HIGHWAY RESEARCH RECORD

Number 469 | Freeway Operational Improvements

6 reports
prepared for the
52nd Annual Meeting

Subject Areas

51	Highway Safety
53	Traffic Control and Operations
55	Traffic Measurements

HIGHWAY RESEARCH BOARD

DIVISION OF ENGINEERING NATIONAL RESEARCH COUNCIL
NATIONAL ACADEMY OF SCIENCES—NATIONAL ACADEMY OF ENGINEERING

Washington, D.C. 1973

NOTICE

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

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

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

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

ISBN 0-309-02255-X
Library of Congress Catalog Card No. 73-22748
Price: $2.40

CONTENTS

FOREWORD

Improved safety and efficiency of operations on freeways were the goals of research reported in the six papers in this RECORD. The nature of the problems investigated clearly supports the increasingly popular concept that freeways must be controlled and operated in order to secure maximum utility and optimum safety from existing facilities. The information presented will be useful to traffic and operations engineers, safety specialists, and others concerned with freeway traffic control strategies.

The need to alert drivers approaching crest vertical curves of traffic stoppages ahead led Dudek and Messer to study automatic means of detecting stoppage waves. After analyzing selected speed and energy parameters, they concluded that both were satisfactory indicators of stoppage waves. A computer algorithm was then structured for automatic control of the warning system.

Wang and May describe the development of a computer program capable of determining desired fixed-time metering rates for a group of freeway on-ramps. This ramp control model permits the user to choose between maximizing total vehicular input and maximizing total freeway vehicle-miles of travel.

Starting with the premise that merging of vehicles from on-ramps into the freeway traffic stream is a very likely cause of congestion and disturbance, Munjal, Hsu, and Lawrence attempted to predict the presence of acceptable gaps into which ramp vehicles can be merged. Gap prediction was of course affected by speed and lane changes, as confirmed by comparison with aerial photographic data used in validation of the theories. The authors conclude, however, that their gap-prediction strategy has good potential application value.

Chatfield examined fatal accident experience by highway system types and also looked at fatal Interstate System accidents as related to Interstate travel data. He reports that sections of systems with higher travel densities typically have lower fatal accident rates and that, for equivalent travel density differences between sections of a highway system, differences in fatal accident rates tend to be greater at lower densities.

Seeking ways to improve fixed-time ramp-metering control strategies, Payne, Meisel, and Teener considered a large number of traffic-responsive ramp control plans in terms of freeway service and delay. Their analysis yielded a set of plans that yield minimum delay for specified levels of freeway service. These plans are also compared to each other to yield a trade-off curve to assist in final selection of a metering plan.

Observing that the investigation of freeway accidents even on the shoulders is a cause of delay and congestion, Pittman and Loutzenheiser devised a system whereby the investigations were moved to concealed off-freeway sites. With cooperation from Houston police, some 60 percent of a year's total of 851 accidents on the Gulf Freeway were investigated at these and other off-freeway sites. They report a benefit-cost ratio of 28:1 considering the value of reduced delay versus site construction and maintenance costs.

DETECTING STOPPAGE WAVES FOR FREEWAY CONTROL

Conrad L. Dudek and Carroll J. Messer, Texas Transportation Institute, Texas A&M University

An experimental warning system has been installed on the inbound control section of the Gulf Freeway as a means of alerting drivers approaching crest vertical curves of stoppages downstream of the crest. Automatic control of the warning system dictated the need to identify measurable traffic parameters that indicate the presence of a stoppage wave. This paper presents an analysis of selected speed and energy parameters as indicators of stoppage waves. The results demonstrate that both the speed and energy parameters perform satisfactorily. Based on the results of the investigation, a digital computer control algorithm was structured for automatic control of the warning system. Recommendations are presented for detector placement.

•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 freeway geometrics 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 or 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 as an approach to reducing the effects of this problem (1). The purpose of the system is to assist the freeway driver approaching crest vertical curves in formulating his expectations of actual downstream traffic flow by alerting him of stoppage waves downstream of the crest.

Three overpasses were selected as 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 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). Although the warning signs can be controlled manually by remote switches located in the control center, automatic operation of the system by a computer is desired. Prior to the installation of the warning signs, double-loop detectors were installed on each lane and located on both sides of the three overpasses to study traffic characteristics relative to stoppage waves, to test automatic control algorithms, and to be used for real-time control. The primary function of the detectors downstream of the overpass is to sense stoppage waves so that the warning sign can be activated. The upstream detectors indicate when the sign should be turned off.

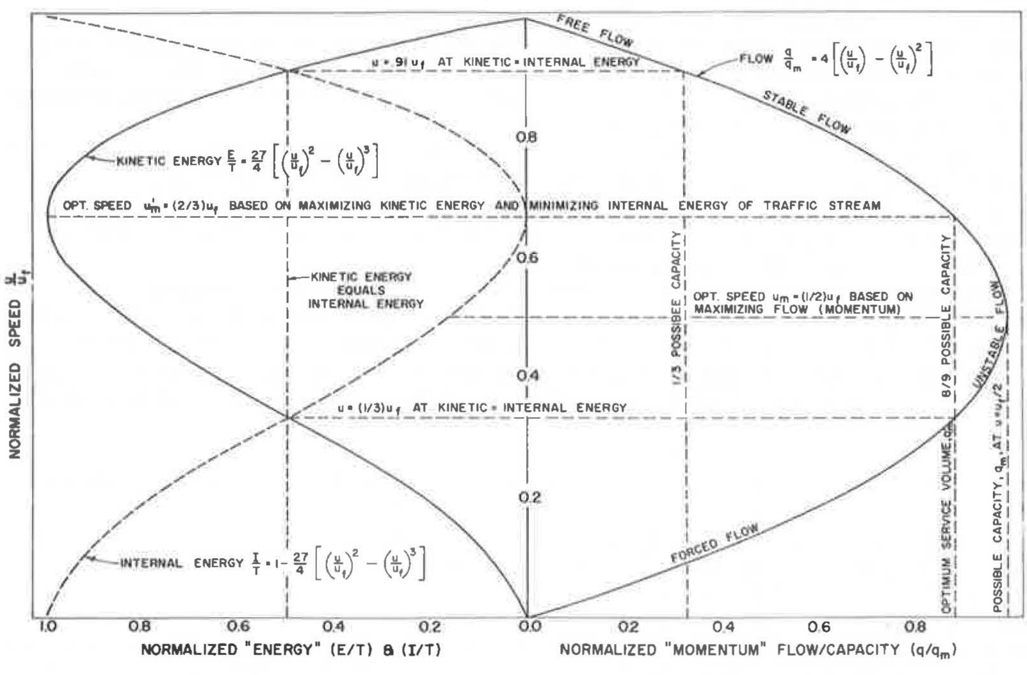
Several researchers have demonstrated the ability to identify major shock waves as they propagate upstream over detectors spaced at considerable intervals along a freeway lane (2, 3, 4, 5). Because traffic incidents can occur anywhere in the system (e.g., immediately downstream of an overpass), it was particularly important to evaluate the ability to detect or predict the passage of stoppage waves propagating across a single detector station. Automatic control of the warning system therefore dictated the need to identify measurable traffic parameters that indicate the presence of a stoppage wave.

Publication of this paper sponsored by Committee on Freeway Operations.

Figure 1. Warning sign with flashers.

Figure 2. Flasher unit at crest of overpass.

Figure 3. Quantitative approach to level of service using total energy-momentum analogy (7).



The selected parameter should minimize the probability that the system will not respond to a stoppage wave (type I error) and should minimize the number of false activations (type II error). This paper presents an analysis of selected speed and energy parameters as indicators of stoppage waves. Also included is the development of a digital computer control algorithm for the pilot system on the Gulf Freeway.

CONTROL VARIABLES

The traffic variables selected for analysis for automatic control of the warning system are speed and kinetic energy. The basic theory and the relation among speed, volume, and kinetic energy have been well documented in the literature (6, 7, 8). If we assume a linear function between speed and density, the normalized relationships of volume q and kinetic energy E_k can be written as a function of speed u:

$$q = k_j \left(u - \frac{u^2}{u_f} \right) \tag{1}$$

$$E_k = \alpha k_j u^2 - \alpha \frac{k_j}{u_f} u^3 \tag{2}$$

where

k_j = jam concentration and
u_f = free speed.

The relation between q and E_k is shown in Figure 3. Optimum service volume, based on maximizing kinetic energy and minimizing acceleration noise, corresponds to a level of flow that is less than capacity. Operating speed, on the other hand, is higher than the speed realized at capacity. The right side of Figure 3 shows that a small increase in demand above the volume at maximum energy tends to greatly increase the density of the traffic stream, accompanied inevitably by a sharp decrease in operating speed.

An examination of the relationship between energy and momentum reveals that the lower intercept of the energy and acceleration noise curves identifies forced flow conditions (level of service F) on the freeway. Flows are below capacity, and storage areas consisting of queues of vehicles form. This type of operation is indicative of stop-and-go traffic stream motion. The transition to the forced flow condition occurs rather rapidly (9). The intercept of the energy and acceleration noise curves occurs when the energy is one-half the maximum energy (E_m') of the stream. Based on this premise, it would appear initially that shock waves could be detected by measured energy less than one-half of maximum energy. This energy level can be referred to as the critical energy, E_c.

$$E_c = \tfrac{1}{2} E_m' \tag{3}$$

Associated with the critical energy parameter is a speed that might be referred to as critical speed u_c, which is equal to one-third of the free speed.

$$u_c = \tfrac{1}{3} u_f \tag{4}$$

Thus, the critical speed parameter might also serve as an initial parameter for evaluation.

It is emphasized that the energy will also be less than one-half maximum energy when the freeway is operating at level of service A. Therefore, it would be necessary to ascertain the level of service by measuring the speed characteristics. One reason for evaluating both energy and speed parameters even though they represent the same operating point in Figure 3 is to determine whether one variable is more sensitive and responsive than the other.

STUDY PROCEDURES

Equipment

Double-loop detectors are positioned on each lane of the inbound Gulf Freeway both upstream and downstream of three overpasses selected as the sites for the prototype safety warning devices. The locations of the three subsystems are shown in Figure 4. Traffic flow data from detectors are transmitted to an IBM 1800 digital computer located in the surveillance and control center. The data are then processed to compute traffic variables that can be used for control and then may be stored on disk, printed, or punched on cards.

Data Collection and Reduction

A computer program was written to collect data from the subsystem detectors, compute the desired traffic flow variables, and store the information at 30-sec intervals. Speed and volume were determined for each lane at both the upstream and downstream stations for the three subsystems. Speed was computed from the travel time of each vehicle between the two detectors. When an incident was observed on the study section, the computer stored the incoming data from the subsystem detectors on remote disk units for later analysis and processing. Simultaneously, a video tape recording was made to provide a visual record of traffic conditions during the incident. This provided the capability for later evaluation of traffic flow that could not be easily accomplished as it occurred. Video tape recordings of incidents were examined, and specific information on the origin of freeway shock waves and the time shock waves were observed to cross individual detectors were noted.

The quantitative computer data were examined, and the traffic flow condition based on speeds and flow rates prior to the shock wave passage was noted. The computer data and the video tape recording were synchronized in time, which permitted comparison of the two types of data.

Several computational time bases ranging from 10 sec to 2 min were considered for the program. Based on a preliminary study of the sensitivity of several traffic variables using different time bases within this time range and the results of freeway control research in different parts of the country, a time base of 1 min with data updated every 30 sec was selected. In other words, the traffic variables were computed for 1 min, and the values were updated every 30 sec by adding the most recent 30 sec of data and dropping the oldest 30 sec.

RESULTS

Critical Energy and Critical Speed

Least squares regressions were performed on kinetic energy-speed data consistent with the basic relationship

$$E_k = b_1 u^2 - b_2 u^3 \tag{5}$$

(where b_1 and b_2 are constants) by using base data collected at each detector station. Statistical tests of the regression coefficients were found significant in all cases at the 0.01 level. In addition, the R^2 values for each regression were all above 0.92, indicating good correlation between kinetic energy and speed.

Once the relationships between energy and speed were established, the maximum energy E_m', critical energy E_c, and critical speed u_c were calculated for each detector station (Table 1).

Detection of Stoppage Waves

Using the E and u as indicators of stoppage waves allowed us to compare, on an individual lane basis, the actual observation of 142 stoppage waves crossing one of the detectors and the time that the critical energy and speed parameters registered the presence of a wave. The observations were made when the freeway was operating at levels of service B, C, and D prior to the occurrence of a stoppage wave. The results of the analysis are shown as performance curves in Figure 5.

The values presented in the figure represent the difference in seconds between the time when the variable on the lane dropped below the critical value and the actual observed time of the stoppage wave moving over the detector. A positive value indicates that the energy or speed dropped below E_c or u_c before the wave was observed to cross the detector station. A negative value represents a late response by the parameter.

The results indicate that critical energy and speed are good parameters for the identification of a stoppage wave under levels of service B, C, and D. Generally, the parameters were able to predict the presence of a downstream stoppage wave. In general, there was little difference in the response between the energy and speed parameters.

Each parameter detected the presence of a stoppage wave either at the time the wave was moving over the detector or several seconds before the wave reached the detector stations in 131 of the 142 cases (93 percent). A total of 141 observations fell within the expected limits of the control logic. Because the 1-min values of energy and speed were updated every 30 sec, it was expected that a stoppage wave in some cases could conceivably pass over the detectors 30 sec before the computed energy or speed fell below the critical value. Therefore, a few late responses as high as 30 sec (-30) might be expected. In only one case did the parameters respond late by more than 30 sec (type I error). The reason for the lack of agreement for the one case could not be ascertained from the data and can only be conjectured at this time. However, the critical energy and speed parameters have been shown to possess a predictive characteristic for stoppage waves.

One-Lane Detection Criterion

Experience has shown that, although there is a degree of sympathy of speed between lanes regardless of volume, stoppage waves do not necessarily move in unison on each lane of a freeway (10). Generally, there are differences in the time that the waves on the individual lanes will reach a certain point on a freeway. An analysis of the relative movement of waves between lanes on the Gulf Freeway was made and is presented later. Because detectors for the safety warning device were placed on each lane, any one of the lanes could serve as the control lane. That is, the system was activated when a stoppage wave was sensed on any one of the lanes. An analysis was therefore made to test the responsiveness of detector stations having detectors in all three lanes to the occurrence of a stoppage wave. Forty-two stoppage waves resulting from incidents were evaluated as they crossed one of the five detector stations. The results of the analysis are shown in Figure 6.

The advance warning of a stoppage wave shown in the figure represents the difference between the time that the energy or speed dropped below the critical value on any one of the lanes and the time that the first stoppage wave was observed to cross one of the detectors. Again, positive values represent advance warning; negative values represent late responses.

The results clearly show that there was essentially no difference in response between the energy and speed parameters. In addition, with a three-lane detection station, adequate advance warning of stoppage waves is achieved within the limitations of the measurement technique. This is accomplished by allowing any one of the three lanes to predict the occurrence of a stoppage wave. In only one case did the wave pass over the detectors before the variable fell below the critical value on any one of the lanes. However, the difference was only 17 sec, well within the limit because of the 30-sec update of the data base.

A review of the data also revealed that, for the incidents studied, the stoppage waves were first detected on either the median or the middle lanes, or both, in 98 percent of the cases. An explanation of this result can be surmised. Because of the traffic leaving the shoulder lane via the off-ramps, the stoppage wave at times is interrupted and, therefore, will take longer to travel upstream. Earlier research on the Gulf Freeway by Drew (10) indicated that there does not seem to be any transverse pattern of failure

Figure 4. Location of test equipment.

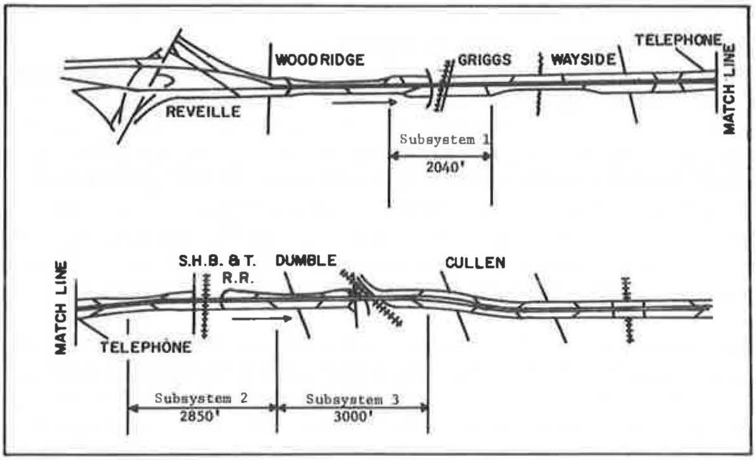
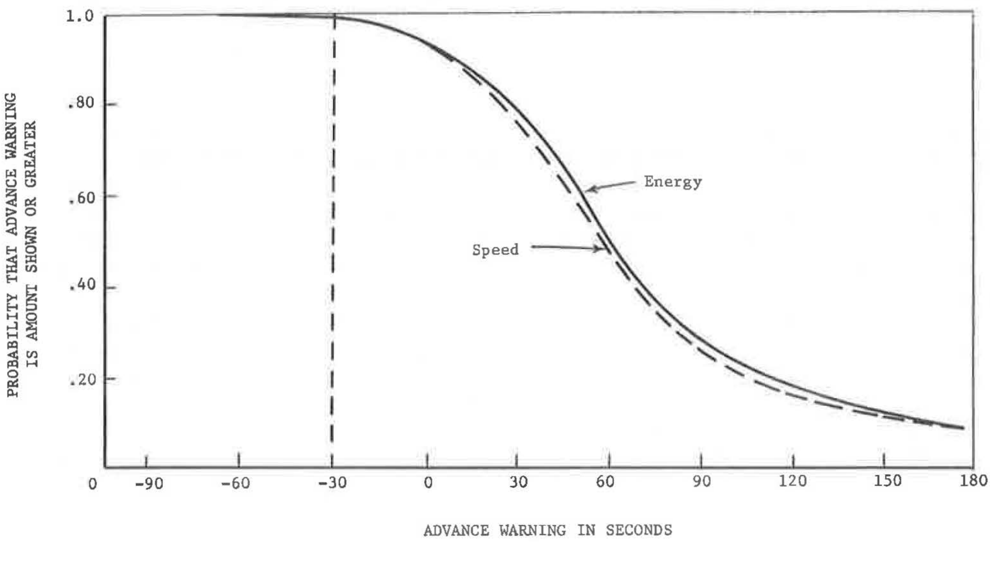


Table 1. Traffic parameters.

Location	Lane[a]	Maximum Energy (1,000 vmph^2)	Critical Energy (1,000 vmph^2)	Critical Speed (mph)
Mossrose	1	56.9	28.45	20.9
	2	68.3	34.15	21.0
	3	67.6	33.80	19.5
Griggs	1	49.2	24.60	19.5
	2	72.0	36.00	21.2
	3	72.7	36.35	20.5
Lombardy	1	75.0	37.50	22.6
	2	79.5	39.75	22.4
	3	79.7	39.85	21.3
Dumble	1	48.0	24.00	22.2
	2	73.0	36.50	21.2
	3	77.9	38.95	21.4
Cullen	1	53.7	26.85	20.8
	2	73.1	36.55	23.0
	3	73.7	36.85	20.1

[a]1 = shoulder, 2 = middle, and 3 = median.

Figure 5. Performance curves for individual lanes.



prompted by the spread of congestion from any one lane to the shoulder lane. He suggested that drivers seem to compensate for turbulence in the shoulder lane.

Type II Errors

The preceding section has shown that, for the incidents studied, a one-lane criterion was acceptable. The results revealed that there were no type I errors. This section discusses the results of an analysis for type II errors, false activations when a stoppage wave does not exist.

One of the assumptions made in selecting energy as one control variable is that the freeway will not be operating at level of service A during the normal periods of control (6 a.m. to 6 p.m.) because of the demands normally experienced on the freeway at these times. However, if for some reason short-period demands become light at a detector station, the energy values could conceivably drop below E_c, resulting in false activation of the safety warning device. Several hours of data, collected during off-peak and peak periods when no incidents occurred within the study section, were evaluated for the possibility of type II errors.

The results revealed that, generally, the operation of the device was satisfactory. However, it was observed that in some instances during the off-peak periods, particularly during the summer months, a reduction in freeway demand would cause false activations by using energy. Although not a frequent occurrence, the data indicated that the type II error was indeed a problem particularly during the summer and, therefore, would require attention. This was particularly true at subsystem 1 (Mossrose-Griggs).

The detector station at Griggs is located about 4,000 ft downstream of a major interchange and immediately downstream of a high-volume off-ramp. It appeared that the influence of the off-ramp coupled with motorists' desires to assume a comfortable headway after merging at the major interchange resulted in intermittent low-volume, high-speed measurement periods, particularly in the shoulder lane. The conditions were sufficiently severe to cause the energy to fall below E_c. This would give the indication of a stoppage wave.

There are at least two approaches that can be taken to circumvent this problem. One approach is to maintain a check of the speed and volume of the downstream detectors. Because the middle lane will carry a higher volume than the other two lanes during the off-peak periods, this lane can be used in the decision process. If the speeds remain above a threshold value, say 35 mph, while the volume in the middle lane stays above a threshold volume, say 8 vehicles per minute, this is an indication of random light flow on the affected lane. The safety warning device then would not be activated.

A second approach is to maintain the same control logic but increase the sampling time base. This would in effect smooth the energy function and reduce the severe peaking characteristic of the variable.

Each of these approaches was analyzed to determine its merits. Off-peak data, which indicated the highest frequency of type II errors, were used as the basis for the analysis. Thus, the approaches were evaluated under the worst noticeable conditions. Data from 414 sampling periods (30-sec periods) collected at the Griggs detector station on July 28 and 30 and August 3 and 4, 1971, were used as the base. The results of the analysis are shown in Figure 7.

The results indicate that using the basic control logic with an increased time base of up to 5 min reduces the frequency of false activations, but the reduction is not sufficient to be acceptable. The results also show that the approach wherein the middle lane at the downstream station is given a volume and speed check to determine the need to activate the safety warning device appears to be an acceptable solution. When a 1-min sampling period is used, a false activation would have occurred less than two-tenths of 1 percent (virtually zero) of the time under the worst possible conditions.

System Stability

It is imperative that the sign continue to operate from the time a stoppage wave is sensed until the wave or waves pass over the upstream detector station. Intermittent on-off operation is not desirable for apparent reasons.

Figure 6. Performance curve for one-lane criterion.

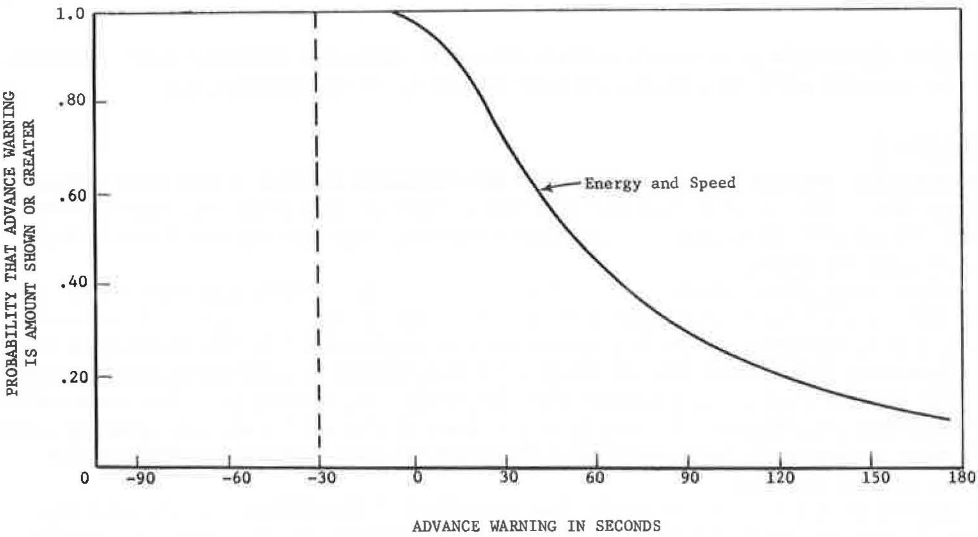


Figure 7. Effect of time base on type II errors.

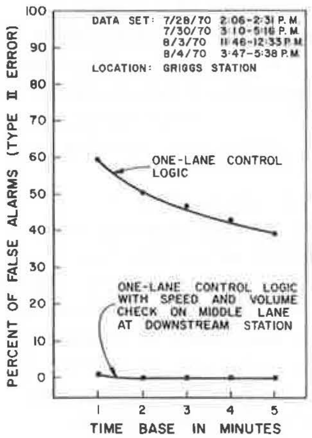


Figure 8. Relationship of stoppage wave differences between first and second waves.

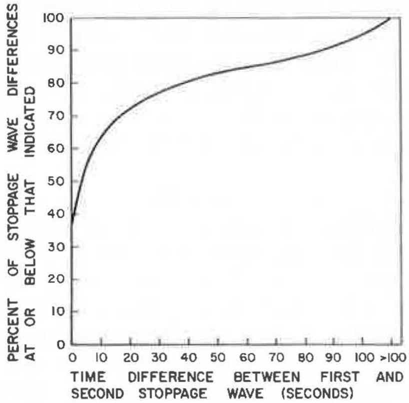
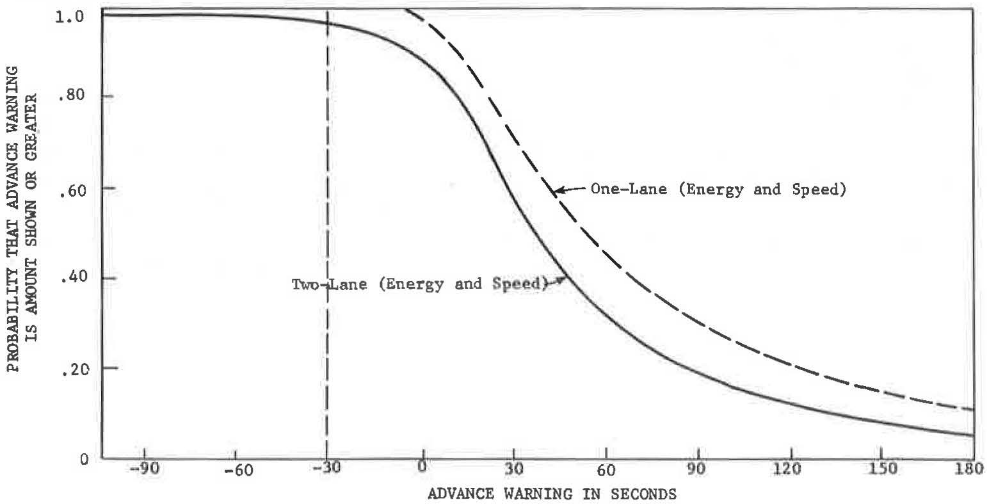


Figure 9. Performance curves for one-lane versus two-lane criterion.



The results of a system stability analysis revealed that, because of the fluctuations of traffic flow, there is some system instability when a 1-min data base is used. It was observed that the variation of traffic characteristics at the downstream detector stations would occasionally cause the safety device to turn off and on intermittently before the stoppage wave reached the upstream detector stations.

One solution to the problem is to require that the unit continue to operate for a fixed time following the initial period that resulted in energy values that called for system activation. An analysis of the data indicates that a hold time of six sampling periods (3 min) would be adequate to compensate for the possible instability of on-off cycling. This minimum time period can be reduced if the stoppage waves propagate over the upstream detectors sooner than 3 min. The minimum time can also be reduced when the sign is activated by slow vehicles (e.g., trucks and funeral processions).

Detector Problems

During the course of this research a rather high frequency of detector malfunctions was noted. This appears to be a problem common to operational freeway control systems. Because of the nature of the electronics associated with automatic traffic detection equipment, a particular detector may possibly become defective and thus transmit erroneous data to the computer or perhaps transmit no data at all. The problem is perplexing because a detector may become defective at any instant in time. Therefore, even though the detection equipment is thoroughly checked prior to control, there is no assurance that every detector will perform satisfactorily throughout the day.

