# HIGHWAY RESEARCH RECORD

**Number 154**

# Traffic Flow Characteristics

4 Reports

Subject Classification

54 Traffic Flow
55 Traffic Measurements

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

Washington, D.C., 1967 Publication 1430

## *Department of Traffic and Operations*

Harold L. Michael, Chairman
Associate Director, Joint Highway Research Project, Purdue University
Lafayette, Indiana

HIGHWAY RESEARCH BOARD STAFF

E. A. Muller, Engineer of Traffic and Operations

COMMITTEE ON THEORY OF TRAFFIC FLOW
(As of December 31, 1965)

Donald E. Cleveland, Chairman
Department of Civil Engineering, University of Virginia
Charlottesville

Donald G. Capelle, Secretary
Traffic Research Engineer, Automotive Safety Foundation
Washington, D.C.

John L. Barker, Vice President, Automatic Signal Division, East Norwalk, Connecticut
Martin J. Beckmann, Department of Economics, Brown University, Providence, Rhode Island
Jack B. Blackburn, Head, Department of Civil Engineering, Kansas State University, Manhattan
A. Charnes, Technological Institute, Northwestern University, Evanston, Illinois
Donald R. Drew, Texas Transportation Institute, Texas A & M University, College Station
Leslie C. Edie, The Port of New York Authority, New York, New York
Robert S. Foote, The Port of New York Authority, New York, New York
Theodore W. Forbes, Department of Psychology and Engineering Research, Michigan State University, East Lansing
Antranig V. Gafarian, Systems Development Corporation, Santa Monica, California
Herbert P. Galliher, Editor, International Abstracts in Operations Research, Department of Industrial Engineering, University of Michigan, Ann Arbor
Denos C. Gazis, International Business Machines Corporation, Thomas J. Watson Research Center, Yorktown Heights, New York
Daniel L. Gerlough, Head, Traffic Systems Section, Planning Research Corporation, Los Angeles, California
Bruce D. Greenshields, Assistant Director, Transportation Institute, University of Michigan, Ann Arbor
Frank A. Haight, Associate Research Mathematician, Institute of Transportation and Traffic Engineering, University of California, Los Angeles
Robert Herman, Theoretical Physics Department, General Motors Research Laboratories, Warren, Michigan
A. W. Jones, Institute for Defense Analyses, Arlington, Virginia
James H. Kell, Consulting Traffic Engineer, Traffic Research Corporation, San Francisco, California
Sheldon L. Levy, Director, Mathematics and Physics Division, Midwest Research Institute, Kansas City, Missouri
Russell M. Lewis, Department of Civil Engineering, Rensselaer Polytechnic Institute, Troy, New York
Mo Chih Li, Department of Civil Engineering, New York University, New York, New York
E. W. Montroll, Documents Control, Institute for Defense Analyses, Arlington, Virginia
Joseph C. Oppenlander, Assistant Professor of Highway Engineering, School of Civil Engineering, Purdue University, Lafayette, Indiana
Richard W. Rothery, General Motors Research Laboratories, Warren, Michigan

Morton Schneider, Tri-State Transportation Committee, New York, New York
Martin C. Stark, Operations Research Analyst, Information Technology Division, National Bureau of Standards, Washington, D.C.
Asriel Taragin, Assistant Deputy Director for Research, Office of Research and Development, U.S. Bureau of Public Roads, Washington, D.C.
Joseph Treiterer, Research Supervisor, Transportation Engineering Center, The Ohio State University, Engineering Experiment Station, Columbus
William P. Walker, Chief, Geometric Standards Branch, U.S. Bureau of Public Roads, Washington, D.C.
Joseph Wattleworth, Texas Transportation Institute, Texas A & M University, College Station
George Weiss, National Cancer Institute, National Institutes of Health, Bethesda, Maryland

## COMMITTEE ON FREEWAY OPERATIONS

(As of December 31, 1965)

Alger F. Malo, Chairman
Director, Department of Streets and Traffic, City of Detroit
Detroit, Michigan

Adolf D. May, Jr., Secretary
Institute of Transportation & Traffic Engineering, University of California
Richmond

John H. Auer, Jr., Principal Research Engineer, General Railway Signal Company, Rochester, New York
John L. Barker, Vice President, Automatic Signal Division, East Norwalk, Connecticut
Louis E. Bender, Chief, Traffic Engineering Division, The Port of New York Authority, New York, New York
Donald S. Berry, Chairman, Department of Civil Engineering, The Technological Institute, Northwestern University, Evanston, Illinois
Harold Cooper, Director, Traffic Division, Michigan State Highway Department, Lansing
Donald O. Covault, Professor of Civil Engineering, Georgia Institute of Technology, Atlanta
James L. Foley, Jr., Commissioner of Transit and Traffic, Baltimore, Maryland
Robert S. Foote, Supervisor, Tunnel and Bridge Research, The Port of New York Authority, New York, New York
Harry Forster, Airborne Instruments Laboratory, Deer Park, Long Island, New York
Edward F. Gervais, Project Director, National Proving Ground for Freeway Surveillance, Control and Electronic Traffic Aids, Detroit, Michigan
Stuart F. Millendorf, Senior Transportation Planner, The Port of New York Authority, New York, New York
William J. Miller, Jr., Director, Delaware River and Bay Authority, New Castle, Delaware
Charles Pinnell, Head, Design and Traffic Department, Texas Transportation Institute, Texas A & M University, College Station
Carlton C. Robinson, Director, Traffic Engineering Division, Automotive Safety Foundation, Washington, D.C.
Harold W. Sullivan, City of Los Angeles Police Department, Los Angeles, California
Asriel Taragin, Assistant Deputy Director for Research, Office of Research and Development, U.S. Bureau of Public Roads, Washington, D.C.
F. C. Tarbox, Traffic Engineer, Ohio Department of Highways, Columbus
James R. Turner, Highway Planning Engineer, Missouri State Highway Commission, Jefferson City

John L. Vardon, Director, Transportation Division, Metropolitan Toronto Planning Board, Toronto, Ontario, Canada
James E. Wilson, Traffic Engineer, California Division of Highways, Sacramento

COMMITTEE ON CHARACTERISTICS OF TRAFFIC FLOW
(As of December 31, 1965)

Joseph C. Oppenlander, Chairman
Assistant Professor of Highway Engineering, School of Civil Engineering, Purdue University, Lafayette, Indiana

Patrick J. Athol, Assistant Director, Illinois Expressway Surveillance Project, Oak Park
C. E. Billion, San Diego, California
Jack B. Blackburn, Head, Department of Civil Engineering, Kansas State University, Manhattan
Martin J. Bouman, Transportation and Traffic Engineer, Balboa Park, San Diego, California
Paul D. Cribbins, Assistant Professor of Civil Engineering, North Carolina State College, Raleigh
Kenneth W. Crowley, Senior Research Analyst, Tunnels and Bridge Department, The Port of New York Authority, New York, New York
J. E. P. Darrell, Minnesota State Automobile Association, Minneapolis
Robert F. Dawson, Civil Engineering Department, University of Connecticut, Storrs
Donald R. Drew, Texas Transportation Institute, Texas A & M University, College Station
H. M. Edwards, Associate Professor of Civil Engineering, Queen's University, Kingston, Ontario, Canada
Wayne T. Gruen, Research Assistant, University of Illinois, Urbana
Roy C. Haeusler, Automotive Safety Engineer, Engineering Division, Chrysler Corporation, Detroit, Michigan
J. E. Johnston, Deputy Director of Highways, Planning & Traffic, Utah State Department of Highways, Salt Lake City
James H. Kell, Consulting Traffic Engineer, Traffic Research Corporation, San Francisco, California
Russell M. Lewis, Department of Civil Engineering, Rensselaer Polytechnic Institute, Troy, New York
Jack C. Marcellis, Traffic Plans Engineer, Nashville Traffic Commission, Nashville, Tennessee
Peter A. Mayer, Assistant Traffic Civil Engineer, Arizona Highway Department, Phoenix
O. J. Reichelderfer, Traffic Engineer, Pennsylvania Department of Highways, Harrisburg
Curtis L. Shufflebarger, Industrial College of the Armed Forces, Fort McNair, Washington, D. C.
Robert P. Shumate, Director, Public Services Group, Indiana University, Bloomington
David Solomon, Principal Research Engineer, Traffic Systems Division, U.S. Bureau of Public Roads, Washington, D.C.
K. A. Stonex, Automotive Safety Engineer, Technical Liaison Section, Engineering Staff, General Motors Corporation, GM Technical Center, Warren, Michigan
William C. Taylor, Traffic Research Engineer, Ohio Department of Highways, Columbus
Kenneth J. Tharp, Department of Civil Engineering, State University of New York at Buffalo, Buffalo, New York
Arthur G. Wake, Director of Traffic Engineering, Bureau of Traffic Engineering, Indianapolis, Indiana
Joseph F. White, Manager, Advanced Methods and Test Design Department, Testing Operations Office, Ford Motor Company, Dearborn, Michigan

# Foreword

It has long been realized that increasing the traffic flow on existing facilities will obviate the need for some new transportation facilities. While this concept cannot and should not be carried to the conclusion of doing nothing about the construction of new facilities, it implies that a great deal more can and should be done to expedite traffic flow on existing facilities.

The first step in reaching this goal is to learn more about the traffic flow characteristics. For example, many of the existing design theories are based on experience and assumptions regarding drivers, vehicles, and roads. Were the research available to substantiate these experiences and assumptions, an increase in traffic carrying capability would be assured. Especially in urban areas, the wise use of public funds requires that highway designers and operators have an accurate and detailed body of knowledge on which to base the expenditure of funds.

The four papers in this RECORD do not represent any particular break-throughs in the field, but they do contribute to the general body of knowledge that is being slowly acquired. The first paper reports on experiments in existing tunnels to control traffic in order to maximize flow. Using data on behavior of traffic in tunnels under high density conditions, the author describes control logics that tend to alleviate undesirable conditions. Electronic equipment is suggested that would smooth out flow.

The second paper is also concerned with the development of density controls for improved operations. The authors have studied traffic through a single-lane road system using the parameters of input flow, section density, output speed, and output flow. They describe a preliminary automatic system for controlling section density by limiting input flow based on measuring output speeds.

The next paper indicates how magnetic loop detectors can be used to facilitate data collection in the case of a computerized system for control of a freeway diamond interchange. It is pointed out that the actual interchange performance can be monitored by the data provided by the detectors measuring count, vehicle size and velocity. The accuracy levels experienced and the limitations of a loop detector system are also discussed.

The last paper compares several hypotheses that have been set forth in recent years concerning the relationships among basic flow characteristics. Using data collected on a midwest freeway, these hypotheses are evaluated. From the standpoint of practical application, all of the presently used hypotheses except one performed well enough to warrant continued use.

This RECORD will be of chief interest to researchers engaged in statistical and mathematical applications of traffic flow characteristics and practitioners in the field of traffic flow theory. It will also be of interest to both operators and researchers in agencies having to do with better flow in tunnels, restricted roadways, and high-density traffic ways.

# Contents

# Control of Traffic in Tunnels to Maximize Flow

LUCIEN DUCKSTEIN, Professor of Systems Engineering, University of Arizona, and Temporary Research Analyst, The Port of New York Authority

In 1961 a preliminary control system based on speed and flow measurements was implemented in the Holland Tunnel South Tube. A red light at the entrance portal introduced gaps into the flow in order to avoid stop-and-go waves and attempt to increase the flows.

This paper presents an inexpensive way of improving the existing Holland Tunnel System; experimental evidence is used to support the proposed change. It also presents several control logics, based on speeds and densities, which can be implemented by either use of a fixed program analog device or of a general analog computer.

A model is derived for general-purpose control of single-lane flow based on finite-state machine theory. Implementation of this system, which is more flexible than the previous one, requires a digital computer.

In the Appendix is a method of estimating section density using one or more point velocities, and the input flow is given and an approximate experimental verification is made. This method can be used as a monitor for a density measuring device based on car-count, which drifts whenever cars change lane or a count is missed.

•THIS STUDY is linked to the material previously presented by Foote (1) and a paper, also presented in this Highway Research Record, by Foote and Crowley (2).

Since 1958 The Port of New York Authority has been investigating whether peak flow through the five tubes connecting New York City and New Jersey could be increased and how the increase could be accomplished. Two of these tubes form the Holland Tunnel, three of them the Lincoln Tunnel. Each tube has two lanes; however, lane changing is normally prohibited so that traffic can be considered as single lane.

A series of experiments in the south tube of the Holland Tunnel showed that peak flow of traffic could be increased by 5 percent if gaps were manually introduced in the stream. This 5 percent increase resulted in a 33 percent decrease of congestion on the approach facilities. An experimental control system was built with fixed-logic elements so that gaps would be introduced automatically in the stream according to a predetermined program. The control box activated a red light at the entrance of the Holland Tunnel each time that certain combinations of local speeds inside the tunnel and flow rate at the entrance of the tunnel were measured. The system had a tendency to overcontrol the traffic; large fluctuations of flow characteristics were also observed. A high density period was followed by a series of red lights which caused a lower density period.

The present study examines how this experimental control system can be improved with a minimum amount of hardware. Other control logics based on density rate and speed are given. They can be implemented by fixed-program systems which are built at a reasonable cost; however, the flexibility of fixed-program control boxes is limited. Control diagrams which can be implemented with analog elements are also given. In addition, this investigation deals with a general finite-state model of a tunnel control system for digital computer control.

---

Paper sponsored by Committee on Theory of Traffic Flow and presented at the 45th Annual Meeting.

A discussion of various density measurement schemes is given in the Appendix. The data used throughout this study were taken in the Holland Tunnel south tube by methods described by Foote and Crowley (2). The Holland Tunnel can be divided into three zones: a 3,100-ft long downgrade, a 3,400-ft long level portion, and a 1,600-ft long upgrade. Point 0 is at the entrance of the tunnel, near the toll booths. The distances between data points 0, 1, 2, 3, 4, 5 and 6 are, respectively, 1,800, 1,300, 1,900, 1,000, 500 and 1,350 ft. Point 5 is at the foot of the upgrade. A bottleneck seems to occur between points 4 and 5. This is why data taken at point 4 are given particular attention.

## NOTATION

A, B, C = categories of calculated densities, constants;
$f_j$ = frequency of a given event in jth section of tunnel;
$h_j$, $H_j$ = threshold values of density in jth section of tunnel;
i = running index for discrete instants of time $t_i$, i = 0, 1, 2, ...,;
j = running index for tunnel section, j = 1, 2, ..., n;
k = density, vehicles per mile;
$k_j$ (m) = density in section j at time m;
$k_{jj'}$ (m) = density between points j and j' at time m;
$L_j$ = length of jth section of tunnel, feet;
m = an integer number of time units;
$M_j$ = data point in jth section of tunnel;
n = number of tunnel sections;
$N_j$ = number of cars passing point j per length of time;
O = origin and entrance portal of tunnel;
Q = flow, vehicles per hour;
$\bar{Q}_j$ (m) = average flow at point $M_j$ at time m;
$t_i$ = a time instant or ith sampling time;
V = speed;
$\bar{V}_j$ (m) = average speed at point $M_j$ at time m;
W, w = threshold values of V;
x = linear distance in tunnel;
HT = Holland Tunnel;
LT = Lincoln Tunnel; and
ST NL = south tube near lane.

## HOLLAND TUNNEL SYSTEM IMPROVEMENT WITH MINIMUM HARDWARE

To improve the control system tested in the Holland Tunnel south tube in 1961-64 (3), one must first examine why the density and velocity exhibit a cyclic behavior. At and after time m the system senses whether or not the following quantities are above a certain threshold:

1. The maximum number $N_4$ of vehicles whose speed was less than 20 mph at point 4 between time m - 1 and m;
2. The average flow $\bar{Q}_4$ at point 4 between time m - 1 and m;

3. The maximum number $N_x$ of vehicles whose speed was less than 20 mph at any point x between 0 and 6, excluding 4 during the time interval (m - 1, m) and

4. The number $N_0$ of cars which have been entering the tunnel since time m, where m = 1, 2, 3, ... minutes.

Using these data, a matrix determines at what time $t_i$, $m \leq t_i \leq m + 1$, a red light is activated at the entrance of the tunnel; at time m + 1, the light is deactivated and the system is reset. Unfortunately, point 4 is 6,000 ft away from the tunnel entrance (point 0), so a car traveling 40 ft/sec takes 150 sec to cover distance 04. Thus, the control actuated at time m for a group of such cars would have been applied at time m - 2.5, say m - 2 to have a round figure. At time m, the control may or may not be beneficial. In fact, it is possible to conceive several extreme cases where the control will act exactly opposite to the way it should (Fig. 1). Due to traffic characteristics, the bad and beneficial effects of the control system have apparently compensated each other, so that no effect on the average tunnel output has been felt. Another argument in favor of this hypothesis is illustrated in Figure 2.

$\bar{V}_4$ (t) and $\frac{2000}{\bar{V}_1 (t - 2)}$ (where 2000 is a scale factor) vary the same way; i.e., the speed at point 1 at time and the speed at point 4, two units of time later, are in opposition of

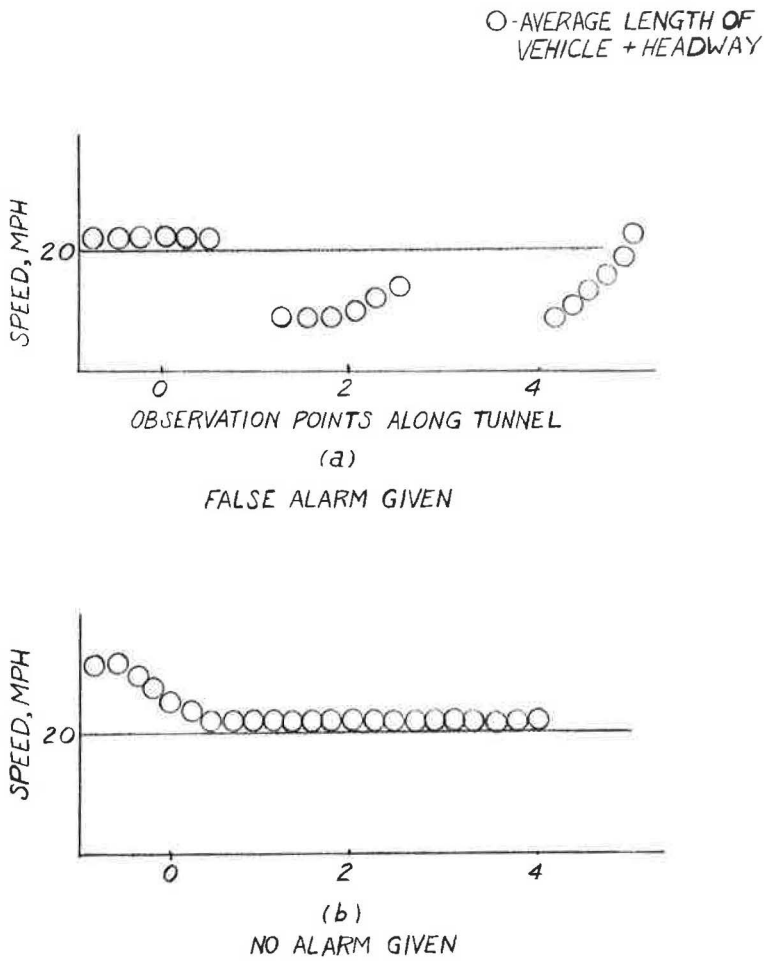



Figure 1. Possible disadvantages of a control system based on speed.

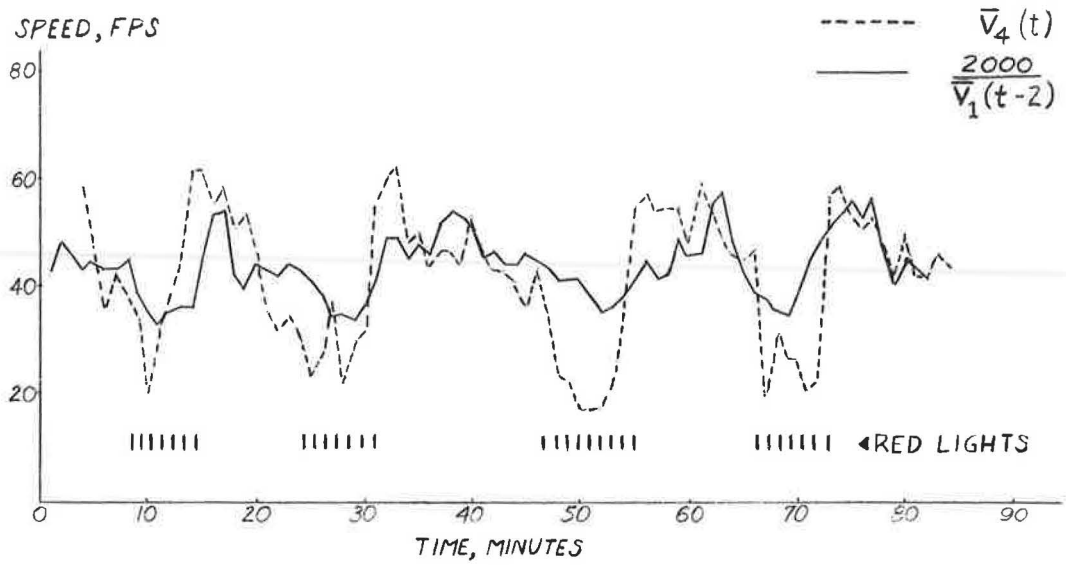



Figure 2. Comparison of $\frac{2000}{\bar{V}_1 (t - 2)}$ and $\bar{V}_4$ (t).

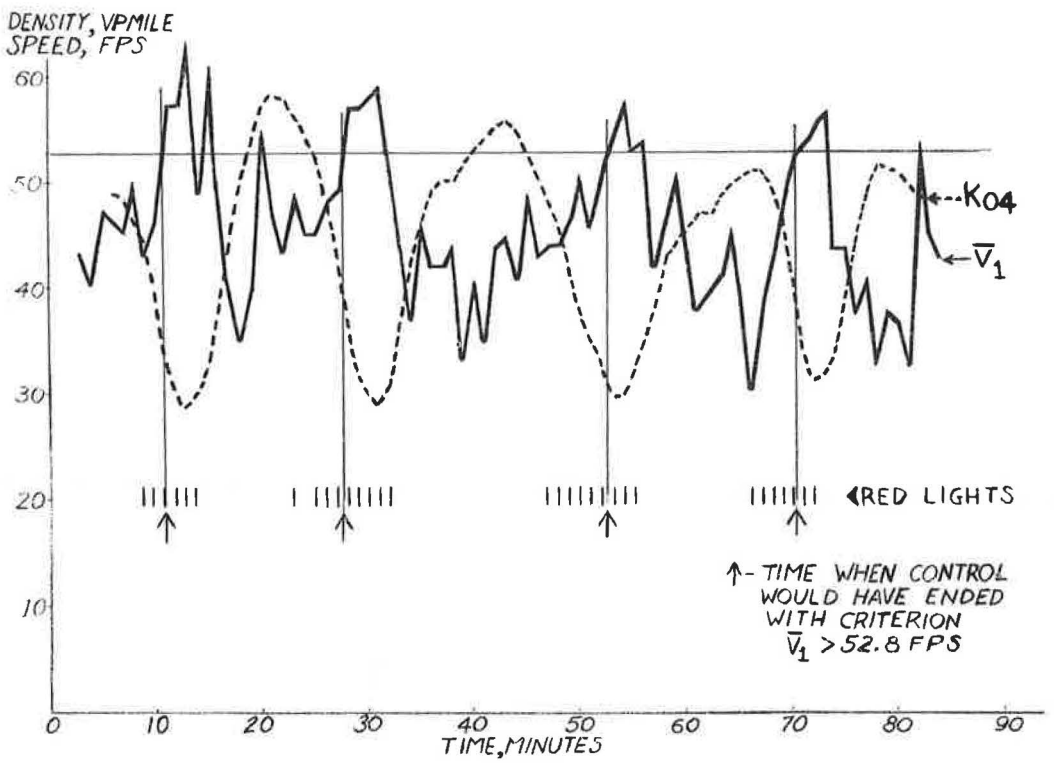



Figure 3. Improving the existing Holland Tunnel control.

phase. This means that the control by intermittent red lights introduces not only gaps into the traffic stream, but also speed waves. Thus, the system has exhibited a tendency to overcontrol the flow.

The existing system would be improved if one of the quantities entering into the control box was sensing the effect of the control after a short time, e.g., 30 sec. As will be seen, point 1, 1,800 ft from the entrance, is an excellent data point for many purposes. Let $\bar{V}_1$ (1 min average) be introduced into the control system as a feedback in the following manner: The red light is controlled as before; if it is activated in the interval (m, m + 1), then $\bar{V}_1$ is examined at time m + 1. If $\bar{V}_1 > 52.8$ ft/sec (or 0.1

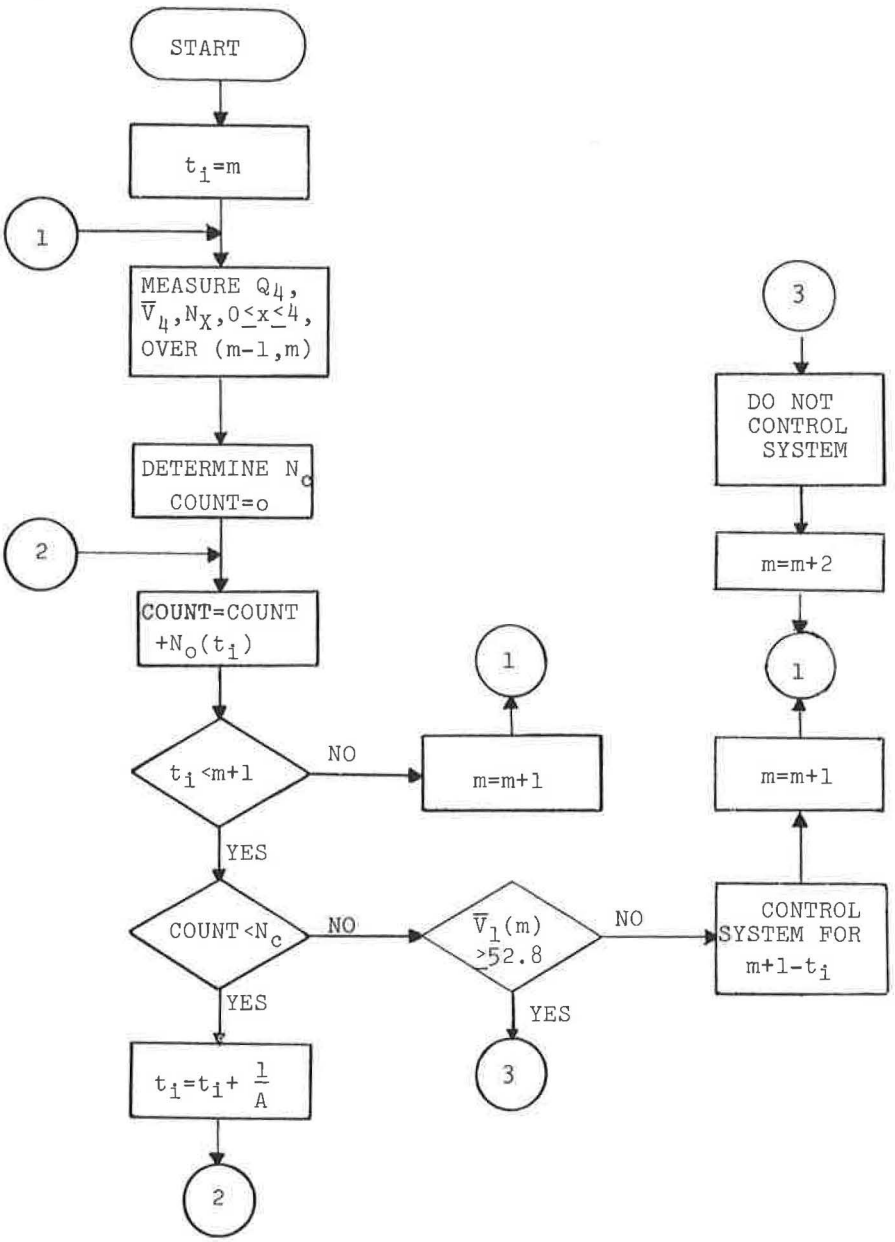



Figure 4. Logic diagram for an improved Holland Tunnel control system.

mi/sec, an empirical value), then the control cannot be activated until the interval (m + 2, m + 3). Otherwise, there is no change in the previous procedure. The vertical arrows in Figure 3 indicate where the sequence of red lights would have been interrupted with the proposed criterion. Figure 4 is a diagram of the modified control system, where 1/A is a small time increment of the order of 2 sec.

Thus, with a minimum amount of additional hardware, the tendency to overcontrol and to generate speed waves would be decreased by an amount to be determined experimentally.

However, an excessive lag between the time when the need for a correction is detected and the effect of the correction is felt is still inherent in such a system. The next step is to try to predict the need for a correction so that it can be applied in due time. To this effect, the input to the control system must include a quantity which can be predicted; i.e., either output flow or section density. Data show that section density is a reliable state variable for a tunnel.

## ANALOG CONTROL SYSTEMS WITH MINIMUM HARDWARE

To clarify the problem, a procedure which a human operator would reasonably follow is analyzed, simplified and mechanized. Density estimation methods are developed and discussed, and a control logic for a tunnel with one bottleneck is given. Finally, control diagrams for a tunnel lane with two bottlenecks are presented.

### A Human Control Analog

A person in charge of maximizing the flow through the given tunnel lane is placed in front of the 18 television pictures of the present LT ST surveillance system. Let this person have the possibility of impeding flow independently through any of the 18 sections at which he can look. A sensible control procedure which he can follow is:

1. Look at all the screens, beginning at the downstream end of the tunnel.
2. If cars are stopped or going very slowly in section j, $1 \leq j \leq 18$, observe whether there exists a sufficient number of sizable gaps upstream from j, so that this disturbance can be absorbed.
3. If the disturbance is likely to propagate backward and impede (i.e., slow down or stop) the traffic upstream from section j; otherwise, continue scanning the screens.
4. Cease control as soon as the disturbance has disappeared.

Such subjective control cannot be programmed for a machine. The sentence "a sufficient number of sizable gaps upstream from j" must be replaced by an inequality between measurable quantities, such as densities. Figure 5 represents an analog to the human control process just described. The quantity $H_j$ is a critical density in sections 1 to j, above which shockwaves have been observed to propagate backward 80 percent of the time.

On the other hand, the system can be made "learning," with a procedure analogous to the one a person would follow. After a while, the person in charge of the control will know where to look for bottlenecks. If the shockwave formation frequency in section j is $f_j$, $\left(\sum_1^{18} f_j = 1\right)$, he will tend to look at screen j with a relative frequency, $f_j$. This system is more efficient than the one in which sections are scanned sequentially, and can be applied to any single-lane traffic situation.

When a tunnel lane possesses one or two fixed bottlenecks, such as the Holland Tunnel south tube near lane and the Lincoln Tunnel south tube near lane, one of the control systems presented later can be used. Since all these systems necessitate a knowledge of section density, k, methods of obtaining k will first be investigated.

