	Wave Speed (mph)[a]
1	14.24		0		76.61		0	
2	8.24	5.90	80	45.0	72.26	14.98	440	8.2
3	21.25	4.62	160	45.0	150.65	24.36	760	9.5
4	9.21	3.35	240	45.0	116.94	-19.20	620	17.4
5	4.56	2.01	320	45.0	7.48	4.32	680	21.2
6	5.12	2.72	380	47.4	8.42	-4.48	220	81.8
7	8.05	2.10	420	51.4	9.26	-4.85	60	360.0

[a]Mean wave speed is determined by distance/time, assuming 1 mile (1.6 km) between stations.

experience similar behavior at some distance apart in time equal to a lag value, k. The cross-covariance function is not necessarily symmetrical about zero as was the autocovariance function (5).

If we have two time series

$$X_{t-n}, \ldots, X_{t-2}, X_{t-1}, X_t, X_{t+1}, X_{t+2}, \ldots, X_{t+n}$$

and

$$Y_{t-n}, \ldots, Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2}, \ldots, Y_{t+n}$$

then $C_{xy}(k)$, the estimate of the cross-covariance coefficient at lag K, is

$$C_{xy}(k) = \frac{1}{n} \sum_{t=1}^{n-K} (X_t - \bar{X})(Y_{t+k} - \bar{Y}) \text{ for } k = 0, 1, 2, \ldots, K$$

and

$$C_{xy}(k) = \frac{1}{n} \sum_{t=1}^{n+K} (Y_t - \bar{Y})(X_{t-k} - \bar{X}) \text{ for } k = 0, -1, -2, \ldots, K$$

where $\bar{X}$ and $\bar{Y}$ are the means of the X series and Y series (5).

Cross covariance may be used to determine the propagation of a pattern within the traffic stream or the similarity in the traffic pattern in different lanes at the same points along the roadway. At a lag of zero for all stations, a definite peak in the cross-covariance function between Lanes 1 and 2 was found for each station. This indicates that the traffic pattern of the two lanes was similar over time at the same points along the roadway.

Table 2 gives certain cross covariance of occupancy values for Lane 2, 2:00 to 3:00 and 4:00 to 5:00 p.m. The maximum absolute value of cross covariance was selected as shown. Cross covariance with negative time lags was used in the free-flow period from 2:00 to 3:00 p.m. to obtain the approximate mean wave speed for the propagation of a disturbance downstream. The cross covariance with positive time lags was used for the 4:00 to 5:00 p.m. period in the hope that it would yield cross-covariance peaks reflecting the speed of propagation of a disturbance upstream in the case of congested behavior.

In the free-flow behavior of 2:00 to 3:00 p.m., it seems clear that disturbances in the traffic stream move through the freeway with the flow of traffic in a predictable manner at roughly the free-flow traffic speed. However, in congested operation, with regard to Table 2 and the cross-covariance analysis, there seemed to be no order in the transmission of disturbances.

If a wave is traveling at 10 mph (16 km/hour)—a little faster than Table 2 gives for the wave speed from Station 1 to 2—then in 880 seconds the wave would only travel about 2.5 miles (4 km). In other words, if the speed of 8 or 9 mph (12.8 to 14.5 km/hour) is realistic, our analysis would not detect the disturbance upstream of Station 2 because our maximum time lag was only 800 seconds.

TRAFFIC-STREAM BEHAVIOR WITHIN FLOW-DENSITY PARAMETER RANGES

For this analysis, flow-density curves were generated with the data, which were broken into hourly segments such that they were divided according to traffic condition, i.e., free flow, congested, and transitional. The analysis yielded three results: (a) a range of distinct linear behavior as shown in Figure 10, (b) a nonlinear range of be-

havior as shown in Figure 11, and (c) combined linear and nonlinear behavior as shown in Figure 12. [Scattergrams of Figs. 10, 11, and 12 were produced by FAKAD, a data analysis package developed by K. McDonald at the University of Essex. The origin of the axes is set to (0, 0). The length of each axis represents an equivalent range of the variables in terms of their standard deviations above and below the mean. The mark 0 on the scattergram indicates one observation at that point, 1 indicates two observations, and 9 indicates ten or more.]

Although the transition effects can make identification of the states difficult, the behavior did recognizably shift in time from one range to the other. Table 3 gives a breakdown of the station behavior. For Station 1, the shifts occurred at approximately 3:40 p.m. and 6:32 p.m. For Station 4, the shifts occurred at approximately 4:40 and 4:52 p.m.

From the time series analysis and flow-density behaviors, the investigation narrowed, seeking parameters that could distinguish between the two operational states.

Development of a Criterion Function

The simplest criterion function is density. In this case, some density value k will distinguish free flow from congested behavior, i.e., the criterion function would be $Z^* = k$.

An advance on this ultrasimplistic approach would be a criterion function of flow and density variables. Multivariate discriminant analysis can provide a linear form of such

Figure 10. Flow-density curve for Station 1, Lane 1, 2:00 to 3:00 p.m.

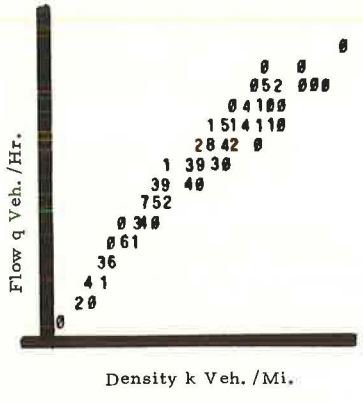


Figure 11. Flow-density curve for Station 1, Lane 1, 4:00 to 5:00 p.m.

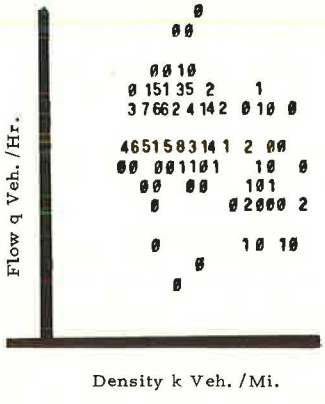


Figure 12. Flow-density curve for Station 1, Lane 1, 3:00 to 4:00 p.m.

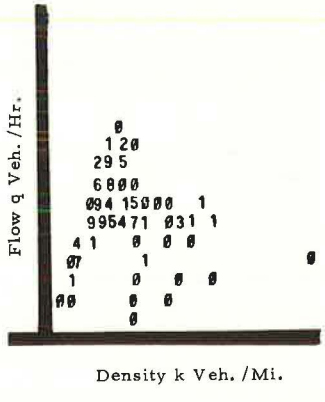
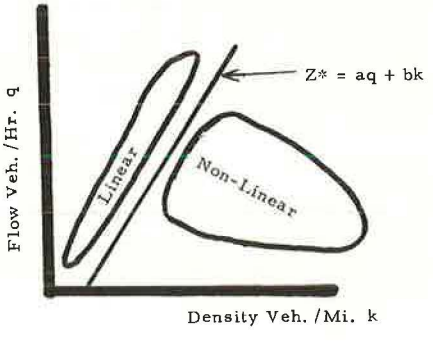


Table 3. Station flow-density behavior.

Time Period, p.m.	Linear Behavior, Station No.	Nonlinear Behavior, Station No.	Combined, Station No.
2:00-3:00	All stations		
3:00-4:00	3, 4, 5, 6, 7		1, 2
4:00-5:00	5, 6, 7	1, 2, 3	4
5:00-6:00	5, 6, 7	1, 2, 3	4
6:00-7:00	3, 4, 5, 6, 7		1, 2
7:00-8:00	All stations		

Figure 13. Criterion function for combined linear and nonlinear behavior.



a relationship (10). In this case, Z*, the value of the discriminant function that best separates the two states, is

$$Z^* = aq + bk$$

where a and b are constants, q is a given flow value, and k is a given density value.

This type of criterion function would appear similar to that shown in Figure 13.

Evaluation of the Density Criterion Function

Analysis of the data for Station 1, Lane 1, from 3:00 to 4:00 p.m. revealed that, for uncongested data, average density

$$\overline{k}_{unc} = 39.4$$

and the standard deviation

$$s.d.(k_{unc}) = 11.99$$

For the congested data,

$$\overline{k}_{con} = 78.11$$

and

$$s.d.(k_{con}) = 27.24$$

These values were determined after the data points were assigned to a congested or uncongested classification after individual analysis.

If the density readings are assumed to be normally distributed, the point at which there is an equal probability that an observation is a member of each group is at a density of 51.3 vehicles/mile. At this point, the probability of misclassifying a data point is 0.1635. Thus, the density criterion function is k* = 51.3.

This criterion was applied to the 3:00 to 4:00 p.m. data set for Station 1, Lane 1, and was found to classify four congested points and eight uncongested points incorrectly. This compares with nine and twenty errors respectively that would be expected with normally distributed data and suggests the data are skewed away from the transition region.

The discriminant analysis was done using BMDØ2T (9). All congested data but one had Z values greater than -0.002 and all uncongested data had Z values less than -0.002. Thus, a suitable criterion function for this particular segment of the data is Z* = -0.002. This discriminant function coefficients a and b were 0.00002 and -0.00082 respectively, giving a criterion function of

$$Z^* = 0.0002 = 0.00002q - 0.00082k$$

which can be rearranged to the form

$$q = 41k - 100$$

This function correctly allocates all but one point. An F statistic value of 131.7 for 2,177 was obtained that is significant at the 99 percent confidence level. The point at which there is a probability of misclassification into either group is at Z = 0.0015. This occurs at 1.13 standard deviations from each mean. Setting Z = 0.0015 results in nine misclassifications, compared to one misclassification when the empirically derived Z = -0.002 is used. This probably means that Z* for each group is not normally distributed.

The same discriminant analysis was also done by using normalized variables, i.e.,

$$z_{11} = \frac{q_i - \bar{q}_i}{s.d.(q)}$$

instead of flow, and

$$z_{21} = \frac{k_i - \bar{k}_i}{s.d.(q)}$$

instead of density, for i = 1, 2, . . . , 180 (for each data point).

The discriminant function for Station 1, Lane 1, using normalized variables for the time period 3:00 to 4:00 p.m. was

$$Z^* = 0.00935\, z_1 - 0.01536\, z_2$$

Thus, density seemed more important than flow in determining whether the freeway was congested or uncongested by a factor of about 1.65. The difference in sign value of z_1 and z_2 meant that a high flow would give a high value of Z, and thus uncongested operation was more likely, whereas a high density would give a low value of Z, and thus congested operation was more likely.

CONCLUSIONS

The strength of this further investigation of flow-density behavior was the data set: the data had a minimum of disturbance due to ramp activity and nonautomobile traffic.

Time-series analysis of traffic-stream occupancy was applied in two ways. Autocovariance reflected the behavior of a single point in the traffic stream over time. The autocovariance indicated random flow-density behavior for occupancy less than approximately 15 percent and varying significant degrees of autoregressive behavior for higher occupancies. The different forms of observed autocovariance functions for various states of traffic-stream occupancy suggest potential use of this technique for detecting behavior under controlled conditions.

Cross covariance analyzed the behavior of two points of the traffic stream over time. This analysis indicated that, under free-flow conditions, disturbances in the traffic stream were propagated with the flow of traffic in a predictable manner at roughly the free-flow traffic speed. Cross-covariance results were not conclusive for congested conditions.

Analysis of flow-density behavior yielded distinct and discontinuous ranges of linear and nonlinear behavior. The existence of these two distinct states of behavior was previously indicated by time-series analysis, particularly the autocovariance functions. Further investigation of criteria to distinguish between the two states of behavior indicated that, although density was the more important parameter, a flow-density criterion function was superior to a simple density criterion function. Multivariate discriminant analysis further suggested that a flow-density criterion function would change over time because of differences in the recovery and breakdown processes.

ACKNOWLEDGMENT

Appreciation is extended to A. G. R. Bullen, P. Doreian, N. Hummon, and J. Vitale for their suggestions. C. Tsai, M. Zubi, and N. Rattay contributed to background data manipulation. Many thanks are extended to Joseph McDermott, Project Director, Chicago Expressway Surveillance Project.

REFERENCES

1. Edie, L. C. Car-Following and Steady-State Theory for Noncongested Traffic. Operations Research, Vol. 9, 1961, pp. 66-77.
2. Drake, J., Schoffer, J., and May, A. A Statistical Analysis of Speed-Density Hypotheses. In Vehicular Traffic Science (Edie, L. C., Herman, R., and Rothery, R., eds.), American Elsevier, New York, 1967.

3. Athol, P. Interdependence of Certain Operational Characteristics Within a Moving Traffic Stream. Highway Research Record 72, 1965, pp. 58-87.
4. Mika, H. S., Kreer, J. B., and Yuan, L. S. Dual Mode Behavior of Freeway Traffic. Highway Research Record 279, 1969, pp. 1-12.
5. Box, G. E. P., and Jenkins, G. M. Time Series Analysis—Forecasting and Control. Holden-Day, San Francisco, 1970.
6. Yamane, T. Statistics: An Introductory Analysis, 2nd Ed. Harper and Row, New York, 1967, pp. 870-874.
7. Marascuilo, L. A. Statistical Methods for Behavioral Science Research. McGraw-Hill, New York, 1971.
8. Kendall, M. A., and Stuart, A. The Advanced Theory of Statistics, Volume 3: Design and Analysis, and Time-Series. Hafner Press, 1966, p. 405.
9. Dixon, W. J. Biomedical Computer Programs, 2nd Ed., No. 2. Univ. of California Press, Berkeley, 1971.
10. Snedecor, G. W., and Cochran, W. G. Statistical Methods, 6th Ed. Iowa State Univ., Ames, 1967.