The consequential effects of a defective detector in an automatic warning system for motorists are apparent. However, there are safeguard features that can be designed into the system to minimize their effects. One approach is to employ redundant detectors. Another approach is to rely on detectors on two lanes to give the alert of a major discontinuity in flow. That is, the safety warning device would not be activated unless the energy or speed on two lanes dropped below E_c or u_c. This approach, in effect, uses information from a detector on a second lane to verify the reliability of the data from the first. To test the feasibility of a two-lane control criterion required that an analysis be made of the relative movements of stoppage waves between lanes and the performance of this concept. These are discussed in the following sections.

Relative Movement of Stoppage Waves

The cumulative frequency of the time difference between the arrival of the first and second stoppage waves at the detector stations is shown in Figure 8. The plot reveals that, in approximately 23 percent of the cases studied, the second wave reached the detector station more than 30 sec after the first wave. Ten percent of the cases resulted in a time difference of 97 sec or more.

It would appear at the outset that a great degree of efficiency might be lost when using a two-lane control criterion. However, because the critical energy and speed parameters did exhibit predictive qualities, the effect of the parameters might compensate for some of the large time differences between stoppage waves. The extent of the change in performance relative to the one-lane criterion was evaluated and is discussed in the following section.

Two-Lane Control Criterion

The response times using a two-lane control criterion are shown in Figure 9. The results indicate that, generally, the system using a two-lane control criterion would respond within the expected limits. In only two cases out of 47 observed did the system respond later than 30 sec after the initial stoppage wave crossed the downstream detector station. In one extreme case, the system would not have sensed the presence of a major discontinuity in flow until 180 sec after the initial stoppage wave reached the detector station. A study of the video tapes revealed that in this one case the response would have been too late to warn motorists approaching the grade.

Comparisons of response times between the one-lane and two-lane criteria for critical energy and speed are also shown in Figure 9. As was expected, the results show that the two-lane criterion is less responsive to stoppage waves than the one-lane control criterion.

Trade-offs must be made in deciding the alternative course of action for an operational system. The one-lane criterion was shown to be acceptable. However, with this type of operation a defective detector on a lane can cause the warning sign to activate erroneously. The two-lane control criterion can compensate for the probability of a detector failure and appears to produce satisfactory results 96 percent of the time.

APPLICATIONS

This paper was concerned with a study of stoppage wave detection methods by using either critical speed or critical energy threshold values. Both parameters were found to be acceptable for application to digital computer control of the warning system.

Detector Location

Because the speed and energy parameters were computed from 1-min data updated every 30 sec, the response to a stoppage wave in some instances would be expected to be late by as much as 30 sec. The detectors located downstream of the overpasses must, therefore, be located a sufficient distance downstream of the critical freeway section to cope with this possibility. Observations in Houston have shown that the speed of stoppage waves will reach as much as 26 fps. Thus, it is possible that a wave could travel at least 720 ft (26 fps × 30 sec = 720 ft) before the logic responds to it. To ensure a factor of safety in the design, we selected a design travel distance of 800 ft. To ensure that the system responds to a stoppage wave before the wave reaches the foot of the vertical curve required that a suggested placement for the basic set of downstream detectors be developed (Fig. 10). It may at times be desirable to place some additional detectors downstream of this basic set to allow for a greater degree of advanced warning. The desirability of these additional detectors would be dictated by the specific problem location.

One function of the upstream detectors is to signify whether the conditions upstream of the overpass are such that a stoppage wave propagating on the far side of the overpass would result in hazardous conditions for approaching motorists. A second function is to turn the sign off once the stoppage wave has passed over the upstream detectors. The detectors should be located downstream from any bottlenecks existing in the immediate area. In some cases, a high-volume ramp may cause some congestion on the shoulder lane. The probability of not being responsive to a stoppage wave in the shoulder lane is reduced by positioning the detectors downstream from the ramp. The results of the analysis on the Gulf Freeway indicated that detectors placed at distances shown in Figure 10 appear to work satisfactorily, barring the influence of a ramp.

Control Logic

Based on the results of this study, a control logic was developed for digital computer control of the warning system. The logic assumes that a one-lane criterion is used at the downstream station and a two-lane criterion is used at the upstream location. Additions were made to permit 24-hour a day operation for 7 days each week. It should be understood, however, that the program is not overly sensitive to incidents occurring during periods of extremely light flow as would be experienced during the early morning hours. This is due to the limitations of detector placements in addition to the fact that the logic responds to the effects of incidents. As long as stoppage waves are present, the program will be responsive. However, the program is not capable of responding to incidents occurring between the upstream and downstream detector stations during these early morning hours. A flow chart for the control logic using energy as a control variable is shown in Figure 11. A comparable program can be structured for speed. The following list refers to the notation used in the figure and represents variables computed on a per-lane basis:

Figure 10. Suggested detector locations.

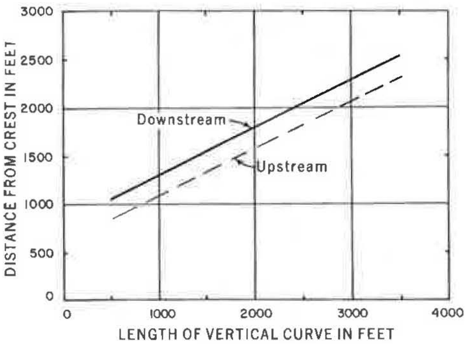
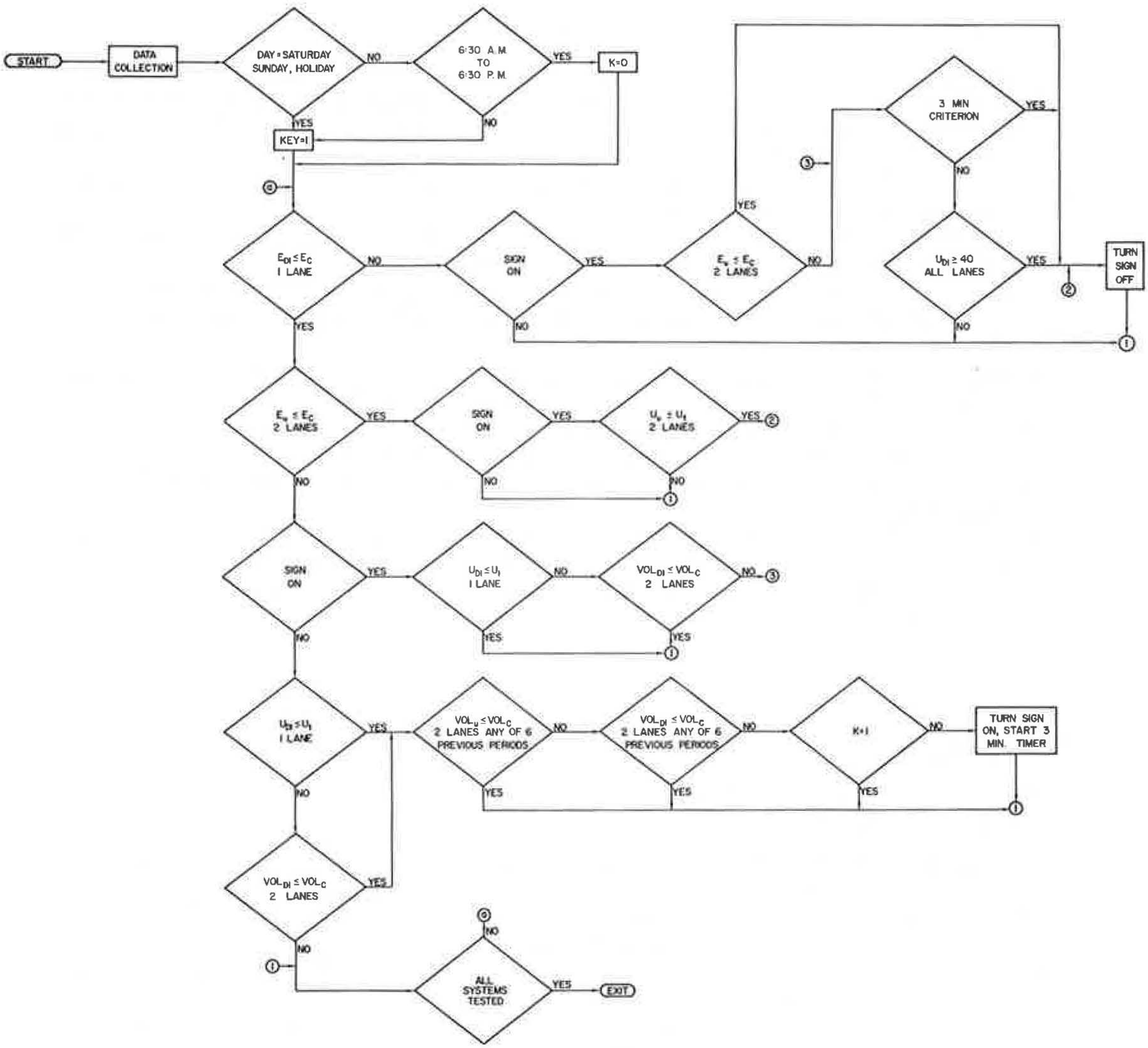


Figure 11. Flow chart of control logic.



E_u = energy upstream
E_{D1} = energy downstream
Vol_c = threshold volume (8 vpm)
Vol_u = volume upstream
Vol_{D1} = volume downstream
u_t = threshold speed (35 mph)
u_u = average speed upstream
u_{D1} = average speed downstream

ACKNOWLEDGMENTS

The authors wish to express their thanks and appreciation to Jim Gibbs, Programmer, Texas Transportation Institute, for writing the computer program of the control model and for his assistance in evaluating the model logic. Roy Loutzenheiser was instrumental in coordinating the data collection for model evaluation.

This paper discusses one phase of a research project entitled "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 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 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, Research Rept. 165-4, May 1972.
2. Barker, J. L. Determination of Discontinuities in Traffic Flow as a Factor in Freeway Operation. ITE Proc., 1961.
3. Auer, J. H., Jr. A System for the Collection and Processing of Traffic Flow Data by Machine Methods. HRB Bull. 324, 1962.
4. Foote, R. S., and Crowley, K. W. Developing Density Controls for Improved Traffic Operations. Highway Research Record 154, 1967.
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, pp. 38-50.
6. Drew, D. R., and Keese, C. J. Freeway Level of Service as Influenced by Volume and Capacity Characteristics. Highway Research Record 99, 1965.
7. Drew, D. R., Dudek, C. L., and Keese, C. J. Freeway Level of Service as Described by an Energy-Acceleration Noise Model. Highway Research Record 162, 1967.
8. Drew, D. R. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968, pp. 298-326.
9. Keese, C. J., and Schleider, R. H. Correlation of Design and Operational Characteristics of Expressways in Texas. HRB Bull. 170, 1958, pp. 1-23.
10. Drew, D. R. Stochastic Considerations in Freeway Operations and Control. Texas Transportation Institute, Research Rept. 24-5, June 1965.

DISCUSSION

Joseph A. Wattleworth, Department of Civil and Coastal Engineering, University of Florida

The authors have presented the development and preliminary testing of a very interesting driver information system. The system detects queues or stoppage waves in the freeway traffic stream and activates upstream beacons to warn the motorists of a forthcoming speed reduction. This system has the potential of accomplishing a great reduction in the frequency of rear-end accidents and other accidents caused by shock wave propagation in a traffic stream.

One area of potential application of a warning system of this type is the situation described by the authors, namely, areas in which there are sight distance restrictions caused by geometric design deficiencies. There is a more general area of potential application, however. This would be the warning of shock waves on any urban freeway during peak periods. When traffic densities are high, sight distance can be severely limited on a freeway of any geometric design. At operating conditions near capacity, headways are small, and this combination produces a potentially hazardous situation if a shock wave develops in the stream.

Another area of application would be freeway or tollroad situations in which traffic volumes are moderately high. In these cases, shock waves would be generated at irregular times and would be propagated through the traffic stream. An advanced warning of this situation would provide a definite safety benefit.

One must commend the authors for proposing a system that is inexpensive and easily understood by the drivers. At a time when most considerations are given to large, expensive, complex, sophisticated, and more glamorous systems, it is refreshing to note the authors' presenting a straightforward solution to a driver communication problem. A system that included a similar concept was proposed for the City of Baltimore after a cost-effectiveness analysis ruled out more elaborate alternatives (11).

The authors presented the results of studies of the practicability of detecting freeway stoppage waves and predicting their arrival at an upstream location. They used both speed and kinetic energy measures and found that they were about equal as far as their detection of shock waves is concerned. The kinetic energy measure was found to produce some false alarms under low-volume conditions, whereas the speed measures did not create this problem. This would lead the discussant to conclude that, for an operational system, speed measures would be preferable to energy measures.

Finally, the authors echo the need for more reliable detectors. Anyone who works in the field of traffic control issues the same plea from time to time. Perhaps someday manufacturers will find it important enough to respond to.

In summary, the authors have described a very good and useful safety warning system for freeways and have adequately developed and tested the concepts. We look forward to a wider application of this type of system.

REFERENCE

11. Courage, K. G., Bissell, H. H., and Wattleworth, J. A. Jones Falls Expressway Surveillance and Control System: Preliminary Engineering Report. Kelly Scientific Corporation report to the City of Baltimore, 1970.

DISCUSSION

Joseph W. Hess, Traffic Systems Division, Federal Highway Administration

The authors should be commended for research and implementation of a system designed to indicate stoppage waves by algorithms by using speed and kinetic energy parameters. Their effort was aimed at some specific problem locations characteristic of the roller-coaster type of design of an outmoded freeway. In this research the stoppage waves were generated out of sight (over the crest) of approaching motorists, and a warning system was needed to reduce the number of rear-end collisions. My experience has been that the stoppage wave development is usually on the upgrade or foot portion of the vertical curve rather than on the downgrade portion past the crest. With this in mind, it would have been helpful for the authors to have given more details on the experimental sites, including both traffic and geometric features. For example, I received the impression that some stoppage waves were generated by excessive demand. However, there were off-ramps in the study sections; exiting traffic was given as a reason for the right lane stoppage waves moving slowly upstream.

The generalized suggested locations of detectors (Fig. 10) are a useful aid. Here again, I would have preferred more detail on the detector placement for the individual study sections and the associated geometrics of the vertical curves.

The authors did not stray from their subject area in presenting this research. Beyond the suitability of the parameters, I am curious about the actual effect on the motorist and the effects on accident statistics. Even a preliminary subjective evaluation would be helpful although I suppose an evaluation of system effectiveness is taking place; in the case of accident statistics, this requires patience over a longer period of time. Past research by others on similar traffic warning devices has been curiously devoid of the human factors element. I do not think it is enough to measure and analyze traffic variables without determination of the devices' alerting effects, which might not show so significant in speed measurements yet might prove important in accident statistics or more subtle aspects of traffic stream flow stability. There have been past weaknesses in not identifying motorists' reactions to the presented information and use thereof. I hope that future research will be more cognizant and attempt to rectify these shortcomings.

At this point, I would like to digress to a topic closely related to the paper presented at this session, that of sight distance on freeways. Our present sight distance criteria might be perfectly adequate for rural freeways but, for those freeways carrying traffic at levels of service C, D, and E, it should be apparent that present sight distance criteria are hardly applicable. Sometimes we see only the back of the car ahead of us. Other times, when at the crest of a vertical curve and looking out over a curving downward section, we can see traffic stream characteristics for a great distance ahead. Perhaps there should be research undertaken to more adequately describe in quantitative terms what the real sight distances are for traffic streams of high density. Conceptualization of a research approach leading to development of applicable sight distance criteria seems to me a very difficult problem in itself.

In summary, I think the authors have done a fine job, and my comments are based more on what they did not tell me rather than what they did tell me. And I am glad I had the opportunity to make a few remarks on sight distance criteria.

AUTHORS' CLOSURE

The authors are appreciative of the excellent reviews by Wattleworth and Hess. Their comments are very appropriate and will be of assistance in developing techniques for improving the safety and efficiency of existing freeway systems.

The prototype safety warning system on the Gulf Freeway has been under digital computer operation since February 1972. We are very confident in the hardware and software operation and feel that the system can be implemented at other locations. As of this writing, however, we have not fully evaluated motorist response to the system. Evaluation studies are in progress, and the results will be available when the studies are completed.

The control algorithm presented in the paper is responsive to stoppage waves generated from both freeway incidents and excessive demand conditions under moderate to heavy flow conditions. Modifications to both the algorithm and detector configuration may be necessary to make the system responsive during extremely low-volume conditions.

Although the speed parameter was shown to be as responsive as the energy parameter, there have been unexpected spin-offs resulting from the latter. First of all, the energy parameter is responsive to slow-moving vehicles such as trucks or funeral processions during the off-peak periods, whereas the speed parameter is not. Secondly, whenever we receive a false alarm during the off-peak periods, we are almost assured that a detector is transmitting erroneous data. We have computer software that allows us to locate detectors that fail completely. However, when a detector fluctuates into a "gray" area without complete failure, this is more difficult to isolate. The energy parameter, however, assists us in spotting those detectors that are operating in the gray area.

It must be emphasized that the computer algorithm for the warning system is not an incident detection scheme. The algorithm is responsive to stoppage waves only and does not necessarily indicate the presence of a freeway incident. We are in the process of developing incident detection logic that will be an add-on to the stoppage wave program. Utilization of this logic in conjunction with travel time prediction techniques (12) will permit real-time evaluation of freeway conditions so that appropriate changes in ramp control strategies can be made and appropriate information can be relayed to the driver for effective diversion within a corridor.

REFERENCE

12. Messer, C. J., Dudek, C. L., and Friebele, J. D. A Method for Predicting Travel Time and Other Operational Measures in Real-Time During Freeway Incident Conditions. Highway Research Record 461, 1973.

COMPUTER MODEL FOR OPTIMAL FREEWAY ON-RAMP CONTROL

Jin J. Wang and Adolf D. May, University of California, Berkeley

Regulating input volume to a freeway system through ramp metering, or ramp closure, maintains traffic flow at an efficient level and improves overall system performance. This paper describes the development of a computer program, LINCON, that can determine the desired fixed-time metering rates for a group of on-ramps to be controlled. The linear programming technique is used to formulate a decision model that is then integrated with a previously developed deterministic freeway simulation model, FREEQ, to become a ramp-control model, RAMPCON. To take into consideration the effect of traffic diversion under control, the decision model was formulated in such a way that, at each on-ramp, the trips with shorter freeway travel distances could divert proportionally more than the trips with longer freeway travel distances. Two objective functions, maximizing total vehicular input and maximizing total freeway vehicle-miles of travel, are considered. The program user has the option of choosing either objective.

•FREEWAY ON-RAMP CONTROL as a potential tool to improve freeway operations did not receive much attention until the early 1960s when increasing congestion on urban freeways became a serious problem. Basically, the idea is that, by regulating the input of the freeway system through ramp metering (or ramp closure in the extreme case), traffic flow on the system can be kept at a more efficient level and thus overall system performance will be improved. Reduction of accidents also produces smoother flow conditions. Many theoretical investigations and empirical analyses have been reported. Recognizing the need to develop a control strategy for a group of ramps as an interrelated system, more research has been directed to the application of mathematical programming techniques to the problem of ramp control. Studies by Wattleworth (1, 2), Payne (3), and Kreer (4) are examples. This paper describes the development of a computer model that can determine the desired metering rates for a group of on-ramps to be controlled. The set of desired metering rates will be called control strategy throughout this paper.

SYSTEMS ANALYSIS

The system of freeway on-ramp control consists of three basic elements: the roadway, the traffic, and the control. The performance of the elements is dependent on each other. Changes in control would cause traffic patterns to change, which in turn would affect weaving section capacities and travel conditions on the roadway. However, roadway characteristics and traffic demands are basic parameters in determining the control strategy. Recognition of this interdependency is most important in studying freeway on-ramp control.

Publication of this paper sponsored by Committee on Freeway Operations.

Objectives and Measures of Effectiveness

Efficiency and safety are the two most important objectives of freeway operations. Efficiency in terms of total vehicles served can be measured by the total input volume. Efficiency in terms of total travel on the freeway can be measured by total vehicle-miles of travel, which tends to favor ramps having drivers with longer average trip lengths. Better utilization of all sections of freeway can also be expected. If travel distance via arterial route is approximately proportional to travel distance via the freeway route, then this criterion will result in smaller diverted vehicle-miles to the arterial street and thus minimize the adverse effect on the arterial system.

Use of total input as the measure of effectiveness tends to favor ramps with shorter average trip lengths. With a smaller number of trips diverted, the impact on the existing trip pattern is reduced. The choice between these two measures of effectiveness is not obvious. In the case where diverted traffic is dispersed evenly throughout the entire arterial network and does not significantly affect the travel on the arterial streets, the total vehicle-mile criterion tends to be better from the point of view of overall corridor traffic. However, if diverted traffic causes significant delay to arterial traffic, the total input criterion might be better. In developing fixed-time control strategies, we cannot use total travel time as the objective inasmuch as the total number of vehicles entering the freeway is not a constant. If the study includes the arterial street traffic where the total number of vehicles can be considered as a constant, then total travel time can be used as the objective.

A safety objective is difficult to measure and almost impossible to predict. Elimination of freeway congestion improves both efficiency and safety. However, further reduction of traffic might improve safety but reduce efficiency. Thus, in on-ramp control, the safety objective is not taken into account explicitly. The constraint that no congestion is allowed on the freeway may be considered as a limit establishing an acceptable level of safety.

Constraints

Because no congestion is allowed on the freeway, one set of constraints is that input demand for each subsection should be less than or equal to the capacity. A modified constraint would be that the demand should be less than the capacity by at least a certain amount. In operation, two more constraints are often necessary. One is the minimum metering rate, which is necessary to prevent excess violation at the metering signal. Sometimes a higher minimum metering rate is used so that accessibility to an area is not unjustifiably reduced. An area without suitable alternative routes is a typical example. The other constraint is the maximum metering rate. This may be limited by the minimum cycle length of the signal or, in certain conditions, to prevent unacceptably long queues on the ramp.

Constraints can also be set up so that certain levels of service can be maintained. For example, it may be desirable to keep the speed on the freeway high. Given that there is a direct relationship between speed and volume, this can be accomplished by requiring the demand to be less than the capacity by a certain amount.

General Framework

Figure 1 shows the general framework in developing ramp-control strategy. A freeway model is needed to determine the traffic performance before and after control, and a decision model is needed to select the best control strategy. The control causes some changes in the origin-destination pattern, which affects the traffic performance and may in turn require modification of the control strategy. This study has attempted to take into account the change of O-D pattern under control by some intuitive assumptions. Better understanding of the nature of traffic diversion is needed to improve the formulation.

THE FREEWAY MODEL AND MANUAL PROCEDURE

The Freeway Model: FREEQ

In 1970, a computer simulation model was developed at the Institute of Transportation and Traffic Engineering in Berkeley. This model, FREEQ, simulates freeway traffic performance under given roadway and demand characteristics. Details of this model are described in a previous report (5). Briefly, the model takes the roadway parameters, which include section length, capacity, ramp location, lane configuration, and design speed, and the load parameters, which are in the form of 15-min origin-destination tables, and computes the traffic performance in terms of speed, volume-capacity ratio, density, travel time, vehicle-hours expended, vehicle-miles expended, actual capacity, and queue length.

Manual Procedure

With the aid of the freeway model, a manual procedure can be used to evaluate ramp control strategies. Essentially, the decision model (Fig. 1) can be performed by the analyst manually. Output of the freeway model is studied, and, by following some rules or judgment, the suitable metering rates are determined. These metering rates are then used to modify the load parameters, which, when entered into the freeway model, determine the traffic performance under the proposed control scheme. The scheme may be modified when necessary. An example of this program can be found in an earlier report (6).

APPLICATION OF THE LINEAR PROGRAMMING TECHNIQUE

Application of the linear programming technique to the problem of freeway on-ramp control was first demonstrated by Wattleworth in 1965 (1, 2). Later work was done by Goolsby and McCasland (7), Messer (8), and Brewer et al. (9) in 1969. The greatest advantage of this technique is its extremely short computation time. This is very important when applied in a real-time traffic-responsive control system. The other advantages are its simplicity in formulation and its systematic methodology for finding the optimum solution for a given objective function and a set of constraints.

Basic Assumptions

1. The demand rates are constant for the time slice under consideration,
2. A steady-state condition is assumed,
3. Traffic diverted from one on-ramp will not enter other on-ramps, and
4. Traffic will not divert from one time slice to another time slice.

In the case of fixed-time control, the time slice is usually no less than 15 min. Except for very long sections of freeway, the second assumption is usually not critical. The first assumption affects primarily the main-line input inasmuch as other ramps are regulated by ramp signals and will produce constant demand rates unless the cumulative demand is less than the cumulative output for some period of time. In the case of traffic-responsive control, the time slice is very short, and the first assumption is not critical as far as the freeway is concerned. The effect of the second assumption can be reduced by breaking up the freeway into segments of 2 to 3 miles bounded by key bottlenecks. Each segment will operate almost independently except that the main-line output of the upstream segment is the main-line input of the downstream segment.

The third assumption is quite unrealistic in many cases, particularly when backtracking to enter upstream on-ramps is favorable. There are two kinds of backtracking. The first kind occurs when the original demand pattern is distorted because some traffic uses downstream on-ramps to bypass a bottleneck. When the control is implemented and the traffic in the bottleneck is flowing smoothly, drivers will use upstream on-ramps as they would have in the case of no bottlenecks. This condition can be corrected by adjusting the original demands to the situation with no congestion on the freeway. These adjusted demands are then used in the formulation. The second kind occurs when upstream on-ramps are not controlled and some traffic finds it is faster to backtrack a little even though the travel distance is longer. This effect can be reduced by extending the control farther upstream. Diversion from upstream on-ramps to a controlled downstream on-ramp is possible but unlikely unless queue delay at the downstream on-ramp is much shorter. The forth assumption is not critical for the beginning and ending time slices of the control. Some trips will start just a little bit earlier or later in order to avoid the control. However, control usually begins before congestion actually occurs and ends after the peak period. Thus, diversion in time at both ends of the control period is not critical from the operations point of view. For other time slices, the diversion is caused mainly by queuing on the ramps, which delays the entry of some vehicles at the end of the time slice. The effect depends on the change of destination pattern from one time slice to another. If the destination pattern is relatively stable, the effect will be small.

Previous Formulation

The basic linear programming model as formulated by Wattleworth is

$$\text{Maximize} \sum_{i=1}^{n} X_i \tag{1}$$

subject to

$$\sum_{i=1}^{n} A_{ik} X_i \leq B_k \tag{2}$$

for k = 1, 2, ..., m,

$$X_i \leq D_i \tag{3}$$

for i = 1, ..., n, and

$$X_i \geq 0 \tag{4}$$

for i = 1, ..., n, where

X_i = desired input rate from on-ramp i (the metering rate),
n = number of on-ramps,
m = number of subsections,
A_{ik} = fraction of traffic from on-ramp i going through subsection k,
B_k = capacity of subsection k, and
D_i = demand rate on on-ramp i.

Equation 1 simply states that the objective of the control is to maximize total input rate from all ramps. Equation 2 is the capacity constraint (total demand for any subsection should not exceed its capacity). Equation 3 states that the input rate for any on-ramp cannot be more than the demand, and Eq. 4 states that the input rates cannot be negative. This formulation can be solved efficiently by the standard simplex method. Four possible modifications were suggested by Wattleworth:

1. B_k may represent the service volume of a desired level of service to be maintained for the freeway traffic.
2. A constraint may be added that will limit the number of vehicles diverted. Expressed mathematically,

$$D_i - X_i \leq Q_i \tag{5}$$

where Q_i is the maximum number of vehicles diverted for ramp i.

3. A different type of constraint on the number of vehicles diverted may be in the form of

$$D_i - X_i = D_{i+1} - X_{i+1} \tag{6}$$

for all i. This constraint will spread all access demand equally over all on-ramps.

4. A merging capacity constraint in the form of

$$P_a \sum_{j=a+1}^{n} A_{jk}X_j + X_a \leq L_a \tag{7}$$

for a = 1, . . . , n - 1 may be added. P_a is the fraction of trips upstream of ramp a, which is in lane 1. L_a is the merging capacity.