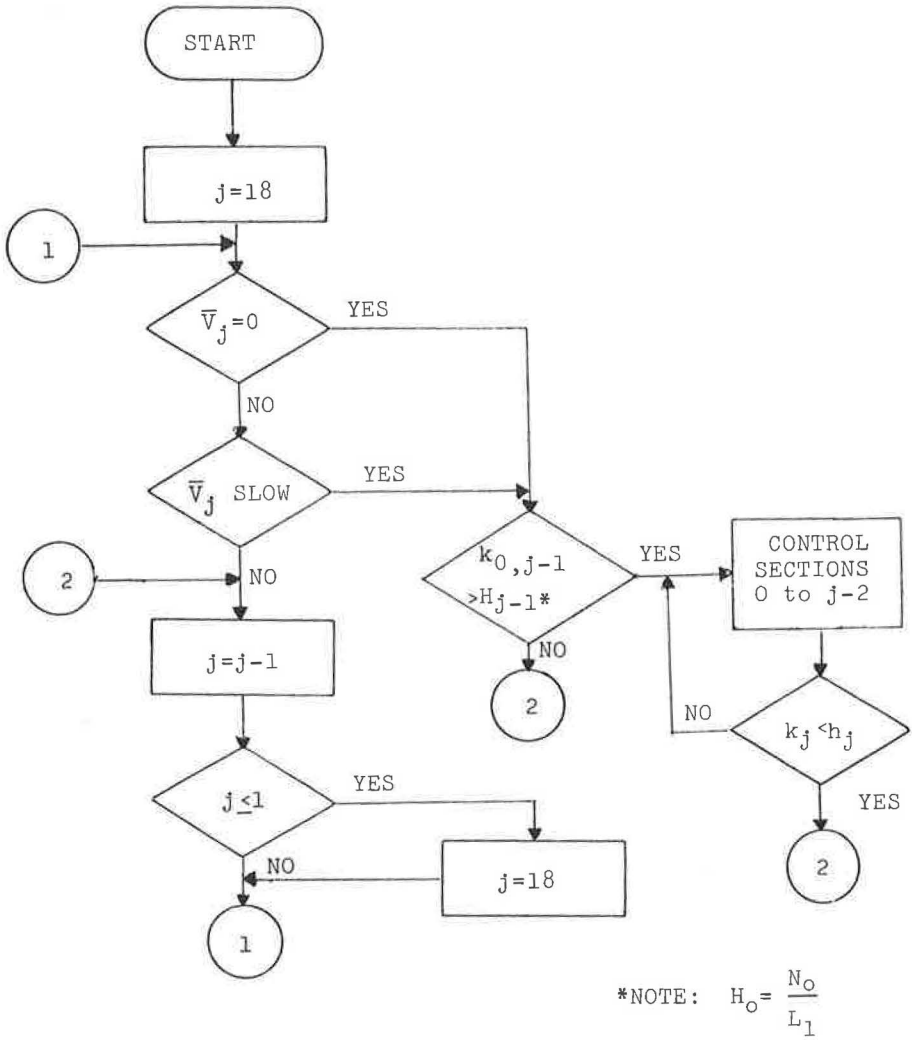



Figure 5. A human control analog.

## One-Bottleneck Tunnel Control

Taking the Holland Tunnel as an example, let the bottleneck start at point 4. Let $H_{04}$ be the density in section 04 at time m, above which the frequency of shockwave occurrence at time m + t, $t \leq 2$, is greater than 0.95.

Let $h_{04}$ be the density in section 04 at the time m, above which the frequency of shockwave occurrence at time m + t, $t \leq 2$, is less than 0.10.

The simplest control procedure consists in keeping $k_{04}$ below $H_{04}$. If $k_{04} > H_{04}$ then the control should impede vehicles from entering until $k_{04} \leq h_{04}$.

The densities $H_{04}$ and $h_{04}$ may be determined from existing data. Naturally, the threshold values 0.95 and 0.10 can be changed. Edie (4) has investigated the generation of stop-start waves and is currently examining the question in further detail. The order of magnitude of $H_{04}$ is 55 veh/mi. It is reasonable to postulate that both $H_{04}$ and $h_{04}$ depend upon the speed at point 4 (or downstream of it). At high speeds, shockwaves may

occur for lower density than at low speeds. A control diagram for the case when the limit densities $H_{04}$ and $h_{04}$ depend upon $V_4$ is shown in Figure 6. Traffic composition, weather and other factors have a definite influence on the threshold values H and h. This is the subject of other studies presently being undertaken at the Port Authority.

## Two-Bottleneck Tunnel Control

Given a roadway tube with bottlenecks A and B, such as the Lincoln Tunnel south tube (5); let this tube be divided into sections OA and AB. Let P and Q be two points

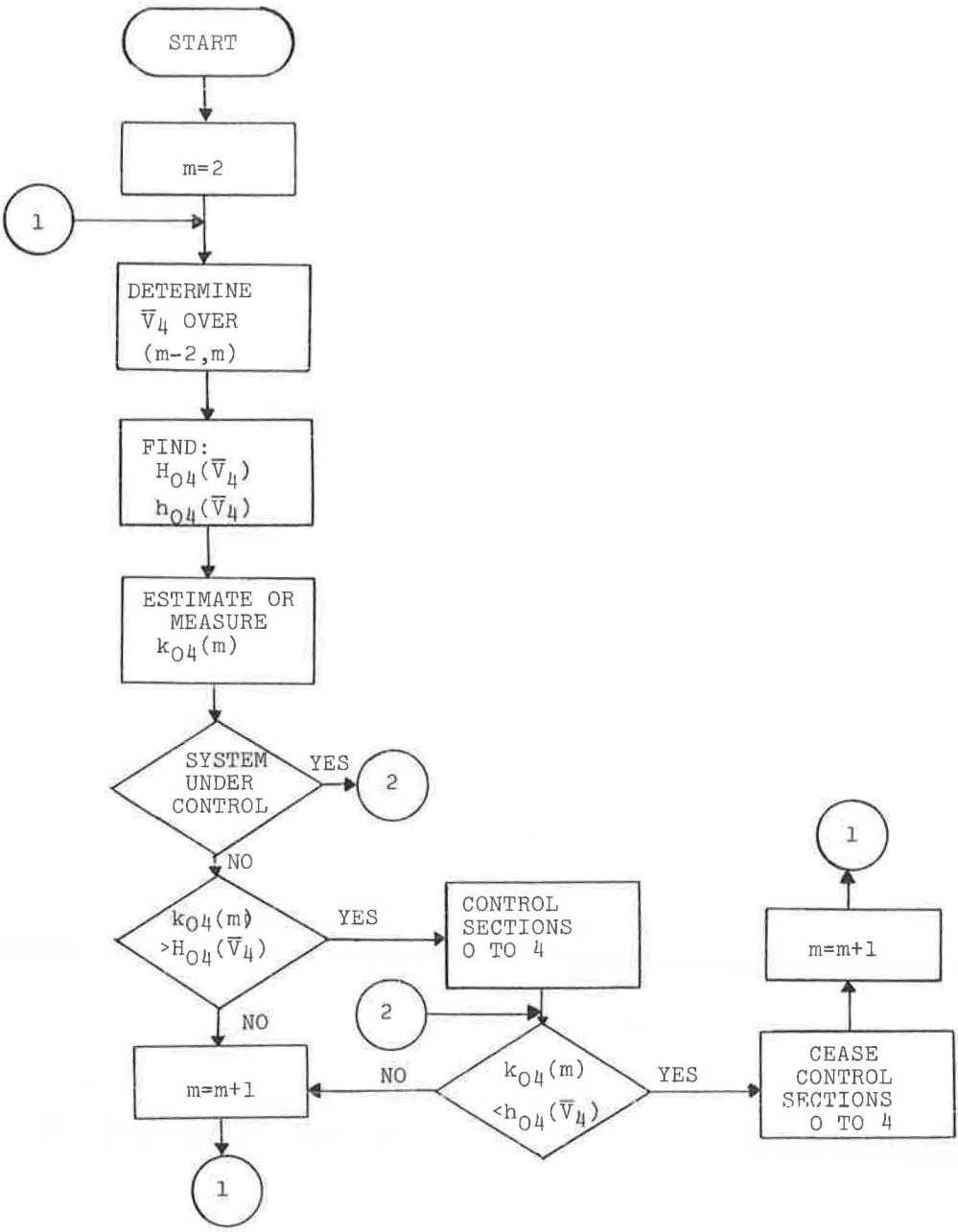



Figure 6. Logic diagram for control of density with varying bounds.

slightly upstream of A and B, respectively. The control diagram suggested for the one-bottleneck tunnel can be applied successively to sections OA and AB. The resulting chart, in which the densities are either estimated using $\bar{V}_P$ and $\bar{V}_Q$ or measured by counting vehicles at O, P and Q is shown in Figure 7.

Various methods to measure or estimate density are discussed in the Appendix. In particular, Figure 11 can be used to estimate $k_{OA}$ or $k_{MB}$.

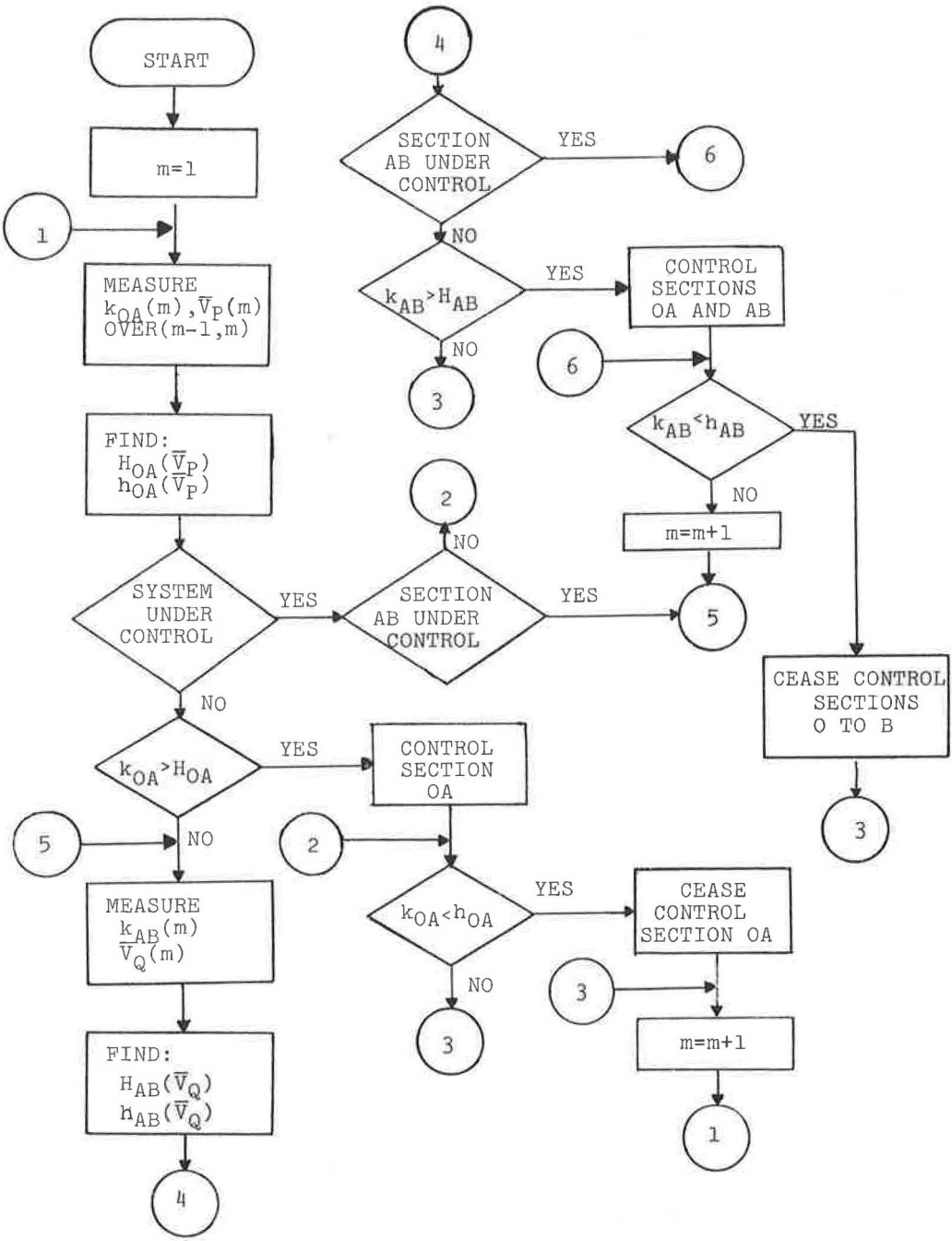



Figure 7. Control of two-bottleneck tunnel using two speeds and two densities.

In the case when the functional relationships H (V) and h (V) are not readily obtainable, but vehicle speed must be taken into account, Figure 8 can be used. Here, an upper limit W and a lower limit w to the speed V must be given in addition to two constant bounds H and h on k. The control system requires only one density measurement $k_1$ and two speed measurements $\bar{V}_1$, $\bar{V}_2$; the second density $k_{AB}(t_i)$ is taken as $k_{OA}(t_i - 1)$. A third speed near the tunnel exit can easily be incorporated into the system (Fig. 8).

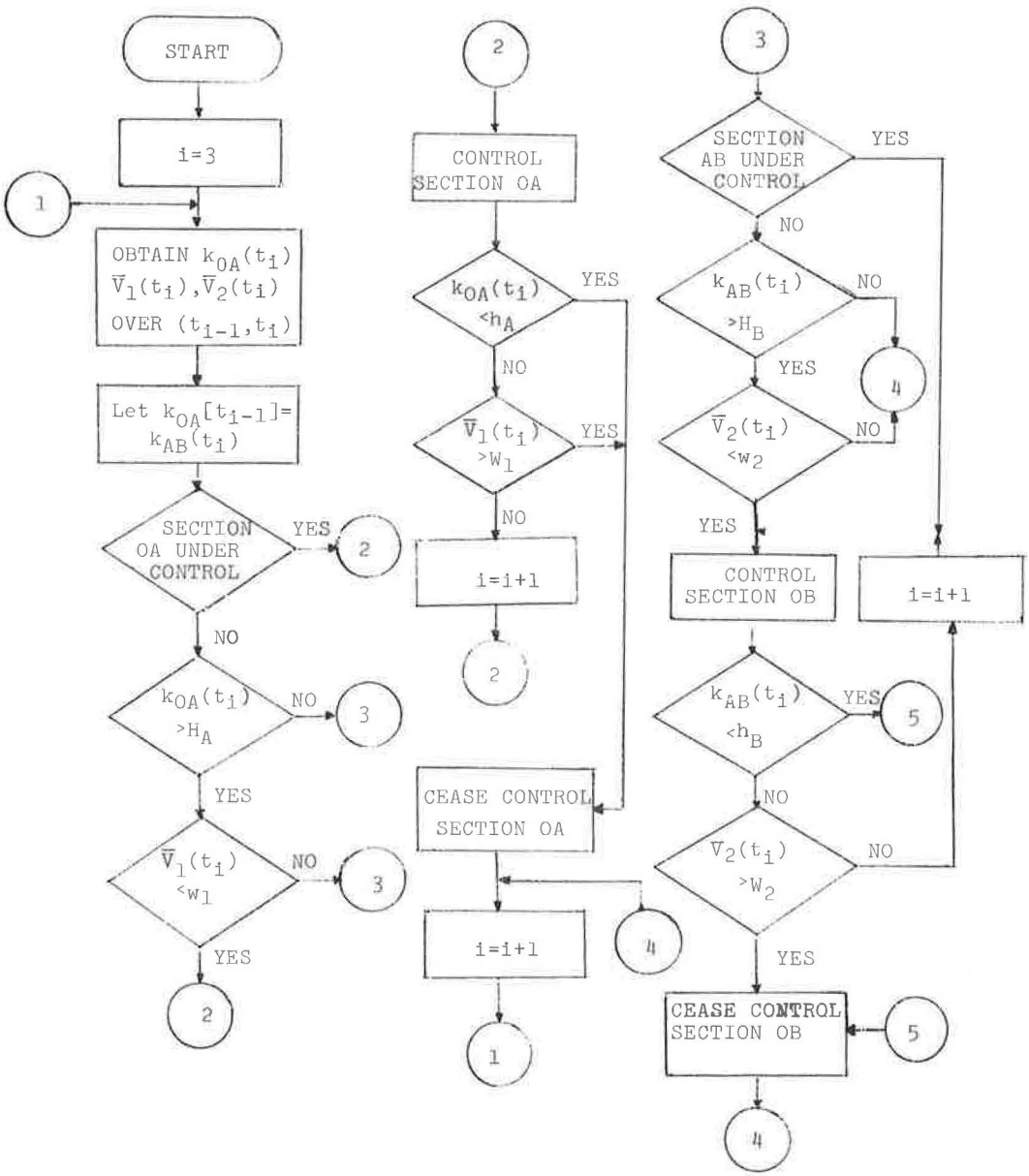



Figure 8. Control of a two-bottleneck tunnel using two speeds and one density.

## DIGITAL-COMPUTER CONTROL

### Finite-State System Representation of a Tunnel Lane

All the control systems presented previously are only moderately sophisticated and can be implemented with solid-state or analog logic. However, an on-line digital computer control can handle a large amount of information with more flexibility in the input-output procedures than a fixed logic or analog system, and is bound to provide for a better control procedure. In addition, experiments, both in the laboratory (simulation) and on-line, are easy to perform with digital devices. A digital control requires no prior knowledge of the dynamics of the system in opposition to an analog control.

The conceptual framework of finite-state machine theory will be used to present the control system. Martin (6) had made several steps in this direction. A good text on the theory of finite-state machines is by Gill (7). The proposed finite-state model is essentially a black box which, when in a certain state, responds at every unit of time to a given excitation (or input) with a predictable output. Here, the output will be a set of control times.

The input to this model consists of the number of cars which are entering the tunnel between time m and m + 1, and of the average speed at several points 1, 2, ..., n in the tunnel. The state of the system is the set of 1-min average densities in sections (0, 1), ..., (1, n).

For every unit of time, a control time of the tunnel entrance is determined by the largest density in the length OX, where X = 1, 2, ..., n. Another control time at point j is determined by comparing speed, density and flow in section (j - 1, j) with corresponding quantities in section (j, j + 1). Here, a very conservative control policy is applied. The control device in section $j > 1$ is activated if and only if the three conditions $\bar{V}_j > \bar{V}_{j+1}$, $k_j > k_{j+1}$, and $k_{j+1}\,\bar{V}_{j+1} \geq Q_0$ are simultaneously satisfied so that no false alarm is likely to occur. Once flow through section j is impeded, it is assumed that flow through all sections upstream of j must be impeded in a similar fashion, so that vehicles do not pack up inside the tunnel.

In addition to an output function, a finite-state machine is characterized by a "next state" function; i.e., the state at time m + 1 is uniquely determined by the state and input at time m. It is clear that the density in section j at time m + 1 can be determined from the speed and density in sections j - 1 and j at time m; the density in section 1 is determined by the entrance rate.

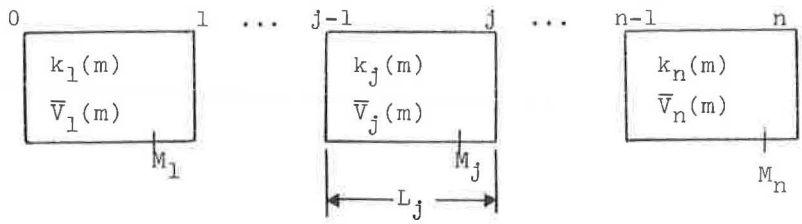
This finite-state machine merely determines a set of control times at fixed time intervals. The implementation of the control is done separately and independently. A considerable amount of flexibility is thus available.

### Finite-State Model

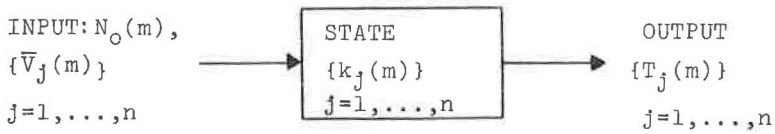
Let the tunnel be divided into n sections of length $L_j$ ending, respectively, at points 1, 2, ..., m. Let the density averaged in the time interval (m - 1, m) in section (j - 1, j) be $k_j$ (m) and the speed averaged in the same time interval at a control point $M_j$ located slightly upstream of j be $\bar{V}_j$ (m). As before, $N_0$ (m) cars have entered the tunnel in the time (m - 1, m) (Fig. 9a).

The finite state control system is represented in Figure 9b and is defined as follows:

1. $I_0$ is the set of non-negative integers $\leq 120$.
2. The input alphabet, state set and output alphabet are Cartesian products of the type $I_0 \times I_0 \times \ldots \times I_0$, i.e., sets of a certain size whose elements are integers between 0 and 120.
3. The input at time m is the ordered (n + 1) - tuple (ordered set of size n + 1), $N_0, \bar{V}_1, \ldots, \bar{V}_n$.
4. The state at time m is the ordered n-tuple (ordered set of size n), $k_1, k_2, \ldots, k_n$.
5. The output at time m is the ordered n-tuple (ordered set of size n), $T_0, T_1, \ldots, T_n - 1$, where Tj is the time in seconds during which traffic through point j is impeded, beginning at time m.
6. The output function, which uniquely relates the output at time m to the input and state at time m, is given by




(a) ACTUAL CONDITIONS AT TIME m




(b) CONDITION OF FINITE-STATE MACHINE MODEL AT TIME m.

Figure 9. Finite-state machine model of tunnel control.

$$T_0(m) = C_0 \left\{ \max_{1 \le x \le m} \left[ \frac{1}{x} \sum_{j=1}^{x} k_j - H(x) \right] \right\}$$

if $\sum_{j=1}^{x} k_j > x\,H(x)$; $T_0(m) = 0$ otherwise,

$$T_j(m) + C_1 (k_j \bar{V}_j - k_{j+1}\,\bar{V}_{j+1})$$

if the three following inequalities are simultaneously satisfied:

$$k_j > k_{j+1},\ \bar{V}_j > \bar{V}_{j+1} \text{ and } k_{j+1}\,\bar{V}_{j+1} \ge Q_0 \quad \text{for } j > 1;$$

all quantities are evaluated at time m; $Q_0$ is a constant of the order of 1,000 veh/hr.

$T_j(m) = 0$ otherwise, where $C_0$, $C_1$ are constants; in addition, the inequalities $T_0 \ge T_1 \ge T_2 \ge \ldots \ge T_j$ must be satisfied.

7. The next state function, which uniquely relates the state at time (m + 1) with the input and state at time m, is obtained by subtracting the flow out of section j from the flow into it:

$$k_1(m+1) = k_1(m) + (N_0 - V_1 k_1)\,\frac{1}{L_1}$$

$$k_j(m+1) = k_j(m) + (V_{j-1} k_{j-1} - V_j k_j)\,\frac{1}{L_j} \quad \text{for } j > 1$$

where all the variables on the right-hand side are taken at time m, and the numbers $k_j$, j = 1, ..., n, rounded off to the nearest integer.

Thus, once the initial state of the system is given, the response of the machine to any input sequence

$$(N_O, \bar{V}_1, \ldots \bar{V}_n)1, (N_O, \bar{V}_1, \ldots \bar{V}_n)2, \ldots$$

is uniquely determined.

Example

To illustrate the finite state machine model, let n = 2. The following quantities and equations are needed to obtain the output ($T_O$, $T_1$) at time m:

$$H(1) = 75, \quad H(2) = 65 \text{ veh/mi}$$

$$T_O(m) = \max \begin{pmatrix} 0 \\ k - 75 \\ \frac{(k_1 + k_2)}{2} - 65 \end{pmatrix}, \text{ with } C_O = 1 \text{ for } T_O \text{ in sec}$$

$$T_1(m) = (k_1 \bar{V}_1 - k_2 \bar{V}_2)\, 10^{-2}$$

if $k_1 > k_2$ and $\bar{V}_1 > \bar{V}_2$

and $k_2 V_2 \geq Q_O$, $Q_O$ - 1,000 veh/hr.

$= 0$ otherwise, with $T_1 \geq T_O$.

The following quantities and relations are used to obtain the next state, i.e., the state at time m + 1:

$$L_1 = 3300 \text{ ft}, \quad L_2 = 4800 \text{ ft}; \text{ time unit} = 60 \text{ sec}$$

$$k_1(m+1) = k_1(m) + \left(N_O - \bar{V}_1 k_1 \frac{60}{5280}\right) \frac{5280}{3300}$$

$$k_2(m+1) = k_2(m) + (\bar{V}_1 k_1 - \bar{V}_2 k_2) \frac{60}{4800}$$

Six columns of the transition table of the machine are shown in Table 1. In the output subtable, three outputs are circled; also the three corresponding states in the next state subtable are circled. The detail of the computation of the circled elements is

a. Output subtable

First column

$k_1, k_2 < 65 \qquad T_O = 0$

$k_1 < k_2 \qquad T_1 = 0$

$\boxed{T_O = T_1 = 0}$

Second column

$$T_O = \max \begin{pmatrix} 0 \\ k_1 - 75 = 15 \\ \frac{1}{2}(k_1 + k_2) - 65 = 0 \end{pmatrix}$$

Hence, $T_O = 15$

TABLE 1

FINITE STATE MACHINE MODEL—EXAMPLE OF A TRANSITION TABLE

OUTPUT SUBTABLE: $(T_o, T_1)$ sec. AT TIME m  NEXT STATE SUBTABLE: $(k_1, k_2)$ AT TIME m+1

| Input at m $(N_o, V_1, V_2)$ | | (15,20,20) | ... | (19,20,20) | ... | (20,25,20) | ... | (15,20,20) | ... | (19,20,25) | ... | (20,25,20) | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| State at m $(k_1\ k_2)$ | | | | | | | | | | | | | |
| | (50, 40) | (0,0) | | (0,0) | | (4,4) | | (44,46) | | | | | |
| | (50, 60) | (0,0) | | (1,0) | | (1,0) | | | | | | | |
| $k_1 \leq k_2$ | (60, 70) | (0,0) | | (0,0) | | (2,1) | | (55,68) | | | | | |
| | (70, 80) | (10,0) | | (10,0) | | (10,0) | | (10,1) | | | | | |
| | (80, 70) | (10,0) | | (10,0) | | (10,6) | | | | | | | |
| | (70, 60) | ( 0,0) | | ( 0,0) | | ( 5,5) | | | | | | | |
| $k_1 > k_2$ | (80, 60) | ( 5,0) | | ( 5,0) | | ( 8,8) | | | | | | (64,72) | |
| | (80, 50) | (15,0) | | (15,0) | | (15,10) | | | | (72,54) | | | |

Since

$$V_1 > V_2, \quad T_1 = 0 \qquad \boxed{T_0 = 15, \quad T_1 = 0}$$

Third column

$$T = \max \begin{cases} 0 & \\ k_1 - 75 = 5 & T_0 = 5 \\ \frac{1}{2}(k_1 + k_2) - 65 = 5 & \end{cases}$$

$$(k_1 V_1 - k_2 V_2)\, 10^{-2} = 20\text{-}12 \text{ hence, } T_1 = 8$$

Since $T_1 > T_0$, one has:

$$\boxed{T_0 = T_1 = 8}$$

b. Next state subtable

First column

$$k_1 (m+1) = 60 + \left(15 - 1200 \times \frac{60}{5280}\right) \frac{5280}{3300}$$

$$\cong 60 - 5 = 55$$

$$k_2 (m+1) = 70 + (1200 - 1400)\, \frac{60}{4800}$$

$$\cong 70 - 2 = 68$$

Second column

$$k_1 (m+1) = 80 + \left(19 - 1600 \times \frac{60}{5280}\right) \frac{5280}{3300}$$

$$\cong 80 - 8 = 72$$

$$k_2 (m+1) = 50 + (1600 - 1250) \times \frac{60}{4800}$$

$$\cong 50 + 4 = 54$$

Third column

$$k_1 (m+1) = 80 + \left(20 - 2000 \times \frac{60}{5280}\right) \frac{5280}{3300}$$

$$\cong 80 - 16 = 64$$

$$k_2 (m+1) = 60 + (2000 - 1000)\, \frac{60}{4800}$$

$$\cong 60 + 12 = 72$$

Operational Procedure

The finite-state model produces a control policy every time unit. The state of the system is given at time zero (density in each section); from there on, entrance flow and speed at the end of each section uniquely determines a control policy to be applied in the next time interval, as well as the next state of the system. Several methods can be used to implement the control policy depending on the time scale chosen. One may

apply it every unit of time so that the policy found at time m + 1 supersedes the one applied in (m, m + 1). One may wish to consult the model every p units of time only. Also, one may choose to execute only selected portions of the recommended control.

Due to round-off error and noise, the state of the model may drift away from the actual state of the system; the model must be reset to zero and its new initial state given. At first, on-line experiments can be made with such a model without actually controlling the flow. The number of correct predictions of shockwave formation and of density value is a measure of the validity of the model. Beyond this, field experiments or computer simulation studies are necessary.

### Extension of the Model

Once experiments have established satisfactory output (control) and next state functions, minimization techniques can be applied to the machine, so that a model equivalent to the original one and generally simpler may be found. Next, one can remark that it is possible to pass from a finite-state machine to a stochastic system using either one of the following two methods: (a) given the machine, define a probability distribution upon the set of inputs; and (b) given an input sequence, define a transition probability from state i to state j for all i, j.

The first method can be used for simulation purposes. The second method yields a Markov chain model which has often been proposed for traffic problems. If the transition probabilities from state $(k_1, k_2, \ldots, k_n)_i$ to state $(k_1, k_2, \ldots, k_n)_j$ can be defined independently of other variables, then the Markov chain model is valid for the system presented here. An on-line experiment can answer this question.

## POINTS TO BE INVESTIGATED

Since the purpose of a tunnel control system is to maximize peak hour flow by avoiding the formation of shockwaves, the conditions under which these shockwaves appear must be accurately studied. Preliminary investigations indicate that at least the following factors are relevant:

1. Density and traffic composition, which may be combined into a weighted density in a manner to be determined experimentally.
2. Intrinsic speed of vehicles which is related to traffic composition (i.e., slow trucks) and drivers "desired speed."
3. Actual speed of vehicles, which is a measured quantity.
4. Environmental parameters, which include geometry and lighting of the tunnel, signalization, presence of police officers, weather, etc.

A study of these four points, which unfortunately are not independent, yields the threshold values H (x), H (V), h (V), w, W and $Q_0$ introduced previously. An adequate definition of density weighted for traffic composition is sufficient to define a first generation control system.

Once shockwave formation is understood, the first step consists in minimizing the environmental factors which cause shockwaves. The next step consists in finding actual methods of impeding traffic in a given section OX. These points, together with the question of hardwave choice, will be discussed elsewhere.

## ACKNOWLEDGMENTS

The author is grateful to R. Foote, K. Crowley and the other members of the Tunnels and Bridges Research Section of the Port of New York Authority for the help received in all phases of this study. A National Science Foundation grant provided the opportunity for initiating work in this subject area and the possibility for continuation of this investigation at the University of Arizona.

## REFERENCES

1. Foote, R. Single Lane Traffic Flow Control. Proc. of the Second Intern. Symp. on the Theory of Traffic Flow, London, 1963. OECD, Paris, 1965.
2. Foote, R., and Crowley, K. Developing Density Controls for Improved Traffic Operations. Highway Research Record 154, pp. 24-37, 1966.
3. Crowley, K., and Greenberg, I. Holland Tunnel Study Aids Efficient Increase of Tube's Use. Traffic Eng., March 1965.
4. Edie, L. Generation and Propagation of Stop-Start Traffic Waves. The Port of New York Authority, Rept. No. RD 65-1, June 1965.
5. Crowley, K. Unpublished research, The Port of New York Authority, July 1965.
6. Martin, W. Description and Control of Single Lane Tunnel Traffic Flow. Research Report No. 3 of the Center for Operations Research, Massachusetts Institute of Technology, March 1963.
7. Gill, A. An Introduction to the Theory of Finite-State Machines. McGraw Hill, 1958.