ALGORITHM FOR A REAL-TIME ADVISORY SIGN CONTROL SYSTEM FOR URBAN HIGHWAYS

S. Kleinman, Product Research Department, Otis Elevator Company, Mahwah, New Jersey; and
R. Wiener, Electrical Engineering Department, City College, City University of New York

To reduce acceleration noise and thereby accident probability, an on-highway traffic-responsive control system is proposed that transmits advance warning of impending slowdowns. In the proposed system, the control system output consists of command settings for advisory speed signs that are spaced along the highway at intervals of 0.1 mile (0.16 km). The system input consists of speed and vehicle-count information from roadbed detectors, also spaced at 0.1 mile intervals. This paper details the derivation of the sign control algorithm that culminates in a formula expressing the speed setting of an advisory sign in terms of the detected speed and vehicle-count data. The control algorithm is designed to influence the trajectory of a vehicle at times of impending slowdowns such that its contribution to the acceleration noise integral is minimized. Examples are presented that show the application of the sign setting algorithm to hypothetical traffic situations.

•ENTRANCE RAMP METERING is the only dynamic traffic-responsive method currently being employed to regulate the flow on heavily traveled expressways. However, entrance ramp metering is limited in that it exerts no control over vehicles already on the roadway. In dense traffic conditions it is well known that the unpredictable response of an inattentive driver can initiate a shock wave that causes upstream traffic to sharply decelerate or come to a prolonged standstill. Such an occurrence, which cannot be substantially mitigated by ramp control, produces severe strain and discomfort among drivers and increases the probability of an accident. A system that transmits advance warning of impending decelerations would be a stabilizing influence, serving to dampen the amplifying effect of car-following in dense traffic. Such a system would improve the driver's level of comfort and would substantially reduce the incidence of rear-end collisions. The reduction in accidents would mean less delay and higher average flows as well as a more reliable trip time.

The type of advance warning system that is presented is a real-time, computer-controlled advisory speed sign system. Such a system is intended for conditions of dense traffic where the demand restricted by ramp control is less than the capacity flow rate. The control system output consists of command settings for the advisory signs located along the highway at intervals of 0.1 mile (0.16 km). Roadbed detectors, also spaced at 0.1-mile intervals, continually monitor and transmit speed and density data to an on-line computer. Processing of this information in accordance with a control algorithm results in command settings for the advisory speed signs. These sign settings are considered to be valid for a 5-second period, after which traffic conditions are reevaluated to determine updated sign settings. For practical reasons, the sign settings are restricted to 5-mph (8-km/hour) increments, and the rate at which signs are changed is such that a driver is not presented with a rapidly varying sequence of sign settings.

Publication of this paper sponsored by Committee on Freeway Operations.

An advance warning advisory sign system is based on the anticipation of the propagation of any significant slowdown to upstream vehicles after the delay due to car-following. Such a system sets the upstream signs to values somewhat lower than the current speeds on the upstream sections. The intended effect is to induce a milder deceleration in advance of the natural car-following response.

Any driver will be predisposed to a certain degree of compliance with the message on the advisory sign, ranging from full compliance to merely a state of preparedness. Noncomplying drivers are forced to decelerate sharply because of either the deceleration of a complying driver further downstream or the propagation of the original disturbance by car-following. Thus, drivers are expected to gain confidence in the sign's message by reinforcement. The driver will adjust his response to the sign such that the ensuing response to the car ahead is not too uncomfortable. By controlling the signs in a consistent manner and by turning off all signs except those where a significant response is necessitated, the highest possible level of compliance will be elicited.

The sign control algorithm, as detailed in this paper, is derived on the basis of a single lane of traffic with no passing. It is likely that a control strategy formulated for a single-lane roadway may be applied to a multilane roadway by the introduction of appropriate averages over the several lanes. A justification of control on the basis of a lane composite is that in heavy traffic the flow on all lanes of a particular section of roadway exhibits a certain uniformity. Thus, it is assumed that the single lane represents one lane of a multilane expressway on which passing opportunities are rare because of dense traffic conditions.

Because it has been shown (1) that a quantitative measure of driver comfort and safety is the acceleration noise,

$$\sigma^2 = \frac{1}{T}\int_0^T [a(t)]^2\,dt$$

the objective of advisory sign control is to minimize the acceleration noise summed over all vehicles. Thus, the sign control algorithm attempts to influence the trajectory of a typical vehicle at times of impending slowdowns such that the contribution of its deceleration to the integral $\int a^2(t)\,dt$ is minimized. Based on a calculus of variations solution to a simplified highway situation, a desired average deceleration is determined for vehicles on each 0.1-mile section of highway upstream of the slowdown. The desired deceleration is expressed in terms of measured section speeds and densities. An algorithm is presented in which sign settings are chosen so that the predicted average vehicle deceleration caused by the signs on each section equals the desired deceleration. The predicted average deceleration is derived from an assumed sign-following law and is a function of a car's speed when it enters the section and also the speed setting of the advisory sign located at the next section boundary.

A formula is derived for A_d, the desired average deceleration, and an expression for the predicted average acceleration A_s (of a vehicle due to a sign) is then determined. Equating the expressions for A_d and A_s yields the value of the advisory speed sign setting as a function of detector outputs.

CALCULATION OF DESIRED DECELERATION

To calculate deceleration, a simple highway situation is considered in which all vehicles are initially traveling at the same constant speed V_0 but at different spacings. The leading vehicle then decelerates to a new steady-state speed V_f. In this case it is assumed that the following vehicles eventually reach an equilibrium speed of V_f at a smaller spacing appropriate to the new speed. The optimum manner in which these vehicles should make the transition from V_0 to V_f, so that the integral $\int [a(t)]^2 dt$ is minimized for each vehicle, is considered.

Attention is directed to a particular following vehicle, and the problem is formally stated as follows: Given two vehicles having initial speeds of $v_1(0) = V_f$ and $v_2(0) = V_0 > V_f$

and initial spacing $x_1(0) - x_2(0) = S_0 > 0$ and given that $v_1(t) = V_f$ = constant, find the acceleration of the following vehicle $a_2(t)$ such that $I = \int_0^T [a_2(t)]^2\, dt$ is minimized subject to the final conditions $v_2(T) = V_f$ and $x_1(T) - x_2(T) = S_f < S_0$. The upper limit is given as T rather than infinity because it is expected that the optimum $a_2(t)$ is a function of the upper limit. Also, the following vehicle may be any of the vehicles upstream of the leading vehicle, not necessarily vehicle 2 in the line.

To simplify the calculation, the following normalized and equivalent problem is solved. Given a vehicle initially having $x(0) = 0$ and $v(0) = V_0$, find the acceleration a(t) to minimize $I = \int_0^T [a(t)]^2\, dt$ and satisfy $x(T) = D$, $v(T) = 0$. This problem is of the form

$$\text{Minimize } I = \int_0^T f(t, x, \dot{x}, \ddot{x})\, dt = \int_0^T [\ddot{x}(t)]^2\, dt$$

with respect to x(t), subject to given values of x and $\dot{x}$ at the upper and lower limits.

According to the calculus of variations (2, 3), the twice-differentiable function x(t) that extremizes I satisfies the differential equation

$$\frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right) + \frac{d^2}{dt^2}\left(\frac{\partial f}{\partial \ddot{x}}\right) = 0$$

Substituting $f(t, x, \dot{x}, \ddot{x}) = [\ddot{x}(t)]^2$ into this differential equation yields the solution

$$\ddot{x}(t) = A + Bt \tag{1}$$

The final solution is found by the following procedure: First, by integrating Eq. 1 and imposing the required values of x and v at the limits t = 0 and t = T, the values of A and B are determined as functions of T, D, and V_0. Second, the integral I is evaluated as a function of T, D, and V_0. Third, the value of T is found that minimizes I.

In the second step of this procedure, it is found that I(T) is a monotonic nonincreasing function, as shown in Figure 1, approaching zero as T approaches infinity. (The derivative of I(T) is zero at $T = 3D/V_0$ and $T = +\infty$.) It is tempting, because of the nature of I(T), to choose T as large as possible so that I(T) is minimized. However, it may be shown that for $T > 3D/V_0$ the acceleration waveform has the shape shown in Figure 2. Such an acceleration waveform is inadmissible because it implies a negative velocity for some values of $t < T$. Thus, the optimum acceleration is obtained by choosing $T = 3D/V_0$, which yields the motion given by

$$a(t) = \frac{2}{9}\frac{V_0^3}{D^2}\left(t - \frac{3D}{V_0}\right) \tag{2}$$

for $0 \le t \le 3D/V_0$ as shown in Figure 3. For $T = 3D/V_0$ the value of the integral I is

$$I = \frac{4}{9}\frac{V_0^3}{D} \tag{3}$$

Note that the minimizing function for the motion constrained by $v(t) \ge 0$ for (0, T) has not been obtained; rather, the minimizing function for the unconstrained problem has been found, and the result has been modified to assure $v(t) \ge 0$ for (0, T).

When the acceleration is assumed to be constant in (0, T) then the solution satisfying the initial and final conditions on spacing and speed is $a(t) = -V_0^2/2D$ for $0 \leq t \leq 2D/V_0$. For this case it is found that $I = V_0^3/2D$.

The constant deceleration waveform yields a value of I only 5.6 percent greater than the value obtained for the ramp function. For this reason and for reasons of mathematical tractability, the constant deceleration waveform is chosen as the desired deceleration waveform for the sign control algorithm.

For the first problem of a vehicle decelerating from V_0 to V_f while its spacing to a downstream car changes from S_0 to S_f, the solution is

$$A = \tfrac{1}{2} \frac{(\Delta V)^2}{\Delta S} \tag{4}$$

where $\Delta V = V_f - V_0$ and $\Delta S = S_f - S_0$.

PREDICTED AVERAGE DECELERATION

The average deceleration of a vehicle in response to a sign is determined from the sign-following model that is assumed to represent the dynamic response. Knowledge of car-following theory leads to a realistic sign-following model in the form of a stimulus-response equation. The sign-following model chosen is given by

$$\dot{v}(t) = \beta[V_s - v(t - \tau_s)] \tag{5}$$

The response is the acceleration at the current time, and the stimulus is the difference between V_s, the speed indication of the sign that the driver sees at time $t - \tau_s$, and $v(t - \tau_s)$, the vehicle's speed at $t - \tau_s$.

To facilitate further analysis of the model to obtain a sign setting algorithm, the model is modified to

$$\dot{v}(t) = \begin{cases} \beta[V_s - v(t - \tau_s)] & \text{for } t \geq \tau_s \\ 0 & \text{for } t < \tau_s \end{cases} \tag{6}$$

This modification has little effect on the dynamics of the response; in most cases the quantity $\dot{v}(t)$ is small for $0 \leq t \leq \tau_s$ because, near the end of the previous section of highway, the vehicle's speed will be nearly equal to the setting of the previous sign.

Values of β and τ_s are chosen to be equal to the corresponding parameters in the analogous car-following law for a very conservative driver. From data in Helly's paper (4), a reasonable choice is $\beta = 0.2$ sec^{-1} and $\tau_s = 2.0$ sec.

The average deceleration of a vehicle traversing a section under control of the sign on that section may be found from Eq. 6 if the initial condition on the vehicle's speed $v(0)$ and the speed setting of the sign V_s are known.

However, for a given initial speed, it is desired to find the value of V_s that yields a specified average acceleration. By an iterative procedure, the value of V_s can be found that yields the specified average acceleration. Although this procedure may be readily implemented on a digital computer, a more direct solution for V_s would save substantial computation time. To this end an approximate solution is proposed. By trying several functions, it is found that

$$A_a = -0.00087[V_0^2 - V_s^2] \tag{7}$$

is a good approximation of the predicted average acceleration in response to a sign. This approximation yields an average absolute error of 5 percent for average section speeds between 40 and 55 mph (64 and 88 km/hour).

SIGN CONTROL ALGORITHM APPLIED TO A HIGHWAY SITUATION

Assumptions

It is assumed that the simple highway situation exists in which a line of vehicles is initially traveling at speed V_0 but at different spacings. The detectors spanned by these vehicles all register the speed V_0. It is assumed that the lead vehicle then decelerates to V_f and some time later passes a detector where the decrease in speed is registered.

At the next analysis time the settings for the advisory signs are computed on the basis of existing traffic conditions. These settings are valid for only 5 seconds, after which time conditions are reevaluated to determine new sign settings.

The basis for determining the sign settings is the temporary assumption that all vehicles upstream of the lead vehicle will eventually be forced to decelerate to V_f. The desired deceleration for these vehicles has previously been determined to be a constant given by $A_d = \frac{1}{2}(\Delta V)^2/(\Delta S)$ over an interval of time $T = 2(\Delta S)/(\Delta V)$. The approach used is to determine the sign settings such that the predicted average deceleration due to the sign equals the desired average deceleration for vehicles traversing an entire section under sign control.

Therefore, to determine the advisory speed setting for the i th sign, attention is focused on an imaginary vehicle located at the (i - 1) th detector that is 0.1 mile upstream of the i th sign. It is this vehicle that responds to the initial setting of the i th sign during the first 5 seconds of its journey on the i th section.

The desired deceleration of the imaginary vehicle A_d is expressed in terms of the measured section speeds and densities. This car's predicted average deceleration A_a is derived from the assumed sign-following law as a function of its speed when it enters the section and the speed setting of the advisory sign located at the next section boundary. Equating A_d to A_a yields the solution for the advisory sign setting in terms of detector outputs.

Computation of Desired Deceleration

A section of highway is shown in Figure 4. The delineators indicate the detector locations, spaced by 0.1 mile. The k th detector stores VD_k, the velocity of the vehicle most recently passing the detector, and ND_k, the number of vehicles currently on the k th section of roadway. The advisory signs are also spaced by 0.1 mile and the speed setting is represented as V_{s_i} for the i th sign.

In accordance with the simple highway situation previously described it is assumed that, at a particular sampling time, $VD_j = V_f$ and that all other upstream detectors register $VD_k = V_0$, where $V_f < V_0$.