Other modifications that have been suggested are listed as follows:

5. In the discussion of the paper presented by Wattleworth, Foote (2) suggested use of the objective in minimizing total excess capacity. Mathematically,

$$\text{Minimize} \sum_{k=1}^{m} \left(B_k - \sum_{i=1}^{n} A_{ik}X_i\right) \tag{8}$$

6. In discussing the same paper, May suggested the use of the objective of maximizing total vehicle-miles of travel.

$$\text{Maximize} \sum_{i=1}^{n} X_i l_i \tag{9}$$

where l_i is the average trip length of all traffic from on-ramp i.

Two additional modifications may be considered.

7. Equation 5 may be expressed as

$$X_i \geq D_i - Q_i = N_i \tag{10}$$

where N_i is the minimum metering rate for ramp i.

8. A maximum metering rate constraint in the form of

$$X_i \leq M_i \tag{11}$$

can be added where M_i is the maximum metering rate.

Alternative Formulation

Wattleworth's formulation implicitly assumes that, for each on-ramp, the destination pattern before and after control is the same. This is reflected by the parameter A_{ik}, which is computed from the destination pattern before the control. In operation, a queue will form on the ramp and cause a certain amount of delay. Some traffic will find it better to use alternative routes. Logically, traffic with better alternative routes or, in terms of travel time, a shorter travel time for the alternative route than for the freeway route will more likely divert first. The exact pattern of diversion is undoubtedly stochastic in nature and depends on the actual origin and destination of each trip and on driver characteristics. As an approximation, it is assumed that trips with shorter freeway trip lengths will divert proportionally no less than trips with longer freeway trip length. This can be expressed as

$$P(i, j) \leq P(i, j + 1) \tag{12}$$

where P(i, j) is the percentage of the original demand between on-ramp i and off-ramp j that still uses the freeway after control. Equation 12 does not guarantee that all short trips will divert first; it only states that short trips could potentially divert proportionally more than long trips. This may appear to be an inconsistent assumption inasmuch as no definite diversion rule is established. However, because of the stochastic nature of the true diversion pattern, we feel that this assumption will produce results closer to reality than the assumption that all short trips will divert first.

The following notations are used in the alternative formulation.

TRIP(i, j) = original demand rate from on-ramp i to off-ramp j,
P(i, j) = percentage of original demand from on-ramp i to off-ramp j that is not diverted,
B_k = capacity (or service volume) of subsection k, and
L(i, j) = freeway travel distance between on-ramp i and off-ramp j.

The linear programming formulation is as follows:

$$\text{Maximize} \sum_{i=1}^{n} \sum_{j=1}^{m} P(i, j)\ TRIP(i, j) \tag{13}$$

where n is the number of on-ramps and m is the number of off-ramps, subject to

$$P(i, j) \leq 1 \tag{14}$$

for all (i, j),

$$P(i, j) - P(i, j+1) \leq 0 \tag{15}$$

for all (i, j),

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \delta \times P(i, j) \times TRIP(i, j) \leq B_k \tag{16}$$

and

$$P(i, j) \geq 0 \tag{17}$$

In Eq. 16, $\delta = 1$ if i is upstream of subsection k and j is downstream of subsection k; otherwise, $\delta = 0$. In this formulation, P(i, j) is the decision variable and TRIP(i, j), B_k, and δ are constant. Equation 13 is replaced by

$$\text{Maximize} \sum_{i=1}^{n} \sum_{j=1}^{m} P(i, j) TRIP(i, j)\ L(i, j) \tag{18}$$

when the selected objective is to maximize vehicle-miles of freeway travel.

In the remainder of this report, the formulation by Wattleworth will be referred to as the formulation with proportional diversion, and the alternative formulation de-

Figure 1. General framework in developing ramp-control strategies.

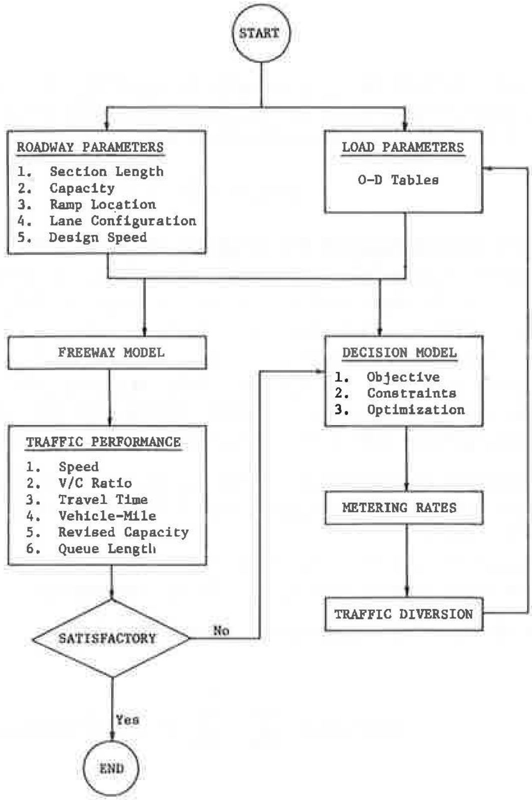
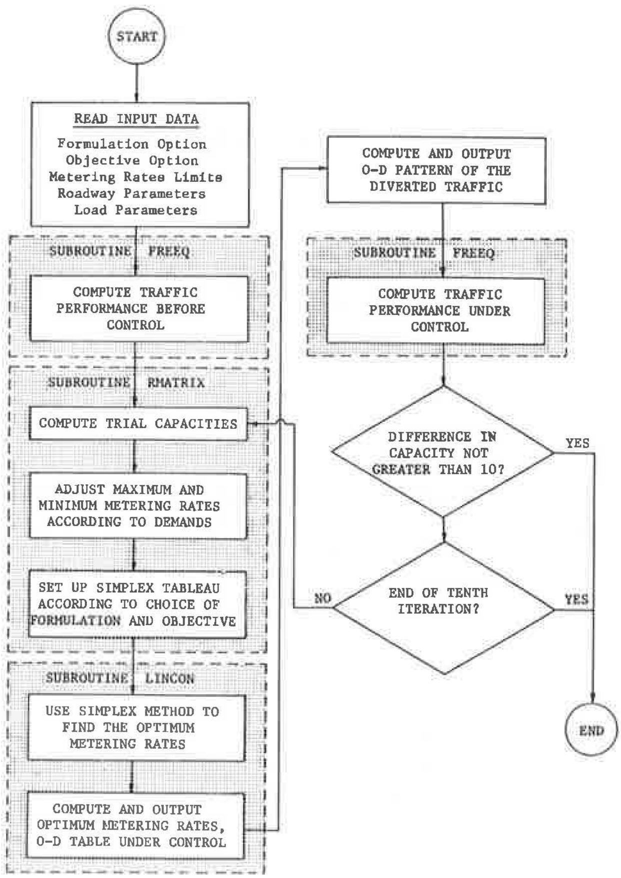


Figure 2. Program RAMPCON.



veloped in this section will be referred to as the formulation with short-trip diversion.

The Computer Program: LINCON

Based on the formulations previously presented, a computer program (LINCON) was prepared. The program includes the four basic modifications by Wattleworth, modification items 6, 7, and 8, and the alternative formulation. Thus, the user has the flexibility of choosing either type of formulation depending on the applicability of the diversion assumption. The user can also choose either total vehicle input or total vehicle-miles as the objective function.

The number of equations in the linear programming formulation is reduced by not entering the maximum and minimum metering rates as constraints. In the program, the maximum metering rates are first adjusted so that they are not greater than the demands. Then, if the demand of a ramp is greater than the maximum metering rate, D_i in Eq. 3 is replaced by the maximum metering rate, or TRIP(i, j) in Eqs. 13 and 16 is modified such that the number of total trips from an on-ramp is equal to the maximum metering rate. Trips with shorter freeway trip lengths are diverted first. The minimum metering rates are also modified so that they are not greater than the demands. Then, these trips are loaded to the freeway as if they are through traffic. If the diversion assumption of the alternative formulation is adopted, long trips are selected first for freeway use. Parameter values for the simplex tableau are automatically computed in the subroutine RMATRX by using the input data.

INTEGRATION OF FREEWAY MODEL AND LINEAR PROGRAMMING

The linear programming model, LINCON, is now used as the decision model and, in combination with the FREEQ model, results in a completely automatic computer model, RAMPCON. The need to integrate these two models has been shown previously. Figure 1 can be considered as the generalized flow chart of this model. A more detailed flow chart is shown in Figure 2.

Iteration Procedure

From the linear programming formulations, it can be seen that it is the change in B_k value that will cause the modification of the metering rates. The modified metering rates may in turn change the value of B_k caused by change in weaving effect. Thus, an iteration procedure is required to obtain the equilibrium solution. This program handles only the case that B_k is the capacity. Minor modification is required if B_k is to be the service volume of a specified level of service.

The value of B_k is affected by the load parameters because of the weaving effect. The step-by-step procedure is as follows:

1. Input roadway parameters and O-D table into the FREEQ model to obtain the traffic performance under no control conditions.
2. Let C_k be the capacity of section K if there is no weaving effect and WE_k be the weaving effect. Compute a trial value of B_k by using Eq. 19.

$$B_k = C_k - \frac{WE_k}{2} \tag{19}$$

3. With the computed B_k value and other necessary data, use LINCON program to compute the optimum metering rates.
4. Revise the O-D tables based on the metering rates, and rerun FREEQ.
5. Compare the previously computed B_k value with the true capacity, which is equal to C_k - WE_k. If the difference is less than 10 for all subsections, terminate the procedure; otherwise, continue to next step.
6. Revise the B_k value by using Eq. 20.

$$B_k = (B_k + C_k - WE_k)/2 \tag{20}$$

Table 1. Summary of test roadway data.

Subsection No.	No. of Lanes	Capacity	Length (ft)	Truck Factor	Ramp	Subsection Description
1	3	5,728	1,660	0.970	O	Central off to Central on
2	3	5,806	1,890	0.970	OD	Central on to Carlson off
3	3	5,520	2,310	0.940		Carlson off to Carlson on
4	3	5,950	1,460	0.980	OD	Carlson on to Potrero off
5	3	5,806	3,800	0.970		Potrero off to Cutting on
6	3	5,880	1,100	0.980	O	Cutting on to grade change point
7	3	5,950	660	0.980	D	Grade change point to Macdonald off
8	3	5,950	1,480	0.980	D	Macdonald off to San Pablo off
9	3	5,728	1,480	0.970		San Pablo off to San Pablo on
10	4	6,850	800	0.980	OD	San Pablo on to Solano off
11	3	5,800	4,690	0.970	D	Solano off to San Pablo Dam off
12	3	5,806	2,190	0.970		Dam Road off to Dam Road on
13	3	5,800	2,320	0.970	OD	Dam Road on to Road 20 off
14	3	5,049	830	0.850		Road 20 off to grade change point
15	3	4,746	1,180	0.793		Grade change point to Road 20 on
16	3	4,700	2,560	0.780	OD	Road 20 on to Mainline Destination

Note: Ramp limit = 1,500 vph.

Table 2. Origin-destination data for test system (15-min volume).

Origin Down	Destination Cross 1	2	3	4	5	6	7	8
1	61	106	56	102	34	121	66	798
2		3	6	9	4	19	7	39
3		7	3		5	14	9	44
4					34	108	40	153
5					14	48	29	152
6							11	55
7								0

Table 3. Summary of test system results.

Formulation		Proportional Diversion		Short-Trip Diversion	
Item	No Control	Maximize Vehicle-Miles	Maximize Input	Maximize Vehicle-Miles	Maximize Input
Total input rate	8,320	7,753	7,753	7,655	7,691
Total output rate	7,087	7,753	7,753	7,655	7,691
Total vehicle-mile rate	6,699	7,703	7,703	7,751	7,735
Computer time (sec)		17	17	34	40
No. of iterations		7	7	6	7
Key bottleneck subsections	6	4, 6, 11	4, 6, 11	4, 6, 16	4, 11, 16

Table 4. Optimum metering rates on test system.

On-Ramp No.	Original Demand Rate (vph)	Optimum Metering Rate (vph)			
1	348	348	348	348	348
2	328	328	328	328	328
3	1,340	512	512	512	403
4	972	925	925	827	972
5	264	264	264	264	264
6	0	0	0	0	0

The B_k value on the right side of the equation is the value used in the previous run of LINCON.

7. Return to step 3.

Test System and Results

The northbound Eastshore Freeway (located in the San Francisco Bay area) from the Central Street off-ramp to the main-line section north of the Road 20 on-ramp was selected as the test system. A previous study (6) has proposed the implementation of ramp control on this section. The roadway data are given in Table 1, and the traffic data in the format of a 15-min origin-destination table are given in Table 2. This O-D table is the projected 1972 demand (assuming BART in operation) and for the time period from 4:30 to 4:45 p.m. A minimum metering rate of 240 vph is selected for all ramps. A maximum metering rate of 1,080 vph is selected for the San Pablo on-ramp and 800 vph for all other ramps.

Tables 3 and 4 give the results of the test system. Results of both formulations and both objective functions show improvements in total output and total vehicle-miles of freeway travel for the controlled system as compared with the uncontrolled system. Although total input for the uncontrolled system is higher than that for the controlled system, there is severe congestion on the freeway. Under the formulation with proprotional diversion, the results for both objective functions are identical. This may not be the case for other systems. The differences among the four conditions do not appear to be significant. This is probably because subsection 6 is the major bottleneck that effectively reduces the demands of the downstream subsections to less than or only slightly above the capacity.

In summary, the computer model developed by combining a freeway model and the linear programming technique can determine the desired metering rates, taking into consideration the changes of freeway subsection capacities when the demand pattern is altered by the metering rates. The freeway submodel also provides data on freeway performance before and after control for better analysis of the control effect. The model was developed basically for the analysis of fixed-time ramp control. Application to the traffic-responsive control is possible when supplemented by a real-time demand forecast and distribution model.

REFERENCES

1. Wattleworth, J. A., and Berry, D. S. Peak-Period Control of a Freeway System—Some Theoretical Investigations. Highway Research Record 89, 1965, pp. 1-25.
2. Wattleworth, J. A. Peak-Period Analysis and Control of a Freeway System. Highway Research Record 157, 1967, pp. 1-21.
3. Payne, H. J., and Isaksen, L. A Systematic Approach to Freeway Traffic Control and Surveillance. To be published.
4. Kreer, J. B., Koenig, H. E., Goodnuff, J. L., Mika, H. S., Panyan, W. D., and Yuan, L. S. Coordinated Control of Urban Traffic Networks. Div. of Engineering Research, Michigan State Univ., technical rept., Dec. 1968.
5. Makigami, Y., and Woodie, W. L. Freeway Travel Time Evaluation Technique. Highway Research Record 321, 1970, pp. 33-45.
6. Aidoo, J. F., Goedhart, R. W., and May, A. D. Bay Area Freeway Operations Study—Final Report: Analytic Techniques for Evaluating Freeway Improvements on the Eastshore Freeway (I-80) Northbound. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, June 1971.
7. Goolsby, M. E., and McCasland, W. R. Freeway Operations on the Gulf Freeway Ramp Control System. Texas Transportation Institute, Texas A&M Univ., Research Rept. 24-25, 1969.
8. Messer, C. J. A Design and Synthesis of a Multilevel Freeway Control System and a Study of Its Associated Operational Control Plan. Texas A&M Univ., dissertation, Aug. 1969.
9. Brewer, K. A., Buhr, J. H., Drew, D. R., and Messer, C. J. Ramp Capacity and Service Volume as Related to Freeway Control. Highway Research Record 279, 1969, pp. 70-86.

PREDICTION AND STABILITY OF FREEWAY GAPS AND ON-RAMP MERGING

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

Merging of vehicles from on-ramp to freeway is the most likely cause of freeway congestion and disturbance besides geometric bottlenecks and freeway accidents. One of the effective ways to control this type of congestion and disturbance is predicting intervehicular acceptable gap sizes by detectors placed in the shoulder lane at some point upstream of the on-ramp. Vehicles from the controlled-ramp site are released and coordinated to merge into the traffic whenever an acceptable gap is detected. This paper gives the analysis on the prediction of gaps and their associated errors for single- and double-loop detectors. Analytical expressions for errors in gap prediction and speed estimation were derived and validated against experimental observation obtained from aerial photographic data and were related to different sensing pulse rates and detector locations for single- and double-loop detectors. The problem of gap prediction was also related to the stability of traffic in the vicinity of an on-ramp. The effects of lane changing and variation in the speed of the individual car between the detector location and the on-ramp merge point were also analyzed from experimental data. A comparison of fixed-metering strategy and gap-prediction strategy is made, and the fixed-metering rate is determined according to the measured occupancy from the experimental data. Results indicate that the gap-prediction strategy has great application value.

•FREEWAY on-ramp control has drawn much attention in recent years. The main purpose of controlling on-ramp flow is to improve freeway traffic operation so that the total demand (on-ramp demand plus through-traffic demand) does not exceed the capacity of the downstream side of the on-ramp.

Two major factors that affect the merging of on-ramp flow to the main-line traffic is gap availability and gap acceptance. Gap availability involves the study of gap distribution. Distribution of time gaps (time headways) has been studied by many researchers (1, 8, 10, 11, 14, 17).

Gap-acceptance studies have also been conducted by a number of researchers (4, 5, 6, 7, 13, 16). Their main efforts have been to study statistical properties of acceptable gaps. The study by Hsu and Munjal (9) differs from the other works in that gap-acceptance study was conducted for all lanes in a freeway, and the relation of a lane changer and its neighboring vehicles was analyzed.

There are several types of on-ramp control strategies. The first is to relate the gap distributions to the occupancy (or speed and concentration) and determine the adequate number of on-ramp vehicles being allowed. The on-ramp flow is thus controlled by a fixed-metering rate according to the occupancy. This technique has been applied in Chicago and Chula Vista, California. The second technique, somewhat similar to the first, is to keep the sum of the on-ramp flow and the flow of the outer lane (the lane into which the on-ramp flow is merging) less than or equal to the outer lane's capacity. This technique was applied in Houston (18). A third approach (3, 12), applied in Houston (2), involves the detection of gaps in the freeway outer lane upstream of the ramp and the release of vehicles from the ramp when an acceptable gap has been detected.

Publication of this paper sponsored by Committee on Freeway Operations.

The first two approaches are macroscopic approaches. On-ramp flow is controlled based on the average traffic condition in the vicinity of the on-ramp. The third approach is a microscopic one in that the individual gap is taken into account. This paper mainly considers problems that occur in the microscopic approach, namely, gap prediction and stability.

A simple way to predict gap length at the nose of an on-ramp (point x) is to install a detector at some fixed distance from x and measure its gap length. A gap is assumed to travel at a constant speed, the space mean speed $\bar{v}_s$. We can thus calculate the time when this gap is available at x. This approach is definitely unrealistic because not every vehicle travels at $\bar{v}_s$. An improved method would be one that considers an individual vehicle's speed. Even with this, the nonconstant speed of a vehicle still degrades the prediction of the arrival time and length of the gap. Furthermore, there are detector measurement errors as well as lane changes occurring between the detector and the on-ramp nose.

We shall describe our prediction method in the next section and analyze the prediction errors (instability).

The prediction errors are functions of the distance between the loop detector and the on-ramp, the sampling rate, vehicle speed, and other factors to be discussed in detail. We shall use aerial experimental data to examine the performances of single- and double-loop detectors and compare them with the analytical results.

Fixed-metering strategy, which is a conventional way to control on-ramp flow, will be compared with the prediction strategy, thus giving us some indication of the latter's application value.

GAP PREDICTION AND ERROR ANALYSIS

If we assume that a freeway has the on-ramp configuration shown in Figure 1, there may be a single-loop detector, D_1, or a double-loop detector, D_1 and D_2, installed distance L from the on-ramp. Detectors in pairs improve the accuracy of measurements but have a higher cost. We shall first analyze the performance of the single-loop detector.

Let us assume that a vehicle has length x_i, which passes the detector D_1, with a time duration of t_i. Its speed, v_i, is

$$v_i = \frac{w + x_i}{t_i} \tag{1}$$

where w is the length of the loop. The occupancy O_c for a time period T is

$$O_c = \frac{1}{T}\sum_{i=1}^{n} \frac{x_i}{v_i} = \frac{1}{T}\sum_{i=1}^{n} \frac{t_i}{1 + \frac{w}{x_i}} \tag{2}$$

where n is the number of vehicles passing D_1 in T.

Because x_i varies from vehicle to vehicle and the single-loop detector cannot measure the vehicle length, we must use $\bar{x}$, the average vehicle length for x_i, in Eqs. 1 and 2. Thus, the estimates of v_i and O_c are

$$\hat{v}_i = \frac{w + \bar{x}}{\hat{t}_i} \tag{3}$$

and

$$\hat{O}_c = \frac{1}{T}\sum_{i=1}^{n} \frac{\hat{t}_i}{1 + \frac{w}{\bar{x}}} \tag{4}$$

where $\hat{t}_i$ is the detector measurement subject to sampling error. It is observed from Eq. 3 that, for vehicles having lengths larger than $\bar{x}$, the estimated speeds are likely smaller than their true speeds, and, for vehicles having lengths smaller than $\bar{x}$, the estimated speeds are likely larger than their true speeds.

For the pair of successive cars i and i + 1 with measured speeds $\hat{v}_i$ and $\hat{v}_{i+1}$, we assume the measured time headway (gap) to be $\hat{\tau}_i$ and true headway to be τ_i. When vehicle i reaches x = 0, the nose of the on-ramp, the gap length is predicted to be

$$\hat{\hat{\tau}}_i = \max\left(\tau,\ \hat{\tau}_i + \frac{L}{\hat{v}_i} - \frac{L}{\hat{v}_{i+1}}\right) \tag{5}$$

where τ is the predetermined minimum headway between successive cars.

When two loops are installed, the length of a vehicle can be measured. Let the detector ON (when it indicates a car is in presence) and OFF (when it indicates the car leaves) times be $\hat{t}_i^{(1)}$, $\hat{t}_i^{(2)}$ for D_1, and $\hat{t}_i^{(3)}$, $\hat{t}_i^{(4)}$ for D_2 when vehicle i passes the two detectors. The speed v_i is estimated by

$$\hat{v}_i' = \frac{w + d}{\hat{t}_i^{(3)} - \hat{t}_i^{(1)}} \tag{6}$$

or alternatively by $\hat{v}_i'' = \frac{2w + d + x_i}{\hat{t}_i^{(4)} - \hat{t}_i^{(1)}}$ where d is the distance between the two detectors (Fig. 1).

The length of the vehicle x_i can be determined by solving

$$\frac{w + d}{\hat{t}_i^{(3)} - \hat{t}_i^{(1)}} = \frac{2w + d + x_i}{\hat{t}_i^{(4)} - \hat{t}_i^{(1)}} \tag{7}$$

That is,

$$\hat{x}_i = \frac{(w + d)\left(\hat{t}_i^{(4)} - \hat{t}_i^{(1)}\right)}{\hat{t}_i^{(3)} - \hat{t}_i^{(1)}} - 2w - d \tag{8}$$

The estimated occupancy using the two loops is

$$O_c = \frac{1}{T}\sum_{i=1}^{n} \frac{\hat{t}_i^{(4)} - \hat{t}_i^{(1)}}{1 + \frac{2w + d}{\hat{x}_i}} \tag{9}$$

The estimated gap length at x = 0 for a double-loop detector has the same form as the single-loop detector in Eq. 5 by replacing $\hat{v}_i$ by $\hat{v}_i'$ and $\hat{v}_{i+1}$ by $\hat{v}_{i+1}'$.

Because the vehicle length can be precisely determined (subject only to sampling error—we have neglected all hardware errors) in the double-loop case, the accuracy in estimating speed and occupancy is higher than in the single-loop case. However, in both cases, the measurements are subject to an additional error source, the sampling error. Figure 2 shows the sampling error in estimation. The turn-on and turn-off times of the detector are y_i sec later and z_i sec later respectively than the true vehicle arrival and leaving times. The errors y_i and z_i can be conveniently assumed independent and uniformly distributed between 0 and r, the sampling interval, and thus have mean $\frac{r}{2}$ and variance $\frac{r^2}{12}$.

Single-Loop Detector

It can be shown (15) that the errors in estimation of speeds and the downstream gap lengths using the single-loop detector are

$$\left.\begin{aligned} E(\hat{v}_i) &= v_i \\ \mathrm{var}(\hat{v}_i) &= \frac{\sigma_1^2}{t_i^2} + v_i^2\frac{\sigma_2^2}{t_i^2} + \frac{\sigma_1^2\sigma_2^2}{t_i^4} \\ E(\hat{\hat{\tau}}_i) &= \tau_i + L\left(\frac{1}{v_i} - \frac{1}{v_{i+1}}\right) \\ \mathrm{var}(\hat{\tau}_i) &= \sigma_2^2 + bL + c^2L^2 \end{aligned}\right\} \tag{10}$$

where σ_1^2 is the variance of vehicle length, $\sigma_2^2 = \frac{r^2}{6}$, the variance of the sampling error, and

$$b = \frac{r^2}{12(w + x_i)} + \frac{r^2}{12(w + x_{i+1})} = \frac{r^2}{12}\left(\frac{1}{w + x_i} + \frac{1}{w + x_{i+1}}\right)$$

$$c^2 = \frac{(w + x_i)^2\sigma_2^2 + t_i^2\sigma_1^2 + \sigma_1^2\sigma_2^2}{(w + x_i)^4} + \frac{(w + x_{i+1})^2\sigma_2^2 + t_{i+1}^2\sigma_1^2 + \sigma_1^2\sigma_2^2 + \sigma_2^2}{(w + x_{i+1})^4}$$

The result in Eq. 10 is based on the assumption that vehicles travel at constant speeds within the distance L. Any variation in speed would cause additional error. Furthermore, if the follower, with a short-time headway, has a higher speed than its leader, it will likely decrease its speed or make a lane change in order not to hit its leader. This is precisely what Eq. 5 says; i.e., a minimum headway τ must be maintained if the follower does not make a lane change. Suppose now that vehicle i has an average speed of $\tilde{v}_i$ within the distance L instead of v_i such that

$$v_i = \tilde{v}_i + \delta_i \tag{11}$$

where δ_i has the mean zero and variance σ_3^2.

It can also be shown (15) that the estimator in Eq. 5 for $\tau < \hat{\tau}_i + \frac{L}{\hat{v}_i} - \frac{L}{\hat{v}_{i+1}}$ has the following properties

$$E(\hat{\hat{\tau}}_i) = \tau_i + L\left(\frac{1}{\tilde{v}_i} - \frac{1}{\tilde{v}_{i+1}}\right) \tag{12}$$

$$\mathrm{var}(\hat{\hat{\tau}}_i) = \sigma_2^2 + b_1L + c_1^2L^2 \tag{13}$$

where

$$b_1 = \frac{\sigma_2^2}{2(\tilde{v}_it_i)} + \frac{\sigma_2^2}{2(\tilde{v}_{i+1}t_{i+1})}$$

and

$$c_i^2 = \frac{1}{\tilde{v}_i^2}\left\{\frac{\sigma_2^2}{t_i^2} + \frac{1}{\tilde{v}_i^2}\left[\sigma_3^2 + \frac{\sigma_1^2}{t_i^2} + \frac{\sigma_2^2\left(\sigma_3^2 - \frac{\sigma_1^2}{t_i^2}\right)}{t_i^2\tilde{v}_i^2}\right]\right\}$$

$$+ \frac{1}{\tilde{v}_{i+1}^2}\left\{\frac{\sigma_2^2}{t_{i+1}^2} + \frac{1}{\tilde{v}_{i+1}^2}\left[\sigma_3^2 + \frac{\sigma_1^2}{t_{i+1}^2} + \frac{\sigma_2^2\left(\sigma_3^2 + \frac{\sigma_1^2}{t_{i+1}^2}\right)}{t_{i+1}^2\tilde{v}_{i+1}^2}\right]\right\}$$

Because $v_i t_i = w_i + x_i$, Eq. 13 reduces to Eq. 10 when $\tilde{v}_i = v_i$; i.e., speed remains constant. For $L > 100$ ft, the first two terms of Eqs. 10 and 13 are negligible. We thus have the approximate results

$$\left.\begin{aligned} E(\hat{\hat{\tau}}_i) &= \tau_i + L\left(\frac{1}{v_i} - \frac{1}{v_{i+1}}\right) \\ \mathrm{var}(\hat{\hat{\tau}}_i) &= c^2L^2 \end{aligned}\right\} \qquad (14)$$

for vehicles maintaining constant speed and

$$\left.\begin{aligned} E(\hat{\hat{\tau}}_i) &= \tau_i + L\left(\frac{1}{\tilde{v}_i} - \frac{1}{\tilde{v}_{i+1}}\right) \\ \mathrm{var}(\hat{\hat{\tau}}_i) &= c_i^2L^2 \end{aligned}\right\} \qquad (15)$$

for vehicles not maintaining constant speed.