# *Appendix*

## SECTION DENSITY ESTIMATION

### Generalities

Measuring the density of vehicles in a tunnel lane section is not as easy as it sounds. If one divides point flow by point speeds, wide errors appear as soon as speeds begin to fluctuate—for example, when a shockwave passes by the observation point. If an input-output count is used, the change in density from sampling time to sampling time is known accurately; however, the counter drifts each time a vehicle changes lane or a count is missed. Such an input-output count must thus be periodically resynchronized. Several methods are available to perform this latter operation:

1. The Port Authority's traffic data collection system permits an accurate measurement of vehicle length. A pattern recognition program can therefore recognize when any given train of vehicles which has entered the tunnel at a marked instant is leaving it. If N cars have entered a tunnel of length L between the passage of the last car of the sequence at, respectively, entrance and the exit portal, the exact value of the density is N/L. Observations have shown that the frequency of a pattern "long vehicle-short vehicle," such as tractor-trailer followed by a Volkswagen, is suitable for resynchronization purposes. This system is soon to be tested.
2. An alternate method for monitoring an input-output count through the use of average speed at a point and number of cars entering the section during a variable time interval is presented. This method should be at least as accurate as dividing point flow by point speed and is proposed only when wave velocities are moderately positive in the section, that is, when no stop-and-go waves are present.
3. An approximate verification of this alternate method is done using a fixed time interval—which is equivalent to using flow rate. The hand computations for the exact method would be too tedious; a verification of the range of validity of the method will be done when an on-line digital computer can be programmed to execute it.

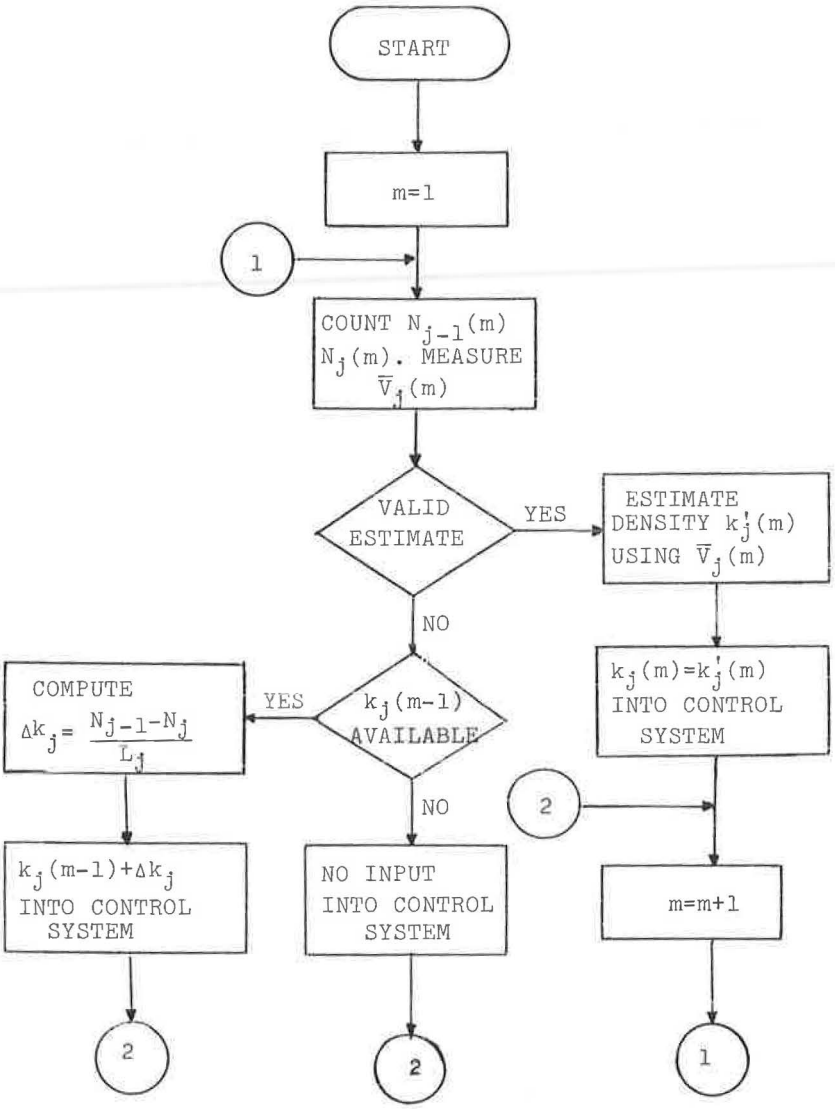



Figure 10. A hybrid method to measure section density.

The logic of a density meter resynchronized every ten or more units of time by a density estimator is shown in Figure 10, where the block "estimate is valid" is to be transformed into an inequality which expresses the fluidity of the flow, i.e., the presence or absence of shockwaves and/or excessively large gaps. The exact method and its approximate verification are presented using the particular example of the Holland Tunnel, where the density $k_{04}$ (or $k_{14}$) between points 0 (or 1) and 4 is to be estimated. The approximate estimate of densities is then compared with the actual value obtained by manual input-output count.

## A Method to Estimate Density in Fluid Flow Cases

The continuity property of flow can be used to derive the following algorithm:

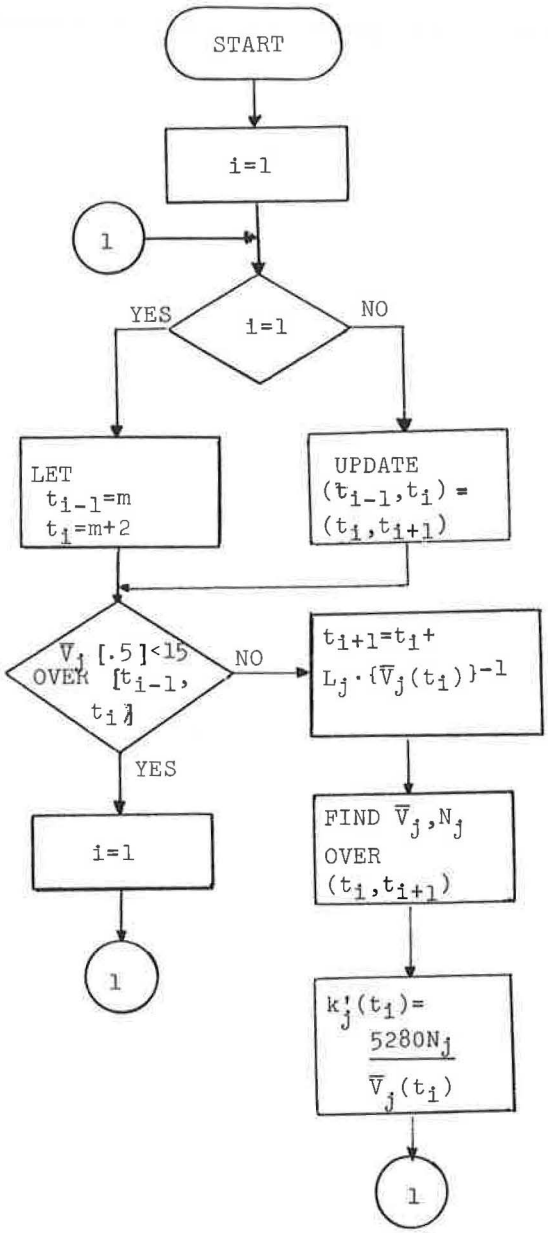



Figure 11. Density estimation by averaging quantities over time intervals of varying lengths.

1. At time $t_i$, average the speed at point 1 (1800 ft from the tunnel entrance), $\bar{V}_1(t_i)$ during the time interval $\left(t_i - \frac{6000}{\bar{V}_1(t_i - 1)}, t_i\right)$, with the boundary conditions: $t_0 = 0$, $t_1 = 2$ min. Note that 6000 is the distance 04 (in feet).

2. Using a time listing of vehicles having entered the tunnel up to time $t_i$, find the input $N_0(t_i)$ during the time interval $\left(t_i - \frac{6000}{\bar{V}_1(t_i)}, t_i\right)$.

3. If the flow was fairly fluid, so that the average speed at point 1 is representative of the speed 04, then the number of cars in section 04 at time $t_i$ is $N_0(t_i)$. Hence the estimated value of $k_{04}$ is

$$k'_{04}(t_i) = \frac{N_0(t_i)\ 5280}{6000} \text{ veh/mi}$$

To illustrate the method, let $\bar{V}_1(t_i - 1) = 45$ ft/sec. Then the average speed at point 1 during the time interval $\left(t_i - \frac{6000}{45}, t_i\right)$ or 133.3 seconds before sampling time is computed and denoted by $\bar{V}_1(t_i)$. Let $\bar{V}_1(t_i) = 52$ ft/sec; let the number of vehicles which have entered the tunnel for the last $\frac{6000}{52} = 115.4$ sec be $N_0(t_i) = 48$. The estimated value of the density 04 at time $t_i$ is $48 \times \frac{5280}{6000} = 42.0$.

4. Figure 11 shows a density estimation method for section (j - 1, j) of a tunnel (Fig. 9a). Discrete time intervals $(t_0, t_1) \ldots (t_i, t_{i+1}) \ldots$ are defined by the recurrence relationship:

$$t_{i+1} = t_i + \frac{L_j}{\bar{V}_j(t_i)}, \quad i > 1, \text{ with } t_0 = 0 \text{ and } t_1 = 1$$

For example, let $L_j = 1800$ ft. The faster the traffic is flowing, the more often k is estimated. If $\bar{V}_j = 60$ ft/sec, k′ is calculated every $\frac{1800}{60} = 30$ sec. If $\bar{V}_j = 30$ ft/sec, k′ is calculated every $\frac{1800}{30} = 60$ sec.

Estimates k′ of k are accepted only if the flow was fairly fluid during the previous time interval. Thus, in Figure 11, the estimate $k'_j(t_i)$ of $k_j(t_i)$ is accepted only if the 0.5-min moving average of $\bar{V}_j$ over $(t_i - 1, t_i)$ was greater than 15 mph (for example). No estimate is computed if this condition is not satisfied.

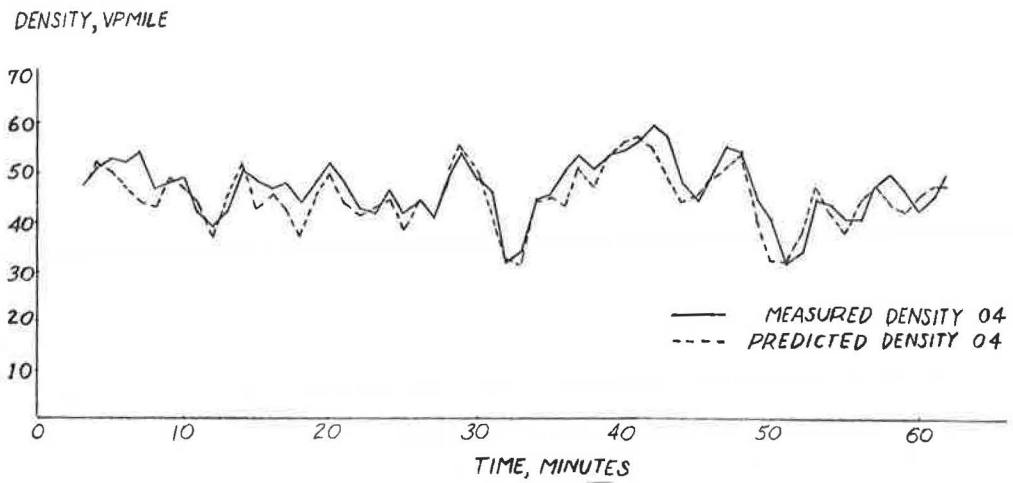



Figure 12. Smooth uncontrolled flow: density estimation, Case I.

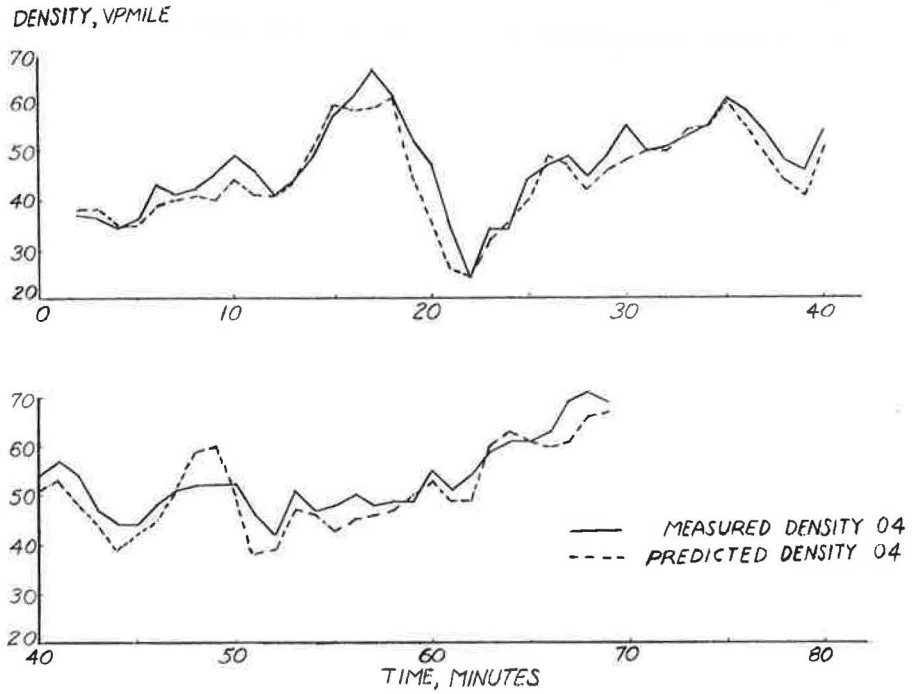



Figure 13. Smooth uncontrolled flow: density estimation, Case II.

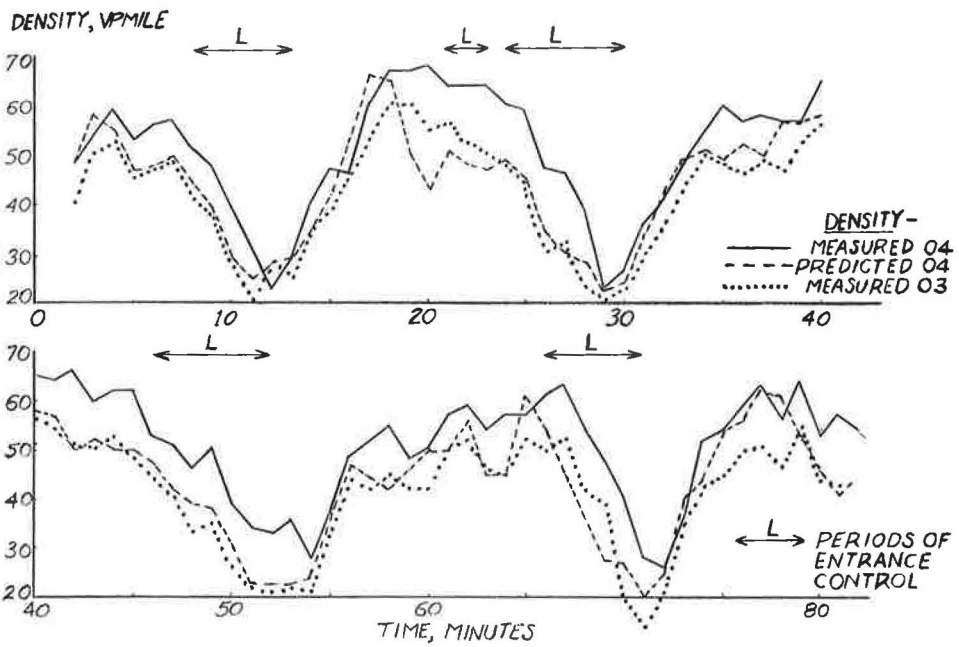



Figure 14. Smooth controlled flow: density estimation.

## Approximate Verification of the Proposed Method

If the speeds are around 50 ft/sec, then the sampling times are close to $\frac{6000}{50} = 120$ sec or 2 min. Thus, a fairly gross approximation to the method proposed above consists in taking a fixed sampling time interval (m -2, m) instead of a variable one $\left(t_i - \frac{6000}{\bar{V}_1(t_i)}, t_i\right)$. This is equivalent to dividing input flow by average speed at point 1, so that the estimated density is given by the formula

$$k'_{04}(m) = \frac{N_0(m)}{120} \frac{6000}{\bar{V}_1(m)} \frac{5280}{6000}$$

Clearly, the error increases every time that $\bar{V}_1$ departs from the value 50 ft/sec. Furthermore, the speed at point 1 is not representative of the speed 04—the former appears to be systematically higher than the latter. Accordingly, the factor $\frac{5280}{6000}$ is dropped from the formula to yield much better estimates of $k_{04}$.

In Figures 12 and 13 density estimated by this latter method is compared with density measured under fluid traffic conditions by manual count for two different uncontrolled flow cases. The estimate is within 10 percent of the actual value. An entrance-controlled flow case is shown in Figure 14; here, the density prediction is quite satisfactory whenever the density is increasing, i.e., after the red light has been deactivated. During the control, large gaps and high speeds cause an underestimation of the actual density; the estimated value of $k_{04}$ is then close to the actual value of $k_{03}$.

## Extension of the Method

An extension of the method for the case of high flow with shockwaves is proposed. Empirical limits on the validity of the method are used to yield a part-time estimation of density. The method is presented in detail through two examples in report TBR5-65 of The Port of New York Authority; only a summarized version is presented here.

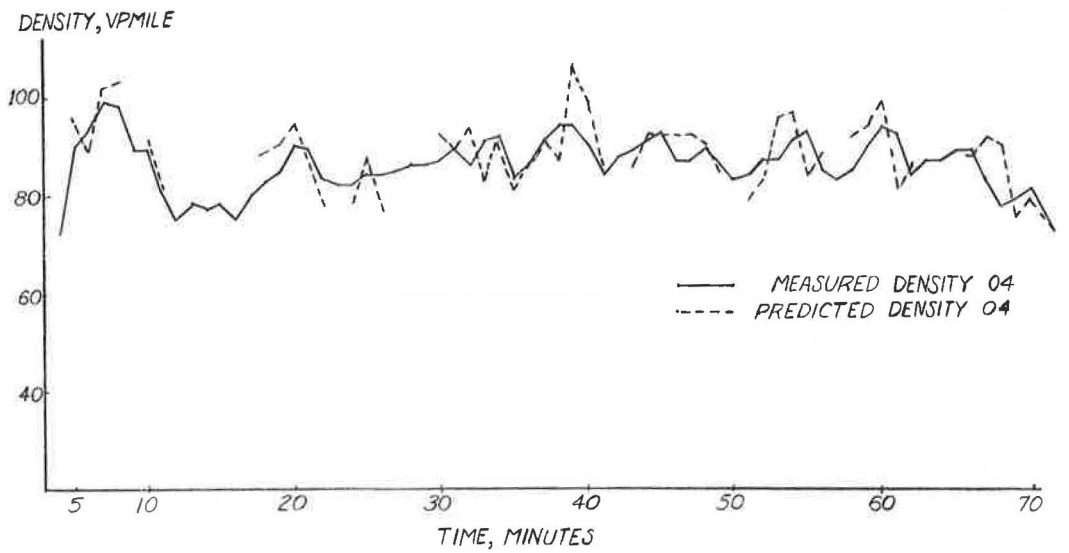



Figure 15. Stop-and-go flow: part-time estimation of density, Case I.

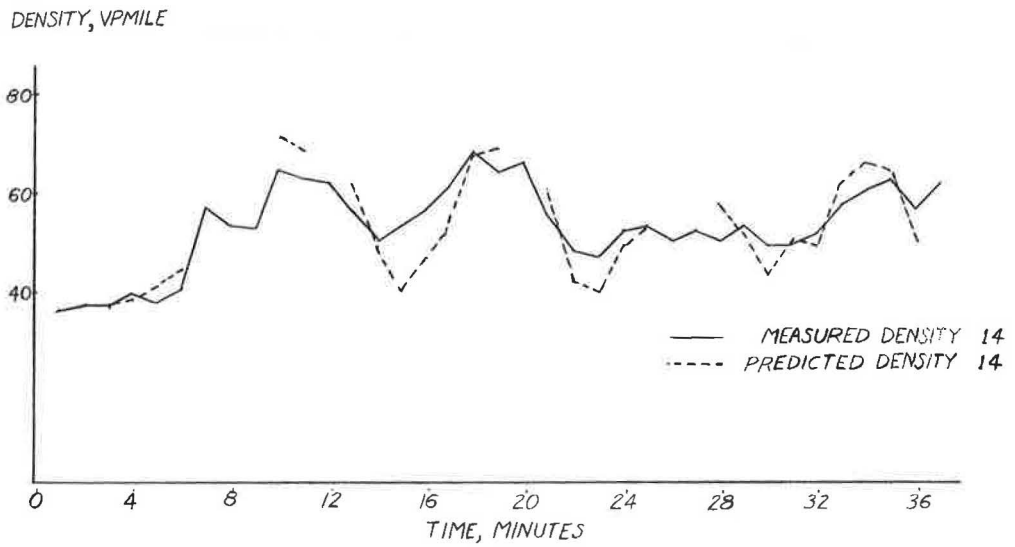



Figure 16. Stop-and-go flow: part-time estimation of density, Case II.

Shockwaves create a wide variation of average point speeds (more than one order of magnitude), whereas the corresponding section density variation is relatively moderate, say ±60 percent. In order to take this experimental fact into account, section densities will be estimated during only that part of the time when speed measurements reflect neither a forward moving wave, i. e., large gaps and high speeds, nor a backward moving wave, i. e., small gaps and very low speeds.

An estimated value $k_{04}$ of $k_{04}$ is first obtained using a weighted average of three virtual densities calculated from speeds at points 1, 2, 3 of the Holland Tunnel ST NL. The comparison between the accepted estimates and the actual values obtained by manual count is shown in Figure 15. The relative error is generally less than 10 percent.

Figure 16 shows a part-time estimation $k'_{14}$ of the density $k_{14}$ using the entrance rate $N_1$ in section 1 - 2 and two speeds $\bar{V}_2$, $\bar{V}_3$. A weighted average of two densities, $k'_{14}$, is obtained by the method used for the case of smooth flow.

It can be noted that the estimation of $k'_{04}$, which is based upon speeds at three points (Fig. 15) is closer than the estimation of $k'_{14}$ which is based upon two speeds only (Fig. 16).

# Developing Density Controls for Improved Traffic Operations

ROBERT S. FOOTE and KENNETH W. CROWLEY
The Port of New York Authority, Tunnels and Bridges Department

Traffic through a single-lane road section with a bottleneck at output is considered as a system involving input flow, section density, output speed, and output flow. Effects of output flow on output speed, output speed on section density, and section density on input flow are shown.

Four case studies are described, each with different patterns and levels of traffic production. Consistent relations among section density, output speed, and output flow are observed. Flow is at a maximum when output speeds are in midrange; output speeds are a delayed inverse function of section density. The effect of an early automatic system for controlling section density by limiting input flow based on measuring output speeds is described. Use of direct measures of section density to stabilize the control system is planned.

•AN INTENSIVE study is being made of traffic flowing through the Holland and Lincoln Tunnels to determine reasons for the significant fluctuations observed in peak traffic production through these expensive roadways, and to enable controlling pertinent variables to raise the overall level of peak traffic production. These studies have shown that increases of a few percentage points in peak traffic production can have a dramatic effect in reducing the duration of congestion (1). An increase of 4 percent in traffic production can result in a 33 percent cut in the duration of congestion (1, 2).

Peak hourly traffic figures in each tunnel lane regularly vary more than 4 percent. Traffic production through these tunnels differs by as much as 50 percent from the capacity of other similar tunnel lanes and by as much as 100 percent from the capacity of open expressway lanes.

One consequence of this research has been to demonstrate that controlling tunnel traffic to maintain fluid movement and prevent congestion can increase traffic production by approximately 5 percent. An automatic system controlling input flow based on traffic conditions inside the tunnel has been found to cause an overall improvement of 2 percent, but significant oscillations in tunnel traffic conditions were evident. The need for an improved control logic led to undertaking more detailed measurements of tunnel traffic behavior over a length of roadway.

This paper reports the conditions observed as a result of these measurements. Four different types of output flow are described. Consistent relations among density, speed and flow are observed indicating that a more stable and effective control logic can be obtained.

## EXPERIMENTAL CONDITIONS

The traffic system consists of a 6,000-ft single tunnel lane having a bottleneck at the output end and changing from 3 percent downgrade to level 3,100 ft from the entrance. In addition to measuring traffic at both the entrance and output points, measurements

Paper sponsored by Committee on Freeway Operations and presented at the 45th Annual Meeting.

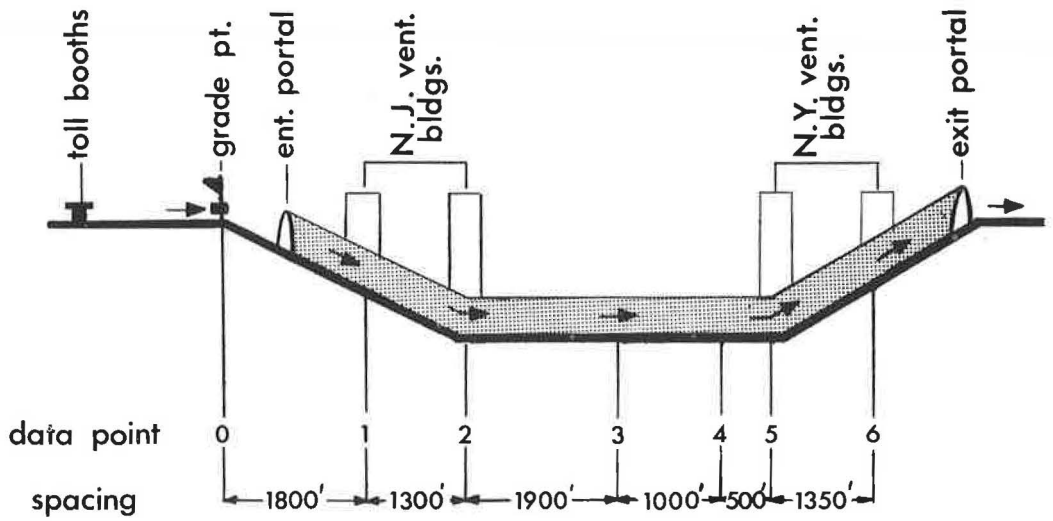



Figure 1. Location of data points: Holland Tunnel, south tube.

were also taken at the intermediate point on the downgrade section, at the grade change point, and on the level section (Fig. 1). Additional data were collected beyond the output section, but with one exception, data from those points are not used.

Data were collected at each tunnel point by a pair of photocells spaced 13 ft apart under the tunnel roadway, completely out of view of passing motorists. As a vehicle passed each pair, four events occurred: (a) the light beam to the upstream photocell was broken, (b) the light beam to the downstream photocell was broken, (c) the light beam to the upstream photocell was re-established, and (d) the light beam to the downstream cell was re-established. The sequence of the second and third events depends on the vehicle length. The occurrence of each event at each pair of cells was recorded on one channel of a stereo-magnetic tape recorder. Data were collected concurrently and multiplexed on the one channel from each of the four tunnel points. Simultaneously, on the other channel a 100-cps time tone was recorded. Therefore, for an hourly traffic flow of 1,200 vehicles, a total of 1,200 times four points times four events—or a total of 19,200 units of information—was recorded. Each of the four cases is based on at least one hour of data, and in one case, on two hours of data.

Data reduction was handled automatically, two cells at a time, by converting magnetic tape information to punched paper tape using the traffic data reduction system developed by the Port of New York Authority (1), based on a similar system developed by General Motors Research Laboratories (3). Punched paper tape was converted to IBM cards and the data were then processed through various error correcting and computational programs, using the IBM 7070 computer.

For traffic passing each point, the computer output states for each vehicle:

1. Time (to the nearest hundredth of a second) at which it entered the trap;
2. Headway time between it and the vehicle ahead;
3. Headway in feet to the vehicle ahead;
4. Velocity in feet per second;
5. Velocity of that vehicle relative to the vehicle ahead in feet per second;
6. Length of the vehicle; and
7. Whether the vehicle was accelerating, decelerating, or maintaining constant speed.

The IBM output also provides a number of computed parameters for each vehicle or group of vehicles, including virtual density, virtual flow, average speed, average density, and average flow.

Measurement of the length of each vehicle passing a detection point is a particularly important element of this system. Observing the sequence of vehicle lengths passing one point and matching with the same sequence of vehicle lengths passing a point downstream later in time, it is possible to state exactly the number of vehicles between the two points at any instant. This measure of section density differs from other measures also labeled "density," which are computed by dividing the number of vehicles passing a point during a fixed time interval by the average speed of that traffic passing that same point. The latter measure indicates the number of vehicles passing over a length of roadway downstream from the measurement point only so long as the vehicles continue to maintain the same speeds as were observed at the measurement point. For control purposes, it is the fact that speeds lessen as traffic passes through the bottleneck section which is of most interest. It is possible to estimate section density through analysis of point densities so long as traffic conditions are fluid, but the method of deriving and averaging speed and flow measures to compute an accurate density measure requires the section measurements as a control (4).

Consideration of the number of vehicles present over the length of roadway used is an important element in distinguishing this study from others (e.g. 5), which have used point density measures. It was of particular interest to evaluate the extent to which section density could be used as a control parameter leading to more stable and higher peak traffic production through the output bottleneck.

## OUTPUT FLOW CONDITIONS

Figure 2 summarizes the four output flow conditions; 5-min moving averages of output flow are plotted against a reference value of 21 veh/min. In Case 1, a morning peak period with 19.8 percent commercial traffic, there is considerable fluctuation above and below the reference value of 21 veh/min. In Case 2, with 0.9 percent commercial traffic (an afternoon peak period), the output flow is remarkably consistent. For 30 min (from 10 min to 40 min after the beginning of the experiment), the output flow was consistently 20 veh/min. At no time did the flow average more than 22 veh/min, and in general, this consistent flow is relatively low.

Much higher output flow was observed in Case 3, also an afternoon peak period with similar composition of traffic as in Case 2 (only 0.4 percent commercial). For most of the time during the 90 min of the experiment, output flow was consistently higher than 22 veh/min. However, during the last 30 min, output flow decreased considerably, breaking the 22 veh/min barrier only for 6 min.

Case 4 (with 1.7 percent commercial traffic) demonstrates the operation of an early automatic system limiting input flow based on output speeds. Severe oscillations in output flow are evident. While the level of output flow at the high values usually exceeded 22 veh/min, the low values were considerably more frequent, and in this particular example the average output flow was quite low.

While the output flow is seen to vary considerably in each of these four cases, analysis was made to determine the extent to which these four different patterns could be explained by a consistent relationship of section density and output speed, as discussed in detail for each of the following cases.

### Case 1: Uncongested Flow

Figure 3 shows Case 1—AM uncontrolled 5-min moving averages of input flow, output flow, section density and output speed. Input flow fluctuates considerably between 17 and 23 veh/min, with no apparent regularity. The trace of output flow shows a pattern very similar to that of the input flow, but occurring 3 to 4 min later. This suggests that traffic conditions are fluid, and therefore, section density will be less than critical.

Section density on a 5-min moving average is at all times less than 50 veh/mi, at some times dropping to 35 veh/mi (Fig. 3). Most of the time the 5-min moving average of output speed was above 30 fps, and on some occasions rose above 45 fps. There also appears to be a tendency for output speed to be an inverse function of section density, displaced by 2 to 3 min.