To determine the speed setting for the i th sign requires that the desired deceleration of an imaginary vehicle located 0.1 mile upstream of sign i be expressed in terms of the detector outputs. The desired deceleration for this vehicle is

$$A_d = \tfrac{1}{2}(\Delta V)^2/(\Delta S) \tag{8}$$

for an interval of time $T = 2(\Delta S)/(\Delta V)$, where $\Delta V = V_f - V_0$ is the difference between the final speed after a slowdown and the initial speed and $\Delta S = S_f - S_0$ is the difference between final and initial spacings from the vehicle initiating the slowdown.

In this example the initial speed is $V_0 = VD_{i-1}$ and the final speed is $V_f = VD_j$, so that

$$\Delta V = V_f - V_0 = VD_j - VD_{i-1} \tag{9}$$

It is assumed that the vehicle that initiated the slowdown has just passed detector j. Thus, the initial spacing equals the distance in feet between detector i - 1 and detector j; i.e.,

$$S_0 = 528(j - i + 1) \tag{10}$$

TRANSPORTATION RESEARCH RECORD 495

Page 68, Equation 10 should read: $S_0 = 528(j - i + 1)$

Figure 1. The integral I(T).

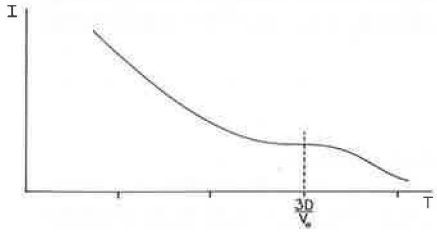


Figure 2. Acceleration waveform for $T > 3D/V_o$.

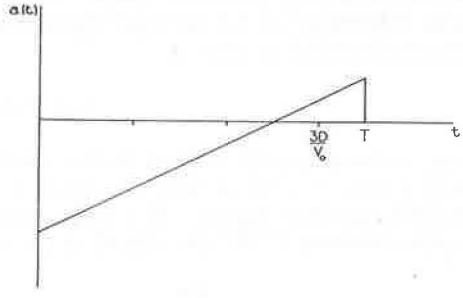


Figure 3. Optimal acceleration, $T = 3D/V_o$.

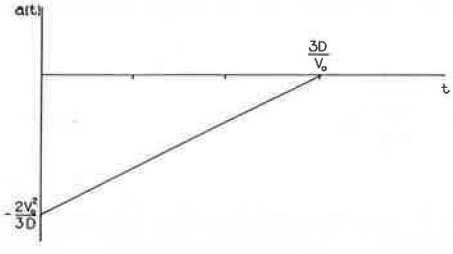


Figure 4. Representation of highway and detector locations.

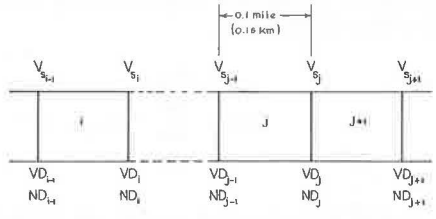


Table 1. Detector information available at t = 135 seconds.

I	NC(I) (vehicles)	VD(I) (mph)
11	4	50
12	5	50
13	5	50
14	5	50
15	3	48
16	4	46
17	4	35
18	6	34
19	2	—

Figure 5. Detector speed outputs for a typical case.

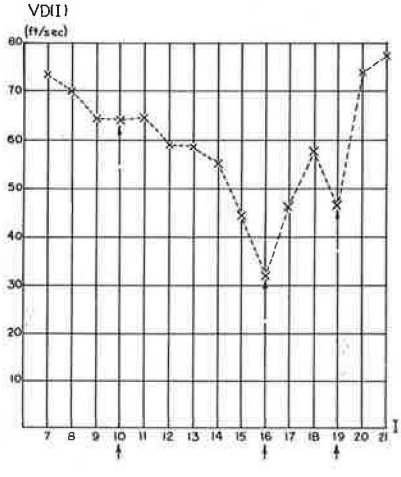


Table 2. Exact sign setting and rounded values.

Sign Number	Speed Setting fps	Speed Setting mph	Rounded Setting (mph)
17	44.0	30.0	30
16	60.0	40.9	40
15	66.1	45.1	45
14	68.3	46.5	45
13	69.4	47.3	45
12	70.2	47.9	Off

The final spacing between the imaginary vehicle currently at detector i - 1 and the vehicle assumed to be initiating the slowdown is determined from the number of vehicles between them and by the final speed. At the conclusion of the slowdown it is desired that each vehicle be at the spacing appropriate to its new speed. For the i th vehicle this desired spacing in feet is

$$DD_i = \alpha_i V_f + 20$$

where α_i is the desired headway factor for the i th vehicle-driver.

The ND_k vehicles on section k initially occupy 528 ft (161 m). After the slowdown they will occupy less space. For these vehicles the sum of the desired final spacings to the front bumper of the car ahead is expressed as

$$S_f(k) = \sum_{i=1}^{ND_k} (\alpha_i \times VD_j + 20) = 20\ ND_k + VD_j \sum_{i=1}^{ND_k} \alpha_i \tag{11}$$

where the sum is over the ND_k vehicles on the section. Because these vehicles initially occupied a 528-ft section,

$$528 = \sum_{i=1}^{ND_k} (\alpha_i v_i + 20) = 20\ ND_k + \sum_{i=1}^{ND_k} \alpha_i \times v_i \tag{12}$$

where v_i is the initial speed of the i th vehicle. The v_i are approximated by the estimated average section speed V_k, determined from the detectors at the section boundaries to be

$$V_k = (VD_{k-1} + VD_k)/2 \tag{13}$$

Replacing v_i in Eq. 12 by V_k from Eq. 13 gives

$$528 = 20\ ND_k + V_k \sum_{i=1}^{ND_k} \alpha_i \tag{14}$$

From Eq. 14,

$$\sum_{i=1}^{ND_k} \alpha_i = \frac{528 - 20\ ND_k}{V_k} \tag{15}$$

Substituting from Eq. 15 into Eq. 11 gives

$$S_f(k) = 20\ ND_k + \frac{VD_j}{V_k}(528 - 20\ ND_k) \tag{16}$$

or, using Eq. 13,

$$S_f(k) = 20\ ND_k + \frac{2\ VD_j}{VD_{k-1} + VD_k}(528 - 20\ ND_k) \tag{17}$$

Thus, the desired cumulative final spacing for the vehicles initially on sections i through j is

$$S_f = \sum_{k=i}^{j} S_f(k) \tag{18}$$

Substitution of the derived quantities into Eq. 18 gives, from Eq. 8,

$$A_d = \tfrac{1}{2} \frac{(VD_j - VD_{i-1})^2}{\sum_{k=i}^{j} \left[20\ ND_k + \frac{2\ VD_j}{VD_{k-1} + VD_k}(528 - 20\ ND_k)\right] - 528\ (j - i + 1)} \tag{19}$$

This is the desired deceleration for a vehicle traversing the i th section in response to a slowdown detected at the j th detector.

Solution for Sign Setting

Expressions for the desired average deceleration A_d and the predicted average deceleration A_a have been obtained. The value of the setting for the i th sign V_{s_i} is found by equating A_d from Eq. 19 to A_a from Eq. 7 with $V_s = V_{s_i}$. The solution for V_{s_i} is

$$V_{s_i} = \left\{ VD_{i-1}^2 + \frac{575\ (VD_j - VD_{i-1})^2}{\sum_{k=i}^{j} \left[20\ ND_k + \frac{2\ VD_j\ (528 - 20\ ND_k)}{VD_{k-1} + VD_k}\right] - 528\ (j - i + 1)} \right\}^{1/2} \tag{20}$$

where VD_j is the speed indication of the downstream detector that is temporarily assumed to equal the final velocity of the upstream vehicles after the slowdown.

Rounding of Sign Settings

For practical reasons the sign settings are rounded to the standard 5-mph increments. In addition, a sign is not turned on if only a minor response to the calculated setting is anticipated.

After computing the exact sign settings via Eq. 20, each sign is examined and those that are off are skipped. If the i th sign is on, it is then ascertained whether a significant response is expected to the sign setting. If $VD_{i-1} - V_{s_i} \leq 2.5$ mph, the i th sign is turned off. If $VD_{i-1} - V_{s_i} > 2.5$ mph, the sign setting, V_{s_i}, is rounded to the nearest 5 mph. If, for the rounded sign setting, the quantity $VD_{i-1} - V_{s_i} \leq 2.5$ mph, then the sign is turned off. If $VD_{i-1} - V_{s_i} > 2.5$ mph for the rounded value, then the rounded value is retained as the current sign setting.

CALCULATION OF SIGN SETTINGS

An example from a simulation run with 80 vehicles is presented. [Details of the simulation are given elsewhere (6).] All vehicles enter the roadway at a speed of 50 mph and continue at this constant speed until t = 120 seconds. At this time the first vehicle decelerates to 35 mph at a rate of -3.0 fps^2.

At t = 125 seconds the slowdown is detected and one sign is set; at t = 130 seconds two signs are set; and at t = 135 seconds five signs are set. The situation at t = 135 seconds is examined in more detail.

The detector information available at t = 135 seconds is given in Table 1. The value of VD(I) is rounded to the nearest integer.

From several computer runs it is found that the acceleration noise is minimum when the constant 575 in Eq. 20 is replaced by a constant between 600 and 625. For the example under consideration, this constant is 625. Consequently, the appropriate equation for determining the setting of the i th sign due to a slowdown detected at VD_j is

$$V_{s_1} = \left\{ VD_{i-1}^2 + \frac{625\,(VD_j - VD_{i-1})^2}{\sum_{k=i}^{j} \left[20\,ND_k + \frac{2\,VD_j\,(528 - 20\,ND_k)}{VD_{k-1} + VD_k} \right] - 528\,(j - i + 1)} \right\}^{1/2} \qquad (21)$$

The data from several simulations indicate that it is preferable not to set the sign that is located directly at the point where the slowdown is detected. It is found that the setting of this sign often retards vehicles in recovering from a slowdown after the vehicle causing the slowdown increases its speed. Therefore, in this example, sign 18 is skipped.

In determining the exact setting (before rounding) for sign 17, the appropriate values to substitute into Eq. 21 are

$$\begin{aligned} j &= 18 \\ i &= 17 \\ VD_{16} &= 66.9 \text{ fps (46 mph)} \\ VD_{17} &= 50.9 \text{ fps (35 mph)} \\ VD_{18} &= 50.9 \text{ fps (35 mph)} \\ ND_{17} &= 4 \\ ND_{18} &= 6 \end{aligned}$$

Substituting these values into Eq. 21 gives $V_{s_{17}} = 44.0$ fps for the setting of sign 17.

To determine the exact setting for sign 14, for example, the values j = 18 and i = 14 should be used in Eq. 21. The value j = 18 should be used in finding V_{s_1} for all values of i less than j.

In the same manner, the exact settings of the other upstream signs are determined as given in Table 2. The results of the rounding algorithm are also tabulated. It is noted that sign 12 is turned off because $VD_{11} - V_{s_{12}} = 50.0 - 47.9 = 2.1$ mph is less than 2.5 mph. That is, the sign is not turned on because the vehicle 528 ft upstream of the sign is traveling at a speed only slightly greater than the computed sign setting.

APPLICATION OF SIGN CONTROL TO A GENERAL HIGHWAY SITUATION

In the simplified highway situation previously described, the speed indications of the detectors are monotonically decreasing in the downstream direction. This is intended to represent only an isolated component of the traffic pattern on an actual highway.

Sampling of the detector speed outputs at a random time generally yields a speed profile exhibiting peaks and valleys. A typical case is shown in Figure 5.

General Procedure for Sign Setting

The heuristic procedure for determining the sign settings is described for a typical case for which VD(I) are as shown in Figure 5. The detector speed outputs VD(I) are scanned, starting at VD(7), to locate the first "local minimum". A local minimum occurs at detector k if both

$$VD(k - 1) > VD(k)$$

and

$$VD(k + 1) \geq VD(k)$$

are satisfied. The three local minima are identified by arrows in Figure 5.

The first local minimum in Figure 5 is VD(10). Sign 10 is not set since it is located at the local minimum. Sign settings for signs 9 and 8 are determined from Eq. 21 using j = 10 and the appropriate values for VD and ND. That is, V_{s_9} and V_{s_8} are computed on the basis of the local minimum of VD(10). Sign 9 is set to 65.9 fps and sign 8 is set to 68.7 fps.

The scan is continued and the second local minimum is found to be VD(16). Sign 16 is not set because the local minimum occurs at the location of this sign. Signs 15 back through 8 are now considered on the basis of a slowdown assumed to originate from the second local minimum at VD(16). Thus, using the value j = 16 in Eq. 21, the computed settings (feet per second) for signs 15 back through 10 are

$$V_{s_{15}} = 38.5$$
$$V_{s_{14}} = 48.2$$
$$V_{s_{13}} = 51.9$$
$$V_{s_{12}} = 59.6$$
$$V_{s_{11}} = 60.8$$
$$V_{s_{10}} = 62.2$$

Sign 9 was considered previously because of the first local minimum of VD(10). However, because the second local minimum is of smaller value, a tentative setting is computed using j = 16. This computation yields 65.4 fps. Because 65.4 is less than the previously computed setting of 65.9 it is concluded that the second local minimum has the stronger influence on vehicles on section 10. Hence, sign 9 is reset to $V_{s_9} = 65.4$.

Similarly, a tentative setting for sign 8 is computed, using j = 16, and the result is 69.0. Sign 8 is not reset because the previously determined setting of 68.7 fps is less than 69.0.

The third local minimum is found at VD(19). Sign 19 is not considered and the setting for sign 18 is found to be $V_{s_{18}} = 47.6$ using j = 19 in Eq. 21.