Double-Loop Detector

In the case of the two-loop detector, we have the same predictor as given in Eq. 5. But in this case, the estimation of speed is more accurate because no assumption on the average vehicle length is used. The resulting speed estimation and the predictor $\hat{\hat{\tau}}'$ have the following properties:

$$\left.\begin{aligned} E(\hat{v}_i') &= v_i \\ \mathrm{var}(\hat{v}_i') &= v_i^2 \frac{\sigma_2^2}{(t_i^{(3)} - t_i^{(1)})^2} \end{aligned}\right\} \qquad (16)$$

$$\left.\begin{aligned} E(\hat{\hat{\tau}}_i') &= \tau_i + L\left(\frac{1}{v_i} - \frac{1}{v_{i+1}}\right) \\ \mathrm{var}(\hat{\hat{\tau}}_i') &= \sigma_2^2 + b'L + c'^2L^2 \end{aligned}\right\} \qquad (17)$$

where $b' = \frac{\sigma_2^2}{(w + d)}$ and $c'^2 = \frac{2\sigma_2^2}{(w + d)^2}$ for vehicles maintaining constant speeds, and

$$E(\hat{\hat{\tau}}_i') = \tau_i + L\left(\frac{1}{\tilde{v}_i'} - \frac{1}{\tilde{v}_{i+1}'}\right) \qquad (18)$$

$$\mathrm{var}(\hat{\hat{\tau}}_i') = \sigma_2^2 + b_i'L + c_i'^2L^2 \qquad (19)$$

where $b_1' = b'$ and $c_1'^2 = \frac{2\sigma_2^2}{(w + d)^2} + \frac{\sigma_3^2}{\tilde{v}_1'^4} + \frac{\sigma_3^2}{\tilde{v}_{1+1}'^4}$ for vehicles not maintaining constant speeds.

For L > 100 ft, Eqs. 17 and 19 can be approximated by

$$\mathrm{var}(\hat{\hat{\tau}}_1') = c'^2L^2 \tag{20}$$

and

$$\mathrm{var}(\hat{\hat{\tau}}_1') = c_1'^2L^2 \tag{21}$$

respectively.

It is easy to see that Eqs. 18 and 19 reduce to Eqs. 16 and 17 respectively when $\sigma_3^2 = 0$; i.e., $\tilde{v}_1' = v_1$.

The errors (standard deviation from the true value) in predicting downstream gaps are just the square roots of the variances in Eqs. 14, 15, 20, and 21 (i.e., cL, c_1L, c'L, and c_1'L respectively) for the four different cases. We note here that the prediction error is defined as the standard deviation of the prediction from the true value rather than the average difference in absolute value of the prediction. In general, the former error is larger because we weigh the differences in squares; in the latter we weigh them linearly. If the error is normally distributed (we can consider the errors in our cases as approximately normal), the standard deviation is $\sqrt{2\pi}$ times the average difference in absolute value. The prediction errors are proportional to the distance L. They are also functions of the loop length, speed, sampling rate, variation in vehicle length for one-loop detector, distance d between loops for two-loop detector, and variation in speeds. The errors increase with increasing L, variation in vehicle length, and variation in speeds but decrease with increasing loop length, distance between loops, vehicle speed, and sampling rate.

To illustrate how the errors in predictions vary, we give a simple example and tabulate the prediction errors. Given loop length w = 6 ft, vehicle-speed variation σ_3^2 = 3.8 ft²/sec² (obtained from the aerial data when L = 600 ft), distance between two loops d = 14 ft (double-loop), average vehicle length in the outer lane $\bar{x}$ = 20 ft, vehicle-length variation σ_1^2 = 100 ft², both the follower and the leader have actual speeds of 80 ft/sec, and lengths of both vehicles are 20 ft, we find the downstream gap prediction errors for L = 600 ft and 1,200 ft and for sampling rate $\frac{1}{r}$ = 10 cps, 15 cps, 30 cps, and ∞ cps (no sampling error in estimating speeds).

Table 1 gives the results. In the case of the single-loop detector, the dominant error source is the increasing of distance L. The sampling rate and speed variations contributed an insignificant number of errors. In the case of the double-loop detector, the sampling rate plays as important a role as the distance L, but, once again, the contribution of speed variation is still negligible.

It is also of interest to compare the speed measurement error for single- and double-loop detectors. We define the error to be the square root of the variances var($\hat{v}_1$), var($\hat{v}_1'$) given in Eqs. 10 and 16 respectively for single- and double-loop detectors. Using the same w, d, $\bar{x}$, and σ_1^2 as in the previous example, we also assume the average speed of a car to be 80 ft/sec. The calculated speed measurement errors for sampling rates of 10, 15, and 30 cps are given in Table 2. We can see that the measurements improve significantly for the single-loop detector. This was the primary reason that the single-loop detector gave very poor gap predictions.

The preceding analysis only tells the error in predicting gap length. An even more important error is that a gap is predicted to be acceptable although it is actually unacceptable. More specifically, this error is the conditional probability

$$e = \Pr\left(\tau_1^{(a)} < \tau_c \,\middle|\, \hat{\hat{\tau}}_1 \geq \tau_c\right) \tag{22}$$

This error is of vital importance because, once we predict a gap greater than or equal to τ_c, a command is sent to the on-ramp metering, and the vehicle is released to merge into the gap. However, the true gap $\tau_i^{(a)}$, being less than τ_c, may cause the vehicle to be unable to merge or make a forced merge and thus create unexpected disturbances. The calculation of e is very much involved and will not be carried out here.

EXPERIMENTAL RESULTS

Aerial data of the three-lane westbound Long Island Expressway in New York were used for this study. The data were taken beginning at 8:15 a.m., Tuesday, June 10, 1969, totaling 464 frames (1 frame/2 sec). A total of 1,113 vehicle trajectories were collected, which corresponds to a flow of 4,381 cars/hour. There were no on-ramps in the road stretch where the data were collected. However, this will not affect our analyses because we can always assume a pseudo on-ramp anywhere on the road with a loop detector placed distance L upstream from it. The aerial data covered about 4,000 ft of road. We assumed that the loops were placed 700 ft from the extreme west of the road stretch, and on-ramps were distance L downstream from it.

A total of 18 analyses were conducted with each experiment having different sampling rates, types of loop detectors, and distances L. They are as follows:

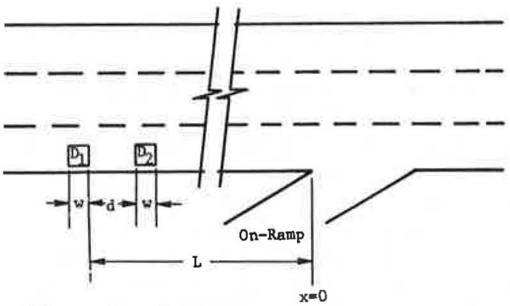
Case	Distance L (ft)	Sampling Rate (cps)
1	600	10
2	1,200	10
3	1,800	10
4	600	15
5	1,200	15
6	1,800	15
7	600	30
8	1,200	30
9	1,800	30

In each of the preceding cases, we have either a single- or a double-loop detector placed for measurement. In each case, we compute from the experimental data that (a) gaps that are predicted to be acceptable are actually acceptable, (b) gaps that are predicted to be acceptable are actually unacceptable, and (c) gaps that are predicted to be unacceptable are actually acceptable.

The acceptable gap is defined to be a gap greater than or equal to the median gap of all drivers who accept after being stopped before they enter the freeway. It was found from the works by Gourlay (7), Glickstein et al. (6), and Pearson and Ferreri (16) that the median accepted gap was from 5 to 6 sec. We shall call the median accepted gap the critical gap.

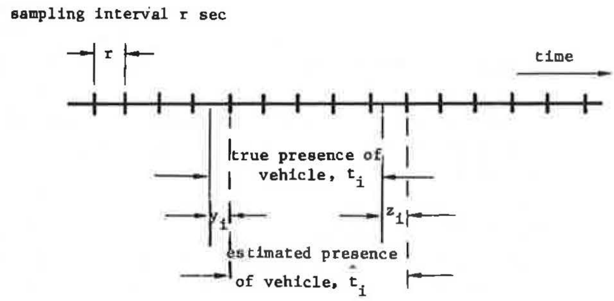
The actual speed of each vehicle when it crossed the loop and the crossing time were obtained from the smoothed aerial data. The actual presence time depended on the vehicle's speed and length. Though it is difficult to determine the exact length of each vehicle from the aerial films, we were able to identify different types of vehicles (as many as seven types) and assign a unique length to each type of vehicle. Therefore, the errors in prediction reflect not only sampling and speed variance errors but also errors from data smoothing and the simplification that vehicles can be only seven possible lengths. We should be aware that the last error source does not affect the prediction when the double-loop detector was used. Experimental results are given in Table 3. The results are much in agreement with our analysis for vehicles changing speeds (Table 1). Once again, we see that there is little difference in the prediction error for different sampling rates in the case of the single-loop detector; the dominant factor is the distance L. In the case of double-loop detectors, both distance and sampling rate affect the prediction error to a large extent. The last column in Table 1 gives the error caused by speed variation alone within the distance L. The three numbers are obtained by using speed from aerial data to predict downstream gap, which is equivalent to using sampling rate = ∞.

ıre 1. Freeway on-ramp configuration.



and D_2 are loop detectors.

Figure 2. Detector measurement error.



ɔle 1. Prediction errors (in sec) under various conditions.

ector	Speed Assumption	Distance Between Loops and On-Ramp (ft)	Sampling Rate (cps) 10	15	30	∞
;le-loop	Constant	600	4.33	4.19	4.10	4.08
		1,200	8.66	8.38	8.20	8.16
		1,800	12.99	12.57	12.30	12.24
	Changing	600	4.37	4.23	4.15	4.12
		1,200	8.74	8.46	8.30	8.24
		1,800	13.11	12.69	12.45	12.36
ble-loop	Constant	600	1.73	1.15	0.58	0
		1,200	3.46	2.30	1.16	0
		1,800	5.19	3.45	1.74	0
	Changing	600	1.75	1.17	0.62	0.22
		1,200	3.50	2.34	1.24	0.44
		1,800	5.25	3.51	1.86	0.66

ɔle 2. Calculated and measured speed
ors (in ft/sec).

ector	Error	Sampling Rate (cps) 10	15	30
gle-loop	Calculated	32.6	31.5	30.9
	Measured	32.2	26.1	24.8
ble-loop	Calculated	13.1	8.7	4.34
	Measured	12.02	7.78	3.96

Table 3. Gap prediction error (in sec) from experiments.

Detector	Distance L (ft)	Sampling Rate (cps) 10	15	30	Aerial Method[a]
Single-loop	600	3.93	3.78	3.78	
	1,200	7.74	7.48	7.50	
	1,800	11.33	10.82	10.90	
Double-loop	600	1.14	0.81	0.44	0.22
	1,200	2.15	1.57	0.87	0.40
	1,800	3.08	2.34	1.39	0.77

[a]Direct prediction from aerial data is equivalent to using double-loop detector with sampling rate of ∞.

ɔle 4. Correct and incorrect gap counts from experiments.

pling e)	Case	Single-Loop Detector X[a]	Y[b]	Z[c]	Double-Loop Detector X[a]	Y[b]	Z[c]
	1	28(37)	22(22)	10(23)	34(46)	15(17)	4(14)
	2	24(32)	23(20)	15(25)	32(46)	26(20)	7(11)
	3	19(26)	19(16)	20(30)	30(42)	23(18)	9(14)
	4	29(39)	22(21)	10(21)	34(50)	12(10)	4(10)
	5	24(33)	19(17)	15(24)	33(47)	17(15)	6(10)
	6	22(28)	18(16)	17(28)	30(44)	24(17)	9(12)
	7	30(40)	18(19)	8(20)	37(58)	7(9)	1(2)
	8	23(33)	18(17)	16(24)	37(52)	13(12)	2(5)
	9	21(29)	17(16)	18(27)	36(49)	18(17)	3(7)
aerial)	L = 600 ft				38(58)	5(4)	0(2)
	L = 1,200 ft				37(55)	11(7)	2(2)
	L = 1,800 ft				38(53)	17(12)	1(3)

: The critical gap used was 6 sec and 5 sec for numbers in parentheses. Using sampling rate = ∞ for ble-loop detector is equivalent to using speeds and time headways of aerial data for prediction.

s predicted acceptable are actually acceptable.
s predicted acceptable are actually unacceptable.
s predicted unacceptable are actually acceptable.

Table 2 also gives the measured speed errors for single- and double-loop detectors at different sampling rates. Again, the percentage of improvement is more significant for the double-loop detector than for the single-loop detector when sampling rate increases. Comparing the results given in Table 2 shows that the double-loop detector measured errors are about the same as the calculated errors, whereas, for the single-loop detector, the measured error is comparable to the calculated error only for the case when the sampling rate is 10 cps. This discrepancy might indicate that the car length and variation in car length that we used were imprecise. In fact, because the aerial data did not provide this information, we used estimations. Nevertheless, the measured errors were in the same order as the calculated errors, and the double-loop detector gave much better measurements than did the single-loop detector.

It is also interesting to examine the conditional error (Eq. 22) from the experiments. The conditional error is given in Table 4, columns Y, for both single- and double-loop detectors. The numbers are results from using 6 sec as the critical gap, and the numbers in parentheses are results from using 5 sec. It is observed that the conditional error decreases when sampling rate increases and, in the case of the double-loop detector, increases when the distance L increases. However, much of the error in b is caused by a lane change in which a car moved into lane 1 from lane 2 and made an acceptable gap unacceptable. In the case of infinite sampling rate, the five gaps in column b, when L = 600 ft and the critical gap = 6 sec, are all caused by such lane changes. Comparing this number with the corresponding number, case 7, when sampling rate was 30 cps, there are only two gaps actually caused by prediction error. The errors caused by lane changes are uncontrollable, but this error is expected to be proportional to the distance L. The other error given in Table 4, columns Y, does not do any harm because this means a vehicle has moved out of lane 1 and thus usually results in a larger gap than predicted. The only loss in this case is that we may not be able to use this larger gap, and part of the gap might be wasted.

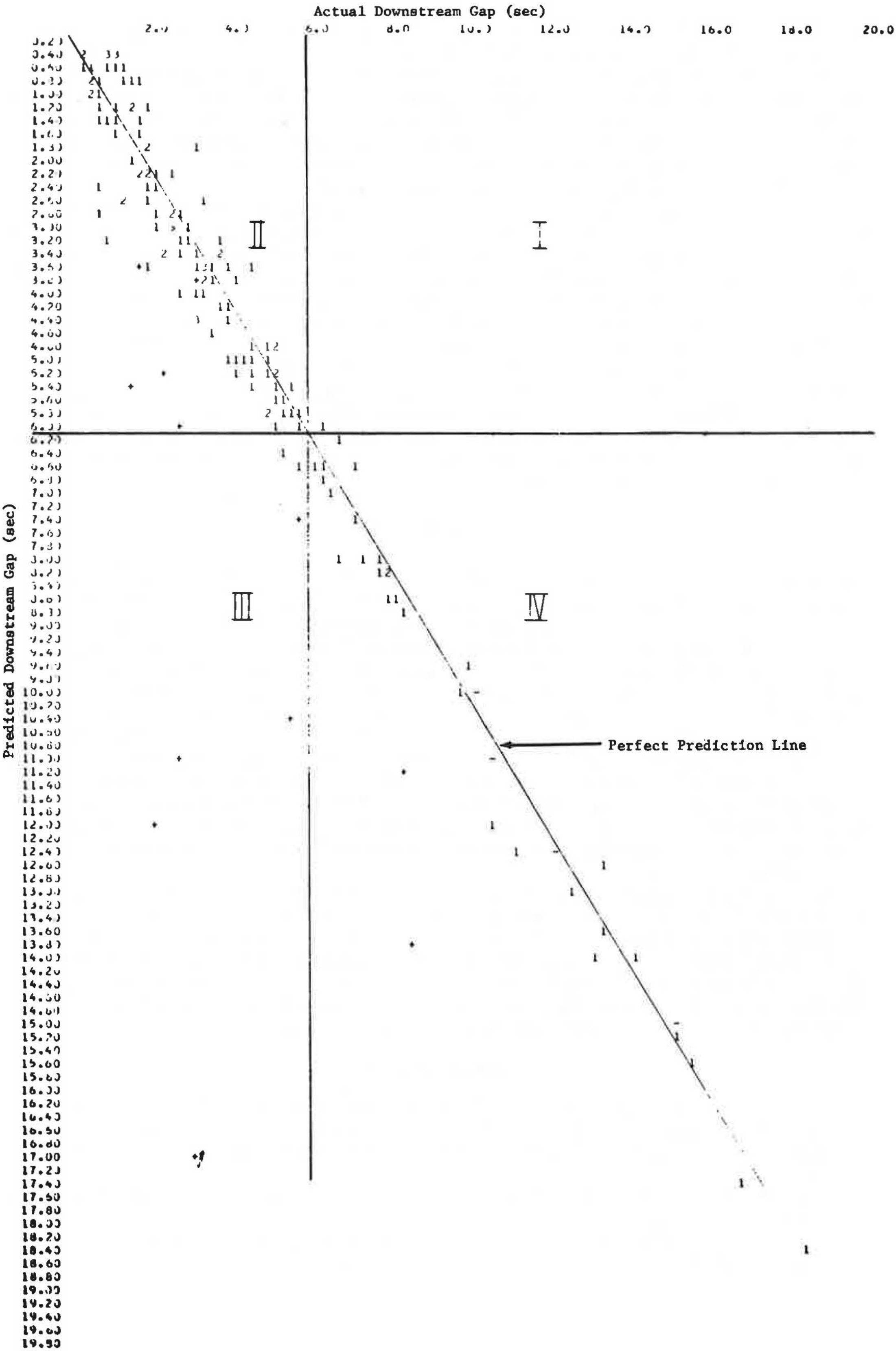
The numbers given in Table 4, columns X, are the valid prediction; i.e., gaps predicted acceptable are actually acceptable. In practice, we want to maximize the data in columns X and minimize those in columns Y and Z.

The differences between actual and predicted gap lengths are illustrated by the results of case 7 (in which sampling rate = 30 cps, L = 600 ft, and double-loop detector is used) shown in Figure 3. In Figure 3, the numerals indicate the number of gaps, "+" indicates a vehicle moving out of the lane and thus creating a gap, and "-" indicates a vehicle moving out of the lane and thus destroying a gap. It should be noted that no gaps greater than 20 sec (either predicted or calculated) are shown in the figure. Let us denote g_a, g_b, and g_c as the gaps that belong to columns X, Y, and Z respectively (Table 4) for a given critical gap length. A vertical line and a horizontal line with coordinates equal to the critical gap length divide Figure 3 into 4 quadrants. The first quadrant gives the gaps g_c, the third g_b, and the fourth g_a. The second quadrant corresponds to the category of correct prediction of unacceptable gaps. It is understood that the more entries we have in the second and fourth quadrants, the better the prediction is.

From the results of Tables 2 and 4, it seems that case 7 gives an adequate prediction. We shall compare these results with fixed metering by using the same critical gap as a basis. Because fixed metering controls the on-ramp by releasing vehicles at a fixed interval based on occupancy or flow of the freeway, we can achieve a fair comparison by examining the variation of gaps per minute and calculating the number of vehicles that could be merged safely. Based on a critical gap of 5 sec for releasing one vehicle and an additional 3 sec for each additional vehicle (releasing 2 vehicles when a gap is 8 sec, 3 vehicles when a gap is 11 sec, etc.), the number of vehicles that could be released in 1 min could be as low as 5 vehicles and as high as 15, with a standard deviation of 1.9 vehicles/min and a mean of 7.8 vehicles/min.

Whenever an acceptable gap is predicted, there might be differences between the predicted number of vehicles and the actual number of vehicles that could safely merge. This difference is called mismatch. For instance, if we have a predicted gap of 15 sec, which is determined to release four vehicles based on the critical gap of 5 sec, but the

Figure 3. Predicted gap versus actual gap.



actual gap is found to be only 10 sec, which can take only two vehicles safely, the number of mismatches in this case is, therefore, 4 - 2 = 2.

When case 7 in Table 4 was used, the number of incorrectly predicted gaps was 11, which was the sum of the two numbers in parentheses of columns Y and Z in case 7 for double-loop detectors. Using Figure 3, we can find the differences between the length of acceptable gaps and actual gaps. We counted a total of 26 mismatches, or 1.7 vehicles/min. This figure, 1.7 vehicles/min, compared to 1.9 vehicles/min found in the fixed-metering strategy seems of little improvement.

However, an examination of Figure 3 indicated that most of the 26 mismatches were caused by the "+," or by those vehicles moving from lane 2 to lane 1 within the distance 600 ft. A further improvement of the prediction could easily be achieved by prohibiting lane changes from lane 2 to lane 1 within that distance L so that mismatches caused by the "+" could be eliminated. This would result in a reduction of 15 mismatches and leave a total of only 11 mismatches, or 0.7 mismatched vehicle/min rather than 1.9.

Moreover, in the fixed-metering strategy, vehicles are released in a fixed interval and may not match the oncoming gap properly; even the number of available gaps is the same as expected in 1 min. On the other hand, when the prediction strategy using double-loop detectors was used, vehicles would be able to merge more properly into the available gap once the prediction is correct and therefore reduce the on-ramp disturbance. This feature is particularly important when an on-ramp has a short acceleration lane and poor sight distance.

CONCLUSIONS

We have analyzed the problem of predicting gaps by using single- and double-loop detectors placed at a fixed distance upstream of an on-ramp. It was found that the error in prediction increased linearly with the distance between the loop detector and the on-ramp. Other factors that caused the prediction error include the sampling rate, vehicle speed and variation, loop length, variation of vehicle length in the case of the single-loop detector, and distance between loops in the case of the double-loop detector The errors increase with increasing variation of vehicle length and variation in speeds but decrease with increasing loop length, distance between loops, vehicle speed, and sampling rate. Sampling rate has little effect on the prediction error in the case of the single-loop detector. However, it is of extreme importance in the double-loop case

Double-loop detectors gave much better prediction than single-loop detectors; the latter was erroneous and not adequate for gap prediction. Aerial data were used for gap prediction for a variety of distances and sampling rates. Predicted gaps were then compared with actual gaps downstream and were found to be in agreement with the analytical results.

Experimental results indicated that a sampling rate of 30 cps and a double-loop placed 600 ft from an on-ramp gave excellent prediction, which corresponded to 1.7 mismatches/min (released either too many or too few vehicles), and a majority of the mismatches were caused by lane changes. If the lane change were prohibited within the 600 ft, it would result in 0.7 mismatch/min. A comparison with the fixed-metering strategy also indicated that the prediction stragegy using double-loop detectors was superior and would cause less disturbance at the on-ramp.

REFERENCES

1. Athol, P. Headway Groupings. Highway Research Record 72, 1965, pp. 137-155.
2. Buhr, J. H., McCasland, W. R., Carveil, J. D., and Drew, D. R. Design of Freeway Entrance Ramp Merging Control Systems. Highway Research Record 279, 1969, pp. 137-149.
3. Drew, D. R. A Study of Freeway Traffic Congestion. Texas A&M Univ., Phd thesis, 1964.
4. Drew, D. R. Gap Acceptance Characteristics for Ramp-Freeway Surveillance and Control. Highway Research Record 157, 1967, pp. 108-134.

5. Drew, D. R., LaMotte, L. R., Wattleworth, J. A., and Buhr, J. H. Gap Acceptance in the Freeway Merging Process. Highway Research Record 208, 1967, pp. 1-36.
6. Glickstein, A., Findley, L. D., and Levy, S. L. Application of Computer Simulation Techniques to Interchange Design Problems. HRB Bull. 291, 1961, pp. 139-162.
7. Gourlay, S. M. Merging Traffic Characteristics Applied to Acceleration Lane Design. Yale Univ., Tech. Rept. 4, 1948, pp. 61-77.
8. Greenshields, B. D., and Weida, F. M. Statistics With Applications to Highway Traffic Analysis. Eno Foundation, 1952, pp. 164-169.
9. Hsu, Y. S., and Munjal, P. K. Fundamental Characteristics of Freeway Lane Changing Behavior. System Development Corp., Tech. Rept. TM-4638/018/00, 1972.
10. Malo, A. F., Mika, H. S., and Walbridge, V. P. Traffic Behavior on an Urban Expressway. HRB Bull. 235, 1960, pp. 19-37.
11. May, A. D. Traffic Characteristics and Phenomena on High Density Controlled Access Facilities. Traffic Engineering, March 1961, pp. 15-18.
12. May, A. D., Jr. Experimentation With Manual and Automatic Ramp Control. Highway Research Record 59, 1964, pp. 9-38.
13. May, A. D., Jr. Gap Availability Studies. Highway Research Record 72, 1965, pp. 101-136.
14. Munjal, P. K., and Hsu, Y. S. Characteristics of Freeway Traffic. System Development Corp., Tech. Rept. TM-4638/017/00, 1972.
15. Munjal, P. K., Hsu, Y. S., and Lawrence, R. L. Prediction and Stability of Freeway Gaps and On-Ramp Merging. System Development Corp., Tech. Rept. TM-U638/019/01.
16. Pearson, R. H., and Ferreri, M. G. Operational Study—Schuylkill Expressway. HRB Bull. 291, 1961, pp. 104-123.
17. Perchonok, P. A., and Levy, S. L. Application of Digital Simulation Techniques to Freeway On-Ramp Traffic Operations. HRB Proc., Vol. 39, 1960, pp. 506-523.
18. Pinnell, C., Drew, D. R., McCasland, W. R., and Wattleworth, J. A. Evaluation of Entrance Ramp Control on a Six-Mile Freeway Section. Highway Research Record 157, 1967, pp. 22-69.

DISCUSSION

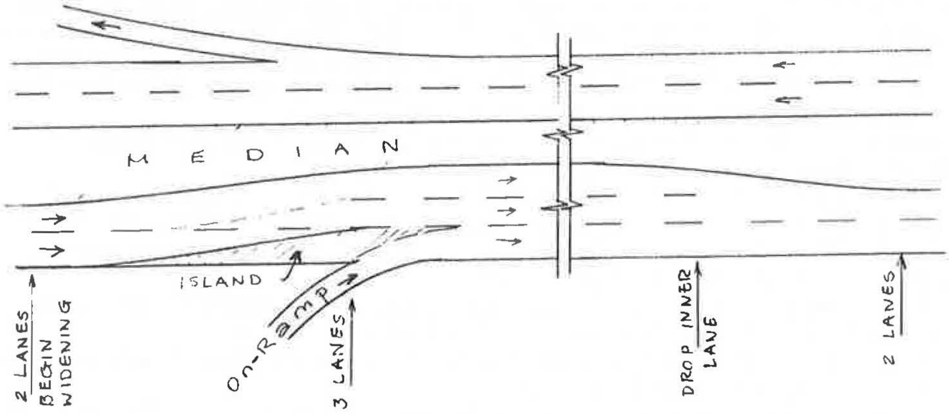
Oscar L. Sebastian, Delaware Department of Highways and Transportation

The increasing problem of on-ramp vehicles merging with freeway traffic has recently been studied, and solutions have been attempted using simulated computer models, electronic devices, and other sophisticated means. Remedial measures have invariably been proposed with emphasis on improving the merging maneuver, the ultimate measure being the provision of an electronic aid to guide merging traffic through acceptable gaps in the lane 1 traffic stream (Fig. 4). How effective this device is depends greatly on two factors that cannot be controlled by engineering tools: driver behavior and vehicle characteristics.

The basic merging maneuver establishes an angular conflict between vehicles, and, because driver behavior and vehicle characteristics are not uniform, this operation often results in the emergence, at this point of the freeway, of two clashing components: the offending driver and the offended driver. Depending on the severity of the conflict, one of the following ensues: slight delay, congestion, or collision.