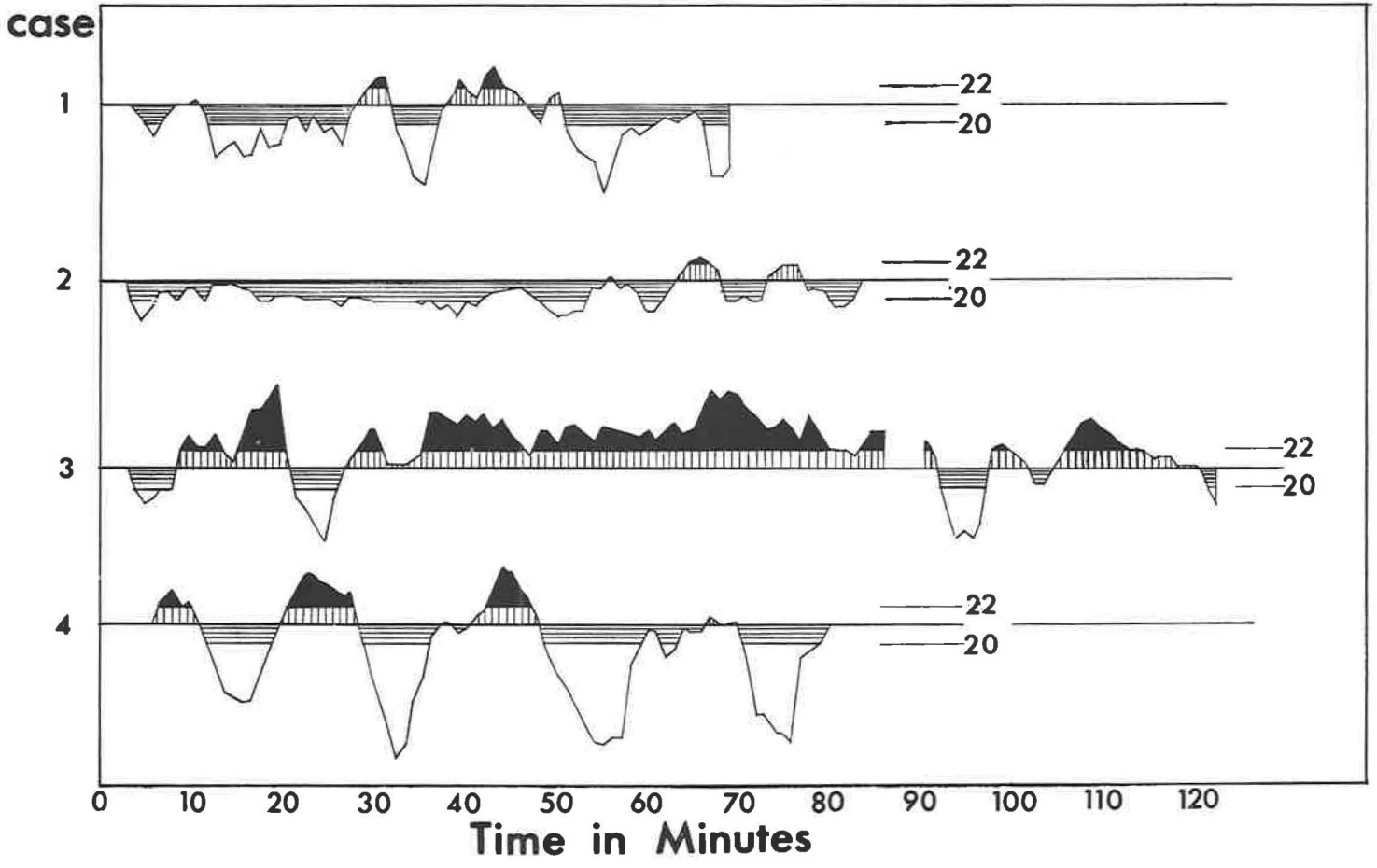



Figure 2. Output flow comparison, 4 cases.

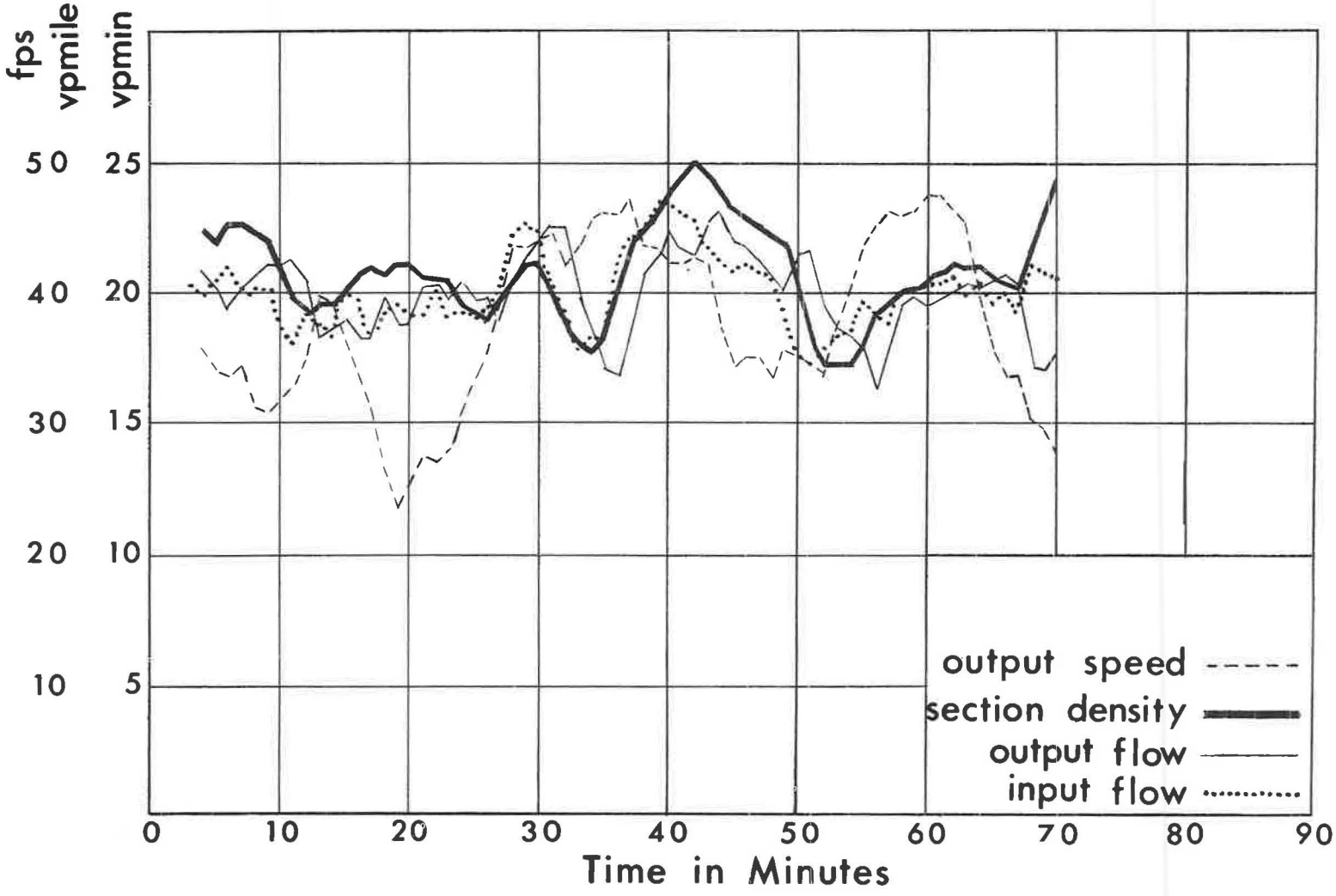



Figure 3. Case 1—AM uncontrolled 5-min moving averages: input flow, output flow, section density, and output speed.

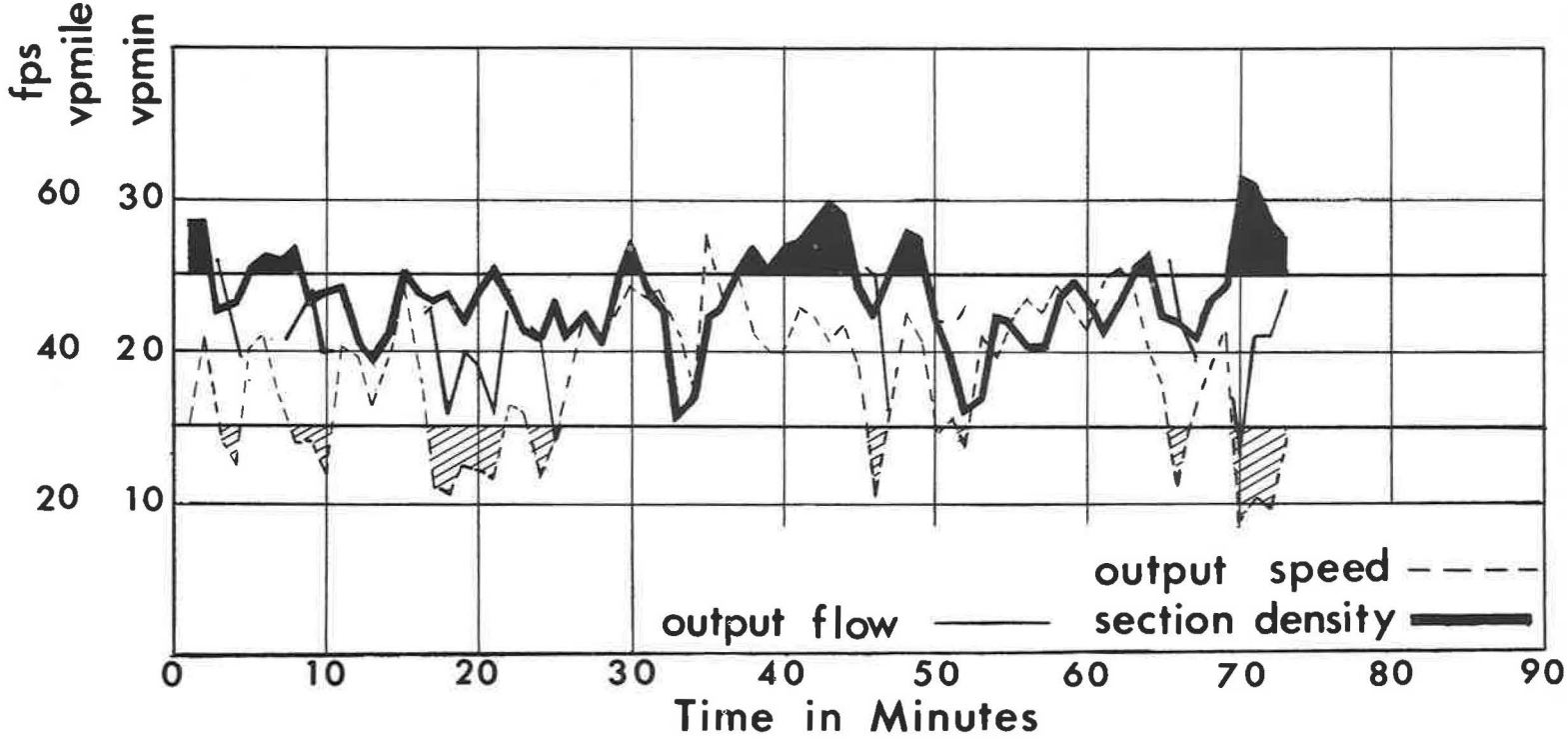



Figure 4. Case 1—AM uncontrolled 1-min data: section density plus output speed, density over critical—velocity below critical, and pattern of output flow.

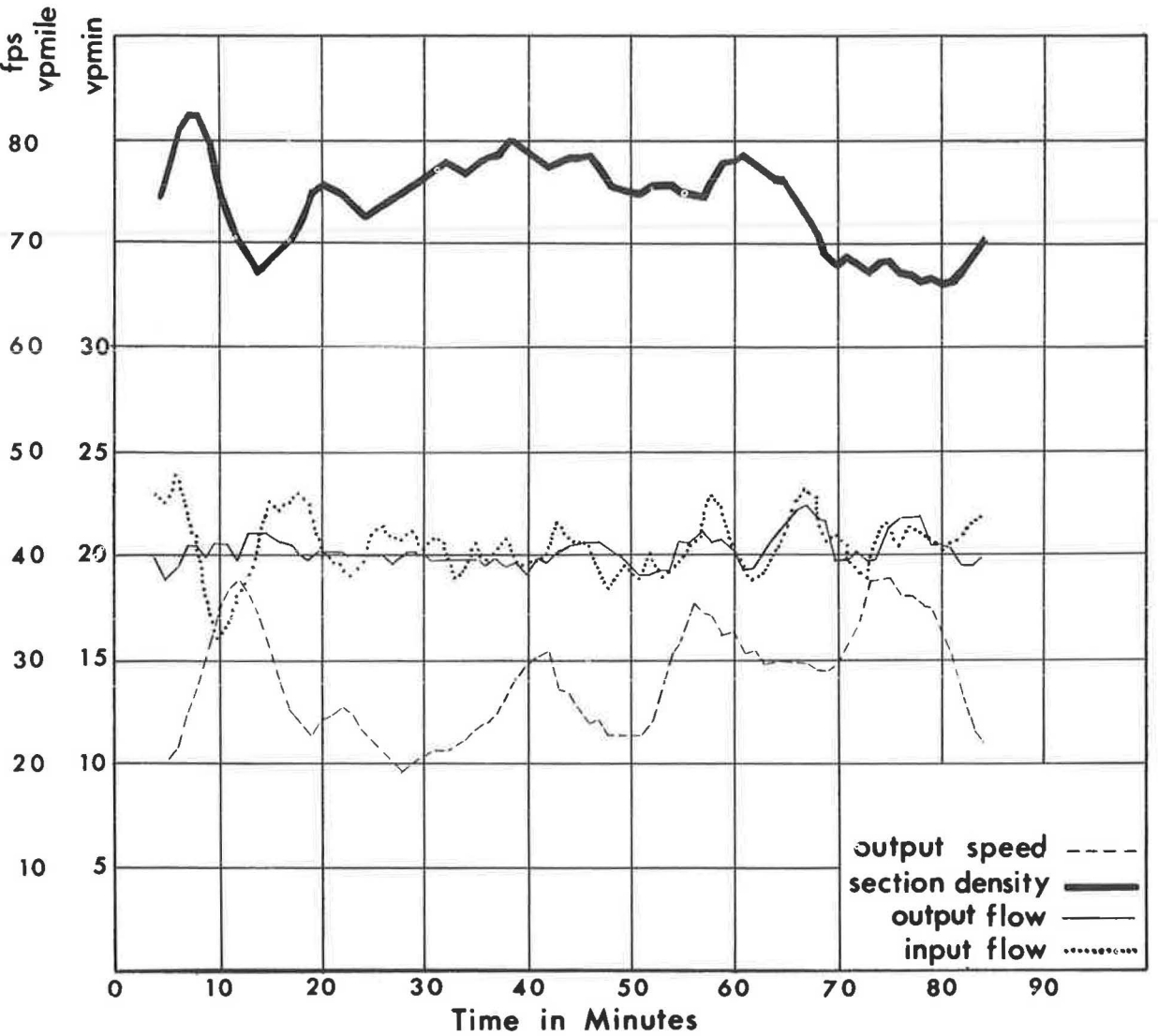



Figure 5. Case 2—PM uncontrolled, high density 5-min moving averages: input flow, output flow, section density, and output speed.

The four traces typify conditions when the tunnel is processing all the traffic entering it without congestion. In terms of vehicles per minute handled throughout the study period, this first case ranks third among the four considered. However, when the comparison is based on the number of vehicle-feet handled rather than the absolute number of vehicles, this case becomes the second most productive. This may or may not be an optimum adjustment for the commercial traffic handled in this example, but it indicates that relatively high flows can be achieved when tunnel traffic conditions are fluid.

Despite the general fluidity of traffic during this example, however, examination of the output speeds averaged over 1-min intervals shows (Fig. 4) several occasions when output speeds drop below 30 fps. Section densities on a minute-by-minute basis were remarkably consistent, but several instances were observed where the 1-min densities exceed 50 veh/mi. Since in previous studies (6, 7) these two levels of 30 fps and 50 veh/mi have been identified as levels at which traffic production might begin to decrease,

the time relationship between these apparently critical density and speed levels was examined next. The shaded and solid areas of Figure 4 emphasize the relationship between high densities and low speeds; in virtually every case a density of about 50 veh/mi was followed in a few minutes by an output speed less than 30 fps.

Considering only those times when output speeds drop below 30 fps, the pattern of output flow can be seen in Figure 4. In nearly every case, output flow decreases while speeds are below 30 fps. However, the pattern occurring at 70 min shows that this is not a completely regular phenomenon since, despite the low output speeds occurring then, section densities were sufficiently high as to provide an increase in output flow.

Case 2: Congested Flow

What probably would have happened in Case 1 if input flow had remained high enough following the end of the experiment to keep section densities above 50 veh/mi is shown in Case 2.

Figure 5 shows an input flow pattern with fluctuations generally between 17 and 23 veh/min, similar to the range in Case 1. However, output flow exhibits none of the fluctuations observed in the input flow. There is remarkable consistency at 20 veh/min, with nearly every flow contained in a band between 19 and 21 veh/min. This consistency in output flow suggests that the road section is operating under pressure—an ample supply of traffic is being processed through the output bottleneck at a uniform and relatively low rate.

The trace of section density (Fig. 5) confirms that there is a high number of vehicles in the tunnel—nearly always above 70 veh/mi. This affects output speeds which are nearly always less than 30 fps, and never reach 40 fps.

To determine whether these high density levels were accompanied by shockwaves as found in previous studies (8), the pattern of speeds throughout the tunnel was examined. Figure 6 shows the 1-min averages of speed at each of the tunnel points, including station 5 just beyond the bottleneck. Zero speed reference is plotted on the ordinate at a location analogous to the geographic location in the tunnel. While the speeds at station 4 are generally in the 20 to 40 fps range, speeds at station 3 drop below 20 fps on several occasions. Similarly, speeds below 20 fps are observed periodically at station 2; at station 1, speeds dropped below 10 fps on some occasions. Furthermore, it is apparent that the low speeds at each of these stations are related to low speeds at the other stations by straight lines of similar and negative slope. Therefore, Figure 6 demonstrates that the congested conditions shown in Case 2 were accompanied by shockwaves generating from the bottleneck section, between stations 4 and 5 at frequent intervals and reflecting back to the tunnel entrance.

Case 3: High Production, Fluid Flow

Exceptionally high flow through this tunnel roadway section is exhibited in Case 3. For 90 min, flow averaged 22.6 veh/min or 1,356 veh per lane hour, compared with the usual output of 1,150. The highest hour flow recorded for this lane in the last two years is 1,402 vehicles—less than 50 vehicles higher than the flow shown in Figure 7.

The pattern of input flow (Fig. 7) exhibits the fluctuation typical of the input flow patterns shown in Cases 1 and 2, but at a generally higher level and exceeding 25 veh/min on several occasions. The output flow pattern appears to follow the input flow pattern generally for the first 40 min, but then becomes more consistent than the input flow pattern. After the first 60 min of the experiment, the input flow pattern exhibited several peaks which were not matched by the output flow pattern. After the first 90 min, the output flow pattern dropped markedly for 5 to 6 min and then returned to the above 20-veh/min level. However, in contrast to the average output flow of 22.6 veh/min for the first 90 min, the last 30 min of this experiment had an average output flow of only 21 veh/min, or 6 percent below the previous flow.

Section densities exhibit a particularly interesting pattern during this experiment (Fig. 7); they were generally below 50 and averaged near 40 veh/mi for the first 60 min of the experiment. Then, however, as input flow consistently exceeded output flow, densities climbed rapidly above the 50 veh/mi critical level and up to a level of 80.

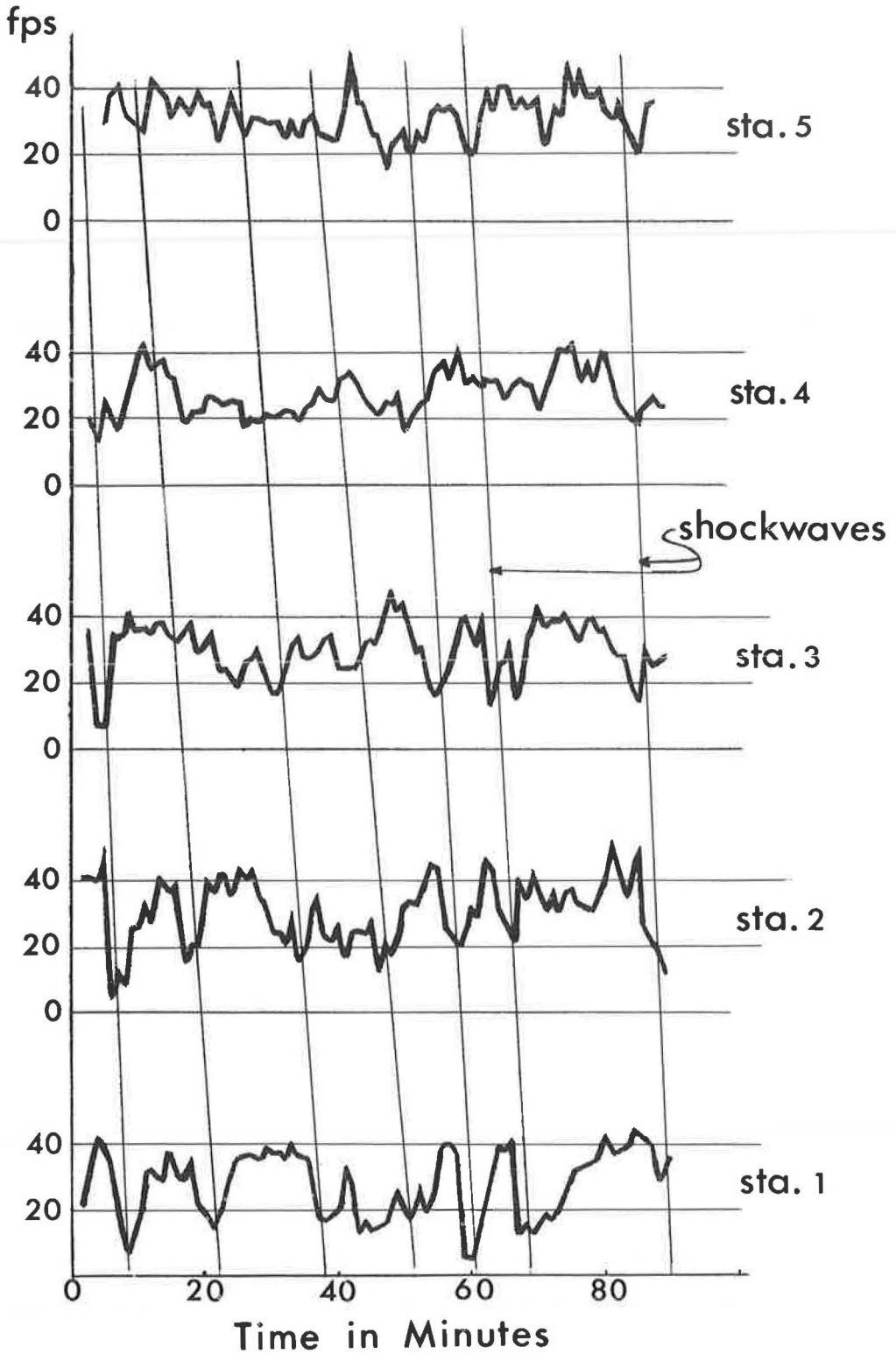



Figure 6. Case 2—PM uncontrolled, high density 1-min data: shockwave occurrence.

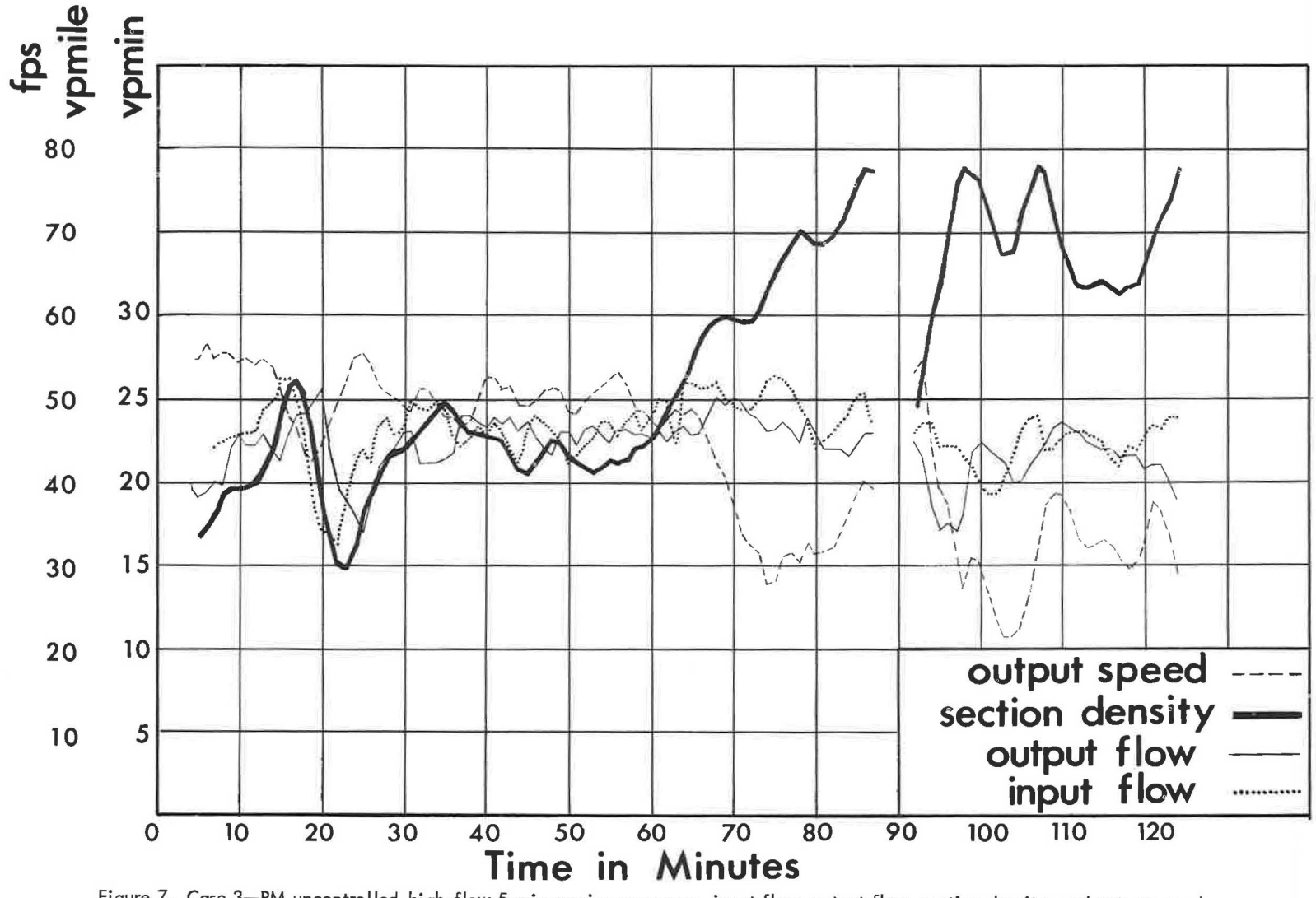



Figure 7. Case 3—PM uncontrolled, high-flow 5-min moving averages: input flow, output flow, section density, and output speed.

The need to change the magnetic recording tape caused a loss of data for the 5-min interval from 87 to 92 min after the start of the experiment. When the measurements resumed there had been a sharp drop in section density. This had to be caused by a drop of input flow relative to output flow. Examining the pattern of densities at the several stations within the tunnel in these critical minutes between 80 and 86, Figure 4 of Crowley and Greenberg (6) shows that a shockwave had generated and was moving back to the entrance. Apparently then, the shockwave caused a sharp drop in input flow during the missing minutes from 87 to 92. But also, when the measurements were resumed, output flow dropped markedly again causing a sharp rise in section density. For the remainder of the experiment, section density remained above 50 and generally near 70 veh/mi.

The pattern of output speeds (Fig. 7) sheds more light on the condition of the tunnel during the missing data. But first, it should be noted that all through the first 60 min of the experiment when output flow was high, output speeds were also high, averaging 50 fps. Then, during minutes 60 through 90 when density climbed from below critical to more than 75, output speeds dropped rapidly and went below the critical level of 30 fps when density went over 60 veh/mi.

Immediately after the measurements were resumed at minute 92, the output speeds were again quite high—more than 50 fps. This suggests that in addition to the shockwave causing a depression on input flow, there may have been some other event which interrupted input flow and allowed densities inside the tunnel to drop below the 50-veh/mi level when data were not being gathered. In any case, the high output speed dropped very rapidly, and for the 10 min between minutes 96 and 106, output speeds were below critical. Output flow was also low, and remained generally lower than during the first 90 min of the measurements.

## Case 4: Controlled Flow

The first three cases generally confirmed the findings of earlier tunnel experiments (1), which showed that output flows could be improved by limiting input flows as necessary to keep output speeds above 30 fps. As described by Foote (1), a control device was built to limit input flow automatically when output speeds drop below 30 fps. On the basis of measurements over several weeks, it was found that the computer did raise the level of output flow, but only by about 2 percent rather than the 5 percent found when control was exercised directly by the experimenters. To study the action of the control device on the traffic stream, measurements of the same type described in Cases 1, 2 and 3 were taken. They showed clearly that the computer during this particular experiment was maintaining an excessively fluid traffic movement through the tunnel.

Of greater immediate interest, however, is the clear demonstration in Figure 8 of the systematic relationship among the four parameters being analyzed. The control system caused oscillations among input and output flows, section densities and speeds, with an approximately 20-min cycle period. Four cycles are observed during 80 min. While these four parameters varied in nearly identical period and amplitude, they were not in phase. Taking minimum output speed at $t = 0$ min, minimum input flow generally occurs at $t = 1.5$ min; minimum section density occurs at $t = 2.25$ min; and minimum output flow occurs at $t = 6$ min. In each case where output speed dropped below 30 fps, output flow decreased.

Input flow fluctuated in a cyclical pattern between highs of 22 to 24 veh/min and lows of 13 to 15 veh/min (Fig. 8). Output flow indicated the same pattern as input flow, generally 4 to 5 min later. Section densities fluctuate on the same cycle, between highs of 52 to 58 veh/mi and lows of 29 to 32 veh/mi. And, output speeds also fluctuate on the same cycle between highs of 53 to 57 fps and lows of 18 to 30 fps.