Sign 17 is not considered because VD(16) is less than the speed at the third local minimum of VD(19). Thus, cars on section 17 will not be forced to decelerate because of the local minimum of VD(19). For the same reason sign 16 is not considered.

Sign 15 has been previously considered because of a local minimum further upstream of smaller value, i.e., VD(16). For this reason the setting of sign 15 is considered valid. Furthermore, additional scanning of signs further upstream is suppressed and the search for the next local minimum is initiated.

SUMMARY

Although a cost-benefit analysis has not been done, several factors indicate that the cost of a surveillance and control system with detectors spaced at 0.1-mile intervals may be justifiable. First, aside from the system benefit of reducing acceleration noise and thereby reducing the rate of occurrence of accidents, the surveillance part of the system would provide essentially instantaneous incident detection and location. Second, with the ever-decreasing cost of large-scale integration in circuit technology and the cost advantages obtained by large-scale use of identical chips, such as found in microprocessors, the system may well be built with all necessary computer capability at the site of each advisory sign, thereby eliminating the high cost of communicating with, owning, and operating a large central computer. Although field maintenance could be a drawback to such a distributed computing system, it is not likely to be expensive because it would be on a plug-in replacement basis.

Although this aspect has not been simulated, an increase of the sign spacing to 0.2 mile (0.3 km) could be tolerated, although an increase in the detectors' spacing would have a more detrimental effect because of the longer average delay before a slowdown would be detected. Of course, as pointed out by Altman (5), on an actual (inhomogeneous) highway, the detectors should be placed more densely in areas of greater intravehicular interference (e.g., near entrance ramps, physical bottlenecks, or permanent visual distractions). The optimum placement of detectors was not considered in this analysis, yet the derivation of the sign control algorithm is easily adapted to accommodate unequal detector or advisory sign spacings.

Compliance with the advisory sign messages is not expected to be a problem as long as the signs are used only when necessary and their messages are consistently reliable.

Each driver will learn to respond to the messages in a manner that maximizes his own comfort. An overall average level of response to the sign's message will be imposed on drivers due to the compliance of drivers further downstream.

The actual model used for a driver's response to a sign is not based on any experimental data but, rather, is patterned after the stimulus-response car-following models in the traffic flow literature. Although the constants in this model may be subject to question, the derivation of the sign control algorithm would be exactly parallel for different values of these constants.

The purpose of this paper has been to show that if a model for the response of a driver to a sign is assumed then an algorithm for controlling the signs along a highway may be readily found. An extensive simulation of the sign control system (6) using the algorithm derived in this paper indicates that even when driver compliance with the advisory signs is poor (i.e., 80 percent of the drivers ignore the sign's message completely) a significant reduction in acceleration noise is obtained. A detailed presentation of the simulation and its results will be made in a future paper.

ACKNOWLEDGMENT

This research was conducted as part of a doctoral dissertation. Partial financial support was provided by the City University of New York and also by the New York State Science and Technology Foundation.

REFERENCES

1. 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.
2. Weinstock, R. Calculus of Variations. McGraw-Hill, 1952, p. 61.
3. Kapur. Transportation Research. Vol. 5, No. 2, p. 69.
4. Helly, W. Simulation of Bottlenecks in Single-Lane Traffic Flow. In Theory of Traffic Flow (Herman, R., ed.), Elsevier, 1961.
5. Altman, S. The design of Automatic Surveillance Systems for Urban Freeways. Polytechnic Institute of Brooklyn, PhD dissertation, 1967.
6. Kleinman, S. A Real-Time Advisory Sign Control System for Urban Highways. City University of New York, PhD dissertation, 1971.

OPTIMIZATION TECHNIQUES APPLIED TO IMPROVING FREEWAY OPERATIONS

Adolf D. May, Institute of Transportation and Traffic Engineering,
University of California, Berkeley

This paper describes optimization techniques that have been developed and applied for evaluating freeway operational improvements such as redesign or ramp control strategies. First, a deterministic macroscopic simulation model is described that predicts the traffic performance on a directional freeway as a function of freeway design and traffic demand patterns. Then two decision models are presented that automatically work with the simulation model to select optimum redesign or ramp control strategies. Finally, a freeway corridor model is described. Emphasis is given to the structure of the model and to the first step in the development of the freeway corridor model, which is a major arterial street model.

• THIS PAPER reports on an investigation of various operational aspects of urban freeways, including the application of optimization techniques for improving freeway operations through redesign or control strategies. The first phase of the study began in October 1967 and continued until December 1968. It inventoried existing traffic conditions on approximately 140 directional miles of freeway in the San Francisco Bay area. It also identified critically congested sections, ascertained their cause, and estimated their effect on traffic operations. In addition, preliminary investigations were conducted to study means of improving identified critical sections and to prepare preliminary estimates of user benefits. Eight interim reports and a final summary report were published (1-9).

The second phase began in early 1969. It analyzed in detail two selected portions of the freeway system by using a systematic analytical procedure for evaluating design and control improvements. This procedure was computerized, and the program was named FREEQ. The results are covered in three project reports (10-12).

The third phase began in January 1972 and extended through June 1973. It studied four major areas: (a) freeway model refinement, (b) systems analysis, (c) control strategies, and (d) network flow. Six interim reports that cover the initial research work completed as of June 1972 were prepared (13-18). Efforts were continued during 1972 and 1973, and four reports (one in each area of research) were published in June 1973 that contained the most significant findings of the six interim reports plus the results of the extended work (19-22).

This paper attempts to summarize the research work completed and emphasizes the application of optimization techniques for improving freeway operations through redesign or control strategies. During the third phase of the study, research efforts considered two levels of activity: (a) only the freeway and (b) the freeway corridor. The first level research activity involves the development of analytical techniques for evaluating freeway design or ramp control improvement plans on a freeway-only basis. The foreground of Figure 1 shows a schematic representation of this freeway evaluation process.

Publication of this paper sponsored by Committee on Freeway Operations.

Figure 1. Optimization techniques for improving freeway operations.

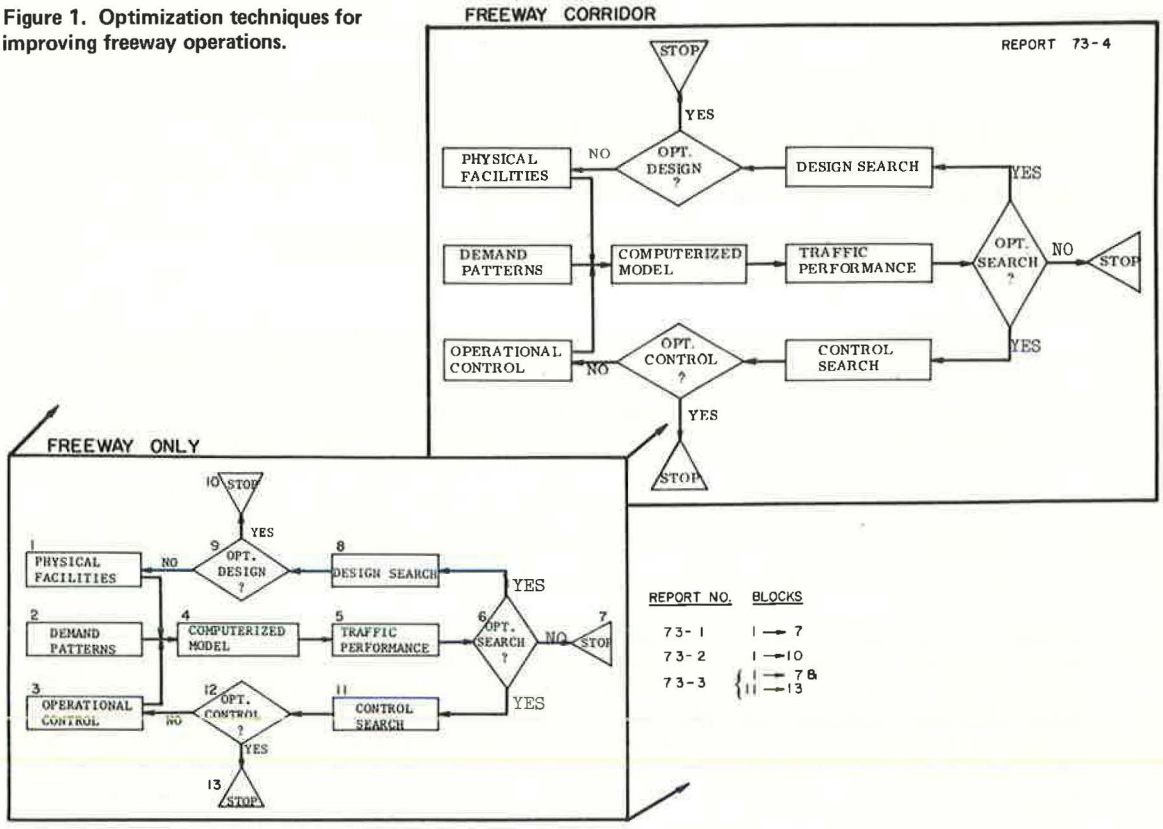
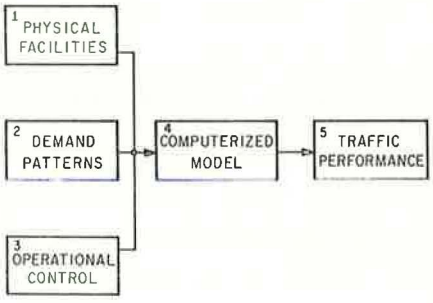


Figure 2. The FREQ3 freeway model.



FREQ3 FREEWAY MODEL

The FREQ3 freeway model (19) is a deterministic simulation model that predicts traffic performance on a directional freeway as a function of freeway design and traffic demand patterns. This is a third-generation version of the FREEQ model and is in essentially final form pending its ultimate incorporation into a freeway corridor model.

A schematic representation of the FREQ3 process is shown in Figure 2. The model consists of three main parts: (a) input (freeway design and traffic demand patterns), (b) FREQ3, and (c) output (freeway traffic performance).

Model Input

Input to FREQ3 consists of freeway design and traffic demand patterns, and Figures 3 and 4 show sample listings of such input data.

The directional freeway section is divided into subsections. A subsection is a portion of the freeway section over which the capacity and demand are essentially constant. In other words, a boundary exists between subsections at each on-ramp and off-ramp because of demand changes and at each location where the freeway capacity changes, such as at lane drops, lane additions, or significant grade changes. Pertinent information that is needed for each subsection includes subsection number, number of lanes, length, "straight-pipe" capacity, truck factor, inflow and outflow station locations, special ramp indicator, and location description. A listing of this input data is shown in Figure 3, and represents a directional freeway that is approximately 10 miles (16 km) long and has been divided into 30 subsections.

The study period is divided into equal time slices. A traffic demand pattern is specified for each time slice in the form of an origin and destination table. It is assumed that the arrival demands during the time slice are uniform. The length of the time slice can be selected by the user; time slices of 15 minutes appear to be reasonable. The number of origin and destination input tables is equal to the number of time slices. These tables may be based on results of origin and destination studies or can be estimated from ramp counts. Figure 4 shows the traffic demand pattern for one time slice that serves as input into the model.

Model Description

FREQ3 is described in two parts: the model structure and its computerized program.

The structure of the FREQ3 model can best be described by considering a distance-time diagram (Fig. 5). The horizontal scale is distance, with traffic flowing from left to right. The plan view of the directional freeway section is shown below the horizontal scale and has been divided into subsections. The vertical scale is time of day and is divided into equal time slices. A cell is defined as having a length of one subsection and a time interval of one time slice.

The design features of the various subsections are translated into capacities, and these capacity values C_{ij} can be thought of as flowing upward through the distance-time diagram. The origin and destination tables (one for each time slice) are translated into subsection demands, and these demand values D_{ij} can be thought of as flowing to the right through the distance-time diagram. Initially it is assumed that there are no bottlenecks, but later this is checked and, if necessary, adjustments are made.

The procedural analysis of FREQ3 begins with the cell in the lower left-hand corner of Figure 5 that represents the first subsection during the first time slice. Then the cells immediately to the right are analyzed until all cells in the first time slice have been analyzed. Then the cell that represents the first subsection during the second time slice is analyzed, followed by the cells immediately to the right. This procedure is continued from time slice to time slice, until all cells are analyzed.

The analysis in each cell consists of comparing the C_{ij} value with the D_{ij} value. Two outcomes are possible: either $D_{ij} \leq C_{ij}$ or $D_{ij} > C_{ij}$. If $D_{ij} \leq C_{ij}$, then this subsection is not a bottleneck, and the flow V_{ij} is equal to D_{ij} is equal to D_{ij}. The flow-capacity ratio can be calculated, and from this ratio, speeds and travel times are estimated. However, if $D_{ij} > C_{ij}$, then this subsection is a bottleneck and a more elaborate analytical procedure is required.

Figure 3. Freeway design input to FREQ3.