In view of the foregoing observation, would the solution not lie in removal of this conflict? If conditions at an interchange reached such a stage as to create either congestion or safety hazards, the answer might lie in providing unhampered access from the on-ramp to lane 1, or the outer lane. This then would divert all research to the study of lane drops because the free entrance of on-ramp traffic would rely on the birth of a new outer lane upstream, and its disappearance farther downstream, of the gore

Figure 4. Proposed on-ramp direct access.



area. In effect, instead of providing an acceleration lane and a merging taper, constructing an outer lane would ensure that the same number of lanes are available to the through traffic, while allowing free access for on-ramp vehicles onto the outermost lane.

The following observations might support this theory:

1. The outer lane (lane 1) invariably carries slower moving traffic, trucks, and buses; consequently, gaps between vehicles are shorter.
2. The overtaking maneuver may be easier to execute; therefore, dropping the innermost lane may be more efficient operation because the acceleration lane with its taper is tantamount to dropping the outer lane.
3. The natural driving tendency is to "keep right, pass on left."
4. The outer lanes are always more traveled than the inner.
5. Where all through lanes are heavily traveled, and if the on-ramp also carries a large volume of traffic, an additional freeway lane would be indicated.

AUTHORS' CLOSURE

The discussion by Sebastian addresses the design of freeway on-ramp location rather than surveillance and control of freeway on-ramps. He removes the basic conflict between the on-ramp and through traffic by creating a new outer lane with the start of an on-ramp and by dropping the innermost lane that is carrying the through traffic. This solution gives a free entrance of on-ramp traffic at the expense of creating increased hazards requiring the through traffic to merge right every time it encounters an on-ramp. This proposed solution is useful in only the unusual cases where the safety hazards (due to on-ramp merging involving an acceleration lane) outweigh the one created by the through-traffic lane drop that requires the through traffic to merge right. Furthermore, observations 3 and 4 of Sebastian are applicable only to rural Interstate freeways, and the comment in observation 1, that vehicle gaps in the outer lane are shorter, is debatable because it has been found (14) that the outer lane always has the longest mean time and space headways. Determination of the extent and characteristics of the lane drop problems has recently been completed by System Development Corporation.

On-ramp merging control under the gap acceptance mode is especially recommended for old urban freeways that are characterized by significant traffic disturbances resulting from very short or no acceleration lane and poor sight distance and where construction of better on-ramp design is either impossible or very costly. Apart from the points of views of congestion and safety hazards, the proposed design would require additional space from median of the freeway near the on-ramp location. This is generally impossible to obtain on the current freeways. On any new construction program, it would be better to have good on-ramp acceleration lanes that use the outlying area of the freeway than to have the proposed design that requires the median area near the on-ramp and results in a wider median area at least the size of a lane width away from the on-ramps. If an acceleration lane is not sufficient and the freeway disturbances at the on-ramp need to be further reduced, consideration should be given to the creation of an auxiliary weaving lane that serves adjacent on- and off-ramps in addition to freeway ramp control.

FATAL ACCIDENTS AND TRAVEL DENSITY

Benjamin V. Chatfield, Office of Traffic Operations, Federal Highway Administration

A study of state highway department reports summarizing fatal accident experience by highway system and individual fatal accidents on the Interstate System in conjunction with Interstate travel data suggests that (a) sections of a highway system with higher travel densities typically have lower fatal accident rates and (b), for equivalent travel density differences between sections of a highway system, differences in fatal accident rates tend to be greater at lower densities.

•AS PART of their continuing effort to make highways safer, highway engineers are concerned with relationships between motor vehicle accident rates and various design and operating characteristics. Knowledge of such relationships is needed to determine whether design or operation modifications can be expected to reduce accident rates and, if so, which modifications are apt to be most effective. The relationships described in this paper are based on a study of the most serious motor vehicle accidents, those involving human deaths.

Reports of more than 5,000 individual fatal accidents on the Interstate System and state highway department summary reports that list over 200,000 fatal accidents on all types of roads and streets suggest that (a) sections of a highway system with higher traffic densities typically have lower fatal accident rates and (b), for equivalent travel density differences between sections of a highway system, differences in fatal accident rates tend to be greater at lower densities. These properties are illustrated by the downward sloping concave upward curve shown in Figure 1.

Precise relationships between fatal accident rates and highway design or operating characteristics are difficult to determine. This is compounded by the fact that other relationships are often overshadowed by that between fatal accident rates and travel density. Descriptions of accident rate density relationships are presented in this paper as an aid in the search for information applicable in the development of methods to reduce fatal accident rates.

TERMINOLOGY

The basic elements used in deriving the relationships reported in this paper are, for each highway system discussed, length of highway system in miles (L), annual volume of travel in millions of vehicle-miles per year (V), and number of fatal accidents per year (N). From these basic elements, "rate" and "travel density" are derived as follows: rate = 100 (N/V); travel density = (1,000,000)(V/L)/(365 days per year).

Thus, "rate" is in terms of fatal accidents per 100 million vehicle-miles, and "travel density" is in terms of daily vehicle-miles per mile of highway.

The introduction of the two derived elements, rate and travel density, permits two-dimensional representation of relationships among the three basic elements. The term density is more commonly used by highway engineers to refer to traffic density, measured in vehicles per mile of highway. Density refers only to travel density, in vehicle-miles per mile, in this paper.

DATA

The data on which this paper is based were submitted to the Federal Highway Administration by state highway departments in three forms. First, each highway depart-

Publication of this paper sponsored by Committee on Freeway Operations.

ment submits annually, on a TA-1 form, a summary of accident and travel data classified by highway system. Portions of this information are disseminated by FHWA in the annual publications (1, 2). The TA-1 information for the 4-year period from 1967 through 1970 was used as the basis for the curves showing state rates and densities.

Second, each state highway department furnishes copies of police reports of fatal accidents on the Interstate System to the Federal Highway Administration. Reports of over 5,000 fatal accidents (about 70 percent of the Interstate fatal accidents that occurred in 1969 and 1970) were used in this study.

Third, each state highway department prepares, on a schedule prescribed by Congress, detailed estimates of the cost of completing the Interstate System. These estimates include data on traffic volumes, number of lanes, urban-rural classification, and so on for short sections of all Interstate highways. Information in the 1970 Interstate estimate was used in conjunction with the police reports mentioned earlier as the basis for urban and rural Interstate rate-density curves for specific numbers of lanes.

There are, of course, defects in the data. The quality of the accident reporting systems in the jurisdictions where reports are generated is not uniform. Estimates of travel volumes are also subject to some variation in reliability. The effect of random variations, e.g., in the number of fatal accidents on the rural portion of a single system in a single state, has been substantially reduced by the use of large quantities of data.

The data appear to be quite adequate to support the general conclusions reported in this paper, but caution should be used in the application of precise readings of the plotted curves.

NOTES ON CURVE FITTING

It is readily evident that, when rate-density coordinates are computed and points are plotted for all states, the points for a few states will have an inordinate influence on fitted curves if each point is given the same weight. For example, although Hawaii's rural Interstate fatal accident rate for 1967-70 is more than four times the national rate, Hawaii's exposure to the opportunity for a fatal accident on this system, in vehicle-miles, is less than 1 percent of the exposure on California's rural Interstate highways. Clearly, Californian experience should have more influence than Hawaiian experience in fitting curves to rate-density points.

One way to overcome the disparity was described in a 1970 article on rate-density relationships (3). This method consisted essentially of grouping states (and parts of states) so that each group had the same exposure in vehicle-miles. A disadvantage of this method is that the effect of small states at the density extremes may be undervalued. In addition, when the number of points is small, the criteria used for grouping the states have a substantial effect on the coordinates of the points and, consequently, on curves fitted to the points.

A least squares technique that incorporates weighting to account for differences in exposure has been used to fit the curves in this paper to rate-density points. This technique involves the minimization of the sum of the products of vehicle-miles of travel and the squared graphical distance from the plotted rate-density point to the curve being fitted.

For convenience in curve fitting calculations, it is helpful to use curves defined by mathematical formulas. Selection of a particular type of mathematical formula is necessarily a matter of judgment. Although some other curve forms might serve equally well, hyperbolic curves were selected to represent the rate-density relationship. These hyperbolas were oriented in every case with one asymptote parallel to the density axis and, measuring counterclockwise to the other asymptote, with an angle greater than 90 deg and less than 180 deg between the two. A density adjustment factor was used to control the relative length of the travel density and accident rate ordinates during the curve fitting process; this adjustment reduces density units to the same order of magnitude as rate units for convenience in calculation.

The general formula for the hyperbolic curves used in this paper is

$$\{[(x/g) - h]\cos\theta - (y - k)\sin\theta\}^2/a^2 - \{[(x/g) - h]\sin\theta + (y - k)\cos\theta\}^2/b^2 = 1$$

Figure 1. Typical accident rate-density relation.

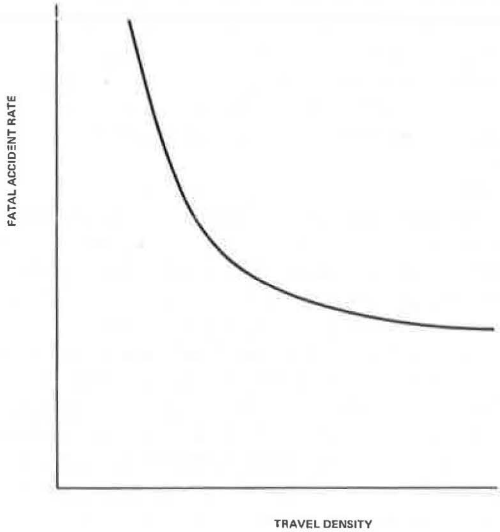


Table 1. Parameters for hyperbolic curves.

Figure	g	a	b[a]	h	k	θ (deg)
3a	500	2.8	15.880	6.970	1.400	-80
3b	2,000	1.0	14.301	12.000	1.205	-86
4a	1,000	1.0	11.430	9.986	1.827	-85
4b	2,000	2.8	80.182	19.980	-0.814	-88
4c	2,000	2.4	45.795	8.000	0.000	-87
4d	2,000	0.5	9.541	19.800	1.310	-87
4e	5,000	1.7	5.929	8.000	1.000	-74
5a	200	2.7	154.683	30.000	3.300	-89
5b	500	1.9	27.171	34.000	1.710	-86
6a	500	5.7	54.232	8.000	0.000	-84
6b	500	1.0	19.081	27.000	2.290	-87
7a	100	6.6	62.795	12.010	0.035	-84
7b	100	0.5	9.541	30.000	3.410	-87
7c	100	5.2	59.436	15.600	0.000	-85
7d	20	2.7	19.212	80.003	3.509	-82
8a	20	2.0	57.273	80.000	1.890	-88
8b	200	1.3	74.477	40.000	1.436	-89
9a	5	5.6	160.363	27.940	1.000	-88
9b(1)	50	0.1	0.225	9.990	3.360	-66
(2)	50	1.6	9.074	9.600	2.640	-80

[a] $b = a(-\tan\theta)$.

where

x = density, in daily vehicle-miles/mile,
y = rate, in fatal accidents per hundred million vehicle-miles,
g = d density adjustment factor, and

other terms are those commonly used in analytic geometry for rotated translated conics. Parameters for the curves illustrated in this paper are given in Table 1.

THE INTERSTATE SYSTEM

Fatal accident data for of the Interstate System are more complete and probably more reliable than data for other segments of U.S. roads and streets. Rate-density points for rural and urban Interstate highways in each state are shown in Figure 2. The rates vary in a random fashion, the magnitude of the variation being dependent on vehicle-miles of travel in that state. In addition to this random variation, each state's rate differs from rates in other states because of differences in highway design and operation and in procedures for collection and classification of data.

In Figure 2, states with rates that are least likely to differ from the national rate solely because of random variations are represented by crosses. Other states are represented by circles. The points were sorted by using a statistical quality control technique conditioned on the assumption that, if all variations were random, the probability should be 0.99 that any individual point would fall within computed limits (4, 5). It is evident that, in both rural and urban areas, the crosses tend to fall above the horizontal lines at low densities and below at high densities. Consideration of both the location and weighting of the data points leads to the conclusion that lines sloping downward to the right would fit the data better.

Because the horizontal lines are graphical representations of the notion that fatal accident rates are independent of travel density, the shift to a downward sloping line is tantamount to rejection of this notion in favor of the hypothesis that sections of a highway system with higher travel densities typically have lower fatal accident rates.

A second hypothesis, which is suggested by the tendency of rates to level off at high densities for both the rural and urban Interstate data (Fig. 2), is that, for equivalent travel density differences between sections of a highway system, differences in fatal accident rates tend to be greater at lower densities.

The two hypotheses are shown graphically in Figure 3 by hyperbolas by using crosses and circles as before. It is apparent from the number of crosses that variation in density does not fully explain variation in rates, yet it appears to be more reasonable to place a higher degree of belief in the hypotheses than in the notion that rate is independent of density.

It we treat the rate-density relations shown in Figure 3 as the "normal" relations between state fatal accident rates and state travel densities, the crosses in Figure 3 may be interpreted as representing states that have "abnormally" high or low rates in consideration of the densities and volume of travel in those states. States with such abnormal rates (as reported by the state highway departments) are listed below.

Urban		Rural	
High	Low	High	Low
Rhode Island	New York	Georgia	Pennsylvania
Virginia	South Carolina	Colorado	Mississippi
Missouri	Florida	Texas	Ohio
Michigan	Louisiana	New Mexico	Illinois
Kansas	Minnesota	Arizona	Iowa
Colorado	Indiana	California	Wisconsin
New Mexico	Oklahoma		Minnesota
Washington	Oregon		North Dakota
			South Dakota
			Nebraska
			Oklahoma

Figure 2. Rate-density on Interstate highways in (a) rural and (b) urban areas.

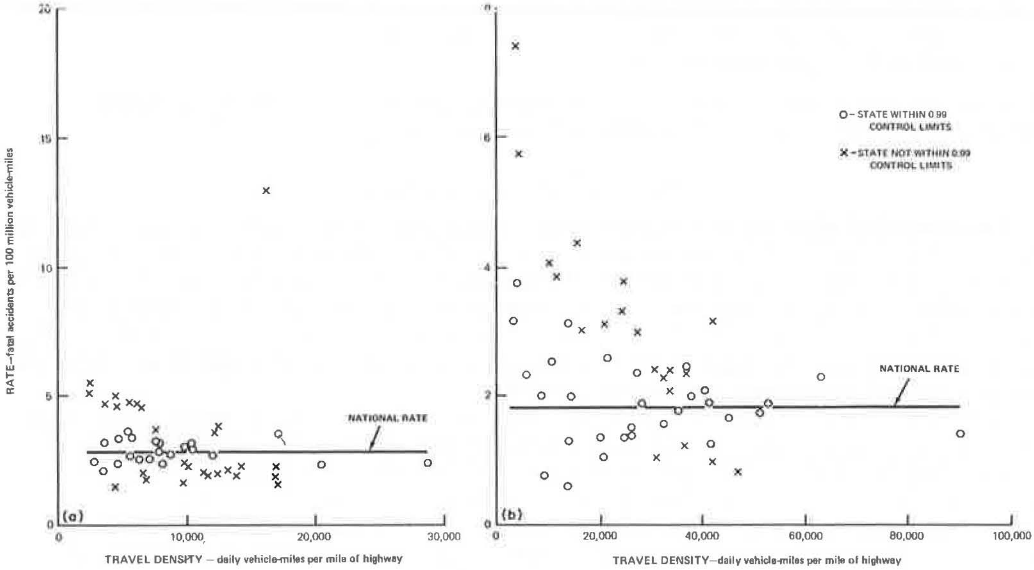
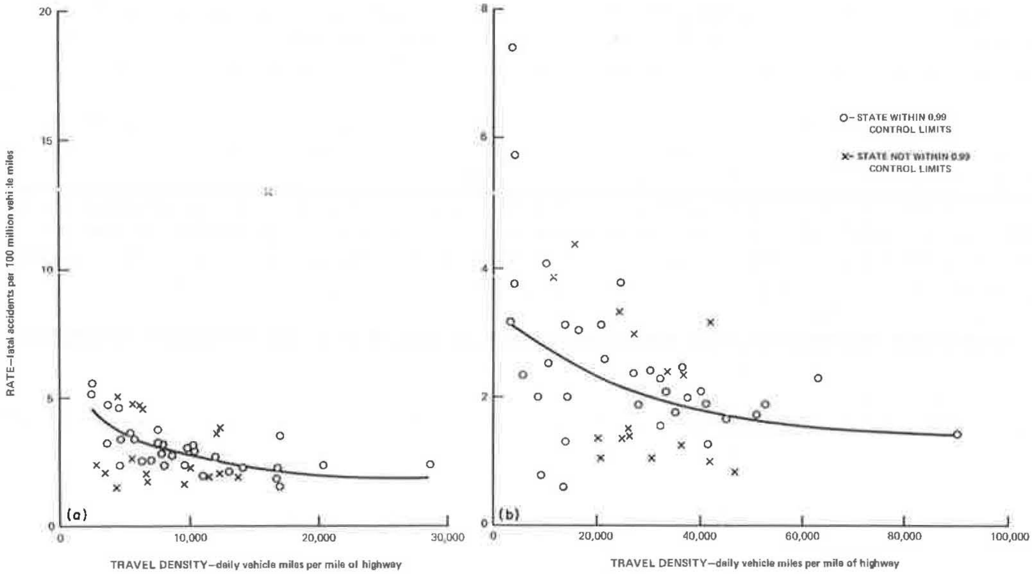


Figure 3. Hyperbolic rate-density relation on Interstate highways in (a) rural and (b) urban areas.



Relations between fatal accident rates and characteristics other than travel density are most likely to be evident in those states where abnormal rates occur. In seeking evidence of such relations, we must realize that some of what appears to be abnormality is known to be a reflection of data collection or classification problems rather than real differences in accident experience.

Figures 2 and 3 are based on data reported by the state highway departments on Form TA-1. To check the hypotheses, data were also taken from reports of individual accidents, combined with information from Interstate cost estimates, and adjusted so that the numbers of highway miles, vehicle-miles, and fatal accidents on completed sections of the Interstate System in each state correspond to the figures reported. For selected types of Interstate highway, estimate sections were arranged in order by density and divided into 40 groups. Rate-density relationships were then derived from the rates and densities of the 40 groups.

Rate-density points and curves are shown in Figure 4 for four- and six-lane rural Interstate highways and for four-, six-, and eight-lane urban Interstate highways. Contrary to the situation in the figures based on state summary data, all points in each of these graphs have the same weight (i.e., represent the same amount of travel in vehicle-miles). Visual inspection of the fit of the curves to the points in Figure 4 is therefore more meaningful than in the preceding figures.

The quality control limits (Fig. 4) indicate that deviations from the relationships represented by the curves tend to be within the normal range for random deviations. However, because only about 1 percent of the observed points would be expected outside of the plotted limits if deviations from the curve were completely random, it is clear that characteristics in addition to density influence fatal accident rates.

The data on which Figure 4 is based strongly suggest the validity of the hypotheses that rates, and the rate of decrease in rates, decrease as density increases.

APPLICATION OF THE HYPOTHESES

The resources available to public agencies for safety improvements are limited. Officials therefore must necessarily choose among competing projects. In a choice between two projects on four-lane rural Interstate highways where the rates and densities before improvement were 3.0 and 2,000 and 3.0 and 20,000, the curve in Figure 4a indicates that, other things being equal, an improvement on the higher density section of highway is more likely to be effective. The rate before improvement (3.0) is substantially below the rate-density curve at the lower density (2,000) and substantially above what is typically experienced at the higher density (20,000).

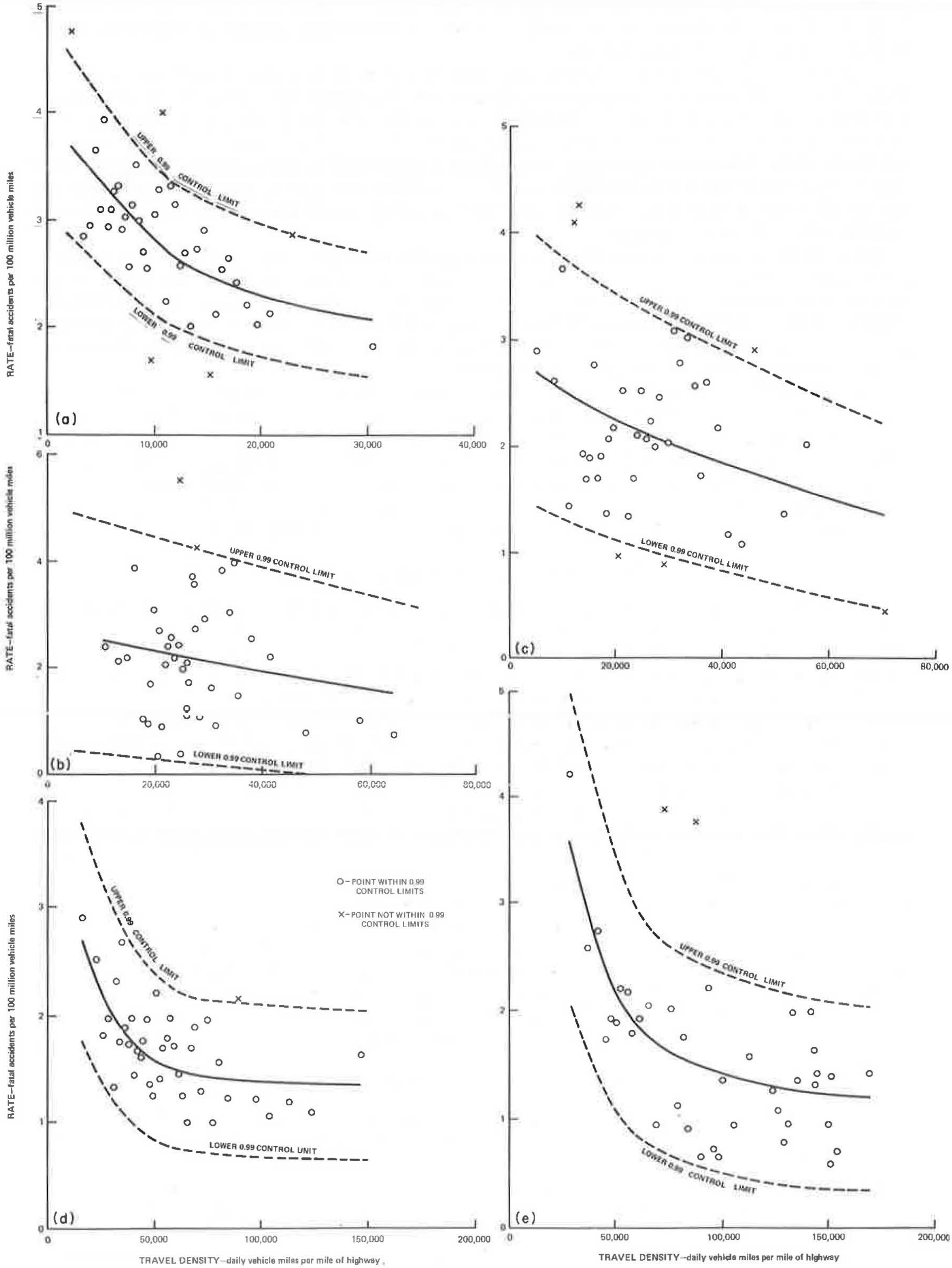
The difference between state and national fatal accident rates is frequently the sole criterion in evaluation of a state's accident experience. When a state rate is significantly above the national rate, there is a tendency to view the situation with alarm; state rates below the national rate may be a source of exultation. A more balanced evaluation of a state's experience would take into account the observation that states with low travel densities normally have high fatal accident rates and vice versa.

Preliminary analysis indicates that rate-density curves remain relatively stable over a period of time. Therefore, for sections of Interstate highway where the travel density is increasing rapidly, the first hypothesis suggests that fatal accident rates will tend to decrease even though no improvements are made. From the second hypothesis it might be inferred that such a decrease would probably be greater on low-density sections of the Interstate System than on high-density sections (Fig. 4).

OTHER HIGHWAY SYSTEMS

The Interstate System is considerably more homogeneous than other major U.S. highway systems. Finished segments of the Interstate System were built within a relatively short period of time with close adherence to national standards; other systems differ markedly from state to state and within individual states. Despite this low level of homogeneity, most of the systems for which TA-1 data are available exhibit rate-density relationships similar to those on Interstate highways. Several states have had difficulty in assigning accident and travel data to the non-Interstate TA-1 categories.

Figure 4. Rate-density relation on (a) four-lane rural, (b) six-lane rural, (c) four-lane urban, (d) six-lane urban, and (e) eight-lane urban Interstate highways.



This is partly due to lack of compatibility among classification systems and partly due to data collection problems. Many of the deviations from what appear to be normal rate-density relationships result from this difficulty rather than the influence of highway design or operating characteristics.

Conventions comparable to those shown in Figure 3 were used to plot data for other systems shown in Figures 5 through 9. To aid in consideration of these data, the average annual number of highway miles, vehicle-miles, and fatal accidents for each system are given in Table 2.

Traveled-way Interstate highways (Figure 5) should perhaps not be treated as a system at all. The total length of traveled-way Interstate highways is the computed difference between the length of finished Interstate highways and the authorized length of the Interstate System. As the Interstate System nears completion, therefore, the length of traveled-way Interstate highways approaches zero. The length decreased from about 18,000 miles in 1967 to less than 13,000 miles in 1970.

In the rural Interstate traveled-way data, the point most at variance with the two hypotheses of this paper is the point at a rate of 10.26 and density of 16,547. This point represents a state that reported no rural traveled-way mileage for 1971. Thus, although this point will remain unchanged when later data are added to the data plotted, its influence will decline and the fitted curve will tend to rotate in a clockwise direction.

For traveled-way Interstate highways in urban areas, the tendency toward lower rates at higher densities seems to show clearly from the points plotted in Figure 5b, but, because of the varying exposure that the points represent (from 48 to 17,626 million vehicle-miles), interpretations based only on the location of the points are unreliable and should be avoided.

The Interstate System is part of the federal-aid primary system, which consists largely of intercity routes. Data for the non-Interstate portion of the primary system are shown in Figure 6. Although differences in density obviously do not fully explain rate differences for this system, the data for both rural and urban areas support the first hypothesis: Rates decrease as density increases. A decreasing rate of decrease is apparent for the urban data but not particularly for rural data.

Federal-aid secondary highways on state systems are shown in Figure 7. The point at the upper right in Figure 7a represents a state that had lower exposure than any other state on this class of highways—about 0.045 percent of the total exposure. Thus, whereas this point does tend to hold up the right end of the curve, its influence is relatively low. Similarly, in Figure 7b, the highest point tends to flatten the right end of the curve but does not have a greater influence because it represents only about 0.2 percent of the total exposure.

For federal-aid secondary highways administered by local jurisdictions, as for those administered by the state, the rate-density curves show the characteristic decrease in rate as density increases. In Figure 7c the high points with rate-density coordinates of (18.75, 859) and (11.76, 1,390), which together represent about 0.5 percent of the exposure on this system, tend to straighten the fitted curve. The one extremely high rate in Figure 7d has the opposite effect.

Rate-density points and curves for state highways not on federal-aid systems are shown in Figure 8. Although the right end of the curve in Figure 8a is obviously held up by the single point on the extreme right, both rural and urban curves are compatible with the two hypotheses concerning the slope of the curves.

Data for rural roads under local jurisdictions are shown in Figure 9a. The characteristic shape of the rate-density curve is evident.

Figure 9b contains data for urban roads or streets under local jurisdiction. The data for this system represent about 44 percent of the total urban exposure in vehicle-miles for all systems. Because of the restrictions used in the curve fitting process, the fitted curve is an L-shaped hyperbola; a U-shaped curve would fit the data better.

If the data for two states, New York and California, are treated separately, a curve fitted to the points for the remaining states exhibits the hypothesized characteristics, as shown by the dashed curve in Figure 9b. Whether the data for these two states should be set aside is debatable. The vehicle-mile exposure in these states is a substantial share of the national exposure: about 22 percent of the vehicle-miles on urban

Figure 5. Rate-density relation for traveled-way Interstate highways in (a) rural and (b) urban areas.