In this automatic control system, input flow is limited whenever six or more vehicles are observed in any minute traveling through the output section at speeds less than 30 fps. The minutes in which input flow is limited because of low output speeds are shown in Figure 8 along the 10 fps coordinate. It is of special interest that the uncontrolled input flow is in every case high enough, when combined with the reduced output flow caused by the excessively fluid traffic, to bring densities rapidly above the critical 50-veh/mi level and cause output speeds to drop rapidly. Stated differently, clearing the tunnel out by excessively limiting the input did not result in a generally stable state with

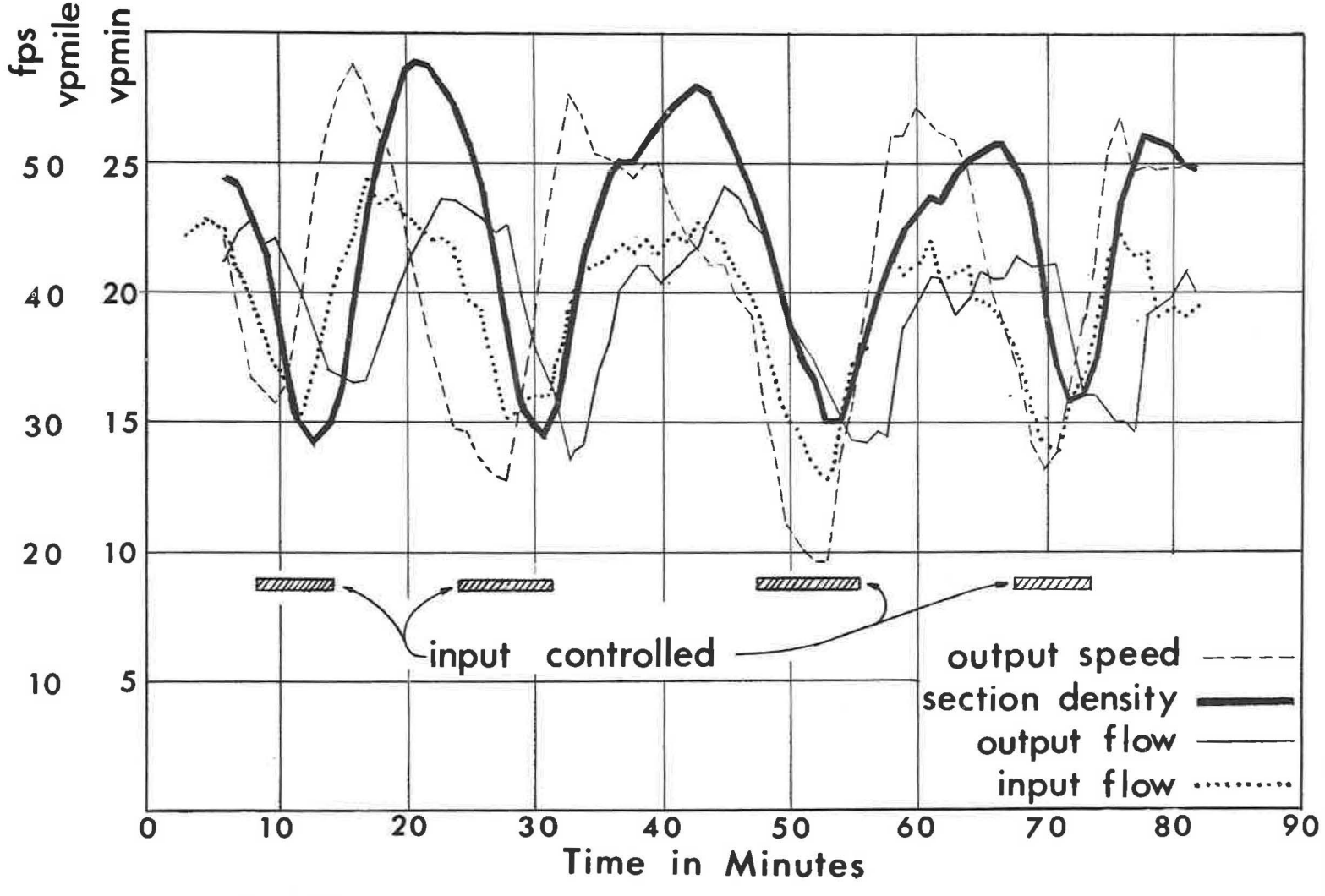



Figure 8. Case 4—PM controlled 5-min moving averages: input flow, output flow, section density, and output speed.

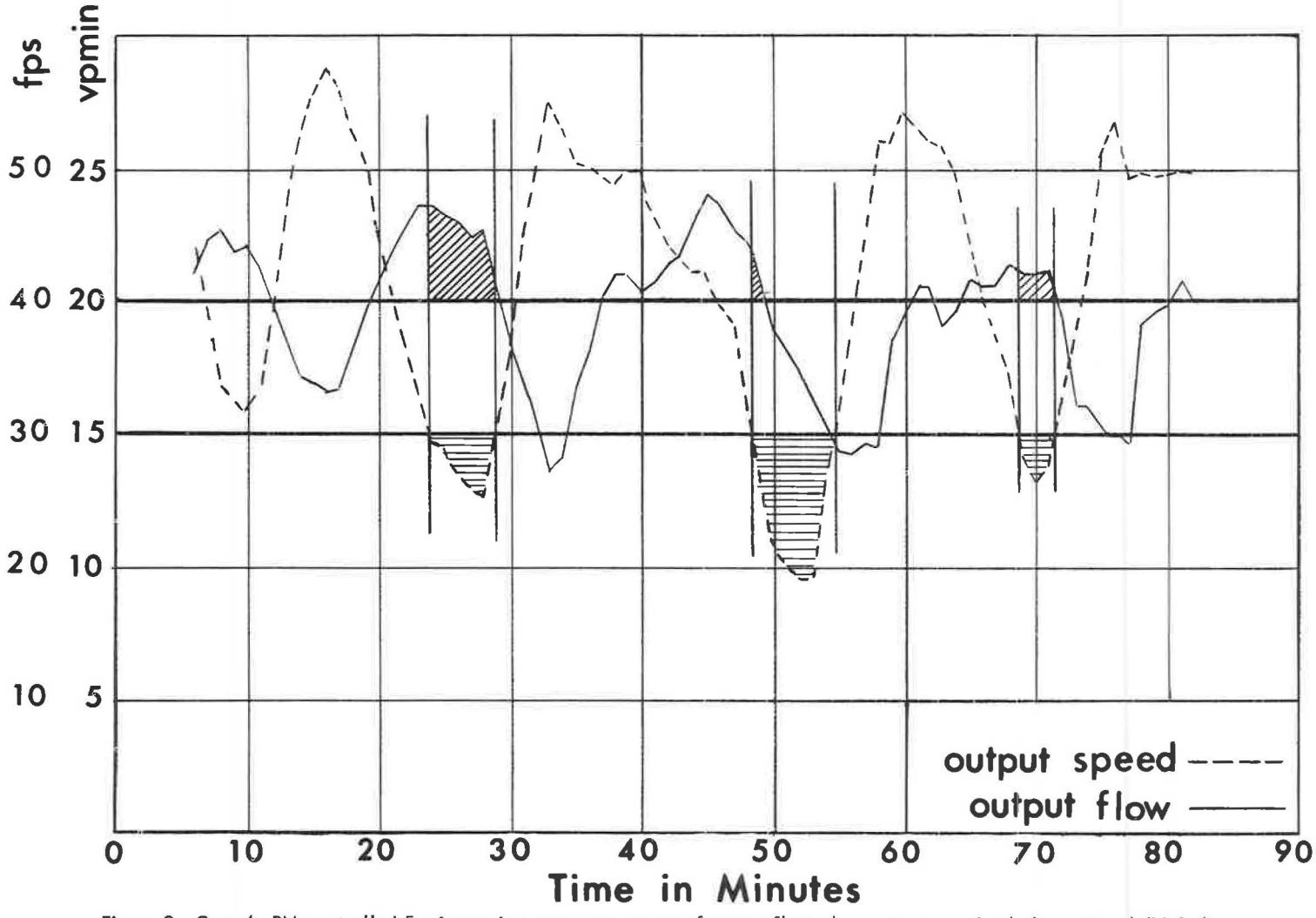



Figure 9. Case 4—PM controlled 5-min moving averages: status of output flow when output speed is below critical (30 fps).

subsequent gradual buildup of densities. Rather, it resulted repeatedly in quite a rapid decay of tunnel conditions and the need generally within a 15-min interval to limit input flow again.

These measurements make clear that the control device was too late in beginning its action on the tunnel traffic stream, and that it continued to restrict traffic for too long a period of time.

The relation between output flow and output speed is of particular interest, since the control of flow is based on the previous finding that output flow is reduced by low output speeds. Figure 9 shows the relationship between these two parameters, and it is clear that in this experiment most of the drop in output flow occurred when speeds were rising rapidly because there was too little traffic reaching the output section. However, the start of the drop in output flow in every case occurred when output speeds dropped below the critical level. At minute 8, output flow is at a high of 23 veh/min, while speed is approaching a low. By minute 11, while speed is passing through the low point, output flow has definitely begun to drop. At minute 24, when output speed dropped again below 30 fps, output flow was at a high of 24. But by minute 30, when the output flow again rose above 30 fps, output flow had dropped to 20 veh/min. At minute 48, when output speed next went below 30 fps, output flow was at 22 veh/min. But, by minute 55, when output speed regained the 30 fps level, output flow had dropped to 15 veh/min. On the last cycle beginning at minute 68, when output speed dropped below 30, output flow was 21 veh/min. By minute 72, when output speeds began to rise, output flow was dropping again.

## CONTROL IMPROVEMENTS

These four cases confirm the desirability of maintaining fluid traffic movement, with speeds above 30 fps and densities below 50 veh/mi, in order to obtain good output flow levels and prevent shockwaves. The cases also show consistently high section densities are followed by low output speeds and a decrease in output flow; therefore, it is clear that a direct measure of section density can be used to predict output speeds. With the ability to predict output speeds, it would be possible to initiate input flow restrictions sooner, and also to end them more quickly when densities have been restored to optimum levels. These steps should eliminate or, at least, greatly reduce the period and amplitude of the oscillations observed in Case 4, and permit maintaining tunnel traffic at near optimum conditions with more stability.

Modifications to the existing controller to limit restrictions on input when speeds midway between the entrance and the bottleneck become too high are discussed by Duckstein (4). Of more fundamental importance however, a device to measure section density continuously as an on-line control parameter is now being developed by the Port of New York Authority staff, and experiments using this control device will be undertaken.

## REFERENCES

1. Foote, R. S. Single Lane Traffic Flow Control. Proc. Second International Symposium on Theory of Traffic Flow, 1963.
2. Foote, R. S., Crowley, K. W., Gonseth, A. T. Instrumentation for Improved Traffic Flow. ITE Proc., 1960.
3. Edie, L. C., Foote, R. S., Herman, R., and Rothery, R. Analysis of Single Lane Traffic Flow. Traffic Engineering, Jan. 1963.
4. Duckstein, L. Control of Traffic in Tunnels to Maximize Flow. Highway Research Record 154, 1966.
5. Weinberg, M. Traffic Surveillance and Means of Communicating with Drivers. NCHRP Report No. 9, HRB, 1965.
6. Crowley, K., and Greenberg, I. Holland Tunnel Study Aids Efficient Increase of Tube's Use. Traffic Engineering, March 1965.
7. Edie, L. C., and Foote, R. S. Experiments in Single Lane Flow in Tunnels. Theory of Traffic Flow, R. Herman, ed., D. Van Nostrand Co., 1961.
8. Edie, L. C., and Foote, R. S. Effect of Shock Waves on Tunnel Traffic Flow. Pub. HRB Proc., Vol. 39, 1960.

# Traffic Flow Data Acquisition Using Magnetic-Loop Vehicle Detectors

STAN STERN, System Development Corporation, Santa Monica, Calif.

One means of acquiring validation data for digital computer simulation of traffic flow in a freeway diamond interchange will be through the use of magnetic-loop vehicle detectors. Actual interchange performance will be monitored by direct measurement of vehicle parameters (count, size, and velocity), leading eventually, through the validation process, to a realistic simulation model.

Techniques were developed to extend existing loop-detector capabilities, permitting direct measurement of vehicle parameters (measurements were available heretofore only on a statistical basis) and experiments were conducted to determine detector/sensor-loop response characteristics.

A practical sensor-loop configuration, which was developed for the validation runs and tested in controlled experiments, provided an accurate picture of the over-all traffic pattern.

Measurement accuracy was investigated, considering the effects of varying vehicle height and of closely following vehicles (tailgating), such as might be expected in heavy traffic. Measurement errors due to tailgating were attributable to recovery-time limitations inherent in the loop-detector circuit.

The techniques described in this paper can be used, employing commercial loop-detectors, for traffic applications in which it is desired to determine vehicle velocity and/or size directly in addition to vehicle counting.

•THIS PAPER discusses one aspect of a traffic flow study being conducted at the System Development Corporation: the development of a technique to obtain direct vehicle parameter measurements using magnetic-loop vehicle detectors.

The measurements will provide data to be used in validating the digital computer simulation of vehicular traffic flow in a diamond interchange between a freeway and an arterial street. The general aim of the research program, which also includes mathematical analysis of traffic models related to the diamond interchange problem, is to gain an understanding of the phenomena involved in traffic flow on surface arteries, freeways, and their interconnecting ramps.

In the validation process, data acquired by the loop-detectors and by synchronized motion pictures taken from a hovering helicopter will be compared in the laboratory with the computer simulation outputs and, if the computer model does not perform realistically, indicated changes will be incorporated into the model and the entire validation process repeated. This iterative procedure is expected eventually to result in a realistic simulation model. The techniques employed to utilize loop-detectors for validation data acquisition, and the results obtained, are the subjects of this paper.

The magnetic-loop vehicle detector was originally developed, and applied in practice, for traffic counting and signal control (1). This device operates in conjunction with an inductive wire loop placed in the traffic lane. The loop leads are brought out to the

Paper sponsored by Committee on Characteristics of Traffic Flow and presented at the 45th Annual Meeting.

roadway and connected to the vehicle detector. The loop excitation signal, generated by an oscillator in the detector, induces a magnetic field surrounding the loop. The passage of a vehicle through the magnetic field is sensed and a relay is activated. The duration of relay closure (which ends when the vehicle leaves the magnetic field) varies directly with the length of the vehicle and inversely with its velocity. A detailed description of the loop-detector is given in Appendix A.

Vehicle velocity and size cannot be determined independently using a single loop-detector. Determination of traffic speed has been obtained in practice on a statistical basis, by assuming an average car length, with the duration of relay closure (pulse width) giving an indication of traffic speed.

To obtain the required computer validation data, it was necessary to develop a technique for applying the loop-detector to measure vehicle parameters directly. This objective was accomplished by employing the detectors in pairs and using a special (but simple) calibration procedure to provide the required velocity measurement accuracy using unmodified commercial detectors.

It is planned to operate loop-detector pairs taped to the roadway at various positions throughout the diamond interchange to measure vehicle count, length (vehicle classification), and velocity. The outputs of the detectors will be sampled periodically by a multiplexer whose output will be recorded by an instrumentation tape recorder. A timing track and synchronizing pulses will also be recorded in addition to a verbal annotation. The recorded data will be returned to the laboratory where it will be reduced from the tape recordings and entered into the Philco 2000 digital computer using an input channel of the RL101 Real Time Input-Output Transducer.

Development of a loop configuration for the freeway validation tests was based on theoretical considerations confirmed empirically by means of controlled experiments, and verified in actual traffic conditions. The loop configuration chosen is based on the dimensions of the diamond interchange lanes.

## THEORY OF PARAMETER MEASUREMENTS

The technique we employed to measure vehicle length and velocity uses a basic sensor-unit consisting of a pair of loops, separated by a known distance (parallel to the direction of traffic), and connected each to its own detector. As a vehicle passes over

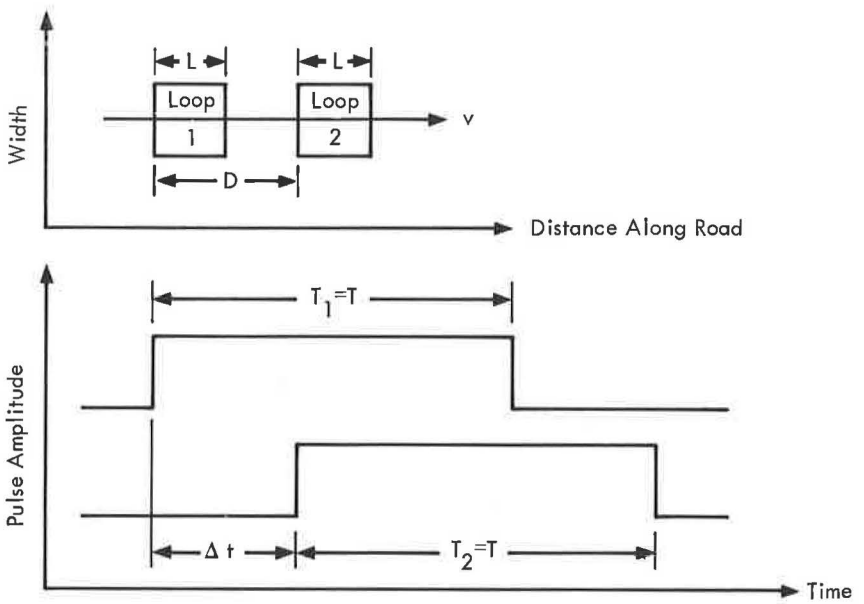



Figure 1. Loop-pair (ideal response).

the loop-pair each detector produces a pulse. Vehicle velocity is computed from the time delay between pulses and vehicle length from the pulse width.

Suppose a vehicle of length a traverses the loop-pair (Fig. 1) with constant velocity, v. The loop length is L and the loop-pair separation is D. The magnetic field extends slightly beyond the loop boundaries and the effective field length L′ depends on the sensitivity of the detector. Each of the two detectors produces a pulse, and the time delay between pulses is denoted by $\Delta t$.

Vehicle velocity is determined by dividing the loop-pair separation by the time delay between pulses.

$$v = \frac{D}{\Delta t} \tag{1}$$

The pulse width, T, is the time required for the vehicle to traverse the field, L′, along its entire length, a. Hence,

$$T = \frac{L' + a}{v} \tag{2}$$

Substituting Eq. 1 in Eq. 2

$$a = D \frac{T}{\Delta t} - L' \tag{3}$$

Equations 1 and 3 express vehicle velocity and length in terms of pulse width and the time delay between pulses. The effective length of field L′ can be taken as equal to the loop length L with little penalty (for vehicles traversing the loop near its center)* since vehicle classification information does not require great precision.

## INITIAL EXPERIMENTS

Initial experiments were conducted using a pair of RCA VeDet vehicle dectors with associated square loops, each consisting of three turns of No. 20 wire. The loop dimensions were 6 ft by 6 ft, as recommended by the equipment distributor.

A photoelectric system was employed as a velocity measurement standard. In operation, a pair of photosensors was placed in positions normal to the vehicle path and parallel to the leading edge of the inductive loops. The light-collimating tube of each of the sensors (which were mounted on tripods) was sighted along the loops' leading edges at a pair of light sources located on the other side of the traffic lane. Figure 2 shows the experimental setup. A vehicle passing over the loop-pair interrupted the light beams from the lamps, activating the photoelectric sensors. The photosensor output pulses were combined by connecting them to the differential inputs of one channel of a two-channel pen recorder, while the vehicle detector pulses were similarly connected to the other recorder channel. The permanent record was analyzed by comparing the two channels, using the optical data as a standard.

This measurement technique was then applied in vehicle velocity tests to determine detector accuracy, using a loop-pair separation of 10 ft and driving the test vehicle over the center of the loops. When each detector was calibrated according to the manufacturer's recommended procedure, the error in the velocity measurement was too great (typically 7 percent), with respect to the optical data.

Further tests conducted to determine the cause of this velocity measurement error revealed that it was due to unequal detector threshold levels, which caused the relative position between loop and vehicle at which the detector activates to be different for the two detectors. Hence, the loop-pair velocity measurement is in error by an amount that depends on the difference in the detector threshold levels. This effect is not sur-

*This is explained under the heading "Adjacent Loop-Pairs as a System."

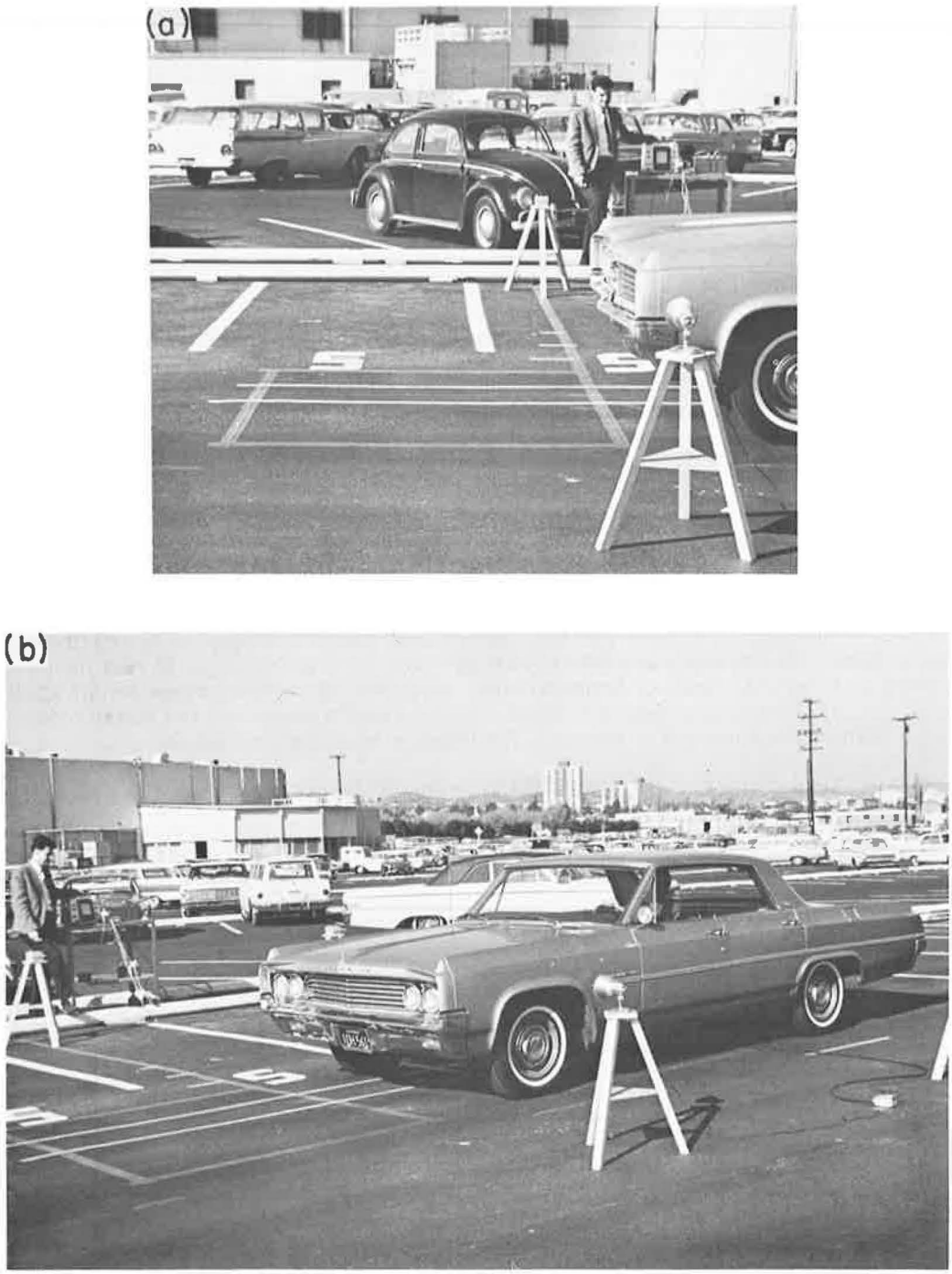



Figure 2. Experimental setup.

prising considering that the detectors were designed for individual vehicle counting, and act independently of one another. No means is provided for detector calibration which uses the loop-pair as a system for velocity measurements. It became apparent, therefore, that a technique must be developed to calibrate a pair of detectors as a system by equalizing their sensitivities. In addition, it was recognized that such a scheme, to be practical, should not employ complex setup procedures.

## CALIBRATION TECHNIQUE

The calibration method developed inserts an external capacitance directly in parallel with the inductive loop. When the capacitance is removed, the tuned circuit experiences a phase shift similar to that experienced when a vehicle enters the loop's magnetic field. The detector sensitivity is adjusted until the relay activates for a narrow range of capacitance values, but not outside that range. The procedure is then repeated for the other detector, similarly establishing its sensitivity setting. The calibration technique is described in Appendix B. This technique was applied in calibrating two pairs of TACDET vehicle detectors. The units tested during this study (RCA VeDet and Fischer and Porter TACDET) are only two of a number of commercially available vehicle detectors, which operate on the same principle. The two units were similar in their operation and circuit configurations. Although no attempt was made to test many different detectors, the techniques described herein should be generally applicable.

Additional velocity tests were conducted using the experimental setup previously described. Typical velocity measurement errors of 2 percent were obtained with respect to the optical standard, using a loop-pair separation of 10 ft. The accuracy of the velocity measurements suffered no observable deterioration during operation at ambient temperature for ½ hr (the time period over which it is planned to acquire validation data). No attempt was made to evaluate long-term accuracy or to test the detectors under a wide range of temperatures. Such testing would be required for applications involving continuous detector usage. These results confirmed the effectiveness of the calibration technique in reducing the velocity measurement error.

## DEPENDENCE OF MEASUREMENTS ON VEHICLE VELOCITY

Experiments were conducted to determine the effects on velocity and pulse-width measurements of varying vehicle velocity. The experiments used a pair of 4-ft square

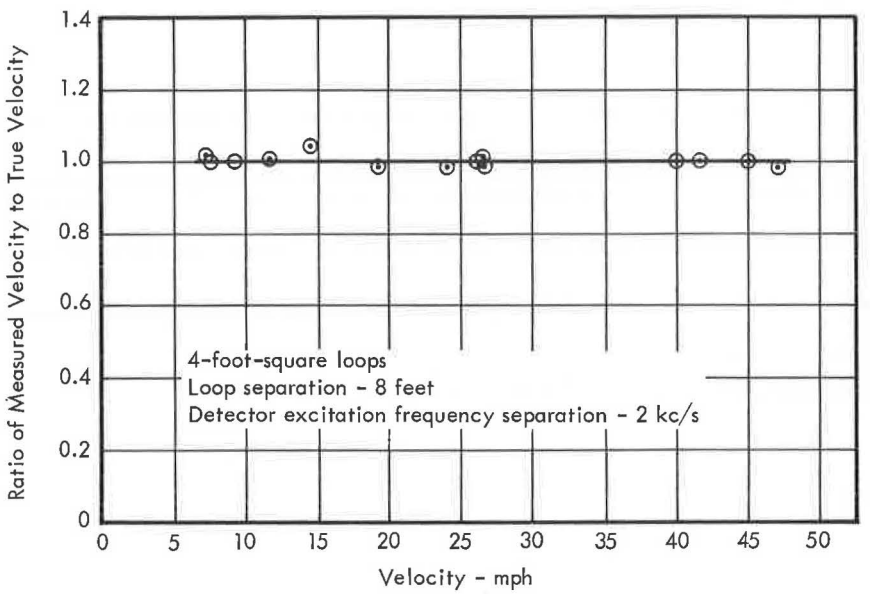



Figure 3. Accuracy of velocity measurement vs velocity.

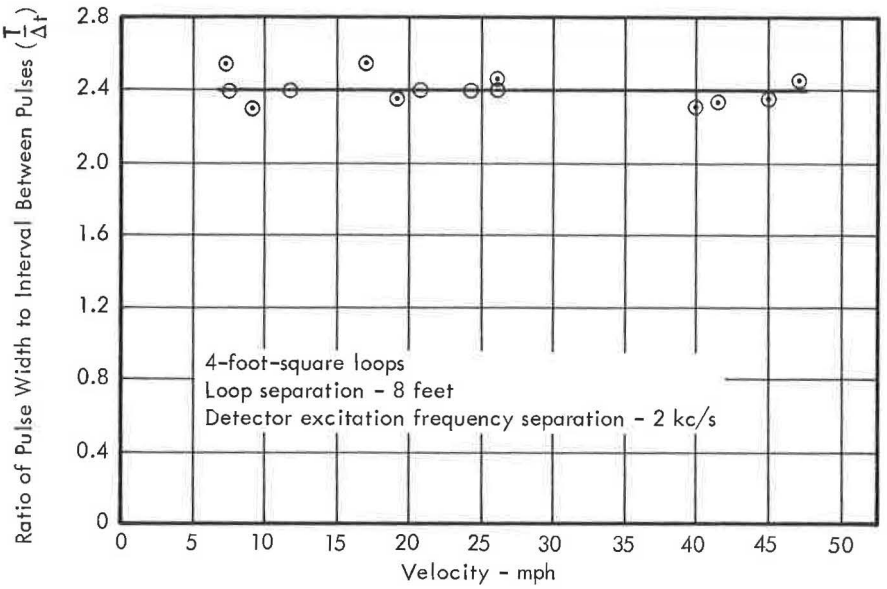



Figure 4. Normalized pulse width vs velocity.

loops separated by 8 ft (from leading edge to leading edge). The vehicle detectors were calibrated and a test vehicle was driven over the center of the loops at various speeds. The photosensor data were used to calculate true vehicle velocity.

Figure 3 plots the ratio of measured velocity to true velocity as a function of true velocity; the accuracy of the velocity measurements is independent of vehicle velocity.

Figure 4 plots the ratio of measured pulse width to the interval between pulses $T/\Delta t$ as a function of vehicle velocity. It is apparent that the ratio $T/\Delta t$ does not vary with velocity. This is consistent with the theory, as can be seen by solving Eq. 3 for $T/\Delta t$.

$$\frac{T}{\Delta t} = \frac{a + L'}{D} \tag{4}$$

Substituting $L' = L = 4$ ft, a (the overall length of the experimental vehicle as given in the manufacturer's specifications) = 15.3 ft, and $D = 8$ ft gives $T/\Delta t = 2.4$, which, it is evident, provides a nice fit for the data points. Hence the measured pulse width agreed with the expected value at all velocities. Christensen and Hewton (1) tested two different vehicle detectors and found that the actual pulse width differed from the calculated (expected) value by a factor that depends on vehicle velocity. The calibration curves they plot (ratio of measured to calculated pulse width vs velocity) for each detector have appreciable slopes. A plot of measured to expected pulse width ratio vs velocity for the vehicle detector tested at SDC would be a horizontal line at ratio unity, permitting the direct determination of vehicle length without correction of the measured data.

These results (Figs. 3 and 4) demonstrate that vehicle velocity and length can be determined directly from the equations previously presented for a vehicle crossing the loops at their centers. The curves reveal no significant effects due to varying vehicle velocity for the velocity range investigated. No special calibration curves are required to provide for different vehicle velocities.

## LOOP CONFIGURATION

Experiments were conducted using various loop sizes and loop-pair separations. The final configuration (Fig. 5) comprises a pair of 4-ft square loops, consisting of 4

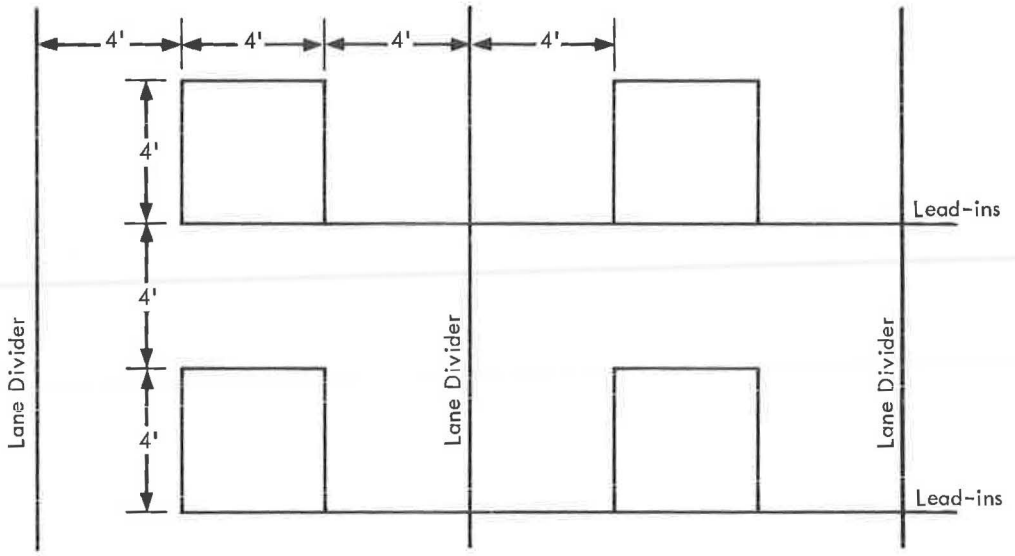



Figure 5. Loop configuration (adjacent 12-ft lanes).