SUB SEC NO.	NO. LANE	CAP.	LENG. FT.	TRK. FACTOR			SUBSECTION DESCRIPTION		
1	5	9434	1630	.943	OD		MAINLINE ORIGIN	TO	POWELL OFF
2	5	9453	1960	.961			POWELL OFF	TO	POWELL ON
3	5	9615	1550	.961	OD		POWELL ON	TO	ASHBY OFF
4	4	7619	1960	.952			ASHBY OFF	TO	ASHBY ON
5	4	7550	500	.952	O		ASHBY ON	TO	500 FEET PT.
6	4	7550	4790	.952	D		500 FEET PT.	TO	UNIVERSITY OFF
7	4	7692	3030	.961			UNIVERSITY OFF	TO	UNIVERSITY ON
8	4	7692	2160	.961	OD		UNIVERSITY ON	TO	GILMAN OFF
9	4	7563	2030	.961			GILMAN OFF	TO	GILMAN ON
10	5	9130	1250	.970	OD		GILMAN ON	TO	BUCHANAN OFF
11	4	7619	900	.952	D	1	BUCHANAN OFF	TO	HOFFMAN OFF (LEFT)
12	3	5671	1320	.961	D		HOFFMAN OFF	TO	PIERCE OFF
13	3	5750	720	.961			PIERCE OFF	TO	PIERCE ON
14	3	5806	2610	.970	OD		PIERCE ON	TO	CENTRAL OFF
15	3	5728	1660	.970			CENTRAL OFF	TO	CENTRAL ON
16	3	5806	1890	.970	OD		CENTRAL ON	TO	CARLSON OFF
17	3	5520	2310	.940			CARLSON OFF	TO	CARLSON ON
18	4	7930	1460	.980	OD		CARLSON ON	TO	POTRERO OFF
19	4	7680	3800	.970			POTRERO OFF	TO	CUTTING ON
20	5	9140	1100	.980	O		CUTTING ON	TO	GRADE CHANGE PT.
21	5	9230	660	.980	D		GRADE CHANGE PT.	TO	MACDONALD OFF
22	4	7930	1480	.980	D	0	MACDONALD OFF	TO	SAN PABLO OFF
23	4	7840	1480	.970			SAN PABLO OFF	TO	SAN PABLO ON
24	5	8840	800	.980	OD		SAN PABLO ON	TO	SOLANO OFF
25	4	7604	4690	.970	D		SOLANO OFF	TO	SAN PABLO DAM OFF
26	4	7760	2190	.970			DAM ROAD OFF	TO	DAM ROAD ON
27	4	7680	2320	.970	OD		DAM ROAD ON	TO	ROAD 20 OFF
28	4	6732	830	.850			ROAD 20 OFF	TO	GRADE CHANGE PT.
29	4	6480	1180	.793			GRADE CHANGE PT.	TO	ROAD 20 ON
30	4	6265	2560	.780	OD		ROAD 20 ON	TO	MAINLINE DESTINATION

ALL RAMPS LIMIT = 1500 EXCEPT OFF-RAMP 6 LIMIT = 2200

Figure 4. Traffic demand pattern input to FREQ3.

```
TIME SLICE 1  FB72N
1.38     =    OCCUPANCY FACTOR
 144 198 227  54 108 252  10  88  25  37  15  40  17  80  44 220
         4     5   5           9  22   5   5   2   9   5  24
             7    20       7  12      19   3   3  13   7  36
                  41           11   7  11       4  16  11  48
              7  31       4   4   8      12   4  17  10  53
                              2   4   2       1   6   3  14
                                 13           3  11   6  31
                                  6           2   8   4  22
                                          17  21  56  20  74
                                              9  34  21 102
                                                     9  46
                                                        44

TIME SLICE 2  FB72N
 1.38
 183 229 232  76  99 325   7  88  46  26  30  69  10  53  28 334
       2   3   2   8  12       2   2   8               3  12   5  36
           5   2  13  23      10   3  17   9  11   9  33  17 103
                  43          19   8  17   8   2   6  25  14  81
              13  16   3  13  10  19  16           6  26  15  86
                                  3               3   9   4  23
                              3  10   6   6   4  14   8  44
                                  3   5   3   3  11   6  33
                                             33  55  32 119
                                             13  39  23 124
                                                     10  56
                                                         44

TIME SLICE 3  FB72N
 1.38
 183 226 214  42 161 314   9 123  51  39  40 112  19  82  33 213
         5    9      14       9      14   5       3   8   8  31
              3      20       9   4   6   7   2   7  21  11  69
                     51       4   7  31   2       9  22  12  73
             16  12       7   7  31   2           7  22  12  75
                                  2               1   6   3  20
                              2   5   2   2   3  10   4  36
                                  2   5           6  16   7  49
                                             33  95  27 115
                                             13  39  19 128
                                                     9  57
                                                        44

TIME SLICE 4  FB72N
 1.38
 203 191 171  45 129 215   4  71      76  33  67   5  27  10 562
              4   6  19   4  16                   4  12   7  40
          2       7  58   2  15  10  17  12  28  10  33  19 103
                     34       4  11  11  20  13  12  41  24 126
             20  83      20  48  21      10   6  20  12  66
                          2                   2   3   9   5  30
                              3   6   9   4  19   7  39
                              7   3       5  14   9  44
                                             34 108  40 153
```

In this more elaborate case four steps are required: First, inasmuch as $D_{ij} > C_{ij}$, then V_{ij} is equal to C_{ij}, not D_{ij}. Second, because $(D_{ij} - C_{ij})$ vehicles are being stored upstream of this bottleneck, the previously predicted downstream demands must be reduced. Third, the backward-moving shock wave is determined, and new flow and travel time situations upstream of the bottleneck are recalculated. Finally the excess demand at this bottleneck during the time slice $D_{ij} - C_{ij}$ is added to the origin and destination table for the next time slice.

This procedure becomes extremely complicated when several bottlenecks occur, and the resulting queues may collide and split at different times and at different locations. Of course, the decreasing queuing situation, which was not described above, is also handled (19).

FREQ3 has been computerized and is written in FORTRAN IV for the University of California's CDC 6400 computer. The computer program consists of the main program, which essentially is a "calling" program, 17 subroutines, and one function. A flow chart of the computer program is shown in Figure 6.

The FORTRAN program deck consists of approximately 2,000 statements. The computer takes approximately 4 seconds to run the FREQ3 program, which includes 10 miles of congested freeway (30 subsections) and a 2½-hour period (ten 15-minute time slices). The computer charge is approximately $1.70.

The computer program results have been calibrated with real-world data obtained from the northbound East Shore Freeway in the San Francisco Bay area.

Model Output

Output from FREQ3 includes speeds, densities, flows, and travel times for each cell; individual trip times and total travel times for each time slice; and total travel times and travel distances for the entire study section during the study period.

The user may request different detail of output information depending on the particular problem. A sample of the basic detail that is provided with every output is shown in Figure 7. This sample output is for one time slice, and one such output is provided for the user for every time slice in the study period.

COST-EFFECTIVENESS EVALUATION OF FREEWAY DESIGN ALTERNATIVES

This section describes an extension of the FREQ3 freeway model, called the FREQ3D design model and demonstrates its application for evaluation, on a cost-effectiveness basis, of alternative design improvements on the northbound East Shore Freeway (20). This is a second-generation version of the FREQ3D design model and is considered to be in essentially final form pending its ultimate incorporation into a freeway corridor model.

A schematic representation of the FREQ3D process is shown in Figure 8. The model consists of the FREQ3 with an iteration procedure for evaluating preselected design improvement alternatives (a design improvement alternative consists of adding a lane at a bottleneck location). The iterative procedure consists of two major tasks: (a) generation of design alternatives and (b) evaluation of design alternatives. Before these two major tasks are described, the structure of the problem will be introduced.

Problem Structure

The directional freeway can be thought of as a sequence of facilities, represented by each subsection, designed to handle arriving traffic. Each subsection acts essentially as a service facility serving a queue. The whole system behaves like a series of queues being processed through a series of service facilities.

Two important features of the system can be identified from this description. First, the travel times are nonlinear functions of the flow-capacity ratio. The result is that improvements in the system are nonlinear functions of investment in capacity. The second important feature is the highly interactive nature of a series of queues. This leads to the realization that the value of improvements in one subsection is not independent of the improvements in other subsections.

Figure 5. Distance-time diagram for FREQ3.

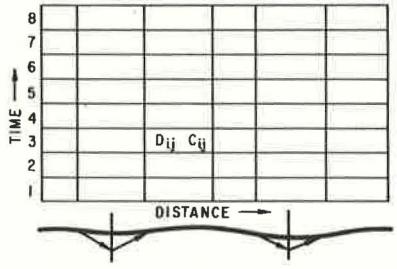


Figure 6. Flow chart of FREQ3 computer program.

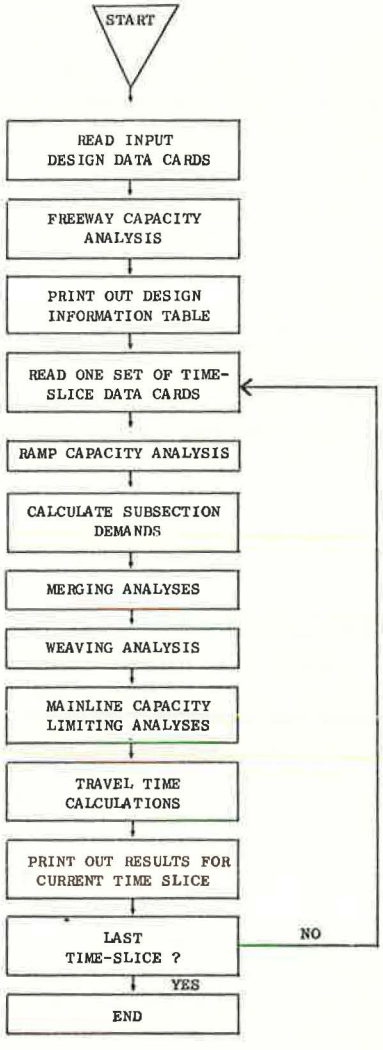


Figure 7. Sample output of FREQ3 computer program.

TIME SLICE 4 FB72C/C 3

OCCUPANCY 1.38

QUEUE SUB SEC	COLLISION ..INPUT DEMAND... ORG	DES	TOTAL	VOL	FWAY CAP	WEAVE EFF	V/C	DENS V/M/L	MPH		TRAV TIME	LENG FEET	QUEUE	N-PRIME
1	7236	812	7236	7236	9434	0	.77	25	56		.33	1630	0	0
2	0	0	6424	6424	9453	0	.68	22	57		.39	1960	0	0
3	448	764	6872	6872	8809	806	.78	24	56		.31	1550	0	0
4	0	0	6108	6013	7619	0	.79	34	43	*	.52	1960	592	95
5	1264	0	7372	7277	7277	273	1.00	51	35		.16	500	0	0
6	0	692	7372	6064	7277	273	.83	45	33	*	1.94	4790	4330	1213
7	0	0	6680	5302	7692	0	.70	51	26	*	1.43	3030	3029	1213
8	1184	196	7864	6566	7399	293	.89	63	26	*	.96	2140	2139	893
9	0	0	7668	6384	7563	0	.84	69	23	*	1.00	2030	2029	833
10	1224	648	8852	7608	7608	1522	1.00	43	35		.40	1250	0	0
11	0	1636	8244	7061	7342	277	.96	32	54		.19	900	0	0
12	0	40	6608	5646	5671	0	1.00	49	38		.39	1320	0	0
13	0	0	6568	5602	5750	0	.97	41	45	*	.18	720	137	10
14	204	504	6772	5806	5806	0	1.00	55	35		.84	2610	0	0
15	0	0	6268	5376	5728	0	.94	32	55		.34	1660	0	0
16	348	284	6616	5724	5806	0	.99	43	44		.48	1890	0	0
17	0	0	6332	5451	5520	0	.99	42	43		.60	2310	0	0
18	240	512	6572	5691	5831	119	.98	38	49		.33	1460	0	0
19	0	0	6060	5249	5806	0	.90	31	55		.78	3800	0	0
20	900	0	6960	6149	6866	494	.90	27	55		.23	1100	0	0
21	0	284	6560	6145	6936	494	.89	27	55		.14	660	0	0
22	0	516	6676	5505	7930	0	.74	26	56		.30	1480	0	0
23	0	0	6160	5460	5728	0	.95	33	55		.31	1480	0	0
24	576	292	6736	6036	6331	519	.95	27	55		.17	800	0	0
25	0	868	6444	5161	5800	0	.99	48	40		1.33	4690	0	0
26	0	0	5576	4960	5806	0	.85	30	55		.45	2190	0	0
27	264	580	5840	5224	5800	0	.90	31	55		.48	2320	0	0
28	0	0	5260	4675	5649	0	.93	28	55		.17	830	0	0
29	0	0	5260	4679	4746	0	.99	35	44		.30	1180	0	0
30	0	5260	5260	4679	4700	0	1.00	41	38		.75	2560	0	0

	CURRENT TIME INTERVAL	CUMULATIVE VALUES
FREEWAY TRAVEL TIME=	394. VEH-HRS= 544. PASS-HRS	1145. VEH-HRS= 1580. PASS-HRS
INPUT DELAY=	0. VEH-HRS= 0. PASS-HRS	0. VEH-HRS= 0. PASS-HRS
TOTAL TRAVEL TIME=	394. VEH-HRS= 544. PASS-HRS	1145. VEH-HRS= 1580. PASS-HRS
TOTAL TRAV DISTANCE=	16015. VEH-MI.= 22100. PASS-MI.	57710. VEH-MI.= 79640. PASS-MI.

As a result of these factors, the most common optimization techniques are not suitable. Linear and piece-wise linear programming are not appropriate because of the nonconvexity of the objective function. Dynamic programming and marginal analysis are not appropriate because of the interdependence between the subsections. Complete enumeration is not possible because of the large number of alternatives (between 2^{30} and 3^{30}).

Fortunately some practical considerations aid in limiting the number of alternatives and, more importantly, in simplifying the structure of the problem. First, all desired improvements cannot be built simultaneously, nor is it likely that previous improvements would be destroyed. Consequently, this limits the procedure to a unique sequence of design improvement stages. Second, it is unlikely that an operating agency would improve a nonbottleneck subsection while a bottleneck subsection exists nearby (even if it were shown to be more cost-effective). Consequently, this significantly limits the number of alternatives to be investigated at each stage of improvement. The net result of these practical considerations is a sequence of stages that give the appearance of a branch and bound mathematical programming technique combined with comparing alternatives at each stage in a marginal analysis manner.

Generation of Design Alternatives

The procedure employed in generating design alternatives consists of (a) identifying bottleneck subsections from the previously selected stage of improvement and (b) formulating a design improvement alternative for each identified bottleneck.