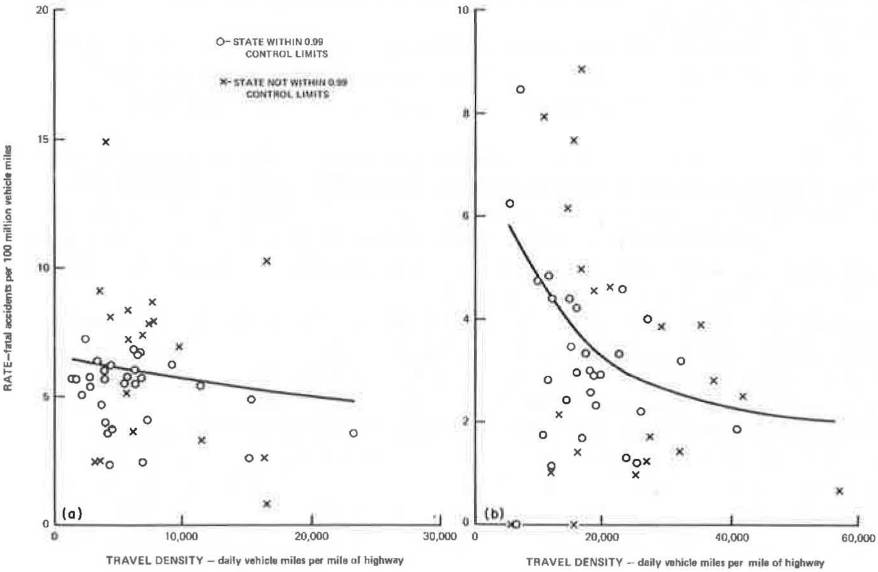


Figure 6. Rate-density relation for other federal-aid primary highways in (a) rural and (b) urban areas.

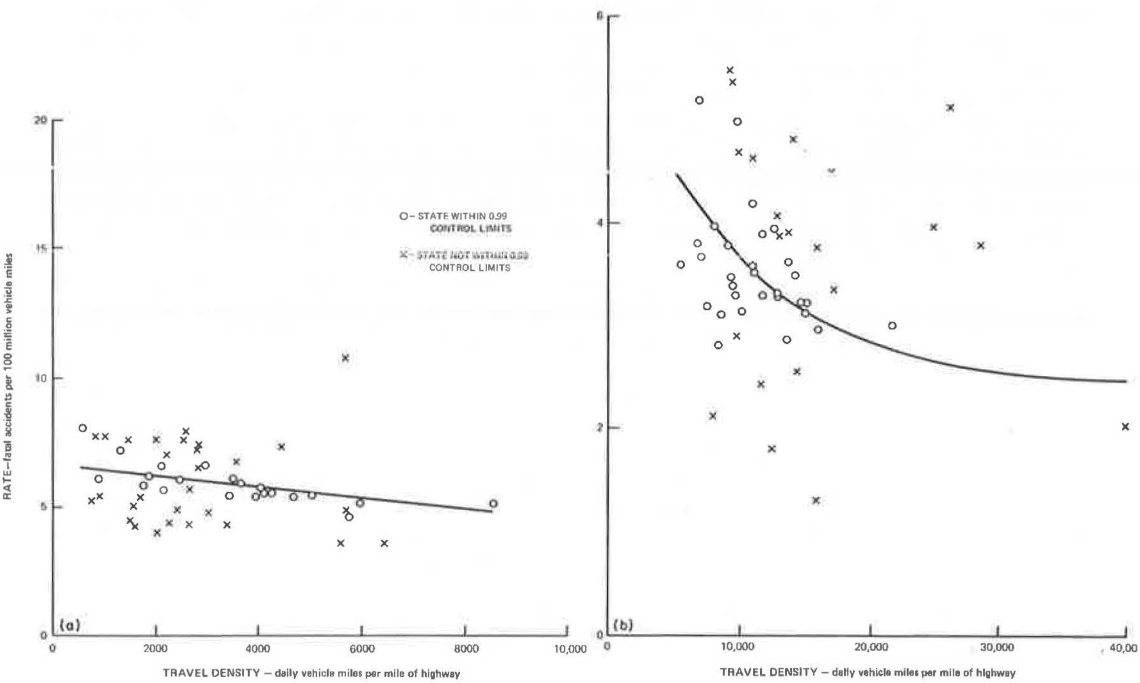


Table 2. Average annual national experience from 1967 to 1970.

System	Location	Highway Miles	Vehicle-Miles (millions)	Fatal Accidents
Interstate system (final)	Rural	21,620	67,366	1,929
	Urban	5,112	68,657	1,261
Interstate system (traveled way)	Rural	12,892	28,713	1,691
	Urban	2,682	23,921	676
Other federal-aid primary	Rural	188,810	187,094	10,984
	Urban	23,760	126,671	4,026
Federal-aid secondary—state	Rural	283,052	86,966	5,926
	Urban	11,254	28,370	978
Federal-aid secondary—local	Rural	324,770	45,506	2,760
	Urban	14,345	24,532	918
Other state highways	Rural	120,030	17,655	963
	Urban	13,295	24,006	658
Local roads and streets	Rural	2,215,920	80,767	4,919
	Urban	455,716	232,900	7,867

Figure 7. Rate-density relation for state federal-aid secondary highways in (a) rural and (b) urban areas and for local federal-aid secondary highways in (c) rural and (d) urban areas.

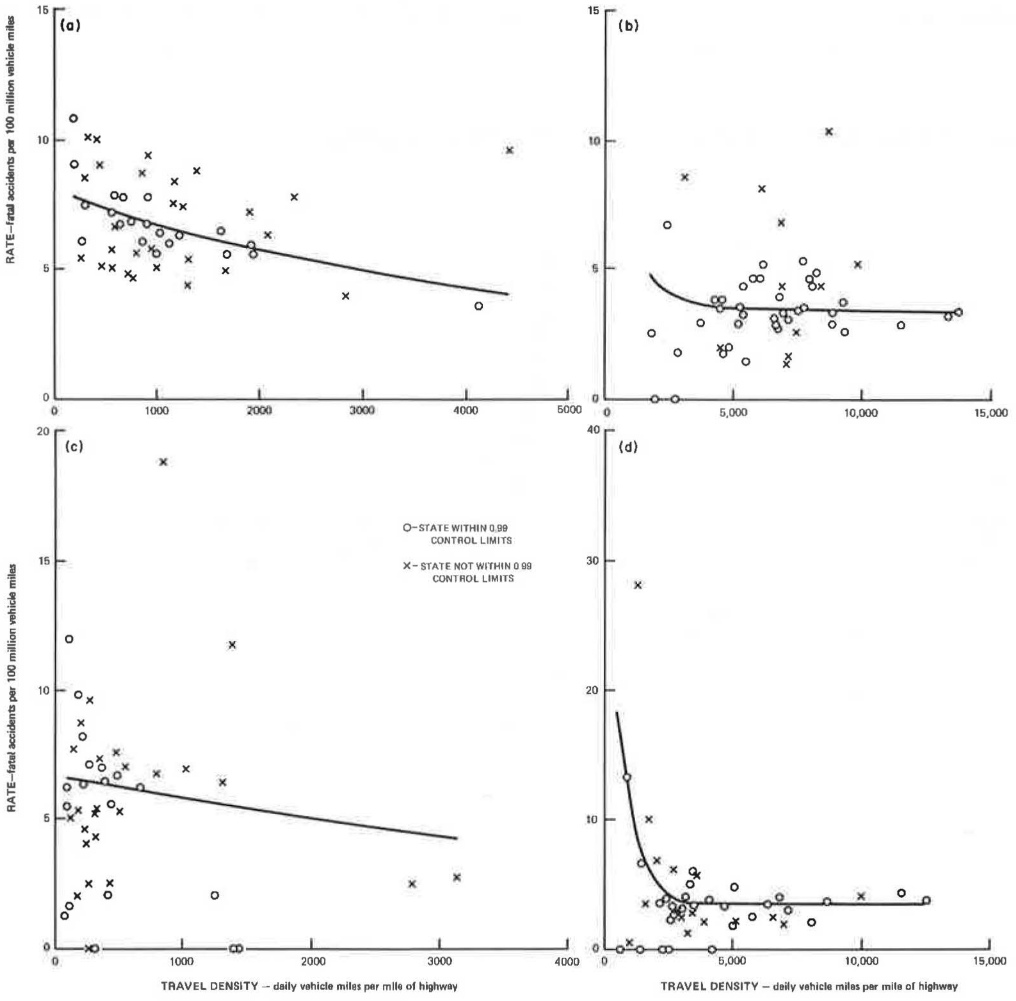


Figure 8. Rate-density relation for other state highways in (a) rural and (b) urban areas.

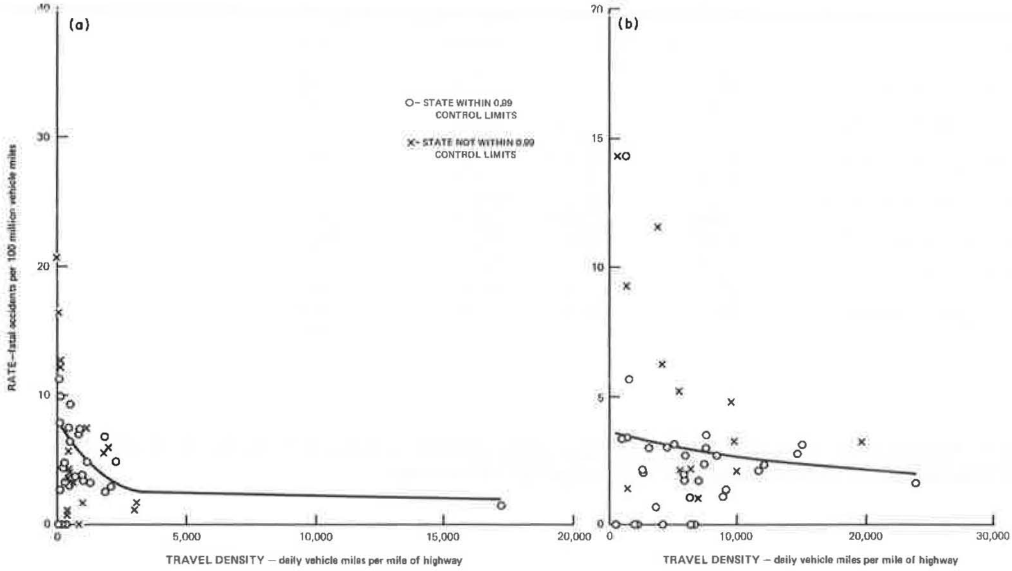
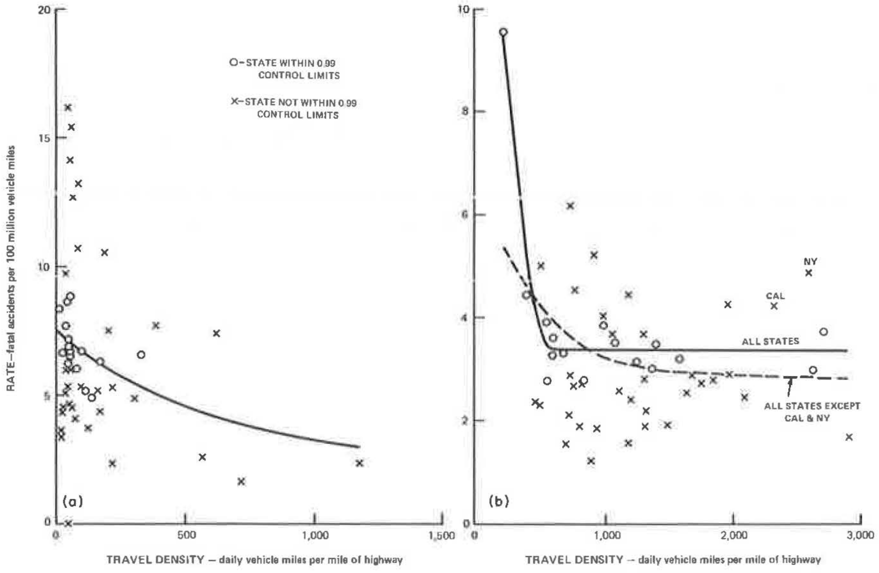


Figure 9. Rate-density relation for local roads in (a) rural and (b) urban areas.



roads and streets under local jurisdiction and almost 10 percent of the vehicle-miles driven on all urban roads and streets in the United States during the 1967 to 1970 period. As a result of a reclassification of certain urban local roads or streets, the addition of 1971 data will shift the California point to the left and slightly downward. The location of the New York point is at best an educated guess; because New York's method of accident classification by highway system is not wholly compatible with the categorization specified on the TA-1 form, particularly for systems not under state jurisdiction, the Federal Highway Administration has distributed New York accidents among systems on the basis of limited information.

On balance, it is considered that the data plotted in Figure 9b do not comprise strong evidence either in support of or in opposition to the two hypotheses presented in this paper.

Taken together, the curves in Figures 5 through 9 offer impressive support of the hypothesis that fatality rates decrease as density increases. To a lesser degree they also support the proposition that the rate of decrease in fatality rate decreases as density increases. Recalling that circles represent points within the range where 99 percent of the points should fall if all deviations of observed rates from rates on the curves were random deviations and that crosses represent points outside of this range, it is evident from the ratio of crosses to circles in Figures 5 through 9 that much of the variation in rates is beyond the range of expected random fluctuation.

SUMMARY

The data on which this report is based are known to be imprecise and, in a few cases, badly flawed. In addition, characteristics other than density vary widely from state to state within the system classifications that have been used, so that exposure on each system is far from homogeneous. Despite these problems, it is clearly evident for most of the systems considered in this paper that high fatal accident rates tend to occur where travel density is low and that, as densities increase, fatal accident rates tend to decrease more gradually.

There is always a question of the desirability of releasing detailed tables or graphs based on imprecise data because of the danger that some users will rely too heavily on such information. The purpose of this paper is to describe what appear to be the general characteristics of the rate-density relationship. It is emphatically not intended that values taken from the plotted curves be treated as highly precise information. The curves should be used with caution.

The information described in this paper strongly suggests the validity of the following two hypotheses:

1. Sections of a highway system with higher travel densities typically have lower fatal accident rates, and
2. For equivalent travel density differences between sections of a highway system, differences in fatal accident rates tend to be greater at lower densities.

REFERENCES

1. Fatal and Injury Accident Rates on Federal-Aid and Other Highway Systems. Federal Highway Administration, U.S. Department of Transportation, 1967 through 1970.
2. Highway Statistics. Federal Highway Administration, U.S. Department of Transportation, Annual.
3. Chatfield, B. V. Relationship of Fatality Rates and Fatal Accident Rates to Travel Densities on the Interstate System. Public Roads, Vol. 36, No. 2, June 1970, pp. 48-49.
4. Morin, D. A. Application of Statistical Concepts to Accident Data. Highway Research Record 188, 1967, pp. 72-79.
5. Evaluation of Criteria for Safety Improvements on the Highway. Roy Jorgensen and Associates and Westat Research Analysts, Inc., Appendix C, 1966.

RAMP CONTROL TO RELIEVE FREEWAY CONGESTION CAUSED BY TRAFFIC DISTURBANCES

Harold J. Payne*, University of Southern California; and
William S. Meisel and Michael D. Teener, Technology Service Corporation

Systems have been installed in several cities in the United States, including Los Angeles, which allow the control of ramp signal lights from a central location. Time-of-day ramp volume schedules have been used with some success. This paper examines a class of traffic-responsive ramp control algorithms for adjusting ramp volumes in response to traffic disturbances, e.g., congestion resulting from lane blockages. A large number of traffic-responsive ramp control plans are considered. Each plan is evaluated in terms of freeway service (vehicle-miles) and delay (vehicle-hours) over a fixed control period by simulating the response of traffic to a lane blockage on a macroscopic model of freeway traffic. The result of the analysis is a set of ramp control plans, each of which yields minimum delay for a specified level of freeway service. The performance measures associated with these plans are plotted against one another, yielding a trade-off curve for final selection of a ramp metering plan.

•TWO TYPES of congestion on freeways can be identified: recurrent and nonrecurrent (1). Recurrent congestion is that caused by daily situations of demand for freeway service in excess of freeway capacity. Nonrecurrent congestion is that caused by unusual circumstances such as accidents or other incidents affecting traffic conditions.

Congestion of both types can be ameliorated by metering the traffic allowed to enter the freeway through on-ramps. Specifically, with standard red-green signals and attendant sensors designed to allow the passage of one vehicle per cycle, the cycle length can be adjusted to achieve any desired ramp volume between a minimum determined by driver impatience and a maximum determined by the signal system and ramp geometry.

Recurrent congestion can be eliminated by scheduling ramp volumes during peak traffic periods to limit the traffic having access to the freeways to levels at or below freeway capacity. An example of ramp scheduling for the inbound Hollywood Freeway in Los Angeles is discussed in a later section.

The purpose of this research is to address the problem of relieving nonrecurrent congestion. The approach studied here is the use of feedback (traffic-responsive) control laws that adjust ramp metering volumes relative to the nominal levels (determined by scheduling ramps as suggested above) in response to traffic conditions measured in real time. There are two significant difficulties in identifying effective feedback control laws: There is no single criterion by which one can measure the effectiveness of a control scheme, and the complexity of the freeway model employed prevents feasible application of optimization schemes. These difficulties are handled here through the use of two novel technical devices. First, the effectiveness of control is measured by a set of separate criteria reflecting most quantitative aspects of interest. The optimization process produces a set of "noninferior" choices for the parameters of the control rule; "inferior" solutions are those that are worse in all criteria than some other feasible control algorithm and are hence not worth consideration. Second, for purposes

Publication of this paper sponsored by Committee on Freeway Operations.

*Mr. Payne is also a consultant to Technology Service Corporation.

of this optimization process, the relatively complex freeway model is replaced by a simpler input-output representation derived from input-output samples of the model, the inputs being the parameters of the control rule and the outputs being the multiple criteria for effectiveness of the control. This procedure has been termed repromodeling (2).

In a later section, the freeway simulation model is detailed for a segment of the southbound Hollywood Freeway. A comparison of simulated freeway traffic to actual data is provided. Measures of freeway performance to be used later in selecting feedback control rules are described; ramp scheduling to eliminate recurrent congestion is specified, and resultant freeway traffic conditions are described. Finally, the means by which traffic incidents are modeled is described, and the effect of incidents on several measures of freeway performance are discussed.

Also, the form of the feedback control rule for ramp metering adjustments is specified.

In other sections, the technique and results of repromodeling to obtain a simplified input-output relationship are described, and a technique of multicriteria parameter selection is employed to identify noninferior choices of ramp feedback control rules. Typical time histories of traffic conditions obtained by exercising the full freeway simulation under the selected feedback control rules are displayed.

Finally, a discussion of results of this study and discussion of suggested future work on the problem of relieving nonrecurrent congestion are presented.

SIMULATION OF FREEWAY TRAFFIC

In this section, we shall discuss the simulation of freeway traffic and in particular the simulation of morning peak-period traffic on a segment of the southbound Hollywood Freeway in Los Angeles.

Simulation Model

A complete development and discussion of the aggregate variable model employed here are discussed elsewhere (3). The intent here is to present the particular form of the model applied to the Hollywood Freeway.

The freeway segment is divided into 16 sections, defined by section boundaries at x_j, j = 1, 2, ..., 10, A, ..., F (Fig. 1). The peak period is divided into uniform time intervals of length Δt = 5 sec. Within the j th section defined by the interval (x_j, x_{j+1}), we define (Fig. 2)

l_j = number of lanes;
Δx_j = section length in miles;
ρ_j^n = section density, the number of vehicles in this section at time $t_0 + n\Delta t$ divided by the number of lanes and the section length in veh/lane/mi; and
v_j^n = section speed, the average speed of the vehicles in this section, in mph.

At the section boundary x_j, we define

q_j^n = volume, rate at which vehicles pass x_j in the time interval $[t_0 + (n - 1)\Delta t, t_0 + n\Delta t]$, divided by the number of lanes, in veh/hour/lane;

and, where appropriate,

$f_j^{ON,n}$ = on-ramp volume, rate at which vehicles enter the on-ramp at x_j in the interval $[t_0 + (n - 1)\Delta t, t_0 + n\Delta t]$, in vph; and
$f_j^{OFF,n}$ = off-ramp volume, rate at which vehicles leave the off-ramp at x_j in the interval $[t_0 + (n - 1)\Delta t, t_0 + n\Delta t]$, in vph.

Equation 1 expresses the conservation of vehicles:

$$\rho_j^{n+1} = \rho_j^n + \frac{\Delta t}{l_j \Delta x_j} (l_{j-1} q_j^{n+1} - f_j^{OFF,n+1}) \tag{1}$$

for n = 0, 1, 2, ..., and j = 1, ..., 10, A, ..., F. Note that we have adopted the convention that a change in the number of lanes is assumed to take place slightly downstream of a section boundary. Consequently, the total freeway volume at x_j is $l_{j-1}q_j$.

Under uniform conditions within a section, the volume, density, and speed are related precisely by

$$q_{j+1}^{n+1} = \rho_j^n u_j^n \tag{2}$$

for j = 1, 2, ..., 10, A, ..., F, and n = 0, 1, 2, ... This will be adopted as our second set of model equations.

The final equation of the model, which shows the dynamic speed-density relationship, is derived from a spatially continuous model by spatial averaging (3):

$$u_j^{n+1} = u_j^n - \Delta t \left\{ \underbrace{u_j^n \frac{u_j^n - u_{j-1}^n}{\frac{\Delta x_j + \Delta x_{j-1}}{2}}}_{\text{convection}} + \frac{1}{T}\left[\underbrace{u_j^n - u_e(\rho_j^n)}_{\substack{\text{relaxation}\\ \text{to}\\ \text{equilibrium}}} + \underbrace{\frac{\nu}{\rho_j^n} \frac{\rho_{j+1}^n - \rho_j^n}{\frac{\Delta x_{j+1} + \Delta x_j}{2}}}_{\text{anticipation}} \right] \right\} \tag{3}$$

for j = 1, ..., 10, A, ..., F, and n = 0, 1, 2, ...

Equations 1, 2, and 3 express three physical processes. The first of these is convection, i.e., the fact that vehicles traveling at speed u_{j-1} in the upstream section will tend to continue to travel at that speed. The second represents the tendency of drivers to adjust their speeds to an equilibrium speed-density relationship. The third is a model of anticipation of changing travel conditions ahead; i.e., drivers tend to slow down if the density is seen to be increasing.

Additionally, boundary conditions and the initial values of the speeds and densities in each section must be defined. The initial densities at slightly before 0630 were obtained from the density chart (Fig. 3), and the initial speeds were obtained from the equilibrium speed-density curve (4) from the given densities. The flow rate, obtained from vehicle counts, at the extreme upstream boundary, q_0^n, is also specified. One "dummy" section at each end of the freeway segment ("0" and "G") has been added with $u_0^n = u_1^n$ and $\rho_G^n = \rho_F^n$.

In the simulations, we have taken T = 15 sec, $\nu = 5\ (\text{mi})^2/\text{hr}$, and

$$u_e(\rho) = 107 - 2.31\rho + 0.0125\rho^2 - 7.4 \times 10^{-5}\rho^3$$

This speed-density relationship is a rescaled version of a least squares fit to data taken from the Harbor and Hollywood Freeways (4). The parameters T and ν were chosen to obtain close agreement with data, as discussed in the next section.

To initiate the calculations, u_j^0 and ρ_j^0 for j = 1, ..., N, must be specified. The separate specifications of q_1^n, u_0^n, and ρ_{N+1}^n for n = 0, 1, ... serve as boundary conditions. On- and off-ramp volumes are provided as inputs. We will see later that an on-ramp regulation scheme can be simulated with this same algorithm by specifying the on-ramp rates as functions of neighboring section variables. Off-ramp volumes are taken as fixed fractions of upstream freeway volumes, i.e.,

$$f_j^{OFF,n} = l_{j-1}\beta_j q_j^n$$

for j = 1, 2, ..., N.

Additional details concerning the simulation are available in other reports (5, 6).

Simulation of Congested Traffic

To illustrate the characteristics of the traffic and depict operations, a traffic density chart (7) based on aerial data from the Hollywood Freeway is shown in Figure 3. This chart shows the buildup of congestion during the morning peak period.

Figure 1. Hollywood Freeway sections.

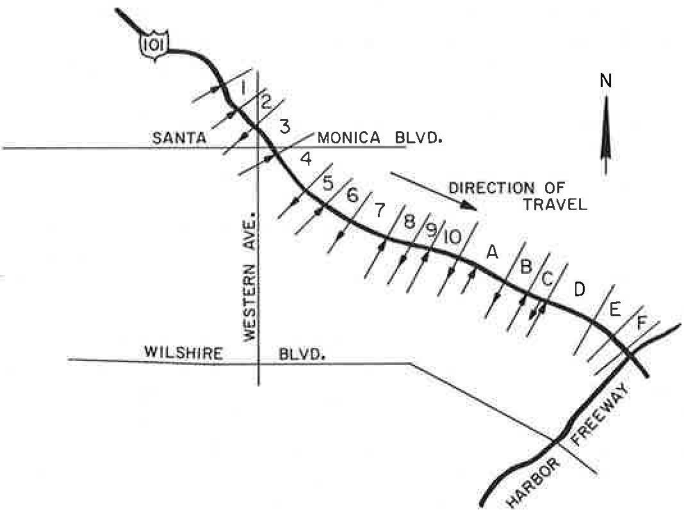


Figure 2. Aggregate variables.

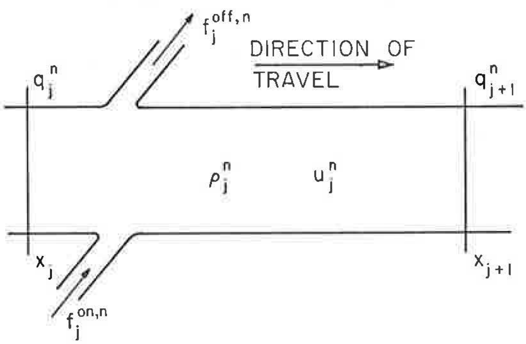
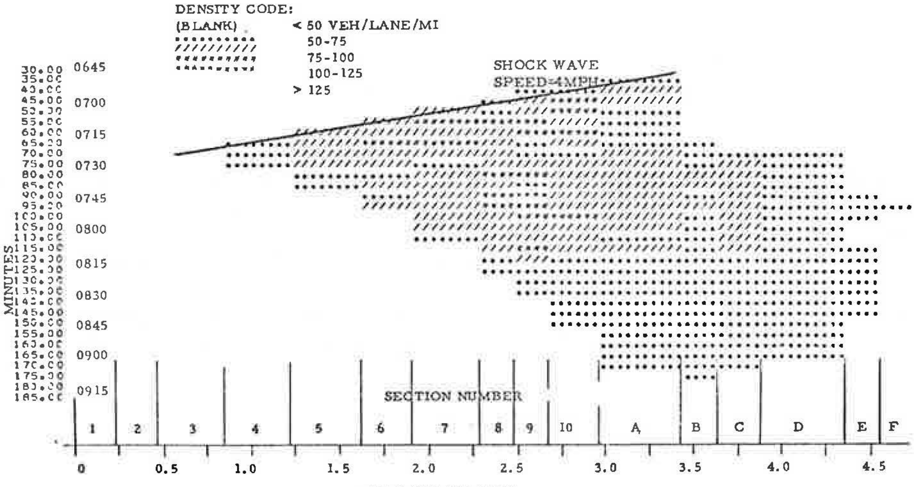


Figure 3. Density chart.



The numbers indicate average density in terms of veh/mi/lane on the section of freeway indicated at the time shown on the vertical scale. The densities are the average of all southbound lanes except where noted. Contour lines have been drawn at the 50, 75, and 100 density levels. Densities between 40 and 50 are high, and, although flow is generally smooth, there is little if any available capacity. Densities greater than 50 generally reflect congested operation with some stop-and-go driving.

The earliest bottleneck appears between 6:45 and 7:15, in section A (just west of Alvarado Street). After 7:15, the bottleneck moves to section D (between the Glendale Boulevard on-ramp and the four-level interchange).