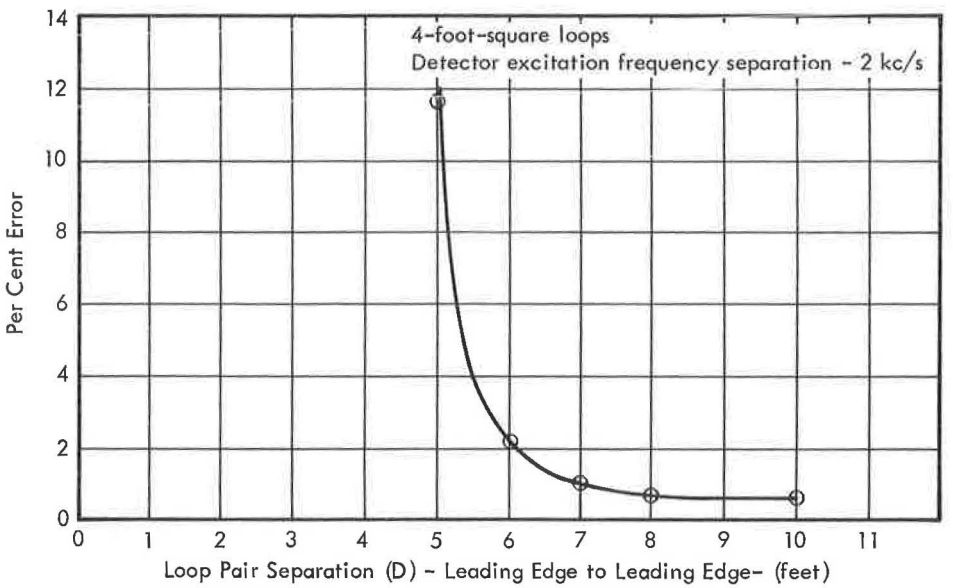



Figure 6. Accuracy of loop-pair velocity measurement vs loop separation.

turns of No. 20 wire, located in the center of each 12-ft lane and separated (leading edge to leading edge) by 8 ft.

The distance between loop-pairs in adjacent lanes (8 ft between adjacent conductors) was found to yield satisfactory results with respect to the numbers of missed counts (vehicles passing undetected between lanes) and double counts (vehicles being detected by adjacent loop-detector pairs as separate vehicles in each lane).

Loops were located at the lane center because parameter measurement is optimized when the vehicles cross the loops at their centers. It is expected that the great majority of vehicles will most often occupy the center of the freeway lanes. Therefore, placing the loops at the center of the lanes will yield the best measurements.

The width of the loops (4 ft) was dictated by the combination of lane-center location in the 12-ft wide freeway lanes and 8-ft adjacent loop spacing. Loop length should be no smaller than loop width to maintain the height of the magnetic field. If the height of the field (which depends on the smaller loop dimension) is too small, the reliability of counting high-slung vehicles (tractor-trailers, etc.) is impaired.

It is desirable that the loop-pair separation be small enough to detect lane-changing vehicles accurately. Close-spacing, furthermore, provides a more nearly instantaneous velocity indication (the vehicle has less time to change its speed in a shorter distance). However, when the loops become too closely spaced, the magnetic interaction between them degrades parameter measurements.

Figure 6 shows the variation in the velocity measurement error as a function of loop-pair separation. Nominal vehicle speed was 20 mph and detector excitation signal frequencies were 2 kc/sec apart. The error was relatively constant for large loop separations but when the separation was reduced to 5 ft (1 ft between adjacent conductors) the error increased to 11.7 percent. Additional data to observe the effect of frequency separation were obtained by operating the detectors 6 kc/sec apart in the 10-ft and 5-ft separation positions. In this case, the change in velocity measurement accuracy was less severe (2 percent error at 5-ft separation) because magnetic coupling was reduced as the circuits were operated farther apart in frequency. An 8-ft separation (leading edge to leading edge) was chosen as a compromise between loop proximity and error minimization.

## ADJACENT LOOP-PAIRS AS A SYSTEM

Experiments were conducted considering adjacent loop-pairs as a system. All 4 lead-in wires for each couple of adjacent loops were run, in proximity, to their respective detectors. Care was taken that adjacent loops and the members of each loop-pair were operated with their detector excitation frequences at least 2 kc/sec apart to avoid interactions. The roadway was divided longitudinally into 1-ft strips and a test vehicle repeatedly driven along the strips, through the system. Figure 7 shows pulse width as a function of vehicle position for a single detector at a vehicle velocity of 25 mph. When the vehicle crossed directly over a loop the pulse width was relatively constant, but as

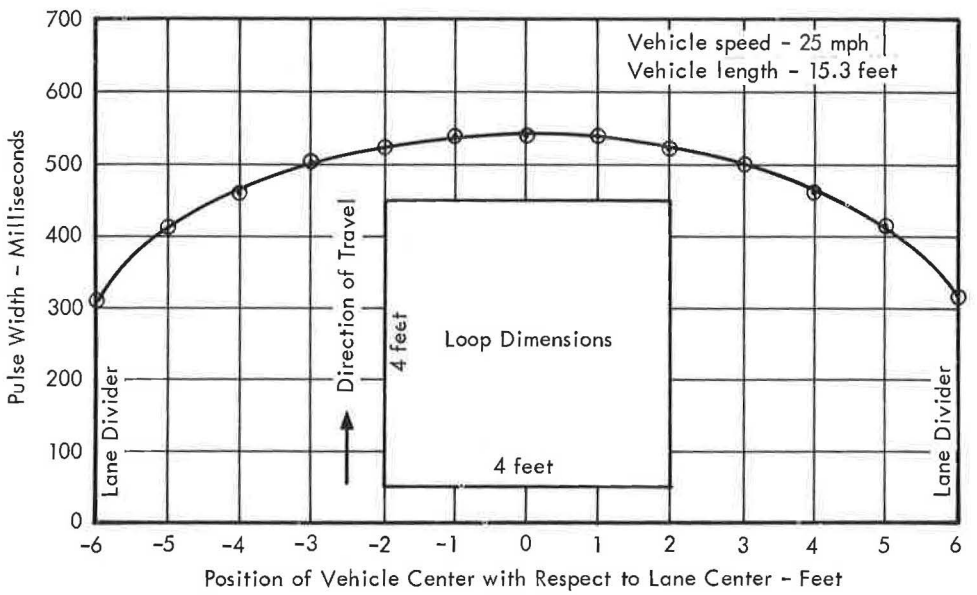



Figure 7. Variation of pulse width with vehicle position in the lane.

it passed farther off the lane-center to the sides of the loop, the magnetic field became narrower and pulse width decreased. (The effective length of field L′ decreased as the vehicle traversed the loop farther off center.) The minimum pulse width occurred when traveling down the lane divider. The mid-lane value of pulse width was consistent with the theory, as can be seen by solving Eq. 2 with L′ = L = 4 ft, a (the overall car length given in the manufacturer's specifications) = 15.3 ft, and v = 36.65 ft/sec (25 mph). The measured pulse width agreed with the calculated value (0.530 sec) within 2 percent. The velocity measurement error with respect to the measurement standard was relatively constant throughout the lane, even approaching the lane divider, because the detectors were so well matched during calibration that the pulses from each detector in the detector-pair remained accurately spaced even though the individual pulse widths varied with vehicle position in the lane. Beyond the lane divider, detector response ceased. Left-hand-lane traffic was not detected by right-hand-lane detectors and vice versa. There was no measurable effect due to magnetic interaction between the closely spaced lead-in wires.

There was a narrow region at the lane divider in which both pairs of adjacent detectors responded. In this region, the vehicle barely covered the adjacent magnetic fields, producing narrow pulses in each of the detectors (corresponding to the end points in Fig. 7). Such a condition will exist, however rarely, for vehicles that occupy a lane-straddling position.

## INTERPRETATION OF DATA

Referring to the previously derived equations, values of D = 8 ft and L = 4 ft give

$$v = \frac{8}{\Delta t} \tag{5}$$

and

$$a = 4\left[2\left(\frac{T}{\Delta t}\right) - 1\right] \tag{6}$$

as the two equations defining vehicle velocity and length. These parameters are determined by measuring the pulse width and the interval between pulses with standard instrumentation techniques and performing the necessary computations.

The first of the two pulses should be used for pulse-width determination, since it corresponds to the interval during which velocity is measured.

## TRAFFIC TESTS

Adjacent loop-pairs were taped to the road surface in each of the 12-ft westbound lanes of Olympic Boulevard in Santa Monica, just outside the SDC facility. The four detectors were calibrated during a brief lull in traffic. Data, including narrative tape recordings, were recorded during several different traffic conditions. Figure 8 shows the test configuration; Figure 9 the vehicle detectors and recording equipment. The recorded data were then analyzed and it was determined that the loop detectors provided an accurate picture of the overall traffic-flow pattern. The following observations were noted.

### Lane Changing

Vehicles changing lanes over the loop-pairs sometimes activated all four of the detectors, and sometimes only three of them. The detectors associated with an active loop-pair reflected the passage of lane-changers by generating pulses of unequal width. The unequal-width pulses corresponded to the different positions at which the two loop-pair members were traversed (e.g., the 5-ft and 4-ft points in Fig. 7). The pulse pattern created by a lane-changing vehicle (a pair of pulses with unequal widths from one detector-pair, accompanied by a single pulse or two likewise dissimilar pulses

Figure 8. Traffic test configuration.

Figure 9. Traffic test equipment.

from the detector-pair associated with the adjacent lane) is characteristic and can be used to recognize lane-changers.

### Lane Straddling

An occasional vehicle traversed the loops riding the centerline too faithfully to be detected as a lane-changer. In this event, a pair of equal-width pulses resulted from the detector-pairs associated with each lane. As previously noted, the pulses resulting from lane-straddling vehicles are of relatively short duration. Furthermore, the right-hand-lane and left-hand-lane pulse-pairs differ from one another in their widths if the vehicle favors one or the other of the adjacent lanes (Fig. 7). Observation of a record of the detector pulses shows a pair of short pulses from one detector-pair accompanied by a pair of even shorter pulses from the adjacent lane's detector-pair. The shorter pulses begin later and end correspondingly earlier than the longer ones. In this case, the characteristic pulse pattern of a lane-straddling vehicle can be distinguished from separate vehicles simultaneously present in each lane.

Any tendency of the lane-straddling vehicle to move horizontally creates an imbalance in the pulse widths of each of the individual detectors, making it difficult to determine whether the vehicle is straddling, or changing, lanes.

### Certain Trucks

The passage of multisection vehicles employing long booms yielded more than one pulse, as though they were separate vehicles following each other closely (tailgating).

### Tailgating

When one vehicle closely followed another, the detector-pairs in all cases detected the presence of both vehicles. Analysis of the data, however, led to the suspicion that the measured lengths of the tailgating vehicles were too small.

## TAILGATING EXPERIMENTS

Experiments were conducted to determine the effects of tailgating on measurements of vehicle length and velocity. Two cars were driven over the center of a loop-pair, both tailgating and individually, at 10-15 mph. A pair of 6-ft square loops separated by 10 ft (leading edge to leading edge) was used. The photoelectric system previously described provided a velocity measurement standard.

Table 1 gives the results of the experiments. The indicated vehicle length was obtained for each vehicle when driven separately over the loops. The vehicles were then

TABLE 1

RESULTS OF TAILGATING EXPERIMENTS

| Run | Vehicle | Measured Vehicle Length (ft) | Length Measurement Error[a] (%) | Total Velocity Measurement Error (%) | Velocity Measurement Error Due to Tailgating (%) |
|---|---|---|---|---|---|
| 1 | A | 16.62 | 0 | -3.8 | — |
| 2 | A | 16.00 | 3.7 | -2.7 | — |
| 1 | B | 13.60 | 10.6 | +6.5 | 10.3 |
| 2 | B | 13.60 | 10.6 | +5.6 | 8.3 |

[a]Indicated vehicle lengths without tailgating (vehicle A = 16.62 ft, vehicle B = 15.22 ft) used for tailgating measurement error comparison.

driven over the loops, car B tailgating car A, in each of two runs. In each run the measured length of the tailgating vehicle differed from its indicated length without tailgating by 10.6 percent.

The error in the velocity measurement for vehicle A (2.7 to 3.8 percent), was due to the slight mismatch in detector sensitivities remaining after calibration. The total velocity measurement error for vehicle B was equal to the algebraic sum of the first error (residual sensitivity mismatch) and the error due to tailgating. The error due to tailgating was then computed from the algebraic difference between the total velocity measurement error and the error due to sensitivity mismatch. The velocity measurement error due to tailgating was found to be 10.3 percent for the first run and 8.3 percent for the second run.

The measurement errors are attributable to recovery time delays in the vehicle detectors. The vehicle length measurement error occurs because the detectors do not have enough time to recover from the passage of the first vehicle before the entrance of the second vehicle. The detectors trigger late, producing shorter-than-normal pulses that are interpreted as a shorter vehicle. Detector recovery time delay is significant despite the fact that the minimum available detector time constant was selected (the so-called pulse mode of operation).

If each detector in the pair had identical operating characteristics, the detectors' recovery time delays would be the same. The velocity measurement, which depends on the time interval between pulses, would then be unaffected. The existence of the additional velocity measurement error (due to tailgating) indicates that there is a mismatch between the detectors' recovery times. These recovery time delay problems are inherent in the circuit design of the vehicle detectors, which were originally intended merely for vehicle-counting applications. To reduce these effects it will be necessary to make circuit modifications.

The time lag between the passage of the rear bumper of car A and the front bumper of car B, the tailgating car, was approximately 0.35 sec in each run as determined by the photosensors. For a typically heavy traffic situation, in which the vehicles trail each other by approximately $1\frac{1}{2}$ to $2\frac{1}{2}$ sec, the measurement errors would be somewhat less, because the detectors would more nearly recover to the quiescent operating state.

The experimental results indicate that the accuracy of both length and velocity measurements will suffer under heavy traffic conditions because of the limited time available for the detectors to recover between the passage of successive vehicles. The measurements can be expected to be approximately 90 percent accurate in extremely heavy traffic, increasing in accuracy for lighter traffic. A higher degree of accuracy in heavy traffic can result only from the employment of improved vehicle detectors. If greater accuracy is required, a detector modification program should be undertaken.

## EFFECTS ON MEASUREMENTS OF VARYING VEHICLE HEIGHT

A question was raised concerning how much the velocity measurement accuracy would be affected by varying vehicle height, such as might occur if a vehicle bounced as it passed over the loops.

Experiments were conducted in which a foreign-made, adjustable-body-height automobile was driven over the center of a loop-pair at a nominal speed of 15 mph with various pre-adjusted body heights. Table 2 gives the results of the experiment. Four different body-height positions were used. The first three are normal driving positions. Position 4, which is normally used only for tire-changing, provided an extreme operating height for the experiment. In each case the pulse width was normalized by multiplying Eq. 2 by the velocity to give the sum $L' + a$.

As the body height was raised, the vehicle entered each loop's magnetic field slightly later and left it slightly earlier, resulting in a decreasing pulse width. This effect was caused by the tapering contour of the loop's magnetic field (variation of $L'$ with vehicle height). Subtracting the length of the loop (6 ft) from $L' + a$ for vehicle length determination shows that the measured vehicle length varies from 15.96 ft ($a_o$) to 14.56 ft, a variation of 8.8 percent in the indicated vehicle length, over the $3\frac{3}{4}$-in. range of body heights. This is indicative of the sort of variation in vehicle length measurement that would be caused by vehicles of differing body heights (i.e., high-slung vs low-slung).

TABLE 2

EFFECTS OF VARYING VEHICLE BODY HEIGHT

| Preset Height Position | Relative Body Height (in.) | Pulse Width Expressed in Feet ($L' + a$) | Indicated Vehicle Length, a (ft) | Assoc. Percent Velocity Measurement Error $\left[\frac{a_o - a}{2D}\right] \times 100$ Percent | |
|---|---|---|---|---|---|
| | | | | 10-Ft Loop Separation | 8-Ft Loop Separation |
| 1 | 0 | $21.96\ (L' + a)_o$ | $15.96\ (a_o)$ | 0 | 0 |
| 2 | + 3/8 | 21.63 | 15.63 | 1.65 | 2.05 |
| 3 | +1 | 20.71 | 14.71 | 6.25 | 7.8 |
| 4 | +3 3/4 | 20.56 | 14.56 | 7.0 | 8.75 |

The experimental velocity measurements were quite accurate since, of course, the vehicle crossed each loop with the same body height. The velocity measurement error that would have occurred if the body height had varied from its minimum value when crossing into the first loop to the new value when crossing into the second loop (or vice versa) approximates the effect that vehicle-bounce would have on the velocity measurement accuracy. The assumption made here is that the vehicle would move with a bounce-period equal to twice the loop-pair separation. The velocity measurement errors (with respect to the minimum-height velocity measurement) that would be associated with the varying pulse widths are calculated by extrapolation from the pulse-width data. The loop-pair used (10-ft separation) would, under the circumstances assumed, measure vehicle velocity with an error of 7 percent for a vehicle bouncing 3¾ in. Further extrapolation for an 8-ft loop-pair separation (chosen for the verification data runs) indicated that the velocity measurement would be in error by 8.75 percent for a 3¾-in. bounce. These are worst-case conditions. Even for a 3¾-in. bounce of the proper bounce-period, the error contribution due to bouncing would be something less than maximum, depending on the exact phase relationship between the bounce-cycle and the loops.

The foregoing represents a somewhat academic approach to the study of the effects of vehicle-bounce on velocity measurement. Important questions must be answered, i.e., how much bounce and what bounce-period can be expected at various speeds in a practical situation, etc., before measurement errors can be fully assessed. In any event, care should be exercised to locate the loops in smooth, flat sections of roadway to minimize vehicle-bounce.

## CONCLUSION

The major advantage of the technique discussed in this paper is that it permits the application of commercially available magnetic-loop vehicle detectors to the measurement of individual vehicle velocity and length. It does so by operating a pair of detectors as a system. This is made practical by the application of a simple, convenient calibration procedure. The measurements are quite accurate in light traffic, but approach 90 percent accuracy when the vehicles follow one another closely (for an unmodified TACDET operating in the pulse mode).

The loop configuration developed is consistent with the diamond interchange application, in which it is planned to use a large number of loop-pairs with their associated detectors to acquire the simulation validation data. For narrower lanes (less than 12 ft wide), the loops should be offset from lane-center to preserve the 8-ft adjacent conductor separation to minimize double counting.

The techniques described are usable for traffic applications in which it is desired to determine vehicle velocity and/or size directly in addition to counting. Some applications that come to mind are the following:

1. 85th percentile determination of speed for the establishment of speed limits.
2. Early warning speed indication for traffic approaching areas of limited distance visibility.
3. Speed limit enforcement.

ACKNOWLEDGMENT

We would like to express our appreciation to the California Division of Highways for their cooperation in making the traffic tests possible.

REFERENCE

1. Christensen, A., and Hewton, J. T. Developments in Conventional Ground-Based Detectors. Proc. of Highway Conf. on Future of Res. and Dev. in Traffic Surveillance Simulation and Control, U. S. Bur. of Public Roads, 1964.

# *Appendix A*

## VEHICLE DETECTOR DESCRIPTION

Figure 10 shows a block diagram of the vehicle detector. The sinusoidal signal from a crystal-controlled oscillator is connected, through an emitter follower, to the inductive loop. Variable capacitors are provided to tune the loop to the oscillator frequency. The voltage across the loop is applied to a phase detector, where its phase is compared with that of the reference oscillator signal. The filtered phase detector output voltage is amplified in a dc amplifier and connected to a two-transistor relay driver amplifier.

When a vehicle enters the loop's magnetic field, it produces a phase shift in the voltage across the loop, which raises the phase detector's output voltage. This voltage is amplified in the dc amplifier, activating the relay through the relay driver amplifier. The dc amplifier is capacitively coupled to the relay drivers to make the circuit insensitive to long-term dc voltage variations. A choice of time constants is provided. The maximum time constant provided permits the detector to continue monitoring the presence of a vehicle stalled in the field of influence, to the exclusion of all other passing vehicles. Minimum time constant should be selected to minimize detector recovery time.

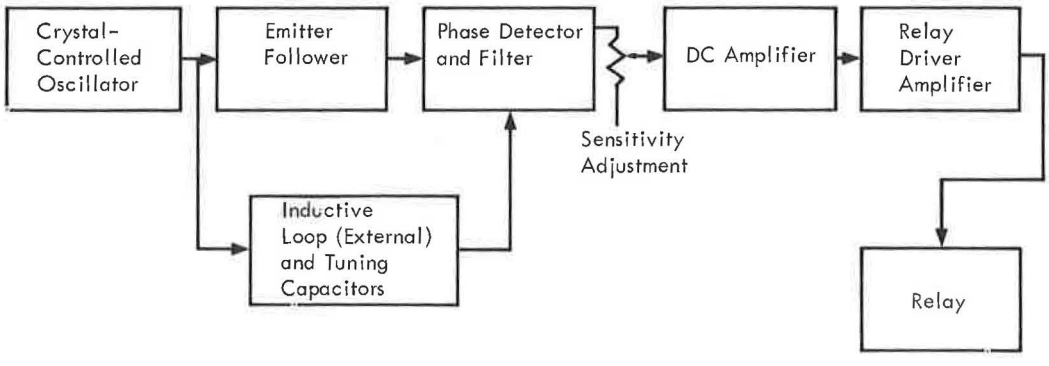



Figure 10. Vehicle detector block diagram.

Plug-in crystals are available in 2 kc/sec steps from 90 to 110 kc/sec.

A sensitivity adjustment provides a method of varying the relative position between loop and vehicle at which the relay activates.

# *Appendix B*

## VEHICLE DETECTOR CALIBRATION TECHNIQUE

The calibration circuit (Fig. 11) is connected to the vehicle detector's 'loop' terminals (the detectors usually provide monitor terminals, which can be used for this purpose). The state of the detector relay is monitored by means of a visual indicator circuit. If not already provided in the detector, a relay monitoring circuit can be constructed by connecting a 6-volt battery and a No. 47 indicator bulb in series with the normally open relay contact terminals. Two slightly different capacitance values (C1 = 25.2 picofarads, C2 = 28.2 picofarads) are used.

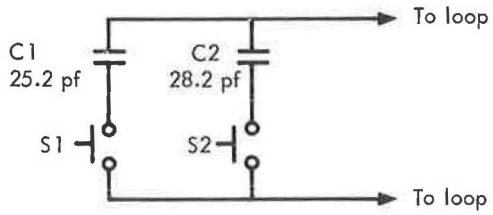



Figure 11. Vehicle detector calibration circuit.

The detector sensitivity is adjusted to a maximum. Pushbutton switch S1 is first depressed, connecting C1 across the loop, and then released, removing it. If the indicator lamp lights when C1 is removed (indicating that the relay has been activated) the detector sensitivity is decreased and the procedure repeated until a sensitivity position is reached for which the lamp does not light. Switch S2 is then operated in the same manner, introducing C2 and a slightly greater phase shift. The relay will now activate unless the detector sensitivity has been set too low. If this is the case, the sensitivity is increased slightly until the lamp lights when S2 is released after being depressed.

The detector is properly adjusted when a sensitivity setting has been found such that the relay activates when S2 is operated and fails to activate when S1 is operated.

When the sensitivity of the first detector is thus precisely set, the same procedure is repeated for the second.

Minimum detector time constant should be used during calibration to permit the detector to stabilize rapidly.

# A Statistical Analysis of Speed Density Hypotheses

JOSEPH S. DRAKE, Research Engineer, Consad Research Corporation;
JOSEPH L. SCHOFER, Graduate Student, Northwestern University; and
ADOLF D. MAY, JR., Associate Professor of Civil Engineering, University of California

•AN UNDERSTANDING of interrelationships among basic characteristics of vehicular traffic flow, such as volume, speed, and density, is of prime importance to the practicing traffic engineer. From the standpoint of design, a knowledge of high-flow-rate characteristics is required for the prediction of highway capacity. Those concerned with traffic operations are faced with the problem of providing an adequate level of service; this calls for an understanding of the entire range of relationships. Development of flow control and ramp metering techniques must be based on these functional interrelationships under high-density conditions. Finally, any efforts toward developing new roadway and vehicular technologies for the purpose of improving flow characteristics will necessarily stem from an understanding of the present relations.

During recent years, a number of hypotheses concerning these interrelationships have been proposed. Some researchers have relied almost completely on the statistical analysis of data for developing functions, while others have begun with a purely theoretical concept, from which relations were derived and later tested.

It is the purpose of this paper to examine currently available hypotheses, using rigorous statistical procedures to test them with a common set of freeway flow data. The degree to which the various functions could be made to replicate the full range of the data was the basis for determining which relations might be better than others, and possibly for deciding that some should be rejected entirely.

Primary reliance was placed on the results of the techniques of statistical analysis. Where these were inadequate, however, sound judgment was applied in order to select the most workable functional relations.

The approach selected was the regression of speed of flow against density, since the other basic relations, volume-density and speed-volume, cannot be transformed into linear functions for all hypotheses. The techniques of linear regression analysis are both simpler and more highly developed than those of the nonlinear analysis. A visual examination of the volume-density and speed-volume relationships derived by conversion from each of the speed-density regression equations was also performed.

## SITE AND DATA SELECTION

To investigate the relation between speed and density properly, it is necessary to sample traffic flow characteristics over the range of all possible densities. There are at least two feasible sampling procedures available: the use of fixed time periods, such as a minute or an hour to represent a data point, or the use of a specific number of consecutive vehicles for each point.

Data collected during the fixed time period represent flow characteristics during that period, but the number of vehicles included in the samples varies over the day. During free flow operations a minute sample could contain so few vehicles as to be statistically unreliable. On the other hand, the use of a short time increment, such as a 1-min period, helps to insure that a nearly instantaneous picture of flow conditions is provided. Vehicles in the sample can be no farther apart than 1 min in time, resulting in a limitation in the possible changes in flow characteristics which might take place during the

Paper sponsored by Committee on Characteristics of Traffic Flow and presented at the 45th Annual Meeting.

sampling period. This would imply that the dispersion of phenomena within a sample would be minimal.

The vehicle-based sample insures that each data point represents the same number of vehicles. Over the day, however, the amount of time included in the sample varies considerably. A series of 100 vehicle samples, taken on the Eisenhower Expressway in Chicago during a 24-hr period, covered time spans ranging from approximately 3 to 50 min. The flow characteristics represented by an average over 50 min may not give a good indication of the true relationships among vehicles on the facility. The behavior of vehicles passing a point during a preceding 50-min period may have little or no effect on current operations.

While neither sampling basis is entirely satisfactory, the 1-min time sample is probably best for examining vehicular interactions with respect to density of flow. In considering the freeway as a system generating some distribution of phenomenological output, as represented by vehicular flow characteristics, the regularity of the time-based sample is more satisfying intuitively than the continuously varying output rate of the vehicle-based sample. The detection equipment can be set to "look at" the expressway every minute and "photograph" its operating characteristics, rather than waiting until sufficient vehicles have passed and developing a time average.

A series of 1,224 one-minute observations was recorded with the pilot detection system of the Chicago Area Expressway Surveillance Project. At the end of each minute the minute volume, time-mean-speed, and occupancy were obtained, and density was computed from the volume and time-mean-speed. Data were collected in the middle lane of the three-lane westbound roadway at Harlem Avenue on the Eisenhower Expressway. The observations were made between 1:00 and 6:00 p.m. on four weekday afternoons under dry weather and normal traffic conditions. Thus, many of the data represented peak hour characteristics, while few were associated with the very lowest density range.

It appears appropriate to describe the location of the study site in relation to adjoining portions of the expressway because of its effect on the flow interrelationships. For example, if a location is studied which is not operating at capacity due either to lack of demand or an upstream bottleneck, high density measurements obviously cannot be obtained. On the other hand, if the study site is located just upstream of a severely limiting bottleneck, mid-range density measurements may be limited or almost nonexistent. Hypothetical speed-volume-density relations describing these conditions are shown in Figure 1.

The study site (Harlem Avenue) is located one-half mile upstream from a bottleneck. The capacity of the downstream bottleneck, however, is only slightly less than the capacity of the study site, and therefore density observations can be obtained over a substantial portion of the density range.

A disturbing feature of the original data set was that points seemed to occur in bunches (Fig. 2), particularly in the 20 to 45 and 65 to 84 veh/mi ranges (1). Early attempts at regression analysis with these data indicated that the concentration of points in a few areas resulted in equations which did not adequately represent the entire range of the relationship. Sample densities varied considerably over the range of the independent variable, resulting in a biased estimate of regions represented less than adequately in the sample.

Once the objective of any research project has been defined, a sampling procedure can usually be designed so as to efficiently collect a relevant set of unbiased data. For an investigation of mean free speed—the Y-intercept of the relation—a sample whose mean density is close to zero could be chosen. Conditions around jam concentrations could be studied by sampling only at the lowest congested speeds. Because this analysis was aimed at testing a number of functional relations over the full range of operating characteristics, it was necessary to give equal consideration to all flow conditions which might occur.

One solution to the problem would have been to return to the field and fill in areas which were under-represented in the original sample. The variable over which the sample was being taken was density, in vehicles per mile. An unbiased sampling pro-

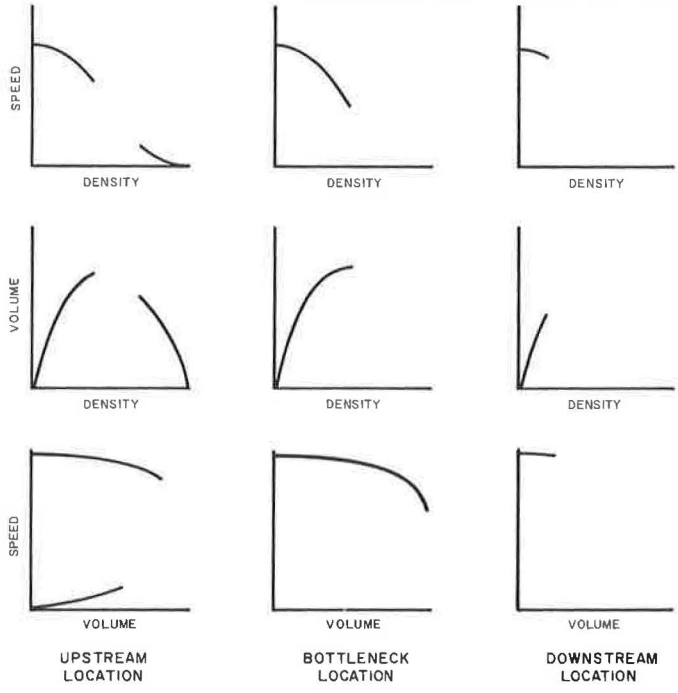



Figure 1. Hypothetical relationships.

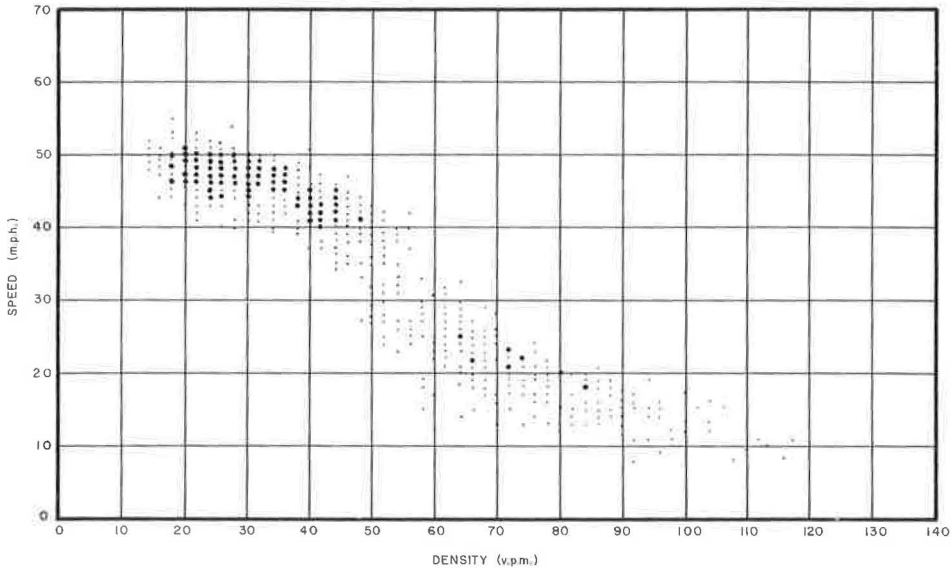



Figure 2. Scatter diagram of total speed-density observations.