The output from the FREQ3 freeway model (this model is actually a submodel in the FREQ3D freeway design model), which represents the previously selected stage of improvement (Fig. 7), is inspected. Any subsection in any time slice that exhibits a flow-capacity ratio of unity is identified as a bottleneck subsection.

Then, a design improvement alternative is formulated for each bottleneck subsection. Only feasible lane arrangements are considered so that good highway design practice is followed. A lane improvement plan is feasible if it satisfies the following constraints:

1. The number of lanes downstream of an off-ramp must either be equal to or be one less than the number of lanes upstream of the off-ramp,
2. The number of lanes downstream of an on-ramp must either be equal to or be one more than the number of lanes upstream of the on-ramp,
3. Only one lane can be added or dropped at a subsection boundary, and
4. The maximum number of lanes permitted in a subsection is six.

Evaluation of Design Alternatives

The procedure used in evaluating design alternatives consists of (a) calculating the annual cost (in dollars) of each design alternative and its measure of effectiveness (savings in passenger-hours) and (b) selecting the design alternative, at this stage, that exhibits the smallest marginal cost-effectiveness (dollars per passenger-hour saved).

The estimated cost of adding a lane to each subsection and parameters such as interest rate, improvement life, and maintenance cost are placed as input to FREQ3D. The model is then used to calculate the annual cost of each design alternative. The freeway model calculates the total passenger-hours expended with and without the design improvement, and the difference in passenger-hours reflects the savings due to the design improvement.

The marginal cost-effectiveness is calculated for each design alternative at this stage, and the design alternative that has the smallest cost-effectiveness is selected. This leads to the next stage of improvement, and a new series of design alternatives is generated.

Numerical Results

The design alternative method was used to calculate the cost-effectiveness curve for the northbound East Shore Freeway. The successive stages of improvements from the existing freeway conditions to noncongested freeway conditions are shown in block-diagram form in Figure 9 and in cost-effectiveness form in Figure 10.

Figure 8. The FREQ3D freeway design model.

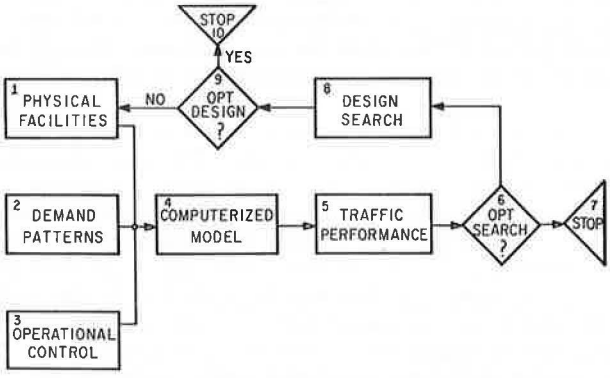
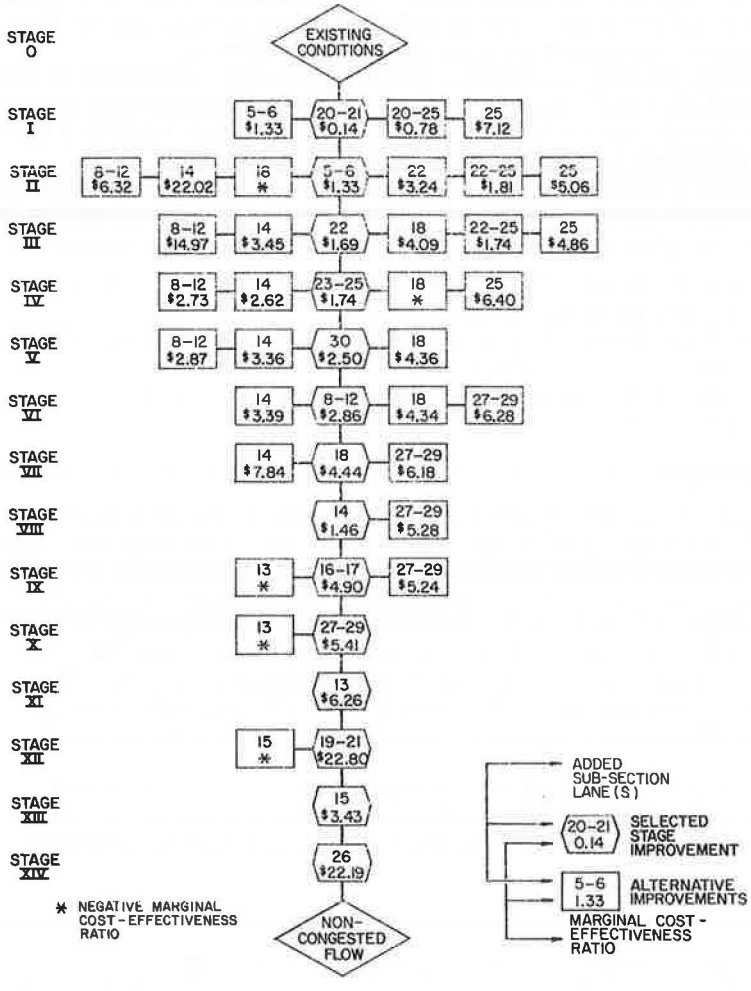


Figure 9. Block diagram of most cost-effective method applied to northbound East Shore Freeway.



Numerous attempts have been undertaken to improve this approximation to the optimal solution (15, 20). One attempt proceeded as in the previously described manner, but at each stage the most effective rather than the most cost-effective design alternative was selected. Another attempt used study teams consisting of experienced engineers with the California and Nevada highway departments who generated and evaluated design alternatives based on their experience and judgment. The final attempt to improve the approximation to the optimal solution used the first as well as the second and third most cost-effective design alternatives. The net results of all three attempts were that the number of iterations was significantly increased and the resulting solutions were almost identical to the solutions of the initially developed procedure.

ANALYSIS OF FREEWAY ON-RAMP CONTROL STRATEGIES

An extension of the FREQ3 freeway model called the FREQ3C freeway control model and its application for developing ramp control strategies on the northbound East Shore Freeway in the San Francisco Bay area are discussed in this section. This is a second-generation version of the FREQ3C control model and is considered to be in essentially final form pending its ultimate incorporation into a freeway corridor model.

A schematic representation of the FREQ3C control model process is shown in Figure 11. It consists of the FREQ3 model and a linear programming decision model (LINCON), which together provide the user with three important outputs for each time slice: (a) freeway traffic performance without ramp control, (b) optimum metering rate for each on-ramp and a traffic diversion table, and (c) freeway traffic performance with the selected ramp control strategy.

An iterative procedure is required because ramp control affects the weaving flows, weaving flows affect the weaving capacities, and weaving capacities affect the optimum capacities affect the optimum ramp control. A convergence algorithm is incorporated into the model to handle the iteration process automatically.

The analytical techniques for ramp control developed in this research work will be presented in three parts: (a) linear programming formulation, (b) various options in using the freeway control model, and (c) numerical results of real-world application.

Linear Programming Formulation

The objective in the linear programming formulation can be either to maximize vehicle-input or to maximize vehicle-miles of travel on the directional freeway

$$\text{Max} \sum_{i=1}^{n} X_i \quad \text{or} \quad \text{Max} \sum_{i=1}^{n} \ell_i X_i \tag{1}$$

where

X_i = desired input rate from on-ramp i or the metering rate,
ℓ_i = average trip length of all traffic from on-ramp i, and
n = number of on-ramps.

The constraints in the linear programming formulation include the following:

$$\sum_{i=1}^{n} A_{ik} X_i \leq B_k \qquad \text{for } k = 1, 2, \ldots, m \tag{2}$$

$$X_i \leq D_i \qquad \text{for } i = 1, 2, \ldots, n \tag{3}$$

$$X_i \geq 0 \qquad \text{for } i = 1, 2, \ldots, n \tag{4}$$

$$S_i \leq X_i \leq T_i \qquad \text{for } i = 1, 2, \ldots, n \tag{5}$$

Figure 10. Cost-effectiveness diagram of most cost-effective method applied to northbound East Shore Freeway.

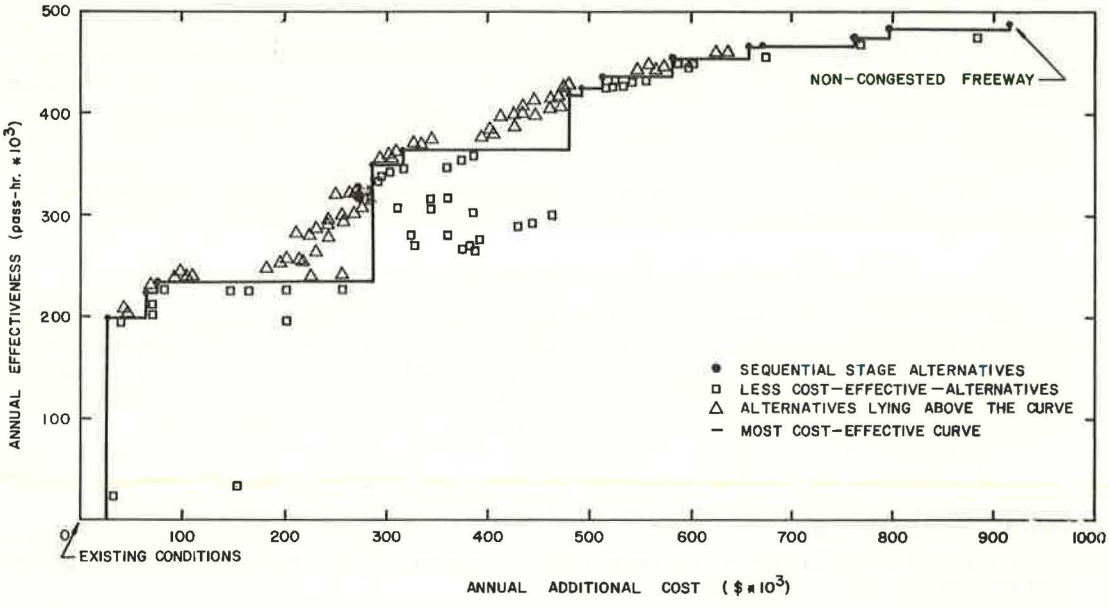
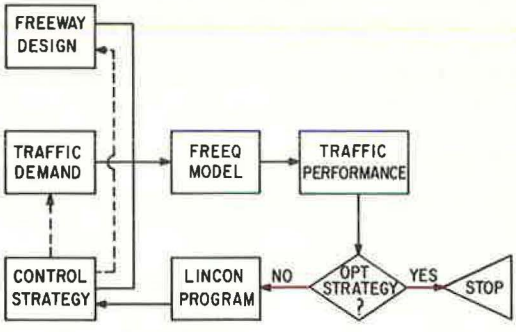


Figure 11. The FREQ3C freeway control model.



where

A_{ik} = fraction of traffic from on-ramp i going through subsection k,
B_k = the capacity of subsection k,
D_i = the demand rate at on-ramp i,
S_i = minimum metering rate,
T_i = maximum metering rate, and
m = number of subsections.

The first set of constraint equations limits the flow demand on every subsection to the subsection capacity. The second set of constraint equations limits the on-ramp flows to the demand rate. The third set of constraint equations limits the on-ramp flows to nonnegative values. The fourth set of constraint equations sets upper and lower limits on the metering rate.

Model Options

Numerous options have been incorporated into the model in order to give flexibility to the user. As indicated in Eq. 1, the objective function can be either to maximize vehicle-input or to maximize vehicle-miles of travel.

At the end of each time slice, the optimum metering rate for certain ramps may be lower than the demand rate. In this situation a queue will exist at the end of the time slice. One of two options is available: (a) It is assumed that no vehicles are diverted and excess demand from one time slice is transferred to the demand of the next time slice (no diversion); or (b) it is assumed that excess vehicles will divert to parallel routes and will not use the freeway (total excess vehicle diversion).

In the event that the ramp demand exceeds the optimum metering rate, some vehicles will wait or be diverted. Two options are available to establish which vehicles will have priority entry: (a) It is assumed that the waiting or diverted vehicles will be determined on a proportional basis—i.e., if 10 percent of the vehicles at a particular ramp must wait or be diverted, then 10 percent of each destination set is selected; or (b) it is assumed that the diverted vehicles will be determined on a trip length basis—i.e., short freeway trips will be diverted before longer freeway trips.

The set of bottleneck constraints can be selected at the capacity level or at a lower service volume level. When the capacity level is selected, there is no slack capacity or safety factor and the emphasis is on production or facility efficiency alone. When a service volume level is selected (service volume is defined as a traffic volume level below capacity that results in higher operating speeds), there is some slack capacity or safety factor and emphasis is diverted between facility efficiency and freeway level of service.

The weaving effect may or may not be considered. If the user considers the weaving effect, capacities (or service volumes) will be estimated more accurately, but an iterative procedure between capacity analysis and ramp control will be required. If the user does not consider the weaving effect, no iterative procedure will be necessary, but there may be some sacrifice in capacity calculation accuracy.

All input demands to the directional freeway are controlled through ramp metering except the main-line input demand. Inasmuch as the main-line input demand may vary from day to day and if the expected demand rate is selected, on approximately 50 percent of the days the actual demand will exceed the expected demand and congestion may occur on the freeway. Therefore, the user may select either the expected demand rate or a slightly higher demand rate that is selected on the basis of the variance-to-mean ratio and desired confidence limits.

Numerical Results

The input to the FREQ3C model includes the basic input requirements to FREQ3 and the specifications of the linear programming formulation. The output for each time slice consists of the freeway traffic performance without ramp control (Fig. 12), the optimum metering rate for each on-ramp and a traffic diversion table (Fig. 13), and the freeway traffic performance with the selected ramp control strategy (Fig. 14).