The chart illustrates the striking shock wave phenomenon characteristic of traffic flow. The wave is shown to start propagating backward from the bottleneck at about 0645. The shock wave speed, which is an important characteristic of the freeway, can be easily calculated as indicated. The result is a speed of 4 mph.

The simulation model detailed above was exercised with average levels of on- and off-ramp and extreme upstream volumes (6) and initial conditions corresponding to Figure 3. The resulting density chart is shown in Figure 4. A comparison of this with the actual freeway density chart (Fig. 3) gives an indication of the validity of the model.

Note the approximate agreement in the buildup of congestion and the shock wave speed. Also note the agreement in the regions of congestion (densities > 50 veh/lane/mi) at various times and the time period of congestion in various sections.

Measures of Traffic Performance

A number of quantitative measures of traffic performance are currently used to describe freeway traffic (8). Two measures that are widely employed will be defined here and used in the remainder of this report.

The service rate (also called throughput or total travel rate) for a freeway partitioned into sections indexed $j = 1, \ldots, N$ is given by

$$\text{Service rate} = \sum_{j=1}^{N} l_j s_j q_{j+1}$$

where

s_j = length,
l_j = number of lanes, and
q_{j+1} = flow rate out of the jth freeway section.

The total service performed by the freeway over the time interval (t_1, t_2) is then simply

$$\int_{t_1}^{t_2} \sum_{j=1}^{N} l_j s_j q_{j+1}(t)dt = \sum_{j=1}^{N} l_j s_j a_{j+1}(t_1, t_2)$$

where $a_{j+1}(t_1, t_2)$ is the total number of vehicles passing the downstream boundary of the jth freeway section in the time interval (t_1, t_2). The service rate and total service are easily computed from the simulation model.

Vehicles traveling at less than some nominal speed V (say 50 mph) are considered to be experiencing delay. A vehicle traveling at speed V_1 in a section of length s will spend a time $s/V_1 - s/V$ in excess of that spent if traveling at speed V. Inasmuch as that vehicle spends a time s/V_1 in this section, the rate of delay is simply $1 - V_1/V$ assuming $V_1 < V$.

Figure 4. Simulated density chart.

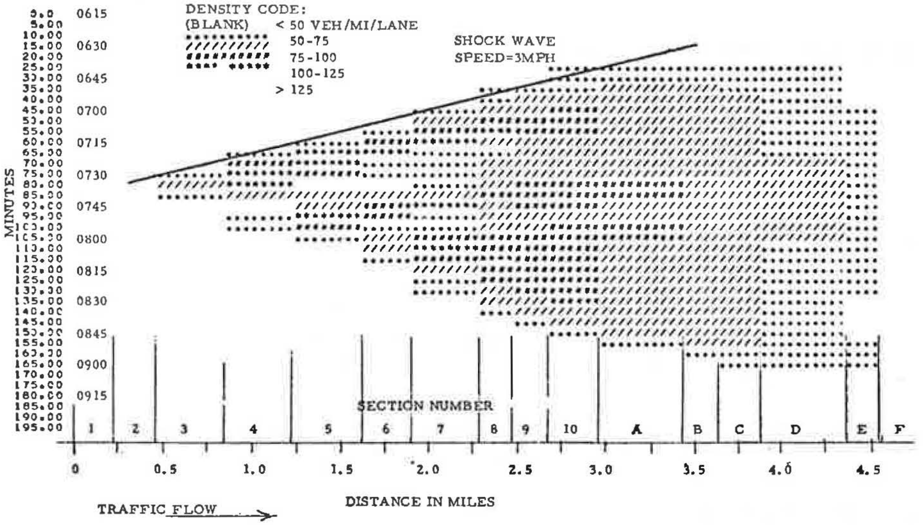
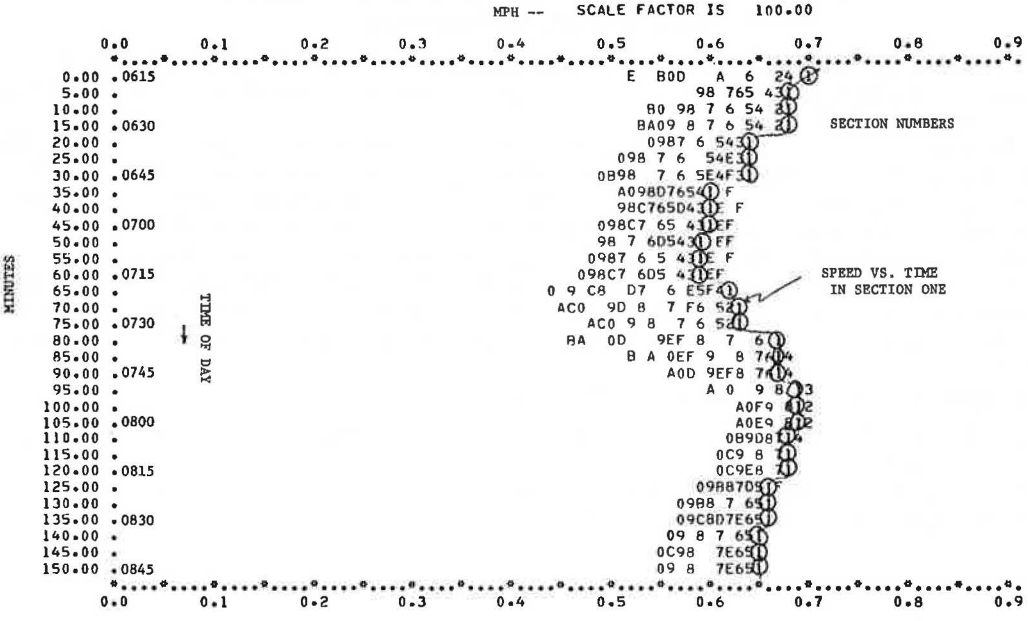


Table 1. Time-of-day on-ramp rates (in vph).

Time of Day (a.m.)	Ramp and Section Number[a,b] Hollywood 1	Sunset 2	Santa Monica 4	Melrose 6	Vermont 8	S. Lake 10	Glendale D
6:15-6:30	288	348	400	468	508	360	320
6:30-6:45	340	368	492	528	360	360	360
6:45-7:00	392	400	592	512	360	360	360
7:00-7:15	448	432	660	860	360	360	360
7:15-7:30	500	460	620	880	680	504	420
7:30-7:45	540	460	548	880	680	960	680
7:45-8:00	540	460	500	828	680	960	1,296
8:00-8:15	528	528	560	820	680	920	1,116
8:15-8:30	528	580	600	820	680	820	952

[a]The Rampart on-ramp was closed.
[b]The rate for the Alvarado on-ramp was 340 between 6:15 and 6:30, after which it was closed.

Figure 5. Simulated speed/time-of-day nominal control.



The total delay rate on the freeway is then

$$\sum_{j=1}^{N} l_j s_j \max\left[0,\ \rho_j(t) - \frac{1}{V}\ q_{j+1}(t)\right]$$

and the total delay for the interval (t_1, t_2) is

$$\sum_{j=1}^{N} l_j s_j \int_{t_1}^{t_2} \max\left[0,\ \rho_j(t) - \frac{1}{V} q_{j+1}(t)\right] dt$$

In the remainder of the paper, V = 50 mph. Again, delay is readily computed from the simulation.

Time-of-Day Nominal Control

A set of time-of-day on-ramp controls is given in Table 1 (6); this set will form the nominal values to be adjusted by the traffic-responsive algorithm discussed later.

Figure 5 shows the simulated section speeds every 5 min during the control period, which result from the proposed time-of-day nominal on-ramp rates. Our purpose is to derive good traffic-responsive algorithms to modify these rates in the presence of a moderate incident partially blocking the freeway.

Simulation of Incidents

An accident or a stalled vehicle generally results in the blockage of one or more lanes until the vehicles can be removed from the freeway. Therefore, we simulate an incident by removing one or more lanes of traffic from service in some section of the freeway. This, of course, results in a reduction of capacity in that section.

In Figure 6, a slight traffic incident is simulated by reducing the number of lanes in section 10 from four to three for the time period 0645 to 0715. Note that, with the proposed time-of-day on-ramp rates, the traffic flow is quite stable, even with the incident; the nominal speeds are all above 18 mph.

Traffic performance measures with and without the incident were calculated. The incident only slightly affects freeway service, whereas delay is increased dramatically.

TRAFFIC-RESPONSIVE CONTROL OF ON-RAMPS TO RELIEVE CONGESTION

The philosophy of the control strategy proposed here is to reduce the number of vehicles in congested sections and to increase the number of vehicles in sections with relatively few vehicles.

The section in which an accident occurs will of course be congested. In addition, the capacity of this section is generally reduced so that, in the section immediately downstream, the density decreases. As time progresses, the shock wave attendant with the congestion propagates upstream so that more sections become congested.

When the accident is removed from the freeway, capacity is restored so that the first section downstream of the accident section soon reaches nominal density. There is, however, no strong tendency for sections once congested to become immediately relieved, so that congestion may persist for a considerable length of time after the accident is removed.

In general terms, the proposed control strategy is to

1. Allow additional vehicles to enter the section immediately downstream of the accident section;
2. Reduce ramp rates into sections on which there is congestion, if possible, to a level that leads to a reduction of density; and

3. When congestion is so severe that ramp control into congested sections is insufficient (because of a lower limit on ramp rates) to achieve a desired rate of reduction of density, reduce ramp rates further upstream to provide help in alleviating the congestion.

The detailed specification of the ramp control strategy is based on the conservation equation of the aggregate variable model:

$$\rho_j^{n+1} = \rho_j^n + \frac{\Delta t}{l_j \Delta x_j} [l_{j-1} q_j^{n+1}(1 - \beta_j) - l_j q_{j+1}^{n+1} + f_j^{n+1}]$$

We have dropped the superscript ON for the on-ramp rate.

Distinct control policies will be in effect for three ranges of density:

Density	Value
Underutilized	$\rho_j^n < \rho_f$
Normal	$\rho_f < \rho_j^n < \rho_c$
Congested	$\rho_j^n > \rho_c$

In normal sections the time-of-day ramp rates will be employed. In underutilized sections, ramp rates are to be increased in an attempt to fill up the section. The choice

$$f_j^{n+1} = l_j q_{j+1}^n - l_{j-1} q_j^n (1 - \beta_j) + l_j \delta \Delta x_j$$

will lead to a rate of increase of density δ. This specification is subject to the constraint $f_j^{n+1} \le f_{j,max}$. We will consider δ in the range of $0 \le \delta \le 200$.

In congested sections, an attempt will be made to reduce densities. The choice

$$f_j^{n+1} = l_j q_{j+1}^n - l_{j-1} q_j (1 - \beta_j) - l_j \alpha \Delta x_j$$

will yield a rate of density reduction α. We will consider α in the range $50 \le \alpha \le 200$.

This specification is restricted by the constraint $f_j^{n+1} \ge f_{j,min}$. If this constraint is violated, we will attempt to provide further help with reductions in upstream ramp rates. For each section, we compute

$$e_j = \begin{cases} f_{j,min} - f_j^{n+1} & \text{if positive} \\ 0 & \text{otherwise} \end{cases}$$

If section j + 1 is congested but section j is not congested ($\rho_j^n < \rho_c$), there is some benefit in reducing that ramp rate:

$$f_j^{n+1} = l_j q_{j+1}^n - l_{j-1} q_j^n (1 - \beta_j) - \gamma e_{j+1}$$

The parameter γ is to be taken in the range $0 \le \gamma \le 1$; for values less than 1, less than the full burden of excess congestion is reflected into this section. Again, we require $f_j^{n+1} \ge f_{j,min}$, and, if this constraint is violated, we set $e_j = f_{j,min} - f_j^{n+1}$.

If section j + 1 is congested and section j is also congested, there is little value in a further reduction in ramp rate inasmuch as the flow out of the j th section is essentially governed by conditions in section j + 1 and is therefore insensitive to the on-ramp flow into section j. In this case, no additional ramp control is to be exerted.

The reflection of excess congestion upstream into reduced ramp rates is to be computed by starting in section N - 1 and sweeping in an upstream direction in the freeway.

The threshold densities ρ_f and ρ_c were fixed for this study to be $\rho_f = 15$, $\rho_c = 50$ veh/lane-mile.

Figure 6. A slight incident: simulated speed using time-of-day nominal control.

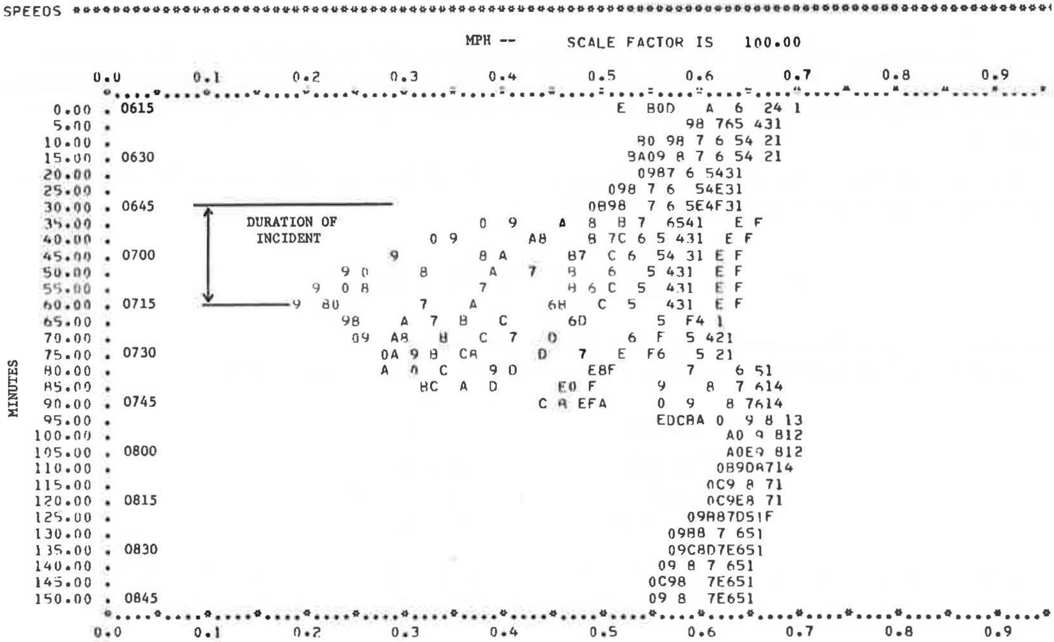


Figure 7. Delay simulated with repromodel.

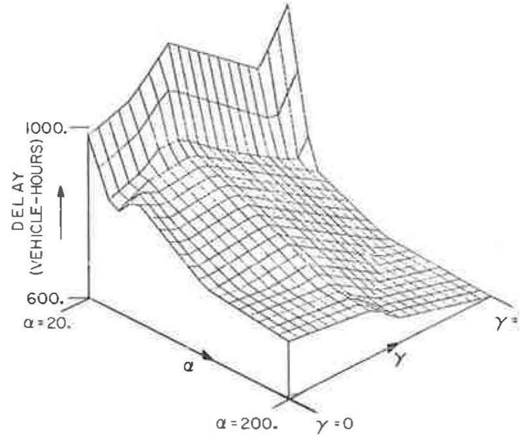
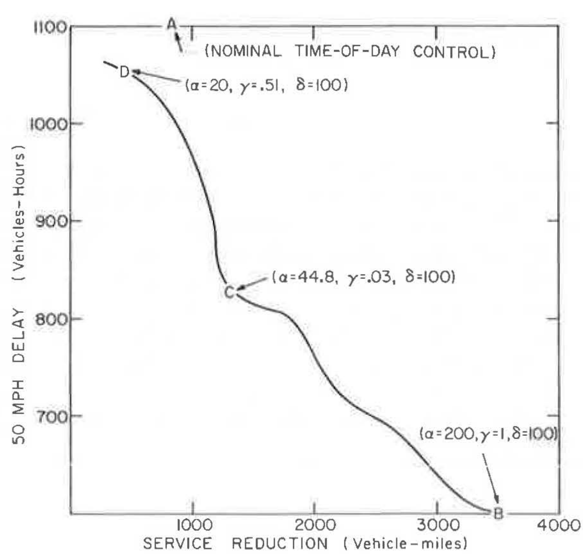


Figure 8. Noninferior solutions (service reduction = 7.101E + 0.4 vehicle-miles; 50-mph delay = 6.038E + 0.2 vehicle-miles).



To summarize, the proposed control strategy is then posed in terms of three parameters:

1. δ governs the rate of filling underutilized sections,
2. α governs the rate of reducing densities in congested sections, and
3. γ governs the extent to which ramps upstream of the congestion region are employed to alleviate congestion.

It was discovered in the analysis that follows that the results were insensitive to the value of δ, which was set at 100 for the remainder of the analysis.

THE REPROMODEL

We desire to identify parameters α, γ, and δ of the traffic-responsive control rule, which reduce freeway delay due to an incident without unduly reducing freeway service. Any set of parameters that yield values of the criterion functions, none of which is better than those yielded by another set of parameters, is an inferior solution. A vector of parameters that are not so dominated is a noninferior solution (9, 10). In the present two-criteria case, a noninferior solution is a set of ramp control parameters that produce minimum delay for a given level of service; each such potential level of service produces a noninferior solution. In the next section, we obtain noninferior solutions for the present problem.

The optimization technique employed to do so requires that the performance measures, freeway delay and service reduction, be available for all allowed sets of parameters. This, in principle, requires that the freeway simulation be exercised for each parameter set evaluated and would entail considerable computational expense if performed directly on the model.

To obviate this requirement, we shall construct an input-output "repromodel" (2) of the freeway simulation based on relatively few simulations with specified parameter sets. The inputs are the parameters of the feedback control rule; the outputs are the performance criteria, freeway delay and service reduction. The repromodel is a continuous multivariate piecewise linear function obtained by a recently developed algorithm (2, 11, 12).

Forty-two runs of the freeway simulation were executed. In each run, an incident was simulated by blocking two lanes of traffic in section 10 from 6:45 to 7:15 a.m., and the resulting measures of traffic performance were tabulated.

The function relating the parameters of ramp control to delay is shown in Figure 7. The relation between the parameters and the two cost functions was repromodeled, resulting in over a 100-to-1 reduction in computation time for each iteration of the model, with small approximation error. This made the multiple-criteria optimization discussed in the next section feasible.

MULTIPLE-CRITERIA OPTIMIZATION

The performance criteria, service reduction in vehicle-miles and delay in vehicle-hours, are not expressed in the same units. Their units are related by a factor with units of speed (miles per hour), but it is not at all obvious what speed should be chosen if one desires to obtain a single criterion by weighting these two. Reduction in delay will generally be earned only at the expense of reduction in service. For example, the choice of $\alpha = 50$ and $\gamma = 0$ rather than of $\alpha = 20$ and $\gamma = 0$ reduces 50-mph delay but also reduces service. However, the choice of $\alpha = 50$ and $\gamma = 0.4$ rather than of $\alpha = 20$ and $\gamma = 0.8$ reduces delay and provides increased service. The last pair of performance measures (and their associated control parameters) is clearly an inferior choice, inasmuch as both performance measures can be improved with another choice of parameters.

We defined a noninferior set of performance measures in the previous section. It should be clear that there are many (a continuum of) noninferior solutions.

A multiple-criteria optimization technique (9, 13, 14) yields noninferior solutions distinct in cost and/or parameter space. Application of this technique yields the noninferior solutions shown in Figure 8.

Figure 9. Speeds in severe incident: a = 200 and γ =1.00.

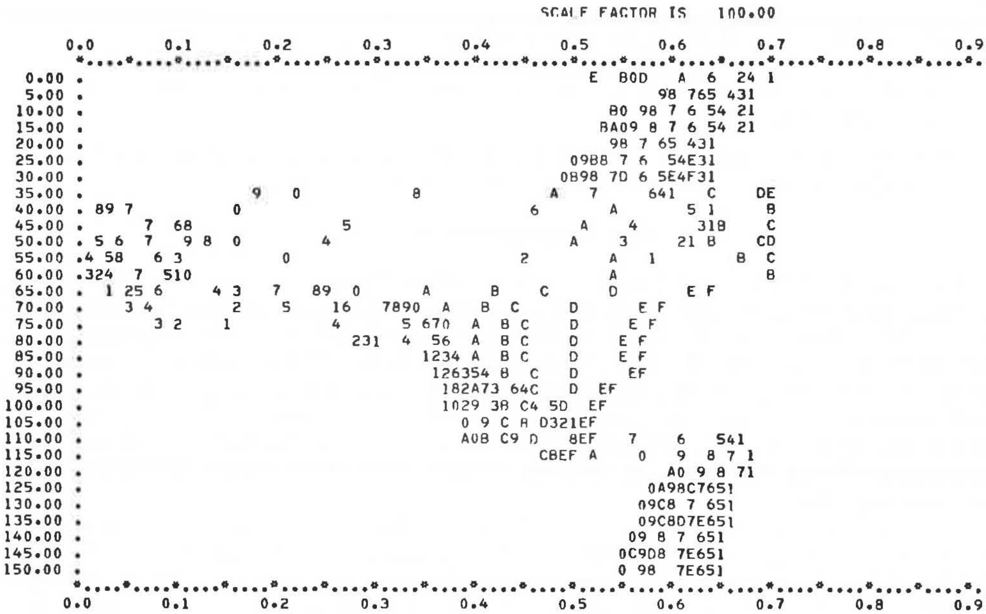


Figure 10. Speeds in severe incident: a = 20 and γ = 0.51 (service = 7.399E + 0.3 vehicle-miles; 50-mph delay = 1.049E + 0.3 vehicle-miles).

The points A, B, C, and D shown in Figure 8 are interesting: Points D and B are extremes, point C is a compromise, and point A is the result of time-of-day control alone, i.e., without the traffic-responsive feedback control rule. Indeed, point A is an inferior solution.

Point B produces minimum delay at the greatest expense of service reduction. This point corresponds to the choice $\alpha = 200$, $\gamma = 1$; this choice produces the greatest reduction of ramp volumes in and upstream of the congested sections. Because the service reduction is less than 5 percent of the total freeway service, good surface street alternates might be able to absorb this burden.

Point C is a preference point; about that point, a large sacrifice in service is required for a small improvement in delay, and a large sacrifice in delay is required for a small improvement in service. This point corresponds to the choice $\alpha = 44.8$, $\gamma = 0.03$; this choice produces moderate reduction of ramp volumes in the congested sections with essentially no reduction of ramp volumes upstream of the congested sections.

Point D results in minimum service reduction and minimum reduction in delay. This point corresponds to the choice $\alpha = 20$, $\gamma = 0.51$; this choice produces relatively small reduction of ramp volumes in congested sections and some reduction of ramp volumes upstream of the congested sections.

The repromodel was used to obtain the spectrum of noninferior solutions; the simulation model was used to examine in detail points B and D. Figures 9 and 10 show the speed profiles for control plans corresponding to points B and D respectively.

DISCUSSION OF RESULTS

It is worthwhile to emphasize the nature of the results produced by this study, specifically, the trade-off curve shown in Figure 8. This trade-off curve indicates the minimum reduction of service attendant with a control plan that will reduce delay to a specified level. The analyst, then, is not in the position of presenting a single "best" control plan; rather, he provides a spectrum of control plans, each of which is a "good" (noninferior) control plan, and for which the consequences, as measured by service and delay, are made evident. A planner or policy-maker can use this trade-off curve, along with subjective judgments about the desired effects of the control plan, to make a decision on the choice of a control plan.

This study can be readily extended in two ways. The first is to simulate a wider range of severity of incidents in several different freeway sections. In this way, one could establish "universally" effective feedback control rules. The mechanism for doing this has already been developed.

The second possible extension is of a more fundamental nature. Origin-destination surveys are a standard part of the effort to establish ramp metering rates for a freeway. Information obtained from such surveys includes the breakdown of on-ramp volumes by destination as a function of time. With input data of this sort, it is possible to contemplate a freeway simulation in which the traffic in each freeway segment is segregated into components by destination. The changes required in the present aggregate variable model to effect this refinement are not extensive.

A "components by destination" model would allow an improved representation of off-ramp volumes. An even more interesting application is the simulation of the effect of information signs used in surveillance systems that can detect incidents. Drivers who react to such information signs would be modeled by an exchange between the components of traffic density.

REFERENCES

1. Marino, R. Freeway Inventory Geometric Bottleneck Congestion. Freeway Operations Dept., California Division of Highways, Feb. 1970.
2. Meisel, W. S., and Collins, D. C. Repro-Modeling: An Approach to Efficient Model Utilization and Interpretation. To be published in IEEE Trans. on Systems, Man, and Cybernetics.
3. Payne, H. J. Models of Freeway Traffic and Control. Simulation Council Proc., Vol. 1: Mathematical Models of Public Systems, Ch. 6.

4. Isaksen, L. Speed Density Relationship for Los Angeles Freeways. Univ. of Southern California, unpublished, March 1971.
5. Payne, H. J., Teener, M. D., and Meisel, W. S. Ramp Control to Relieve Freeway Congestion Due to Traffic Disturbances. Technology Service Corp., Rept. 034, April 1972.
6. Payne, H. J., and Thompson, W. A. Time-of-Day Ramp Control Strategies for the Hollywood Freeway. Dept. of Electrical Engineering, Univ. of Southern California, USCEE Rept. 415, March 1972.
7. Traffic Density Chart and Discussion for Route 101 Southbound (Hollywood Freeway), PM 1.66/6.52, AM Peak Period. Freeway Operations Dept., California Division of Highways, July 31, 1968.
8. Measures of Quality of Traffic Service. HRB Spec. Rept. 130, 1972.
9. Beeson, R. M., and Meisel, W. S. The Optimization of Complex Systems With Respect to Multiple Criteria. Proc. 1971 Systems, Man, and Cybernetics Conf., Joint National Conf. on Major Systems, Oct. 1971.
10. Zadeh, L. Optimality and Non-Scalar Valued Performance Criteria. IEEE Trans. on Automobile Control, Vol. AC-8, 1963, pp. 59-60.
11. Meisel, W. S. Computer-Oriented Approaches to Pattern Recognition. Academic Press, New York, 1972, Appendix B.
12. Teener, M. D., Meisel, W. S., and Collins, D. C. A Test of an Algorithm for Continuous Piecewise Linear Approximation. Technology Service Corp., Rept. TSC-033, March 1972.
13. Meisel, W. S. Trade-Off Decisions in Multiple-Criteria Decision-Making. In Selected Proceedings: Seminar on Multiple-Criteria Decision-Making, Univ. of South Carolina Press, Oct. 1972.
14. Collins, D. C. Practical Applications of Multiple-Criteria Decision-Making. In Selected Proceedings: Seminar on Multiple-Criteria Decision-Making, Univ. of South Carolina Press, Oct. 1972.

REDUCTION IN FREEWAY CONGESTION BY USAGE OF ACCIDENT INVESTIGATION SITES

Mary Ann Pittman and Roy C. Loutzenheiser*, Purdue University

Accident investigations conducted on the freeway shoulder cause freeway congestion and delay to motorists. If the accidents are investigated off the freeway at a site concealed from motorists, congestion and delay will be reduced, and traffic flow will return to normal more rapidly. Sixteen accident investigation sites were designated along a 6-mile section of the Gulf Freeway in Houston. Eight of the sites are located on city streets adjacent to the freeway; two are located on city streets under the freeway; and the other six are on unused space within the freeway right-of-way. Houston police officers began using the sites on July 12, 1971. Data were collected for 1 year through supplementary accident report forms that each investigating officer filled out. During the first year of operation, 851 accidents were reported in the study area, and the sites were used for 339 investigations (40 percent usage). In addition, another 176 investigations were conducted at other off-freeway locations (21 percent). Benefits in terms of delay saved from usage of the investigation sites and other off-freeway locations amounted to $203,000. Construction costs were prorated, and the annual cost and the maintenance costs were estimated at $8,000. For the first year of operation, the benefit-cost ratio was 28:1. Analysis showed that the sites under the freeway had a higher usage rate than those located on city streets.