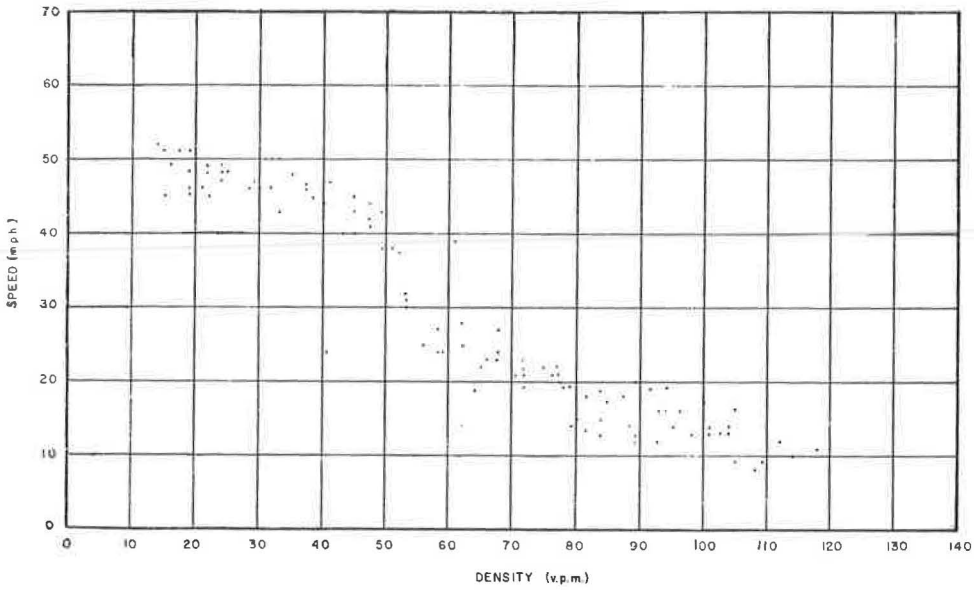



Figure 3. Scatter diagram of selected sample of speed-density observations.

cedure would require that each possible density value should have an equal probability of being in the sample. To collect data by turning on the detectors for a period of 5 hours per day automatically biases the analysis against regions of the relation whose densities rarely occur during the sampling hours.

Because of the additional expense entailed in collecting data under conditions similar to the original sample in order to fill in the sparse areas, another technique was devised. The set of 1,224 one-minute observations, each on a single punch-card, was arranged in the order of increasing density and divided into ranges of approximately 5 veh/mi. The number of observations falling in the most sparse 5-veh/mi range was determined, and a like number of data points was randomly sampled from each of the other ranges. This resulted in a sample of 118 observations, with each 5-veh/mi range containing an equal number of points. The procedure resulted in a considerable degree of uniformity of sample density throughout the range of the independent variable, as shown in Figure 3, particularly when compared with Figure 2. Some characteristics of the sample are as follows:

| | |
|---|---|
| Lowest density | 14.2 veh/mi |
| Highest density | 118.4 veh/mi |
| Average density | 61.9 veh/mi |
| Lowest speed | 8.2 mi/hr |
| Highest speed | 52.3 mi/hr |
| Average speed | 29.6 mi/hr |

## ANALYSIS OF RELATION BETWEEN TIME MEAN SPEED AND SPACE MEAN SPEED

Data were collected by using motion and presence detectors which measured speed, lane occupancy and volume. Individual speeds were averaged over each minute, resulting in a time-mean-speed value. Density was computed based on the fundamental relation, volume = speed × density.

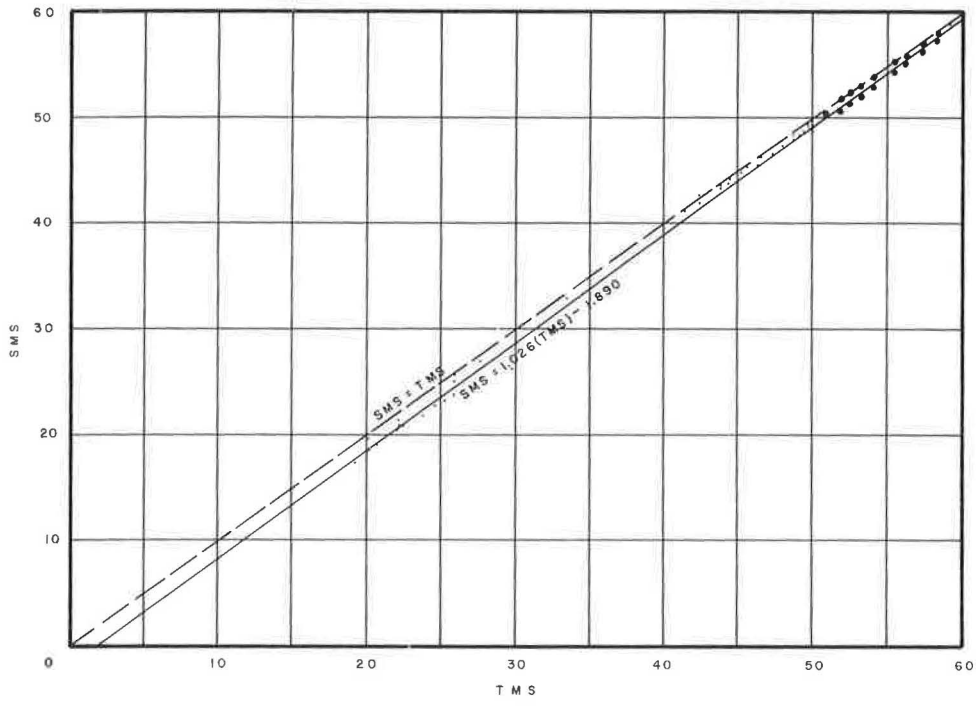



Figure 4. Time-mean-speed and space-mean-speed relationship.

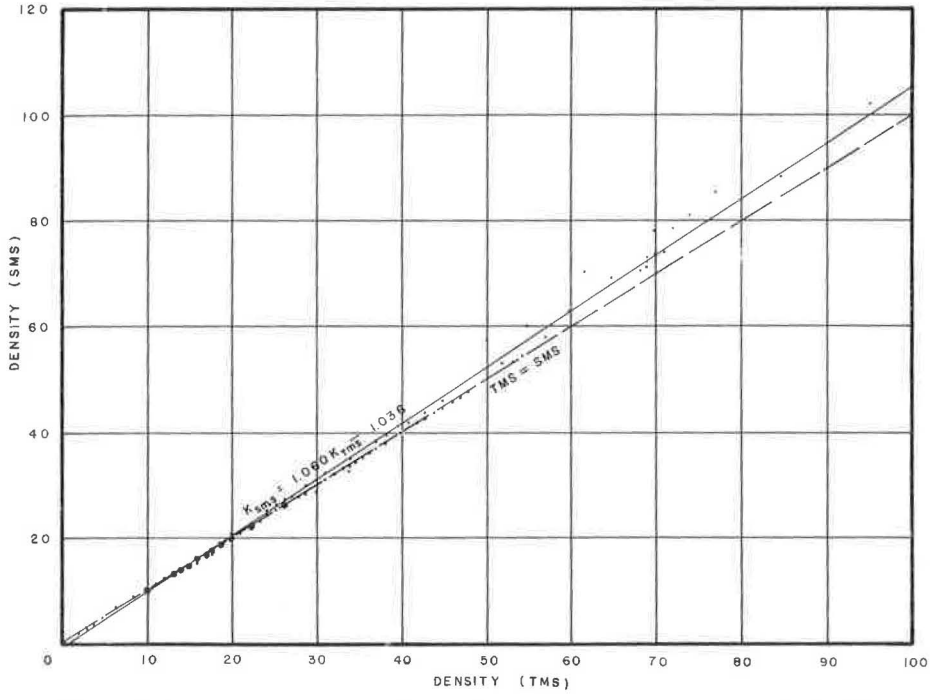



Figure 5. Time-mean-speed density and space-mean-speed density relationship.

Wardrop (2) has shown that this relation holds only for space-mean-speed, the speed computed from the mean of vehicle travel times over a specified distance; he demonstrated that the two speed measurements could be related as follows:

$$TMS = SMS + \frac{(\sigma SMS)^2}{SMS}$$

The use of the collected time-mean-speed data, then, introduces a bias into the values of both speed and density. To investigate the magnitude of this bias and possibly develop a procedure for correcting the data points, a second study was conducted at the same location, in which 224 groups of 100 consecutive vehicles were analyzed. Time-mean-speed, space-mean-speed, standard deviations of both time- and space-mean-speeds, and densities calculated from both speeds for each group were determined.

A regression analysis of time-mean-speed and space-mean-speed resulted in the following equation:

$$SMS = -1.88960 + 1.02619\ (TMS)$$
$$r^2 = 0.99834$$
$$\sigma = 0.38404$$

The nearness of the slope of this relation to 1.00 indicates the close correspondence between the two speed measures (Fig. 4). The maximum difference between the two speed measures occurs at zero speed and is 1.9 mph. As speed increases the difference becomes smaller, and at 72 mph the time-mean-speed and space-mean-speed would be equal. In addition, there was no significant difference in the means of the two speed measurements at the 99 percent confidence level.

A similar investigation of the relation between the densities computed from time-mean-speed and space-mean-speed gave the following regression line:

$$K_{sms} = -1.03638 + 1.06018\ (K_{tms})$$
$$r^2 = 0.99712$$
$$\sigma = 0.94337$$

The two measurements were again found to be quite close, with a maximum difference on the order of 1 veh/mi (Fig. 5). There was no significant difference in the means of the two parameters at the 99 percent confidence level.

Tests of the slopes of both of these equations revealed that they were significantly different from 1.000 at the 95 percent confidence level; i.e., the data indicate the existence of a real difference between space-mean-speed and time-mean-speed, a result not unexpected considering the theoretical work of Wardrop.

In spite of these results, it was felt that any reduction in error achieved through conversion of data from a time-mean-speed to a space-mean-speed basis would be unimportant when compared to the variation introduced by the inaccuracies of the data collection equipment. For example, the range of error in speed measurement over a 1-min sample has been estimated at ± 1 mph; for an individual vehicle this broadens to 4 or 5 mph.

The conclusion from this phase of the analysis was that time-mean-speed and the corresponding density would be used for the investigation, since this would introduce only a minimal amount of error into the results of the study.

## PRESENTATION OF HYPOTHESES

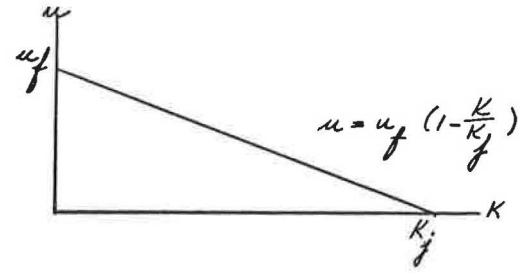
Seven proposed speed-density relationships were selected for examination. Some have no theoretical background, being based primarily on the researcher's observation

of a particular data set. Others have considerable support; four have been shown to be directly related to specific car-following rules (3, 4).

## Linear Forms

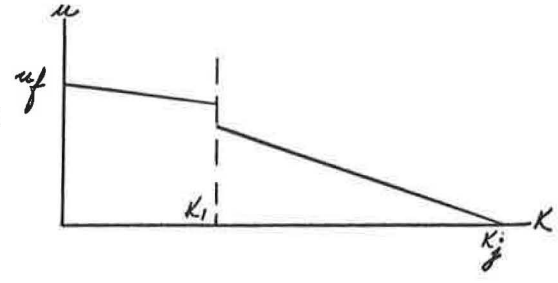
Early work in the analysis of speed-density relationships, notably that by Greenshields (5), was devoted to the investigation of a continuous linear form, as shown below:

HYPOTHESIS I

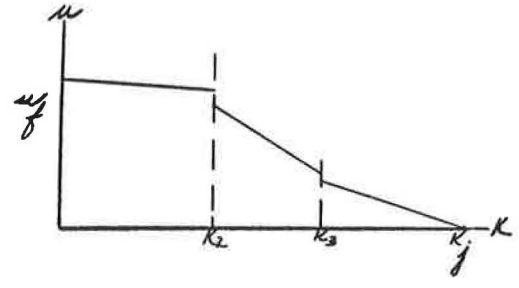



Prompted by the failure of the continuous relation to fit data at all density levels, some researchers (6) have devoted attention to the possibility of the existence of two or three distinct zones, each characterizing a different driver behavior pattern. These hypotheses follow; additional parameters are introduced.

HYPOTHESIS II



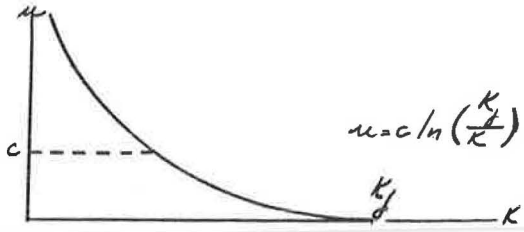

HYPOTHESIS III




The assumption of linearity was adopted because of its simplicity, and although there is no strong theoretical basis, the function may be derived from the car-following rule in which sensitivity of the following vehicle is inversely proportional to the square of the inter-vehicle spacing.

## Greenberg's Exponential Curve

In 1959, Greenberg (7) postulated another speed-density form, based on a hydrodynamic analogy. Treating the traffic stream as a perfect fluid, he combined the equations of motion and continuity for one-dimensional compressible flow and arrived at the following form:

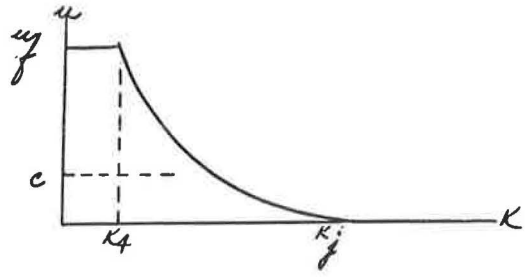



The parameter c is defined as the speed at which volume experiences its maximum with respect to density.

Gazis, Herman and Potts (3) have shown that this relationship can be independently derived from their microscopic car-following theory for the case in which the sensitivity of the following vehicle is inversely proportional to the spacing between vehicles.

Because this equation is not defined at k = 0, it will be tested in the following more realistic form; again, another parameter is introduced:

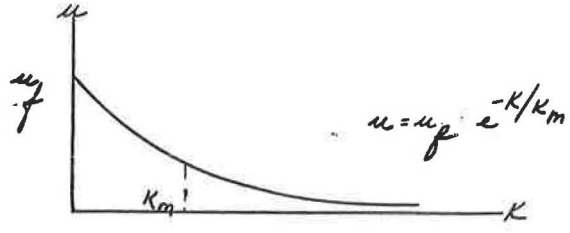
HYPOTHESIS IV




Underwood's Transposed Exponential Curve

Apparently disturbed by the failure of Greenberg's curve to remain finite at zero density, Underwood (8) suggested that perhaps the infinity asymptote should be along the density scale, since almost any jam always has some finite movement:

HYPOTHESIS V

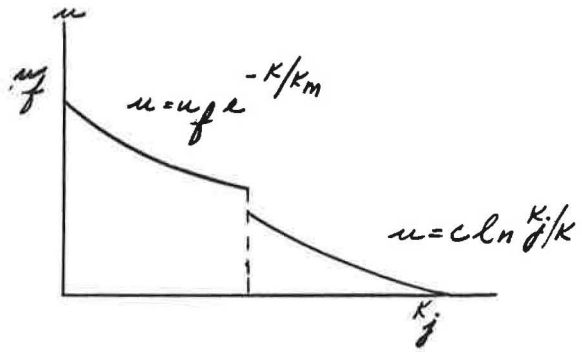



The parameter $k_m$ is defined as the density at which volume experiences its maximum with respect to density. This relation can be traced to the car-following rule where sensitivity is directly proportional to the speed of the following vehicle and inversely proportional to the square of the spacing.

Edie's Discontinuous Exponential Form

The frequent occurrence of a discontinuity in empirical volume-density curves in the vicinity of optimum density and the poor correlation of low-density data with Greenberg's hypothesis led Edie (9) to postulate in 1960 that a different car-following law applies to noncongested operation, namely, that the sensitivity is proportional to velocity and to the reciprocal of the square of spacing between vehicles. Consequently, he derived the following hypothesis:

HYPOTHESIS VI

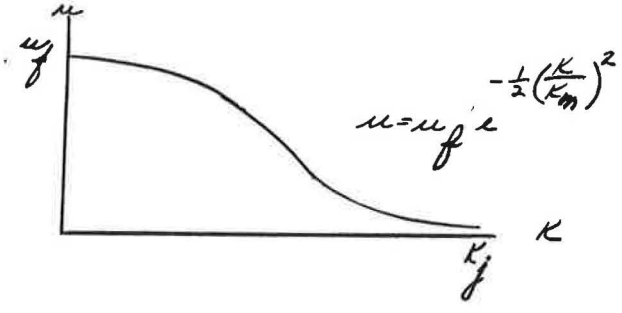



This form is merely a combination of the Greenberg and Underwood relations.

Further Suggestions

May has pointed out that speed-density data from the Eisenhower Expressway in Chicago tend to exhibit concavity at low densities. Guided by these observations, the following bell-shaped curve has been suggested:

HYPOTHESIS VII




This hypothesis has no theoretical foundation.

## DISCUSSION OF STATISTICAL ANALYSIS TECHNIQUES

The basic purpose of this research was to make decisions regarding the relative merits of the seven speed-density hypotheses as applied to the study data. The essence of any experimental design lies in the development of rigorous tests which will permit exacting binary decisions to be made. The researcher must rely on his own intuition to decide what is expected of his hypothesis; he must then translate these expectations into testable form.

The attitude taken in the formulation of such tests in this research endeavor was based on the traditional philosophy of rejection. In general, such an approach translates a general hypothesis into several working hypotheses which express the various expectations of the researcher. These working hypotheses are structured in a test framework and are designed to maximize the "falsifiability" of each hypothesis. The failure of any one working hypothesis to meet the researcher's expectations necessarily constitutes rejection of the general hypothesis, and suggests a restudy of the theory underlying that hypothesis.

The use of such an approach was particularly warranted in this study, for two reasons.

1. Traditional applications of regression techniques tend to ignore opportunities for verification. The goodness-of-fit statistic ($r^2$) is generally quoted, and frequently, a test is conducted to shed some light on the significance of the regression in reducing variation in the dependent variable. Viewed in this limited context, regression analysis

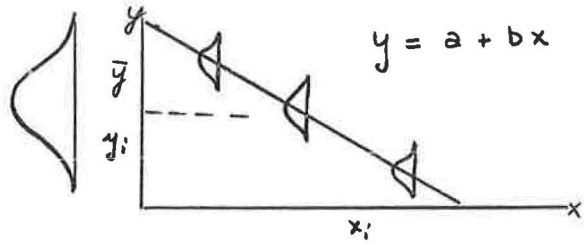
tends to become a self-fulfilling prophecy. The statistical design in this study has been approached with the intention of formulating tests not only of the significance of the regression considered by itself, but also of the calibrated regression model's ability to predict independently derived parameters.

2. The basic purpose here was to compare the seven models with each other. Regression theory presently offers no technique for comparing directly two regression models applied to the same sample. The only alternative, then, is to develop tests which may be applied against each hypothesis without favor, and to insist that the failure of any one hypothesis to meet all such tests necessarily constitutes elimination of that hypothesis from further consideration. This idea is consistent with the notion of comparing hypotheses in terms of their application over the entire domain of density.

## Fundamental Concepts of Regression Analysis

Before introducing those statistical concepts which have been developed to cater to the peculiar nature of this research, it is worthwhile to review the traditional setting of regression analysis. Suppose measurements of a variable y are recorded and considerable variation in y about its mean $\bar{y}$ is observed. Suppose that y is influenced by an independent variable x according to some relation; if it is possible to isolate the effects of x on y the dispersion of y might be considerably reduced. Linear regression analysis is a convenient means for accomplishing this desired isolation.

The following diagram, which depicts this situation, provides a convenient review of terminology and concepts which are vital to what follows.




The diagram illustrates the manner in which the allowance for variation in x reduced the variation in y. Such allowance reduced the difference between an individual observation, $y_i$, and the mean, $\bar{y}$, to the difference between $y_i$ and precited $Y_i$. The equation $Y = a + bx$ is fitted to the data according to the criterion of minimizing the sum of the squares of deviations about the regression line. More formally, the following equations are solved for the parameters a and b:

$$\frac{\partial}{\partial a}\left[\sum_{i=1}^{n}(y_i - Y_i)^2\right] = 0$$

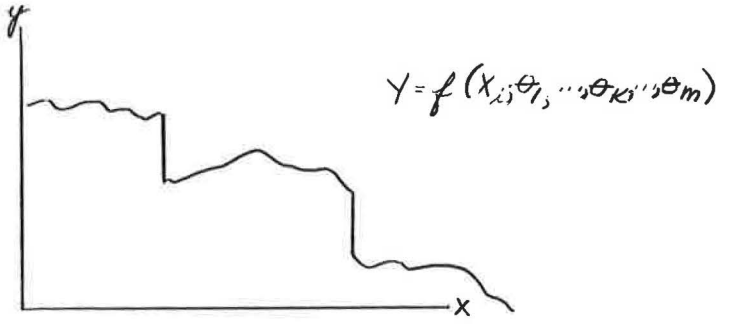
$$\frac{\partial}{\partial b}\left[\sum_{i=1}^{n}(y_i - Y_i)^2\right] = 0$$

In this fashion, the sum of squares (of deviations) about the mean is decomposed into two components: the sum of squares due to the regression, and the resulting sum of squares about the regression. In more formal terms,

$$\Sigma(y_i - \bar{y})^2 = \Sigma(Y_i - \bar{y})^2 + \Sigma(y_i - Y_i)^2$$

## Composite Statistics for Discontinuous Regressions

The following diagram shows the same type of problem in more general terms. Some general relationship y = f(x) is to be fit to a set of observations. The relationship is not susceptible to textbook approaches, because of its complexity. Not only is it nonlinear, but it also suffers discontinuities. For the purely nonlinear case, with no discontinuities, it is frequently feasible to perform some transformation upon y and/or x which reduces the relationship to linear form. Given discontinuities, however, two fundamental questions arise: Can a discontinuous relationship be described in terms of single statistics? For example, can one $r^2$ value be quoted for the entire relationship? And, can one decide whether or not discontinuities exist in the true relationship between y and x?




Ideally, these questions might be answered by fitting the function to data according to the least squares criterion described for the linear case. This would be accomplished by minimizing the expression

$$\sum_{i=1}^{n} [y_i - f(X_i; \theta_1, \ldots, \theta_k, \ldots, \theta_m)]^2$$

with respect to each parameter $\theta_k$.

Even for very simple functions, such techniques lead to mathematics too cumbersome to handle. The alternative approach would be to treat each continuous regime in its own right, and then treat the result as a whole. More specifically, the relationship between Y and X for each regime would be fitted (after transformations, if necessary) to its associated data according to its own least squares criterion. The results may then be reassembled, and single composite statistics may be computed for the entire discontinuous regression in terms of the values of $y_i$, $Y_i$, and $\bar{y}$. This approach assumes that minimizing the sum of squares about the regression line for each regime is equivalent to the more desirable but insurmountable task of minimizing the sum of squares about the entire discontinuous regression. Such an approach was applied in this research.

## Testing of Multi-Regime Hypotheses

Having devised a method for quoting composite statistics for discontinuous regression models, the question regarding how one might decide whether the true relationship between y and x is discontinuous or not still remains. The approach to this problem was derived from earlier work by Quandt (10) on the treatment of two-regime linear regression hypotheses. He discusses two problems: (a) the estimation of the parameters of a two-regime hypothesis, and (b) the possible methods for testing whether the data do in fact obey two separate regimes. The treatment of the first problem, i.e., locating the breakpoint between regimes, precludes the exercise of any tests.

The location of the breakpoint t for two linear regimes is ascertained by maximizing the likelihood function of the entire sample, formed by multiplying together the (normal) frequency functions of the individual error variances, $\sigma_1$ and $\sigma_2$. Upon application of conventional maximization procedures for obtaining best estimates of parameters $a_1$, $a_2$, $b_1$, $b_2$ and $\sigma_1$ and $\sigma_2$, the following expression results:

$$L(t) = -(\tfrac{1}{2} + \ln\sqrt{2\pi})\, T - t \ln\hat{\sigma}_1 - (T - t) \ln\hat{\sigma}_2$$

where T is the total number of observations, t is a discrete variable representing the serial rank of each observation (with respect to x), and $\sigma_1$ and $\sigma_2$ are functions of t:

$$\hat{\sigma}_1^2 = \frac{1}{t} \sum_{i=1}^{t} (y_i - \hat{a}_1\ x_i - \hat{b}_1)^2$$

$$\hat{\sigma}_2^2 = \frac{1}{T-t} \sum_{i=t+1}^{T} (y_i - \hat{a}_2\ x_i - \hat{b}_2)^2$$

The problem then, is to find the value of t which maximizes L(t). Because t is a discrete variable, and anticipating several local maxima, the only feasible method for determining an optimal t, t*, is a systematic search procedure. This requires the performance of a complete regression analysis on each regime, for every selected value of the breaking point, t.

Having established the value of t*, Quandt (11) examined various alternatives for testing the hypothesis that the data obey two separate regimes. For this study, his F-test recommendation was adopted. Essentially, the resulting regression equation for regime I is applied to observations for $i \leq t$, and then to observations for $i > t$. The quantities

$$\eta_{I\text{-}I} = \frac{1}{\sigma^2} \sum_{i=1}^{t} (a_1 + b_1\ x_i - y_i)^2$$

$$\eta_{I\text{-}II} = \frac{1}{\sigma^2} \sum_{i=t+1}^{T} (a_1 + b_1\ x_i - y_i)^2$$

are independently distributed as $\chi^2$ with t - 1 and T - t - 1 degrees of freedom, respectively. Their ratio therefore follows the F-distribution with the given degrees of freedom. A similar ratio is obtained by applying the resulting regression equation for regime II to observations in each regime, and forming the appropriate F-ratio in terms of $\eta_{II\text{-}1} + \eta_{II\text{-}2}$. If either F-ratio exceeds F-critical, the hypothesis of one continuous regime will be rejected. The power of this test depends on how close the true breakpoint t* is to the endpoints of the sample.

To meet the demands of hypothesis III, Quandt's work was extended to the case of three regimes, with the following likelihood function resulting:

$$L(t) = -(\tfrac{1}{2} + \ln\sqrt{2\pi}\ T - r \ln\hat{\sigma}_1 - (s - r) \ln\hat{\sigma}_2 - (T - s) \ln\hat{\sigma}_3$$

The appropriate tests were formulated by applying each of the three regression equations to its own regime and comparing it to the remaining two regimes. Six F-ratios

were developed; the failure of any one ratio to exceed F-critical constituted the basis for rejection of the three-regime hypothesis.

### Tests for All Hypotheses

Regression analysis offers several statistical tests which may be exercised according to the purpose of the investigation. Two such tests were adopted for this study: the F-test for significance of the regression and the t-test for determining whether the slope differs from zero. For continuous regression models, these two tests are equivalent. For discontinuous hypotheses, however, the t-test was applied to the slope of each regime, while the F-test was formulated in terms of the entire regression, according to the methods discussed previously for quoting composite statistics for multi-regime hypotheses.

The use of such tests in themselves does not constitute much verification power. To promote independent verification, the predicted values of mean free speed for each hypothesis were tested for significant difference from an independent measurement of that parameter.

## ANALYSES OF RESULTS

### Break-Point Analysis

The first stage of the analysis was the establishment of the break-points for the four discontinuous hypotheses according to the technique described. For the three hypotheses which required only one break-point, separate regression analyses were performed for each of the possible ways in which the observations could be broken into two groups, using a minimum separation of 5 veh/mi between alternative break-points. The resulting likelihood functions are shown in Figures 6, 7 and 8, for hypotheses II, IV, and VI, respectively.

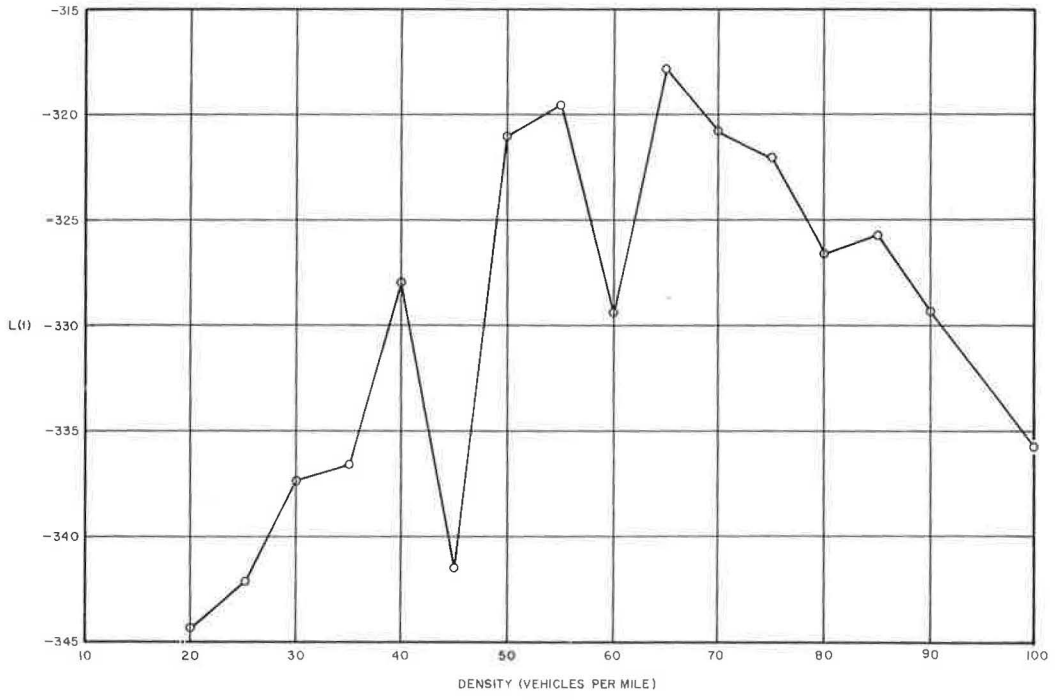



Figure 6. Likelihood function for hypothesis II (2 Regime Linear).

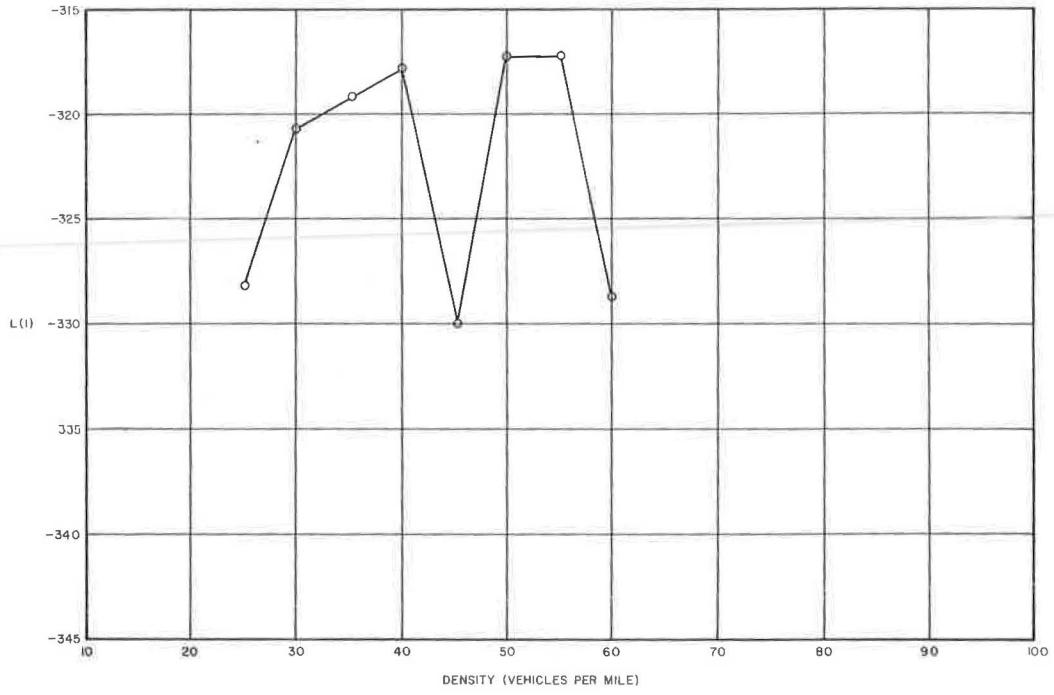



Figure 7. Likelihood function for hypothesis IV (Greenberg).