Figure 12. Freeway traffic performance without ramp control.

F072N DATA - TIME SLICE 6

TIME SLICE 5 OF 5
GROWTH PERIOD 0 OF 0
OCCUPANCY 1.38

ORIGIN - DESTINATION TABLE (VEHICLES PER HOUR)

ORIGINS DOWN / DESTINATIONS ACROSS

0	1	2	3	4	5	6	7	8	9	SUM
1	536	356	420	212	372	180	584	248	2496	5404
2	8	0	40	16	40	20	64	28	232	448
3	0	20	0	0	0	24	68	32	264	408
4	0	0	0	0	16	12	40	20	156	244
5	0	0	0	0	12	152	388	128	576	1256
6	0	0	0	0	0	88	244	116	920	1368
7	0	0	0	0	0	0	0	32	276	308
8	0	0	0	0	0	0	0	0	220	220
SUM	544	376	460	228	440	476	1388	604	5140	9656

ROW SUMS = TOTAL INPUT DEMANDS, COLUMN SUMS = TOTAL OUTPUT DEMANDS

QUEUE COLL. SECTION 12 T2= .025
QUEUE OUT OF SECTION 1

SUB SEC	NO. LNS	SSEC LENGTH	O-D DATA DEMANDS ORG.	DES.	DEM.	ADJUSTED VOLUMES ORG.	DES.	VOL.	FRWY CAP.	WEAVE EFF	V/C	DENS. V/M/L	SPEED MPH		TRAVEL TIME	QUEUE-LENGTH	STORAGE RATE
1	3	720.	5404.	0.	5404.	4539.	0.	4589.	5750.	0.	.80	80.0	19.1	**	.43	720.	815.
2	3	2610.	448.	544.	5852.	448.	539.	5037.	5806.	0.	.87	76.3	22.0	**	1.35	2610.	769.
3	3	1660.	0.	0.	5308.	0.	0.	4498.	5728.	0.	.79	80.5	18.6	**	1.01	1660.	769.
4	3	1890.	408.	376.	5716.	408.	373.	4906.	5806.	0.	.84	77.7	21.0	**	1.02	1890.	769.
5	3	2310.	0.	0.	5340.	0.	0.	4533.	5520.	0.	.82	75.3	20.1	**	1.31	2310.	769.
6	3	1460.	244.	460.	5584.	244.	456.	4777.	5877.	73.	.81	80.8	19.7	**	.84	1460.	769.
7	3	3800.	0.	0.	5124.	0.	0.	4320.	5806.	0.	.74	84.6	17.0	**	2.54	3800.	769.
8	3	1100.	1256.	0.	6380.	1256.	0.	5334.	5334.	546.	1.00	50.5	35.2		.36	0.	0.
9	3	660.	0.	228.	6380.	0.	181.	4947.	5404.	546.	.92	37.5	44.0	*	.17	169.	388.
10	3	1480.	0.	440.	6152.	0.	352.	4765.	5857.	93.	.81	40.4	39.3	**	.43	1480.	388.
11	3	1480.	0.	0.	5712.	0.	0.	4413.	5728.	0.	.77	63.3	23.2	**	.72	1480.	388.
12	4	800.	1368.	476.	7080.	1368.	423.	5780.	6096.	754.	.95	55.7	25.9	**	.35	800.	316.
13	3	4690.	0.	1388.	6604.	0.	1153.	5357.	5357.	443.	1.00	50.7	35.2		1.51	0.	0.
14	3	2190.	0.	0.	5216.	0.	0.	4204.	5806.	0.	.72	24.6	56.9		.44	0.	0.
15	3	2320.	308.	604.	5524.	308.	504.	4512.	5800.	0.	.78	26.6	56.4		.47	0.	0.
16	3	830.	0.	0.	4920.	0.	0.	4008.	5049.	0.	.79	23.7	56.3		.17	0.	0.
17	3	1180.	0.	0.	4920.	0.	0.	4008.	4746.	0.	.84	23.9	55.9		.24	0.	0.
18	3	2560.	220.	5140.	5140.	220.	4228.	4228.	4700.	0.	.90	25.4	55.4		.52	0.	0.
TOTAL		33740.											TOTAL		13.88		

		QUEUE LENGTH VEHICLES	DELAY VEH-HRS
ON-RAMP 1	INPUT POINT	404.61	75.69
	MERGING POINT	0.	0.
	TOTAL	404.61	75.69
ON-RAMP 5	INPUT POINT	0.	0.
	MERGING POINT	261.08	57.71
	TOTAL	261.08	57.71
ON-RAMP 6	INPUT POINT	0.	0.
	MERGING POINT	.15	.02
	TOTAL	.15	.02

	CURRENT TIME INTERVAL				CUMULATIVE VALUES			
FREEWAY TRAVEL TIME=	273.	VEH-HRS=	376.	PASS-HRS	1076.	VEH-HRS=	1485.	PASS-HRS
INPUT DELAY=	133.	VEH-HRS=	184.	PASS-HRS	225.	VEH-HRS=	311.	PASS-HRS
OUTPUT DELAY=	0.	VEH-HRS=	0.	PASS-HRS	0.	VEH-HRS=	0.	PASS-HRS
TOTAL TRAVEL TIME=	406.	VEH-HRS=	560.	PASS-HRS	1301.	VEH-HRS=	1796.	PASS-HRS
TOTAL TRAV DISTANCE=	7526.	VEH-MI.=	10386.	PASS-MI.	37026.	VEH-MI.=	51096.	PASS-MI.

TRAVEL TIME FOR ONE TRIP IN MINUTES

0	1	2	3	4	5	6	7	8	9
1	1.78	3.81	5.96	9.02	9.45	10.52	12.04	12.94	13.88
2	1.35	3.38	5.53	8.59	9.02	10.10	11.61	12.52	13.45
3	0.	1.02	3.17	6.23	6.65	7.73	9.25	10.15	11.09
4	0.	0.	.84	3.90	4.33	5.41	6.92	7.82	8.76
5	0.	0.	0.	.53	.95	2.03	3.54	4.45	5.38
6	0.	0.	0.	0.	0.	.35	1.86	2.77	3.70
7	0.	0.	0.	0.	0.	0.	0.	.47	1.40
8	0.	0.	0.	0.	0.	0.	0.	0.	.52

END OF SIMULATION FOR ABOVE CRITERION

Figure 13. Optimum metering rates and diversion table.

MEAN OD DEMAND-RAMP QUEUE(HOURLY RATE)**FETZN DATA - TIME SLICE 6

		DESTINATION								
ORIGIN	TOTAL	1	2	3	4	5	6	7	8	9
1	5404.	536.	356.	420.	212.	372.	180.	584.	248.	2496.
2	448.	8.	0.	40.	16.	40.	20.	64.	28.	232.
3	408.	0.	20.	0.	0.	0.	24.	68.	32.	264.
4	244.	0.	0.	0.	0.	16.	12.	40.	20.	156.
5	1256.	0.	0.	0.	0.	12.	152.	388.	128.	576.
6	1368.	0.	0.	0.	0.	0.	88.	244.	116.	920.
7	308.	0.	0.	0.	0.	0.	0.	0.	32.	276.
8	220.	0.	0.	0.	0.	0.	0.	0.	0.	220.
SUM	9656.	544.	376.	460.	228.	440.	476.	1388.	604.	5140.

THE STRATEGY IS TO BE FREE OF RECURSIVE CONGESTION WITH* .90 *PROBABILITY
TAKING INTO CONSIDERATION THE FLUCTUATION OF THE MAINLINE INPUT
THE MAINLINE INPUT IS ASSUMED TO BE NORMALLY DISTRIBUTED WITH VARIANCE= 1.00 *MEAN **********
CAPACITIES TAKE 5 ITERATIONS TO CONVERGE**********

O-D PATTERN UNDER RAMP CONTROL, IN HOURLY RATES

	METERING/OR	DESTINATION								
ORIGIN	INPUT RATE	1	2	3	4	5	6	7	8	9
1	5498.	545.	362.	427.	216.	378.	183.	594.	252.	2539.
2	250.	0.	0.	0.	0.	0.	0.	0.	18.	232.
3	408.	0.	20.	0.	0.	0.	24.	68.	32.	264.
4	244.	0.	0.	0.	0.	16.	12.	40.	20.	156.
5	662.	0.	0.	0.	0.	0.	0.	14.	113.	536.
6	519.	0.	0.	0.	0.	0.	0.	15.	39.	466.
7	240.	0.	0.	0.	0.	0.	0.	0.	0.	240.
8	220.	0.	0.	0.	0.	0.	0.	0.	0.	220.
SUM	8041.	545.	382.	427.	216.	394.	219.	731.	473.	4653.

MAXIMUM METERING RATES ARE *MAINLINE* 720. 720. 720. 720.1080. 720. 720.
MINIMUM METERING RATES ARE *MAINLINE* 240. 240. 240. 240. 240. 240. 240.
V/C BUFFER= .010 CAPACITY BUFFER= -0. VPH

****MEASURE OF EFFECTIVENESS****	CURRENT TIME INTERVAL ******	****** CUMULATIVE VALUES *****
TOTAL INPUT	2010. VEH	9844. VEH
RAMP INPUT	636. VEH	3195. VEH
TOTAL VEHICLE FREEWAY TRAVEL	8304. VEH-MILE	39611. VEH-MILE
RAMP VEHICLE FREEWAY TRAVEL	2130. VEH-MILE	9596. VEH-MILE

TOTAL TRIPS DIVERTED,IN HOURLY RATE= 1709. WITH THE FOLLOWING O-D PATTERN

		DESTINATION								
ORIGIN	TOTAL	1	2	3	4	5	6	7	8	9
1	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
2	198.	8.	0.	40.	16.	40.	20.	64.	10.	0.
3	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
4	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.
5	594.	0.	0.	0.	0.	12.	152.	374.	15.	40.
6	849.	0.	0.	0.	0.	0.	88.	229.	77.	454.
7	68.	0.	0.	0.	0.	0.	0.	0.	32.	36.
8	0.	0.	0.	0.	0.	0.	0.	0.	0.	0.

Figure 14. Freeway traffic performance with ramp control.

```
FREEWAY DATA - TIME SLICE 6

TIME SLICE  5 OF  6
GROWTH PERIOD  0 OF  0
OCCUPANCY  1.36

                                ORIGIN - DESTINATION TABLE (VEHICLES PER HOUR)

ORIGINS                              DESTINATIONS ACROSS
DOWN
   0     1     2     3     4     5     6     7     8     9    SUM

   1   545   382   427   215   378   183   594   252  2539  5498
   2     0     0     0     0     0     0     0    17   232   249
   3     0    19     0     0     0    23    67    31   263   407
   4     0     0     0     0    16    12    40    20   156   244
   5     0     0     0     0     0     0    14   112   535   662
   6     0     0     0     0     0     0    14    38   465   519
   7     0     0     0     0     0     0     0     0   240   240
   8     0     0     0     0     0     0     0     0   220   220

SUM    545   382   427   215   394   219   730   473  4652  8037

ROW SUMS = TOTAL INPUT DEMANDS, COLUMN SUMS = TOTAL OUTPUT DEMANDS
```

SUB SEC	NO. LNS	SSEC LENGTH	O-D DATA DEMANDS ORG.	DES.	DEM.	ADJUSTED VOLUMES ORG.	DES.	VOL.	FRWY CAP.	WEAVE EFF	V/C	DENS. V/M/L	SPEED MPH	TRAVEL TIME	QUEUE-LENGTH	STORAG RATE
1	3	720.	5498.	0.	5498.	5498.	0.	5498.	5750.	0.	.96	33.4	54.8	.15	0.	0.
2	3	2610.	250.	545.	5748.	250.	545.	5748.	5806.	0.	.99	45.3	42.3	.70	0.	0.
3	3	1460.	0.	0.	5203.	0.	0.	5203.	5728.	0.	.91	31.3	55.4	.34	0.	0.
4	3	1890.	408.	382.	5611.	408.	382.	5611.	5806.	0.	.97	35.3	53.0	.41	0.	0.
5	3	2310.	0.	0.	5228.	0.	0.	5228.	5520.	0.	.95	31.7	55.0	.48	0.	0.
6	3	1460.	244.	427.	5472.	244.	427.	5472.	5920.	24.	.92	33.0	55.2	.30	0.	0.
7	3	3600.	0.	0.	5045.	0.	0.	5045.	5806.	0.	.87	30.2	55.7	.78	0.	0.
8	3	1100.	662.	0.	5707.	662.	0.	5707.	5772.	108.	.99	44.1	43.1	.29	0.	0.
9	3	660.	0.	216.	5707.	0.	216.	5707.	5842.	108.	.98	38.6	49.3	.15	0.	0.
10	3	1480.	0.	394.	5492.	0.	394.	5492.	5950.	0.	.92	33.1	55.2	.30	0.	0.
11	3	1460.	0.	0.	5097.	0.	0.	5097.	5728.	0.	.89	30.6	55.5	.30	0.	0.
12	4	800.	519.	219.	5616.	519.	219.	5616.	6434.	416.	.87	25.2	55.6	.16	0.	0.
13	3	4690.	0.	731.	5397.	0.	731.	5397.	5800.	0.	.93	32.6	55.2	.97	0.	0.
14	3	2190.	0.	0.	4666.	0.	0.	4666.	5806.	0.	.80	27.7	56.2	.44	0.	0.
15	3	2320.	240.	473.	4904.	240.	473.	4906.	5800.	0.	.85	29.3	55.9	.47	0.	0.
16	3	830.	0.	0.	4433.	0.	0.	4433.	5049.	0.	.88	26.6	55.6	.17	0.	0.
17	3	1180.	0.	0.	4433.	0.	0.	4433.	4746.	0.	.93	26.8	55.1	.24	0.	0.
18	3	2560.	220.	4653.	4653.	220.	4653.	4653.	4700.	0.	.99	36.6	42.3	.69	0.	0.
TOTAL		33740.											TOTAL	7.34		

	CURRENT TIME INTERVAL			CUMULATIVE VALUES		
FREEWAY TRAVEL TIME=	159.	VEH-HRS=	220. PASS-HRS	739.	VEH-HRS=	1020. PASS-HRS
INPUT DELAY=	0.	VEH-HRS=	0. PASS-HRS	0.	VEH-HRS=	0. PASS-HRS
OUTPUT DELAY=	0.	VEH-HRS=	0. PASS-HRS	0.	VEH-HRS=	0. PASS-HRS
TOTAL TRAVEL TIME=	159.	VEH-HRS=	220. PASS-HRS	739.	VEH-HRS=	1020. PASS-HRS
TOTAL TRAV DISTANCE=	8304.	VEH-MI.=	11460. PASS-MI.	39611.	VEH-MI.=	54663. PASS-MI.