•MOVEMENT OF VEHICLES on urban freeways has become an important part of a metropolitan area. Motorists usually find uninterrupted flow and few hazards on a freeway. However, freeway incidents such as accidents or stalled vehicles cause congestion on the freeway and delay to motorists. When such an incident occurs, one or more lanes are blocked resulting in a bottleneck situation and reduction in freeway capacity. Normally, an accident causes more freeway congestion than a stall because it requires police investigation. The degree of congestion and delay caused by an accident depends on the length of time that the accident vehicles block a lane and are visible to motorists. Police usually investigate accidents on the freeway shoulder, thus extending the time period during which motorists are distracted by the accident vehicles.

POSSIBLE SOLUTION

The premise of this study is, if an accident investigation is made at a location not visible to freeway motorists, congestion and delay will be reduced and the traffic flow will return to normal more rapidly. This paper presents the more important findings of the first year of operation of a system of accident investigation sites (AIS). The sites, located on a section of the Gulf Freeway (I-45S), are concealed from freeway motorists and are used by the police to make their accident investigations.

The Texas Transportation Institute (TTI) with the cooperation and assistance of the Texas Highway Department, District 12, designed and evaluated the AIS system. The AIS study was carried out in cooperation with the Houston Police Department (HPD) and

Publication of this paper sponsored by Committee on Traffic Law Enforcement.

*Mr. Loutzenheiser was with the Texas Transportation Institute when this study was conducted.

the City of Houston. More complete details of this study have been published by the Texas Transportation Institute (1).

SOME PREVIOUS RESEARCH

This study is an outgrowth of earlier research of accidents in moving freeway lanes. In 1963, Wilshire and Keese (2) conducted a study on the effects of traffic accidents on freeway operation and the methods of accident investigation. In their conclusions they stressed the importance of clearing the freeway of all visible signs of the accident as quickly as possible. Lynch and Keese (3) evaluated the average time elapsed between the time of the accident and the time when the damaged vehicles were moved from the roadway. They recommended that studies be conducted to devise procedures for more rapid removal of accident vehicles. In 1969, Goolsby (4) recommended the designation and construction of accident investigation sites on the Gulf Freeway. His study showed that on the average a minor accident, occurring during peak periods, affects traffic flow for 41 min, and, of this, 24.5 min are spent in police investigation. When the accident investigation is conducted at a site off the freeway, the accident affects traffic flow for only 16.5 min. Goolsby (5) further determined that a minor accident blocking one lane of a six-lane facility reduces capacity by 50 percent even though the number of lanes is only reduced by 33 percent. Also, if the damaged vehicles are moved to the freeway shoulder, the main-lane capacity is still reduced by 33 percent because of the "gapers-block" phenomenon.

PILOT STUDY SYSTEM

The Gulf Freeway was selected for the study because of the research and surveillance facilities located there. The Surveillance and Control System, used by THD and TTI, consists of inbound entrance-ramp signals, two digital process control computers, and a closed-circuit television system. Designed and built in the late 1940s, the Gulf Freeway is a six-lane facility with a theoretical capacity of 6,000 vph in each direction of flow. The six main lanes are complemented by an adjacent noncontinuous frontage road, and a slip type of design is used for the ramps.

Location of Sites

Sixteen accident investigation sites were chosen along a six-mile section of the Gulf Freeway from Dowling Street to Broadway Street because of their accessibility from the freeway and concealment from freeway motorists. A site was located downstream of each exit-ramp (Fig. 1). The minimum preparation for all sites was the installation of direction signs and NO PARKING signs. Direction signs consisted of a sign(s) on the service road directing people to the site and a sign designating the site. NO PARKING signs were posted at each site to ensure available space for the investigation and accident vehicles.

The investigation sites were grouped into three types by location: on a city street, on a city street under the freeway, and on unused space within the freeway right-of-way. The first two types have the advantage of low cost, whereas the second and third types are usually more accessible. Figure 2 shows typical layouts of the investigation sites.

Because most sections of the Gulf Freeway are at-grade with the service road and city streets, many locations within the freeway right-of-way are visible to motorists. Therefore, eight sites were located on city streets adjacent to the freeway. Besides being downstream of an exit-ramp, these were on streets with light traffic flow. The only expense for preparation was $35 per site for signs.

At one freeway overpass, the crossing city streets carry a minimum of traffic flow; therefore, two accident investigation sites were located on these streets under the freeway. Available space under the overpass could have been used; however, to reduce costs, the city streets were selected. The necessary costs were $35 per site for installation of signs.

Design of Sites Requiring Construction

A typical accident investigation involves five vehicles: one police car, two damaged vehicles, and two wreckers. If it is assumed that each vehicle requires a 10- by 20-ft space to park, a typical site should contain at least 1,000 ft^2 of space. The six constructed sites have a surfaced area of approximately 30 by 85 ft, or 2,250 ft^2. The extra area provides a lane for driving.

One of the constructed sites is located in an open area off a city street. The ground, near a preexisting luminaire, was graded and paved. This construction amounted to $3,200, and an additional $35 was spent on installation of signs.

The five sites constructed under the freeway were also graded and paved, and guardrails were placed between the pavement and the bridge supports for protection. To discourage local use of the sites, the access road between the service road and the site did not provide smooth curves for turning into the sites. All of the construction work amounted to about $3,200 per site. In addition to direction and NO PARKING signs, it was necessary to add two clearance signs. NO THRU TRAFFIC signs were also installed to discourage motorists from using the sites for U-turns. Cost of the various signs amounted to $115 per site. Because existing street lighting did not provide sufficient illumination, additional lighting was mounted under the overpasses. Installation of the lighting increased the construction costs at each site by about $2,800.

Of the 16 investigation sites located on the 6-mile section of the Gulf Freeway, four sites are accessible from either the inbound or outbound direction, six sites are accessible to inbound traffic only, and six sites are accessible to outbound traffic only. Therefore, a site is located an average of every 0.6 mile for either the inbound or outbound direction. Of the six sites requiring extra construction, four sites are accessible from both directions, whereas the other two sites are accessible from one direction only.

Study Procedures

HPD officers began using the sites on July 12, 1971. Prior to this date, booklets identifying the location of the investigation sites were distributed to the police officers. At that time, they were also given supplementary freeway accident report forms to be filled out at each freeway accident. To provide a basis for the total city, officers investigating accidents on all freeways in Houston were requested to fill out the forms; therefore, freeway accidents were reported 24 hours a day, 7 days a week. These forms were revised in mid-August after representatives of TTI, THD, and HPD decided that the information provided on the original forms was confusing about location of the accident and location of the investigation. By mid-September, the revised forms were being used by a majority of the officers. Each investigating officer was requested to include the following information on the forms: date, time, location of accident, location of investigation, why investigation site not used, length of investigation, and officer's name.

DATA ANALYSIS

Analysis of the accident investigation sites included four major areas: usage rate, benefit-cost ratio, impact on freeway operation, and evaluation of individual sites. The usage rate was evaluated according to time of day, month, and direction of travel. Estimated delay time saved was used to determine benefits of the system. Other benefits derived from the added safety and convenience of the sites were discussed, but a monetary value was not estimated. In addition to the decrease in time during which capacity was reduced on the freeway, the impact of accident experience before and during the study was analyzed. Analysis of individual sites provided information on establishing additional criteria for an AIS system.

Use of AIS

During the first year of operation, 851 police report forms were received. In 61 percent of these, the officer indicated that he had used an AIS or some other location off the

Figure 1. Locations of investigation sites on Gulf Freeway.

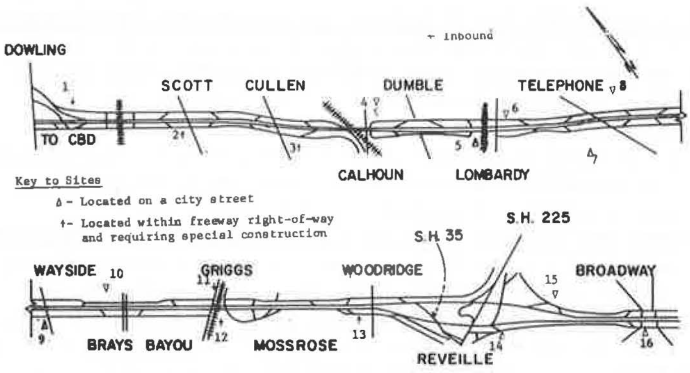


Figure 2. Typical layouts of investigation sites.

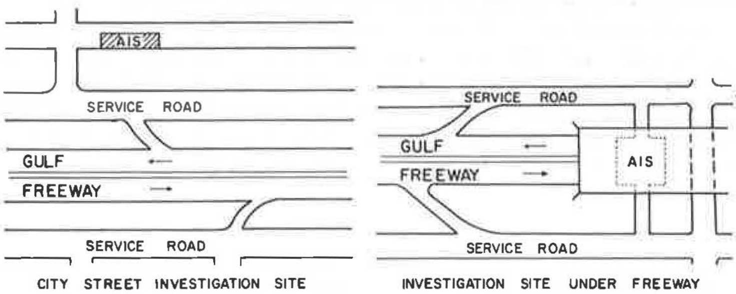


Table 1. Frequency of AIS usage.

Information	Number	Percent
Police report forms received	851	100
Use of AIS	339	40
Use of other off-freeway sites	176	21
Investigation on shoulder	336	39

Table 2. AIS usage during peak and off-peak periods.

	Peak Periods (weekday)		Off-Peak Periods	
Item	6 to 9 a.m.	3 to 6 p.m.	Daylight[a]	Nighttime[b]
No. of Accidents	152	186	321	192
No. of investigations at AIS (percentage)	75 (45)	78 (42)	132 (41)	54 (28)
No. of investigations at other off-freeway sites (percentage)	23 (15)	39 (21)	70 (22)	44 (23)
Percentage of AIS and other off-freeway site usage	64	63	63	51

[a]From 9 a.m. to 3 p.m. on weekdays and from 6 a.m. to 6 p.m. on weekends.
[b]From 6 p.m. to 6 a.m. daily.

freeway to conduct the investigation. These off-freeway locations included service roads, city streets, or parking lots. Table 1 gives the frequency of the site usage.

Data given in Table 2 compare the frequency of usage for the peak and off-peak travel periods. The morning and evening peak-period usage for the AIS was 45 percent. The usage rate for the daylight off-peak was 41 percent and for the nighttime 28 percent. One apparent reason for the lower usage rate at night is that the lighter traffic flow does not produce congestion.

The monthly usage rates of the AIS showed a general increasing trend. Except for the first 2 weeks, the usage rate increased from 27 percent to about 50 percent. A 48 percent usage rate during the first 2 weeks was probably due to the initial efforts of starting the study. The combined usage rates of the AIS and other off-freeway sites varied between 53 and 74 percent, with no increasing trends observed.

The frequency of AIS usage related to direction of travel was similar. The investigation sites on the inbound side of the freeway were used 44 percent of the time, and on the outbound side they were used 43 percent. Such usage was expected because 10 sites were accessible to inbound traffic and 10 sites to outbound traffic.

A total of 115 officers reported accidents in the study area during the first year of operation. The usage rate for a police accident investigator was obtained by dividing the number of times investigation sites were used by the number of accidents investigated. Twenty-eight officers investigated only one accident, and their usage rate (18 percent) was much lower than that of other accident investigators (40 percent).

Comments From Officers

To obtain first-hand opinions on the value of the AIS system, 18 Houston police officers were interviewed in June. Each officer had investigated more than 10 accidents in the study area during the previous year, and their usage rates varied from 14 to 68 percent. Most of the officers agreed that the AIS system improved traffic operations during an accident investigation. When queried on the conditions under which they would not move the accident vehicle off the freeway, they cited the following situations: when a fatality or possible fatality has occurred, when a crime has been committed, or when photographs or measurements are needed at the scene. Several of the officers said that they hesitate to move the vehicles when too many cars are involved and when an accident site is some distance away. Because the AIS system is a new concept, one officer stated that sometimes he forgot the investigation sites were available.

One of the problems encountered by the officers was that they had to explain to the motorists how to get to a site. Also, motorists were not aware that they could move their vehicles off the freeway before the police arrived. During the last quarter of the study year, wrecker drivers were instructed by the police department to move noninjury accident vehicles to a site as soon as possible. Several officers pointed out that this procedure caused problems if the wrecker driver failed to report where he had relocated the vehicles.

The officers agreed that using a site made their jobs easier because of the more relaxed atmosphere there. The sites provided a place concealed from freeway motorists and with reduced noise levels. Under-freeway sites provided an added convenience of sheltering police and motorists from inclement weather. In general, the theme that the officers related in the interviews was to inform the motorists of the locations and purpose of the sites. Most officers preferred using the under-freeway sites because they are more accessible. Placing some type of communication system at the sites was suggested by a majority of the officers.

Benefit Analysis

The anticipated benefits of the AIS system were improvement in safety and convenience, reduction in delay time, and reduction of secondary accidents. Benefits derived from the safety and convenience that the investigation sites provide were difficult to evaluate quantitatively. Eliminating the 25 min for the actual investigation on the freeway results in only 16 min during which traffic flow would be affected. Thus freeway operation is restored to normal more rapidly, making it possible for emergency and other vehicles to reach their destination more quickly.

Use of the accident investigation sites also decreases delay to freeway motorists inasmuch as "gapers-block" or "rubbernecking" is eliminated after the vehicles are removed from the freeway. Usage of the sites also reduces the hazards to persons involved in an accident investigation.

Reduction in Delay—Use of the AIS system and other off-freeway locations reduced the number of vehicle-hours of delay significantly. Time-delay graphs were developed to estimate the total hours of delay saved during the first year. Initially, time-flow graphs were used to develop the time-delay relationships.

To provide a conservative estimate, we made the following assumptions: all accidents blocked only one lane, accident vehicles were moved from the freeway in 15 min, and no injuries were incurred by occupants of the accident vehicles. The time-flow graph shown in Figure 3 illustrates the effects of such an accident occurring at 7:00 a.m. on the inbound Gulf Freeway at Telephone Road. The demand curve was based on normal operational data, and the reduced volume curves (5) were plotted by using the following three-lane flow rates: accident vehicles on freeway, 2,750 vph; accident vehicles on freeway shoulder, 4,030 vph; and service volume during normal peak, 5,560 vph. The area between the demand and service volume curve is the delay in vehicle-hours that motorists will experience. The 15 min of freeway blockage produced a fixed delay of 690 vehicle-hours. Additional delay is a function of the investigation procedure, of which three cases are presented.

In case 1, it was assumed that the accident vehicles were moved to an AIS or other off-freeway site. Thus, no additional delay occurred, and freeway operation was normal by 8:15. For case 2, the investigation was conducted on the freeway shoulder and required 20 min. This procedure caused a total delay of 1,470 vehicle-hours. A 40-min investigation on the shoulder (case 3) produced 2,170 vehicle-hours of delay. Similar graphs were drawn for hypothetical accidents occurring at various times during the day at three additional locations. Because of the light flow rates, delay times between 7:00 p.m. and 6:00 a.m. were nearly zero.

Time-delay graphs consisting of three curves of delay versus the time of day were plotted for accidents occurring near the four locations. Only the 13-hour period from 6:00 a.m. to 7:00 p.m. was summarized on each graph. Figure 4 shows the time-delay graph for accidents occurring at the Telephone Road overpass. For example, if an accident occurred on the inbound freeway over Telephone Road at 7:30 a.m., the amount of delay to freeway motorists is 460 vehicle-hours if the investigation is conducted off the freeway. If the investigation is conducted on the freeway shoulder and takes 20 min, the amount of delay is 1,000 vehicle-hours. Therefore, 540 vehicle-hours of delay are saved by moving the vehicles off the freeway. Similarly, a 40-min investigation on the freeway causes 1,480 vehicle-hours of delay. The delay saved in this instance would be 1,020 vehicle-hours if the investigation is conducted at an off-freeway site.

There was no significant difference in delay for the three cases during the daylight off-peak periods (9 a.m. to 3 p.m.) because traffic demand usually did not exceed the reduced capacity caused by an accident investigation on the shoulder. Thus, for this study, delay time saved was computed for accidents occurring during the peak periods only. From September 13, 1971, to July 9, 1972, the estimated delay time saved by the 93 uses of the investigation sites was 29,250 vehicle-hours. An additional 8,100 vehicle-hours were saved by investigations conducted at other off-freeway locations. Data prior to mid-September were not included in the analysis because the information on the original forms was insufficient for this analysis.

In 1969, researchers (6) determined that one vehicle-hour of travel on the Gulf Freeway was worth \$2.92. If we assume a compounded increase of 5 percent per year and increased occupancy from 1.0 to 1.2 persons per passenger vehicle, the value of one vehicle-hour in 1972 would be \$4.50. By using this updated value, the monetary savings can be calculated. The total delay saved for the 43-week period was 37,350 vehicle-hours, which represents an annual savings of \$203,000.

Reduction in Accidents—Restoring freeway operations more rapidly also aids in the reduction of secondary accidents that occur as a result of shock waves. Data for the analysis of secondary accidents were obtained from records in the surveillance office television room during peak periods. During the year prior to the AIS system 15 of 212 accidents were classified as secondary, whereas with the use of the AIS the secondary accidents decreased to 8 of 179 accidents. Thus, the total number of peak-period accidents decreased by 33, and the number of secondary accidents decreased by 7. Secondary accidents, therefore, represented 21 percent of the reduction in peak-period accidents.

Data obtained from the City of Houston showed that, on a 24-hour basis, 1,046 accidents occurred in the study area during the year prior to the study. Since the AIS system was installed, there were 851 accidents, a reduction of 195 accidents. If it is assumed that the probability of occurrence of a secondary accident is the same for peak periods and off-peak periods, then about 41 secondary accidents were prevented (that is, 21 percent of 195 accidents).

Burke (7) in 1970 determined the costs for various types of accidents. By assuming a 5 percent per year compounded increase, the cost involved for a property damage accident in 1972 would be $307 per vehicle. It was further assumed that all secondary accidents involved only two cars; therefore, the annual savings due to a reduction of 41 secondary accidents was approximately $25,000.

Comparison of Benefits and Costs

The construction cost for the AIS system was determined as follows: 10 sites at $35 each, one site at $3,235, and five sites at $6,115 per site. Total construction costs for all sites amounted to approximately $34,200. Maintenance for the AIS system was minor for the first year. No cost figures were available, so a very conservative estimate of $200 per month was made. An estimate of maintenance costs for the first year was, therefore, $2,400.

To determine the annual cost of the AIS system, we multiplied the initial construction costs by a uniform series capital-recovery factor and added the sum to the annual maintenance costs. The capital-recovery factor was based on a conservative interest rate of 10 percent for only 10 years. The annual cost was about $8,000, whereas the benefits of the system due to delay saved and reduction in secondary accidents totaled $228,000. Thus,

$$\text{Benefit/cost} = \frac{\$228{,}000}{8{,}000} = 28.5$$

Evaluation of Individual Sites

An analysis of the usage rate for each site was made. This usage rate was obtained by dividing the number of times a site was used, obtained from the supplementary police forms, by the number of accidents that occurred near it, determined by subjective analysis. No accident was considered for more than one site, and, when there was a question of which was the nearest site, the accident was omitted from analysis.

The sites located under the freeway, including the two on city streets, had a combined usage rate of 53 percent, whereas the usage rate for the sites located on city streets was 35 percent. The rates at individual sites varied from 12 to 64 percent. Of the seven sites that had usage rates greater than 50 percent, only two are on city streets. These two are the only city-street sites immediately downstream of an exit ramp. To reach the other city-street sites, motorists must drive farther. Thus, there is a definite trend to use sites that are located under the freeway or directly adjacent to it.

An analysis of the nighttime usage of the accident investigation sites was made to determine whether the sites were being used at night and whether the additional cost for lighting was justified at the five sites. Unfortunately, the number of accidents near each site was too small in most cases to provide a valid analysis. Most sites had a decrease in the usage rate at night. The nighttime usage rate for sites under freeway overpasses, where lighting was installed, was 41 percent as compared to 52 percent for 24 hours. For the other sites, the usage rate decreased from 39 percent to 22 percent.

Figure 3. Time-flow relationship for a one-lane blocked, noninjury accident over Telephone Road inbound.

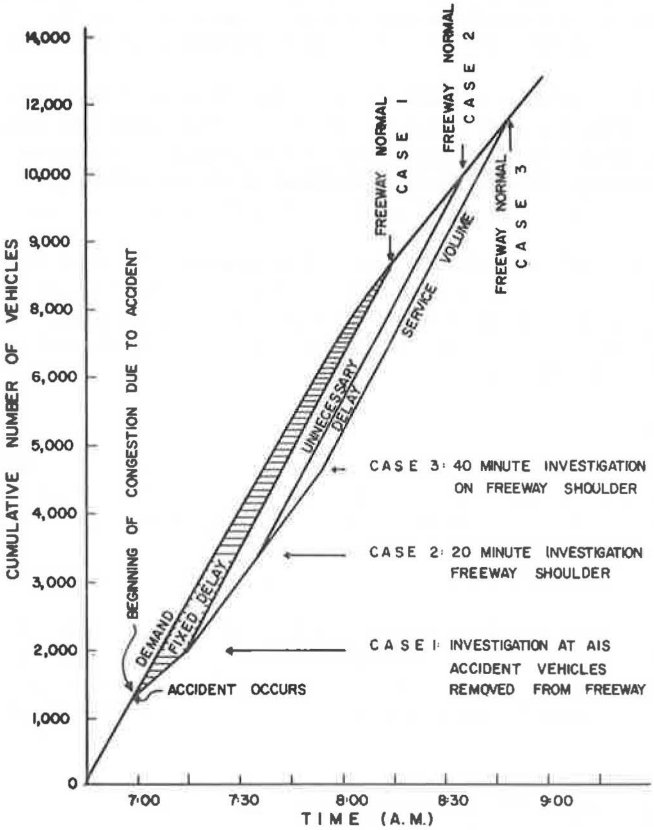


Figure 4. Time-delay relationship for a one-lane blocked, noninjury accident over Telephone Road inbound.

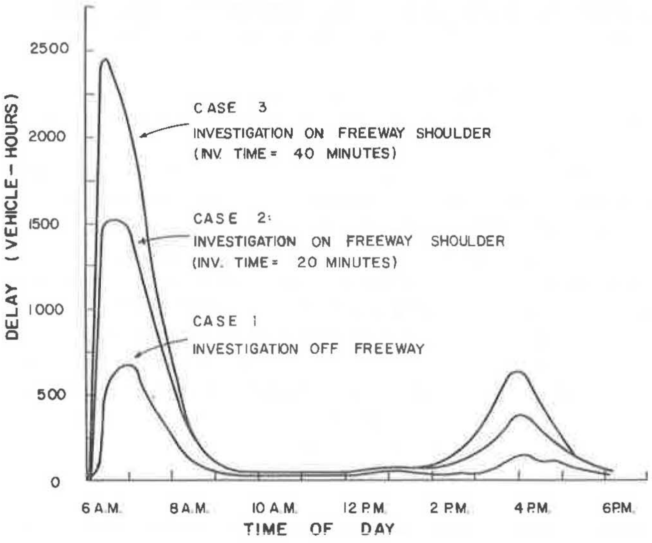
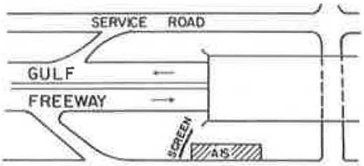


Figure 5. Modification of accident investigation site.



Discussion of Analyses

After a year of experience, the AIS system on the Gulf Freeway has proved satisfactory, based on design and location of sites. The basic design of the sites on unused freeway rights-of-way was sufficient; however, the use of the site as a U-turn roadway continued to be a minor problem. A low curb at the entrance to the site could be used to discourage improper use. Location of the entrance and exit of the site directly opposite a driveway or street is undesirable. The sites located on city streets should be at least 30 ft wide to allow traffic to pass the site in both directions during an investigation. A street narrower than 30 ft should have NO PARKING signs on both sides of the street.

The installation of lights at a site may not be justified based on the added cost. The purpose of the lighting is to illuminate the area and not to provide light for completing the investigation forms (officers use flashlights). Therefore, additional lighting should be limited to sites that have a high usage rate and no city lights.

The most used sites were those under the freeway overpasses. Several sites were located at places with low accident rates and may be unnecessary. However, the cost of installation was low, and other sites were difficult to reach. In general, a site should be located so that it is accessible from the freeway and easy to find. Locating the site so that it is out of view of freeway motorists should take secondary consideration inasmuch as screens (metal or foliage) could be installed. Where possible, sites should be constructed adjacent to the service road as shown in Figure 5.

SUMMARY

The usage rate for the accident investigation sites on the Gulf Freeway was 40 percent during the first year of operation. Although this was lower than was anticipated, it is felt that the program has been a success. The AIS system is a new concept for handling accidents, and, therefore, it should be expected that, through an educational and managerial process, the usage rate will increase. That is to say, as policemen and motorists become more familiar with the purposes and benefits of the AIS, the usage rate will increase. Expansion of the AIS system to all freeways in Houston is being proposed.

In addition to the use of the AIS, another 21 percent of the accident vehicles on the Gulf Freeway were moved to locations off the freeway. Analysis showed that normal delay, encountered by freeway motorists driving past an accident investigation on the shoulder, was eliminated when the accident investigation was conducted at the investigation sites or other off-freeway sites. The benefits derived from usage of the investigation sites or other off-freeway sites were valued at $228,000, whereas the cost of installation and maintenance of the AIS was less than $8,000. Therefore, the benefits of accident removal exceeded installation costs by a ratio of 28:1.

The initial design of sites proved to be satisfactory. The following criteria have been established for an acceptable accident investigation site: easily accessible, well-marked, concealed from freeway motorists, located near high-accident areas, low construction costs, at least 1,000 ft^2 of space, and sufficient lighting. These criteria should be considered in establishing an AIS system on other freeways.

REFERENCES

1. Pittman, M. A., and Loutzenheiser, R. C. A Study of Accident Investigation Sites on the Gulf Freeway. Texas Transportation Institute, Research Rept. 165-1, 1972.
2. Wilshire, R. L., and Keese, C. J. Effects of Traffic Accidents on Freeway Operation. Texas Transportation Institute, Bull. 22.
3. Lynch, F. L., and Keese, C. J. Restoring Freeway Operation After Traffic Accidents. Texas Transportation Institute, Bull. 28.
4. Goolsby, M. E. Accident Reporting and Clearance Procedures on the Gulf Freeway. Texas Transportation Institute, Research Rept. 139-1, 1969.
5. Goolsby, M. E. Influence of Incidents on Freeway Quality of Service. Highway Research Record 349, 1971, pp. 41-46.

6. McFarland, W. J., Adkins, W. G., and McCasland, W. R. Evaluation of the Benefits of Traffic Surveillance and Control on the Gulf Freeway. Texas Transportation Institute, Research Rept. 24-22, 1969.
7. Burke, D. Highway Accident Costs and Rates in Texas. Texas Transportation Institute, Research Rept. 144-1F, 1970.

SPONSORSHIP OF THIS RECORD

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

Committee on Traffic Law Enforcement
Gregory L. Goodson, Arizona Highway Patrol, chairman
Russell Arend, Gregory E. Austin, Gil W. Bellamy, Walter E. Boles, Carlie Bowmer, Dale Carson, Andrew E. Colclough, Thompson S. Crockett, Olin K. Dart, Jr., Edmund F. Fennessy, Jr., William H. Franey, John O. Hidinger, Adam G. Johnson, Edward R. Klamm, Leslie H. Leach, Keith L. McRoberts, Martin M. Puncke, Kenneth J. Tharp

Committee on Freeway Operations
Robert S. Foote, Port Authority of New York and New Jersey, chairman
Patrick J. Athol, John H. Auer, Jr., John L. Barker, Roye G. Burnfield, E. N. Burns, J. R. Doughty, Conrad L. Dudek, Robert E. Dunn, A. C. Estep, Paul H. Fowler, Anthony J. Freed, John J. Haynes, Joseph W. Hess, Gerhart F. King, Adolf D. May, Jr., Stuart F. Millendorf, William J. Miller, Jr., Donald E. Orne, John G. Sample, Gerald W. Skiles, T. Darcy Sullivan, A. Taragin, Joseph Treiterer, Joseph A. Wattleworth, Robert H. Whitson, Wolfgang Woischnig, Richard D. Worrall, H. Nathan Yagoda

K. B. Johns and James K. Williams, Highway 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, 1972.