TRAVEL TIME FOR ONE TRIP IN MINUTES

0	1	2	3	4	5	6	7	8	9
1	.85	1.60	2.37	3.59	3.90	4.36	5.33	6.24	7.34
2	.70	1.45	2.22	3.44	3.75	4.21	5.18	6.09	7.19
3	0.	.41	1.18	2.40	2.71	3.17	4.14	5.05	6.15
4	0.	0.	.30	1.52	1.82	2.29	3.26	4.17	5.27
5	0.	0.	0.	.44	.75	1.21	2.18	3.09	4.19
6	0.	0.	0.	0.	0.	.16	1.13	2.04	3.14
7	0.	0.	0.	0.	0.	0.	0.	.47	1.57
8	0.	0.	0.	0.	0.	0.	0.	0.	.69

```
END OF SIMULATION FOR ABOVE CRITERION
```

PROGRESS TOWARD A FREEWAY CORRIDOR MODEL

The proposed structure of the freeway corridor model and the first-generation version of a major arterial street model (MART1) are described in this section. This is considered to be a second-level research activity because it involves the development of analytical techniques for evaluating design improvements or control strategies on a freeway corridor basis. The illustration in the background of Figure 1 is a schematic representation of this freeway corridor evaluation process.

Structure of Freeway Corridor Model

A freeway corridor consists of a directional freeway with at least one parallel major arterial street and with cross-arterials connecting the major arterials to the freeway. Geographically, the corridor can be thought of as being 5 to 10 miles (8-16 km) long and 1 to 2 miles (1.6-3.2 km) wide.

The structure of the freeway corridor model is similar to the freeway model (Fig. 1). The input is expanded from freeway design parameters to freeway corridor design parameters, and the freeway demand patterns are expanded to freeway corridor demand patterns. The freeway simulation model becomes the freeway corridor simulation model, and the freeway design and control decision models become the freeway corridor design and control models.

Obviously, the freeway corridor model will be much more complex. The handling of freeway corridor demands will be difficult, and route selection or traffic assignment will be required. The control alternatives will include intersection control as well as freeway ramp control. The complexities of demand and capacity analysis will be significantly increased. In addition to the added complexities, the quantity of analysis is grossly expanded, and the computer time and user ease must be maintained within certain limits.

It is anticipated that the freeway corridor modeling will be undertaken in stages. First, a major arterial street model will be developed. The second stage will be to extend the major arterial street model to a major arterial corridor model that would include the parallel and crossing arterial streets. At this point this developed model manually combined with the freeway model can be employed for simulating a freeway corridor.

The next stage would be to combine the major arterial corridor model with the freeway model to automatically simulate a freeway corridor with the assumption that a preselected routing of the demand pattern is specified. Then, the preselected routing would be replaced by a built-in traffic assignment subroutine. Finally, the freeway corridor design and control decision models (extensions of the freeway design and control decision models) would be developed and incorporated with the freeway corridor model.

Major Arterial Street Model

The major arterial street model can be thought of as the freeway model with two important extensions. First, the concern now is with two directions of flow, not one, and the two directions of flow may interfere with one another. Second, signalized intersections are introduced, and the question of signal timing and intersection queuing and delay must be handled. The developed major arterial street model is called the MART1 model.

Input to MART1 consists of design parameters, traffic demand patterns, and intersection control parameters (splits, cycle length, and offsets). The MART1 model is a deterministic macroscopic simulation model. It has been computerized, and is written in FORTRAN IV for the University of California's CDC 6400 computer. The FORTRAN program deck consists of approximately 1,200 statements.

The output from the MART1 model includes capacity, demand, volume, queuing characteristics, and average delay for each approach to each intersection, plus volume and travel time for each link. Also, total vehicle-hours on the system are presented. All of these results are provided for each time slice during the time period of interest, as with FREQ3. Some sample output is shown in Figure 15.

Figure 15. Sample output of MART1 computer program.

INT	ARM	L.T. TYPE	CAPCTY	L.T. CAPCTY	L.T. DEMAND	THR. DEMAND	R.T. DEMAND	QUEUE	L.T. QUEUE	QDELAY	L.T. QDELAY	DELAY	L.T. DELAY
1	1	1	1396.	0.	27.	206.	50.	0.	0.	0.	0.	14.1	0.
1	2	1	1045.	0.	5.	150.	38.	0.	0.	0.	0.	9.5	0.
1	3	1	1933.	0.	12.	316.	29.	0.	0.	0.	0.	14.1	0.
1	4	1	921.	0.	61.	203.	26.	0.	0.	0.	0.	9.9	0.
2	1	1	1889.	0.	24.	180.	9.	0.	0.	0.	0.	11.0	0.
2	2	1	707.	0.	17.	37.	20.	0.	0.	0.	0.	11.8	0.
2	3	1	1890.	0.	5.	294.	21.	0.	0.	0.	0.	11.4	0.
2	4	1	754.	0.	28.	79.	21.	0.	0.	0.	0.	12.1	0.
3	1	1	2105.	0.	21.	170.	10.	0.	0.	0.	0.	6.7	0.
3	2	1	753.	0.	19.	56.	27.	0.	0.	0.	0.	17.3	0.
3	3	1	2227.	0.	15.	295.	16.	0.	0.	0.	0.	7.0	0.
3	4	1	786.	0.	14.	41.	8.	0.	0.	0.	0.	17.1	0.
4	1	1	2371.	0.	2.	149.	19.	0.	0.	0.	0.	5.9	0.
4	2	1	408.	0.	2.	1.	8.	0.	0.	0.	0.	20.3	0.
4	3	1	2384.	0.	9.	312.	6.	0.	0.	0.	0.	6.2	0.
4	4	1	377.	0.	8.	1.	13.	0.	0.	0.	0.	20.3	0.
5	1	1	1606.	0.	20.	131.	27.	0.	0.	0.	0.	10.8	0.
5	2	1	1766.	0.	1.	293.	55.	0.	0.	0.	0.	12.5	0.
5	3	1	1751.	0.	14.	309.	33.	0.	0.	0.	0.	11.3	0.
5	4	1	1765.	0.	1.	222.	27.	0.	0.	0.	0.	12.2	0.
6	1	1	1754.	0.	37.	192.	9.	0.	0.	0.	0.	12.2	0.
6	2	3	1008.	992.	29.	57.	65.	0.	0.	0.	0.	10.5	9.8
6	3	1	1835.	0.	7.	279.	41.	0.	0.	0.	0.	12.5	0.
6	4	1	920.	0.	23.	131.	8.	0.	0.	0.	0.	10.7	0.
7	1	1	1762.	0.	7.	58.	22.	0.	0.	0.	0.	9.7	0.
7	2	1	1111.	0.	9.	53.	37.	0.	0.	0.	0.	13.0	0.
7	3	1	1836.	0.	27.	365.	14.	0.	0.	0.	0.	10.5	0.
7	4	1	1018.	0.	53.	95.	24.	0.	0.	0.	0.	13.2	0.

TOTAL VEHICLE HOURS 871.1 TOTAL PASSENGER HOURS 1202.1
CUM. TOTAL VEHICLE HOURS 2620.8 CUM. TOTAL PASSENGER HOURS 3616.8

ACKNOWLEDGMENT

This freeway operations study was sponsored by the California Division of Highways in cooperation with the U.S. Department of Transportation.

REFERENCES

1. May, A. D., and Gillfillan, W. E. Bay Area Freeway Operations Study, First Interim Report. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Jan. 1968.
2. Ybarra, W. A., and May, A. D. Bay Area Freeway Operations Study, Second Interim Report. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, July 1968.
3. Salam, Y. M., Gillfillan, W. E., and May, A. D. Bay Area Freeway Operations Study, Third Interim Report. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Dec. 1968.
4. Jacobs, R. R., and Gillfillan, W. E. Bay Area Freeway Operations Study, Fourth Interim Report. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Dec. 1968.
5. Jacobs, R. R. Bay Area Freeway Operations Study, Fifth Interim Report. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Feb. 1969.
6. Hom, R. E., and Gillfillan, W. E. Bay Area Freeway Operations Study, Sixth Interim Report. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, March 1969.
7. Jacobs, R. R. Bay Area Freeway Operations Study, Seventh Interim Report. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, March 1969.
8. Gillfillan, W. E. Bay Area Freeway Operations Study, Eighth Interim Report. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, March 1969.

9. Bay Area Freeway Operations Study, Summary of the Comprehensive Phase. Institute of Transportation and Traffic Engineering, Univ. of California, March 1969.
10. Makigami, Y., Woodie, L., and May, A. D. Bay Area Freeway Operations Study, Final Report, Part I: The Freeway Model. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Aug. 1970.
11. Allen, B. L., and May, A. D. Bay Area Freeway Operations Study, Final Report, Part II: On the San Francisco-Oakland Bay Bridge. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Aug. 1970.
12. Aidoo, J. F., Goedhart, R. W., and May, A. D. Bay Area Freeway Operations Study, Final Report, Part III: On the Eastshore Freeway (I-80) Northbound. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Aug. 1970.
13. Faravardeh, M. Freeway Operations Study—Phase III, Interim Report No. 1: A Suggested Freeway Corridor Model. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, July 1972, unpublished.
14. Wang, J. J., and May, A. D. Freeway Operations Study—Phase III, Interim Report No. 2: Analytical Techniques for Ramp Control. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, June 1972.
15. May, A. D., and James, J. H. Freeway Operations Study—Phase III, Interim Report No. 3-1: Cost-Effectiveness Evaluation of Freeway Design Alternatives. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, July 1972.
16. Torpp, O. C., and May, A. D. Freeway Operations Study—Phase III, Interim Report No. 3-2: Extended Analytical Techniques for Evaluating Freeway Operations Improvement. Institute of Transportation and Traffic Engineering, Berkeley, June 1972.
17. Blankenhorn, R. C., and May, A. D. Freeway Operations Study—Phase III, Interim Report No. 4-1: FREQ2—A Revision of the 'FREEQ' Freeway Model. Institute of Transportation and Traffic Engineering, Univ. of California, June 1972.
18. Yeh, S. F., and May, A. D. Freeway Operations Study—Phase III, Interim Report No. 4-2: Investigation of Capacity and Flow-Speed Relationship for Freeway Model. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, June 1972.
19. Stock, W. A., Blankenhorn, R. C., and May, A. D. Freeway Operations Study—Phase III, Report 73-1: The FREQ3 Freeway Model. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, June 1973.
20. Eldor, M., and May, A. D. Freeway Operations Study—Phase III, Report 73-2: Cost-Effectiveness Evaluation of Freeway Operations Study—Phase III, Report 73-2: Cost-Effectiveness Evaluation of Freeway Design Alternatives. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, June 1973.
21. Wang, J. J., and May, A. D. Freeway Operations Study—Phase III, Report 73-3: Analysis of Freeway On-Ramp Control Strategies. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, June 1973.
22. Orthlieb, M., Roedder, S., and May, A. D. Freeway Operations Study—Phase III, Report 73-4: Progress Toward a Freeway Corridor Model. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, June 1973.

SPONSORSHIP OF THIS RECORD

GROUP 3—OPERATION AND MAINTENANCE OF TRANSPORTATION FACILITIES
Lloyd G. Byrd, Byrd, Tallamy, MacDonald, and Lewis, chairman

Committee on Communications
Edward F. Weller, General Motors Research Laboratories, chairman
Edith Bairdain, Frank Barnett, Robert L. Cosgriff, Conrad L. Dudek, Robert Earl Fenton, John R. Freeland, William L. Garrison, Robert C. Harp, Arthur C. Johnson, Peter A. Mayer, Robert H. Parker, Donald G. Penterman, George Petrutsas, John J. Renner, Donald W. Richards, Lyle G. Saxton, S. J. Stephany, Glenn E. Wanttaja, Ivor S. Wisepart, Stanley Woolman

Committee on Freeway Operations
Robert S. Foote, Port Authority of New York and New Jersey, chairman
Patrick J. Athol, John L. Barker, Elmer N. Burns, J. R. Doughty, Conrad L. Dudek, A. C. Estep, Paul H. Fowler, Joseph W. Hess, E. Vinson Hoddinott, Hugo O. Liem, Alger F. Malo, Adolf D. May, Jr., William R. McCasland, Joseph M. McDermott, Stuart F. Millendorf, Ann Muzyka, Donald E. Orne, Eugene F. Reilly, Robert A. Reiss, David H. Roper, Robert E. Shields, T. Darcy Sullivan, A. Taragin, Joseph Treiterer, Joseph A. Wattleworth, Robert H. Whitson, H. Nathan Yagoda

Committee on Traffic Flow Theory and Characteristics
Donald G. Capelle, Alan M. Voorhees Associates, Inc., chairman
Robert F. Dawson, University of Vermont, vice 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, Transportation Research Board staff

Sponsorship is indicated by a footnote on the first page of each report. The organizational units and the chairmen and members are as of December 31, 1973.