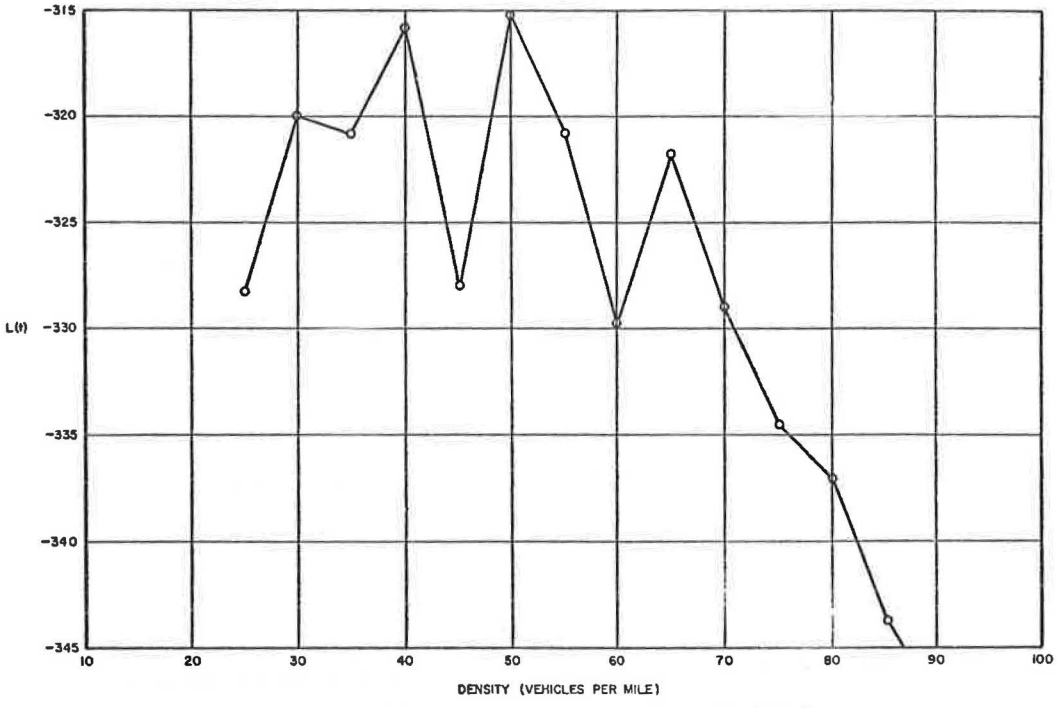



Figure 8. Likelihood function for hypothesis VI (Edie).

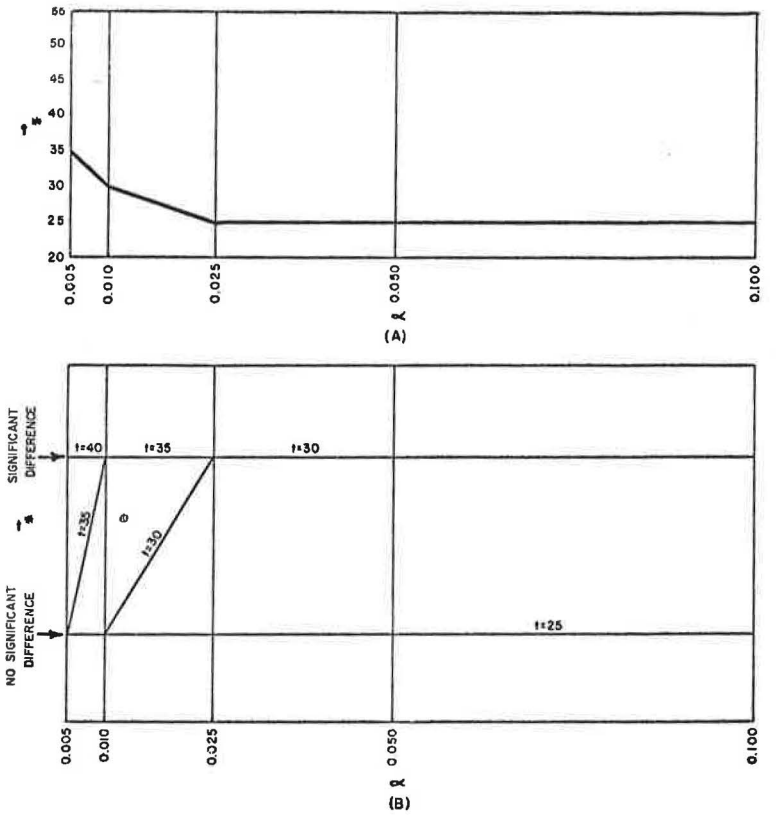



Figure 9. Comparsion of slope for low-density regime of hypothesis IV to zero for various confidence levels.

The likelihood functions for the Edie relationship and the two linear regimes hypothesis behaved quite similarly. While the optimal break-points of these two functions differed appreciably (65 veh/mi for hypothesis II and 50 veh/mi for hypothesis VI), both likelihood functions showed two local peaks in the middle range of density, and these peaks did not differ much in value for each hypothesis. Furthermore, an abrupt drop in the function occurred between these peaks, suggesting that L(t) was quite sensitive to t.

This apparent sensitivity had no important bearing on the use of the likelihood function analysis because the optimal density value was the sole matter of interest. Had the analysis investigated L(t) more accurately, e.g., for intervals of 2.5 veh/mi, it is conceivable that the maximum values of L(t) might have been found at such intermediate points, and that these maximum values might have been considerably greater than the values derived in Figures 6 and 8. The goal of this phase of the investigation was not to determine the global maximum value of L(t), but merely to establish the proper break-point. The resulting location of this point must be interpreted with a maximum tolerance of ± 2.5 veh/mi.

The break-point analyses for hypotheses II and VI differed in one important respect. The Edie hypothesis required a different transformation for each of its two regimes in order to permit the use of linear regression analysis. The low-density regime was regressed in terms of $Y = \ln u$ and $x = k$, while the high-density regime was regressed in terms of $Y = u$ and $x = \ln k$. The resulting error variances were dimensionally incompatible and could not be used directly in computing L(t). To avoid this problem the low-density regression line (for each value of t) was converted to its corresponding nonlinear form, and the error variance (in terms of u) was computed in terms of the difference between the observations and the calibrated curvilinear relationship. Such an approach lacked

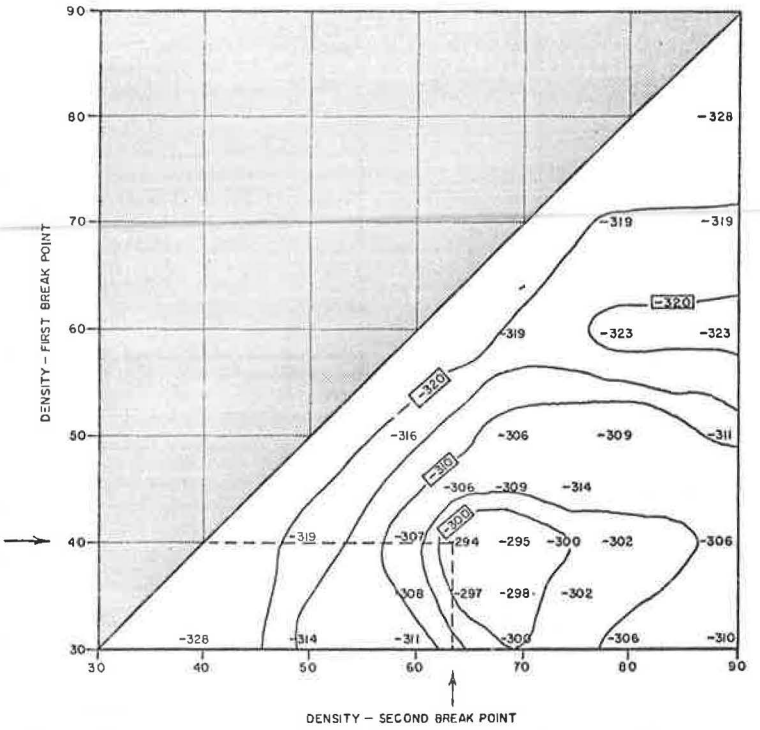



Figure 10. Likelihood function for hypothesis III (3 Regime Linear).

some rigor, since the regression line was determined not by minimizing the sum of squares about the curvilinear form (the mathematics of which are quite cumbersome), but by a minimization of the squared residuals in the transformed relationship.

Hypothesis IV (the modified Greenberg form) was treated in a similar fashion, except that the condition of zero slope was imposed on the low-density regime. The likelihood function given by the break-point analysis, again evaluated at 5-veh/mi intervals, is shown in Figure 7. Three local maxima were observed (k = 40, 50, 55), with the global maximum occurring at 55 veh/mi. The differences between the various peaks were slight.

In the process of computing separate regression lines for each 5-veh/mi increment, t-values were computed to test the slope of the low-density regime vs zero. Figure 9a shows the results of these tests at five confidence levels. For a break-point $t \geq 40$ veh/mi, the slope was significantly greater than zero for all levels, while for t = 25 veh/mi there was no significant difference at any level tested. For t = 30 veh/mi there was no significant difference at levels 0.005 and 0.010.

Figure 9b was derived from these results and shows for any given confidence level, the maximum value of t for which the slope was not significantly greater than zero. For the levels considered, the curve could not be greater than 35 veh/mi.

To establish the break-point value for this hypothesis, the likelihood function was considered together with the interpretation of Figure 9b, which provided a basis for imposing a "hat" of zero slope on the Greenberg hypothesis. Considering the values of t which yielded the peak values of L(t), and choosing a confidence level of 0.005, the optimal break-point was selected to be 35 veh/mi. At any greater value, the slope became significantly greater than zero. At any lower value, the likelihood function fell appreciably below its maxima. The result of this compromise indicated that "significant" interaction between vehicles began at an average spacing of about 150 ft. (While the original purpose of quoting results for various confidence levels was to provide the

reader freedom to exercise his own intuition, the peculiar nature of this analysis forced a decision at the 0.005 level.)

The location of the optimal break-points for hypothesis III (three linear regimes) was accomplished in the same manner as the analysis of hypothesis II, except that it required an extension to three dimensions. Denoting the two break-points by r and s (r separating the low- and middle-density regimes), L(r, s) was computed for combinations of r and s for 10-veh/mi increments in either variable. Then r and s were varied in 5-veh/mi increments in the vicinity of the coordinates (40, 70) where the maximum appeared to be located. The analysis in terms of this finer subgrid yielded the (approximately) optimal break-points or r = 40 veh/mi and s = 65 veh/mi (Fig. 10).

The "spiked" nature of the two-dimensional likelihood functions shown in Figures 6, 7 and 8 suggests that the decision to use a 5-veh/mi subgrid only in one section of the (r, s) plane might have been a dangerous one. It is conceivable that an abrupt increase might have been discovered anywhere on the (r, s) plane had a finer analysis grid been used throughout. The surface represented by the L(r, s) values was characterized by very slight undulations, however, implying that no such spikes may have existed. Furthermore, the fact that the three-dimensional analysis resulted in the same break-point for its high-density regime as did the two-dimensional analysis for hypothesis II lends credence to the strategy selected.

## Tests for Distinctly Separate Regimes

Having established the break-points for the four discontinuous hypotheses, the F-tests were conducted to compare each such "calibrated" hypothesis to the supposition of strict continuity over the entire density range. (Quandt points out that the F-test used here is sensitive to the location of the true break-points with respect to the extreme ends of the data; the hypothesis of continuity becomes less susceptible to rejection as the optimal break-point approaches either extreme of the range of the density observations.)

Table 1 gives the results of these tests. At the 0.025 and 0.010 confidence levels, the two linear regimes of hypothesis II did not differ significantly from the supposition that the regression line for the low-density regime prevailed over the entire range of density. Tests for all other hypotheses revealed significant differences of varying magnitudes, indicating that the assumption of separate regimes was in fact justified. In view of the rejection policy adopted for this research, hypothesis II was eliminated from further consideration, although it was subjected to the remaining tests.

TABLE 1

RESULTS OF F-TESTS FOR DISTINCTLY SEPARATE REGIMES

| Hypothesis | Test | Calculated F | F-Critical | α |
|---|---|---|---|---|
| II | I on II | 1.555 | 1.414 | 0.100 |
| | | | 1.552 | 0.050 |
| | | | 1.696[a] | 0.025 |
| | | | 1.876[a] | 0.010 |
| | II on I | 33.52 | 1.876 | 0.010 |
| III | I on II | 33.10 | 2.34 | 0.010 |
| | II on I | 5.388 | 2.36 | 0.010 |
| | II on III | 3.477 | 2.22 | 0.010 |
| | III on II | 64.07 | 2.06 | 0.010 |
| IV | I on II | 529.7 | 2.25 | 0.010 |
| | II on I | 3.744 | 2.00 | 0.010 |
| VI | I on II | 25.46 | 1.92 | 0.010 |
| | II on I | 3.729 | 1.84 | 0.010 |

[a]At the 0.025 and 0.010 confidence levels, hypothesis II's two linear regimes do not differ significantly from one linear regime. All other tests for all other hypotheses reveal significant differences of varying magnitudes, indicating that separate regimes as suggested by each of these hypotheses in fact exist.

TABLE 2

SUMMARY OF REGRESSION ANALYSES

| Hypothesis | Equation | Coefficient of Determination ($r^2$) | Standard Error ($S_e$) | F-Ratio Test of Significance-Values of F | Slope vs Zero Values of t | $U_f$ Prediction Values of t | Multi-Regime vs Single | Mean Free Speed ($U_f$) | Jam Density ($k_j$) | Optimum Density ($k_m$) | Optimum Speed (c) | Maximum Flow ($q_{max}$) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I<br>Greenshields | $U=58.6 - 0.468k$ | .896 | 4.648 | 1005 | 31.8 | 1.57 | — | 58.6 | 125 | 62.5 | 29.3 | 1830 |
| II<br>2-regime linear | $U=60.9 - 0.515k$ ($k \leq 65$)<br>$U=40 - 0.265k$ ($k \geq 65$) | .685 | 4.158 | 250 | I 11.2<br>II 11.7 | 7.46 | a | 60.9 | 151 | 59.2 | 30.4 | 1800 |
| III<br>3-regime linear | $U=50 - 0.098k$ ($k \leq 40$)<br>$U=81.4 - 0.913k$ ($40 \leq k \leq 65$)<br>$U=40.0 - 0.265k$ ($k \geq 65$) | .590 | 3.556 | 167 | I 6.9<br>II 35.8<br>III 11.7 | 22.6 | b | 50.0 | 151 | 44.6 | 40.7 | 1815 |
| IV<br>Modified Greenberg | $U=48.0$ ($k \leq 35$)<br>$U=32.8 \ln \frac{145.5}{k}$ ($k \geq 35$) | .866 | 3.867 | 745 | I 2.1<br>II 28.8 | 26.3 | b | 48.0 | 146 | 53.7 | 32.8 | 1760 |
| V<br>Underwood | $U=76.8e^{\frac{-k}{56.9}}$ | .901 | 5.076 | 1050 | 32.4 | 40.4 | — | 76.8 | — | 56.9 | 28.3 | 1610 |
| VI<br>Edie | $U=54.9e^{\frac{-k}{163.9}}$ ($k \leq 50$)<br>$U=26.8 \ln \frac{162.5}{k}$ ($k \geq 50$) | .681 | 3.550 | 245 | I 5.3<br>II 15.4 | 8.59 | b | 54.9 | 162 | 50.0 | 40.5 | 2025 |
| VII<br>Bell Curve | $U=48.6e^{-0.00013k^2}$ | .884 | 4.571 | 872 | 29.5 | 21.2 | — | 48.6 | — | 62.0 | 29.5 | 1830 |

[a]No significant difference.
[b]Significant difference.

6.85 2.36 2.33
Critical values at .01 level.

## Tests for Significance of Entire Regression

Table 2 gives the F- and t-values for the several tests performed on each hypothesis. The fifth column gives the F-values (essentially a measure of the ratio of explained to unexplained variance) for significance of the entire regression in terms of the extent to which the ratio exceeded unity. The critical values of F for each of the confidence levels considered were:

| Confidence Level | F-Critical |
|---|---|
| 0.10 | 2.75 |
| 0.05 | 3.92 |
| 0.025 | 5.15 |
| 0.01 | 6.85 |

All hypotheses were highly significant.

## Tests for Slope Greater Than Zero

The regression equations estimated for each hypothesis were tested to determine whether or not the slopes were significantly different from zero. Nonlinear hypotheses were so tested in the context of the appropriate transformation to linearity. Each regime of the discontinuous relationships was tested separately. For entirely continuous hypotheses, this test was redundant to the F-test. The sixth column in Table 2 gives the values of t obtained in this test. Critical t-values for the levels of confidence considered were:

| Confidence Level | t-Critical |
|---|---|
| 0.10 | 1.29 |
| 0.05 | 1.66 |
| 0.025 | 1.98 |
| 0.01 | 2.36 |
| 0.005 | 2.62 |

The only hypothesis to exhibit a slope not significantly different from zero was the modified Greenberg equation (low-density regime). However, this condition was forced on the hypothesis as described previously.

## Prediction of Mean Free Speed

An independent analysis of mean free speed for the study section provided a possible means for independent verification of the "calibrated" speed-density regressions. The ability of each hypothesis to predict the (lane 2) mean free speed for the facility was examined through the use of a t-test for significant difference between the predicted values and the independently measured value. The substudy, which consisted of sixteen 100-vehicle samples all having average densities of less than 10 veh/mi, yielded an estimated mean free speed of 57.9 mph with a standard deviation of 4.9 mph.

All hypotheses except the Greenshields relationship failed this test (see t-values in column seven, Table 2). The critical t-values were as follows:

| Confidence Level | t-Critical |
|---|---|
| 0.10 | 1.28 |
| 0.05 | 1.64 |
| 0.025 | 1.96 |
| 0.01 | 2.33 |
| 0.005 | 2.58 |

Hypothesis I exhibited a significant difference only at the 0.05, 0.025, etc., confidence levels. Ignoring the critical t-values, the actual t-values suggested that the Edie rela-

TABLE 3

STATISTICS FOR INDIVIDUAL REGIMES OF DISCONTINUOUS HYPOTHESES

| Hypothesis | Regime | $r^2$ | $s_e$ |
|---|---|---|---|
| II | I | 0.679 | 5.343 |
| | II | 0.717 | 2.373 |
| III | I | 0.125 | 2.122 |
| | II | 0.561 | 6.066 |
| | III | 0.717 | 2.373 |
| IV | I | 0.093 | 2.292 |
| | II | 0.874 | 4.214 |
| VI | I | 0.299 | 3.858 |
| | II | 0.767 | 3.424 |

tionship and the two linear regimes hypothesis (previously rejected) were considerably less inaccurate than the other hypotheses (except Greenshields) on a relative basis.

In one respect, an examination of mean free speed as performed here is little more than a "curious extrapolation," since no car-following rule is intended to apply at very low densities. Indeed, the breakpoint analysis for hypothesis IV suggested that no such (statistically significant) interaction occurred at densities less than 35 veh/mi. Since in the final analysis some type of car-following rules theoretically underlies, or could probably be discovered to underlie all hypotheses tested, the results of these mean free speed tests were not surprising. The very poor performance of the modified Greenberg hypothesis in predicting mean free speed, however, was quite disappointing since exacting procedures were employed to force a condition of "no interaction" on its low-density regime.

A truly comprehensive methodology would have imposed similar "hats" on all other hypotheses. The Greenberg equation was so modified primarily to permit a finite mean free speed prediction for purposes of comparative testing. Had a zero-slope regime been forced on other hypotheses, perhaps a variety of mean free speed predictions would have resulted, depending on the behavior of the likelihood function for each hypothesis. The complexity of the likelihood function analysis increased considerably with number of regimes, and time did not permit such extensions.

It should be observed that two hypotheses (V and VI) were formulated to include mean free speed explicitly as a parameter. In view of the observations discussed above, interpretation of this parameter as mean free speed per se in any accurate sense implies perhaps too rigorous a faith in such relationships.

In summary, from the standpoint of verifying theory, the test for mean free speed prediction was not meaningful. An improvement might have been to impose "hats" on all hypotheses. From the standpoint of practice, if one is interested only in predicting mean free speed, but is restricted to measurements at higher ranges of density, the Greenshields form would appear to be most appropriate.

### Measures of Goodness-of-Fit

Table 2 gives the values of the composite coefficient of determination ($r^2$) and the composite standard error of estimate ($s_e$) for all hypotheses. The quotation of such composite statistics for discontinuous hypotheses was intended to facilitate comparisons among all relationships considered. (Values of $r^2$ and $s_e$ for individual regimes of each discontinuous relationship are given in Table 3.)

Table 2 gives the consistent tradeoff between $r^2$ and $s_e$: the higher the $r^2$ value, the higher the standard error. Because of this consistent tradeoff, these statistics taken together at face value gave no meaningful comparative information. One interesting result, however, was the tendency for these two composite statistics to decrease with the number of regimes hypothesized.

## INTERPRETATION

### Results

Although the tests described followed quite religiously the rejection philosophy underlying the entire research effort, they failed to give very meaningful results after all. It should be emphasized that this failure in no way implies that similar rigor should not be applied in other research endeavors. To develop a rigorous test structure within which binary conclusions might be permitted is one matter; the tests described served

this criterion well. But the structure itself was not sufficient to permit meaningful comparisons, simply because most hypotheses behaved either extremely well or extremely poorly and consistently so for each test. There were always considerable variations among the F- and t-values for each hypothesis. But compared to critical values for conventional confidence limits, these differences were negligible. The tests could have exhibited considerable falsifiability had critical values been applied for levels of confidence on the order of $10^{-5}$. The rejection of a hypothesis at such a confidence level presents a difficult problem in interpretation, for the difference between $10^{-5}$ and $10^{-6}$ is not likely to be as intuitively evident to the researcher as that between 0.01 and 0.005.

While it might appear from these results that classical tools of statistical inference are not necessarily the key to successful experimentation, it should be realized that more meaningful tests could have been developed had it not been for the fact that independent estimates of parameters other than mean free speed were not available. Reference is made specifically to jam density and maximum flow. Measuring mean free speed in an independent fashion was (and is, in general) a much simpler task than acquiring independent statistical estimates of either $k_j$ or $q_m$. The only way to estimate jam density would have been to measure the average density over a mile-long section of lane 2 (or some sufficiently great fraction of that distance) when the average flow rate was negligible, and to conduct such measurements with frequency sufficient to yield a meaningful sample size (such an analysis of the facility studied here would have required advanced knowledge of the occurrence of accidents). Similarly, an independent estimate of maximum flow is not easy to determine.

Because the various hypotheses endured these tests with so little differentiation, there remained considerable latitude for judgment on more directly intuitive grounds. While the reader is encouraged toward self-interpretation, it seems appropriate to point out the more obvious deficiencies in the various hypotheses outside of the contest of a strict test structure:

1. The value of mean free speed (76.8 mph) predicted by the Underwood curve was considerably high, particularly since this relationship was developed partly to permit a finite estimate of this parameter (as opposed to the original Greenberg relationship).
2. Previous experience with the operation of the study location and inspection of speed-volume diagrams indicated an optimum speed in the vicinity of at least 40 mph. Except for the relationship advanced by Edie and the three linear regimes hypothesis, the various values for optimum speed were rather low.
3. The jam density value (125 veh/mi) predicted by the Greenshields hypothesis was extremely low.

Considering these observations in a "rejection" context, it became evident that the Edie form (and perhaps the three linear regimes hypothesis, ignoring its relatively poor prediction of mean free speed) warranted further attention. Three further observations were of interest in this connection:

1. These two hypotheses yielded the highest estimates for maximum flow (1,845 and 2,025 veh/hr for hypotheses III and VI, respectively);
2. They furthermore yielded the highest (finite) estimates for jam density (151 and 162.5 veh/mi for III and VI, respectively); and
3. Although they resulted in the two lowest values for $r^2$, they yielded the two lowest values for the standard error of estimate.

The fact that these two hypotheses gave the highest estimates for maximum flow is significant. That a flow of 1,845 to 2,025 veh/hr is possible on the middle lane of a three-lane urban freeway is not an unreasonable contention. Indeed, within some upper bound (say, 2,100 veh/hr), a high value is most appealing since this parameter estimates the maximum flow which might possibly be sustained over a meaningful time period.

With regard to these two maximum flow values, the maximum in each case was not a local maximum, but rather a boundary value of the low-density regime. In particular, the extension of the freeflow curve of hypothesis VI yielded an "optimal" density (of interest as a parameter only) of 163.0 veh/mi. This extremely high value, which

actually exceeded the jam density value given by the forced-flow curve, indicated a definite discrepancy in comparison to Edie's results for Lincoln Tunnel data (9). Allowing the two regimes to overlap, he determined a break-point "range" of 75 to 100 veh/mi (presumably by inspection), and an optimum density of 90 veh/mi. While it is generally agreed that optimum density is lower for a freeway than for a tunnel, the parameter representing the corresponding local maximum location in this research was almost twice as high as the value determined by Edie.

Inspection of the sample observations (Fig. 2) shows that the form of each regime was somewhat sensitive to the location of the break-point. Consequently it would not be surprising to find, using a similar likelihood function analysis, that $k_m$ for some other urban freeway were considerably lower than 164 veh/mi. This is particularly true since the study location was immediately upstream from a section with slightly lower capacity. For the data given in this research, however, a very vigorous methodology was exercised in locating the break-point; the 163.9 figure tests upon quite sound statistical procedures. To have assumed an overlap between regimes might have led to considerably different results, but his assumption could not have been treated in any fashion except by inspection for "best fit." Whether an overlap is desirable from a theoretical standpoint is debatable.

The second observation listed—that hypotheses III and VI also predicted the highest (finite) values for jam density—is also significant. In general, the (finite) $k_j$ values predicted by all hypotheses seemed quite low. In terms of extremes, it is conceivable that a density of roughly 300 veh/mi is potentially measurable (bumper-to-bumper conditions for one mile). More realistically, it is generally agreed that the jam density for a freeway lane is less than for a tunnel lane. But even the value estimated by hypothesis VI (162.5 veh/mi) was far below Edie's result for the Lincoln Tunnel (265 veh/mi). Of course, neither value has been checked against an independent estimate; thus, whether the tunnel value is too high or the freeway values are too low is yet to be resolved. (At k = 265 veh/mi, both hypotheses V and VII predicted a value of speed which was not significantly greater than zero.)

Finally, with regard to the third observation listed, the fact that hypotheses III and VI yielded the two lowest values for the standard error of estimate was somewhat appealing. After all, the end purpose of a regression analysis is to predict; this point of view would tend to favor the use of $s_e$ over $r^2$ as a figure-of-merit.

In summary, the results tended to support these two hypotheses above all others considered (according to the interpretations). From the standpoint of logical theoretical consistency, the Edie hypothesis certainly excels in comparison to the three linear regimes alternative. From the less elegant standpoint of application, however, all hypotheses (except the two linear regimes) performed well enough to warrant continued use.

The converted speed-volume and volume-density plots of the regression lines of the various hypotheses are shown in Figures 11 through 31 in the Appendix. In these forms, the three linear regimes, and particularly the Edie formulation, are somewhat more satisfying in the maximum flow range.

## Methodology

Several novel techniques were employed in this research. Of interest is the computation of representative composite statistics for discontinuous and/or nonlinear relationships. These techniques should be considered for continued use and improvement. The state of regression theory has lagged the need for treating complicated relationships such as those analyzed in this research. While the techniques used here suffered from certain limitations, they seem to offer the best available means for conducting such a comparative analysis.

The most important result of this research, as far as experimental techniques are concerned, was the successful implementation of the break-point analysis suggested by Quandt. The break-points resulting from these analyses agreed quite well with visual inspections of the sample scatter diagram. Such analyses were time-consuming only because of the magnitude of the particular study. The times of computer runs and

complexity of analyses were not at all overburdening. The only disadvantage of the technique was that the large scope of the study did not permit a very precise exploration of the likelihood functions; consequently, the results based on such break-point locations were not as accurate as might be desired.

The validity of the break-point technique was checked by applying it to a set of data for which three linear regimes had been hypothesized and located by a factor analysis scheme (12). The resulting break-points given by the likelihood function analysis were virtually identical to the results of the factor analysis.

Finally, it seems appropriate to emphasize one point: the question of establishing a test structure to permit meaningful statistical inference. Establishing a test structure in terms of confidence limits is a necessary but insufficient condition for providing a meaningful yes-or-no decision. The sufficient condition is that the test be designed to give such binary decision power within the range of intuitively comprehensible confidence limits.

## CONCLUSIONS

The important results regarding the comparison of alternative hypotheses are listed below. It should be emphasized that these conclusions reflect the authors' intuitions to some degree.

1. The Greenshields hypothesis prediction of mean-free-speed was much superior to that of all other hypotheses. The importance of this prediction is slight from a theoretical standpoint since no following-rule is presumed to apply at such low-density conditions.
2. The supposition of two linear regimes was shown to be insignificantly different from one linear regime.
3. Although its performance on the mean-free-speed test was relatively poor, the three linear regimes hypothesis gave relatively good estimates of optimum speed, maximum flow and jam density. Its standard error was the second lowest of the seven hypotheses; its coefficient of determination was lowest.
4. The modified Greenberg hypothesis exhibited only fair performance with respect to all parameters examined. Its values for maximum flow, mean-free-speed, and optimum speed were slightly low, but not distinctly so. The break-point analysis for this form implied that the threshold of interaction between vehicles on the facility was about 35 vehicles/lane-mile, a spacing of 150 ft.
5. Except for having the highest $r^2$ value, the Underwood form gave poor results. While no statistical test offered such a distinction, the extremely high value of mean-free-speed was quite disturbing, particularly since this relationship was developed partly to permit a finite estimate of this parameter. Furthermore, both the optimum speed and maximum volume values were low.
6. The Edie formulation gave the best estimates of the fundamental parameters. While its $r^2$ was the second lowest, its standard error was the lowest of all hypotheses. The analysis of this research yielded a somewhat different form than that given in Edie's own work, in that the optimum density value was found to occur at the boundary between the free flow and congested regimes, rather than at the apogee of the free flow curve.
7. In no respect was the bell curve outstanding in its predictive ability. A relatively low estimate of mean-free-speed was anticipated by its very formulation. However, the actual mean-free-speed of the facility appeared to be considerably higher than predicted by this hypothesis.

From a methodological point of view, the following results of this study were of interest:

1. A maximum likelihood technique for locating optimal break-points in multi-regime regression analyses was implemented successfully.
2. Novel techniques of estimating goodness-of-fit for nonlinear and/or discontinuous regressions were developed and used.
3. The original intention of this research was to approach the problem of comparing alternative hypotheses by means of a rigorous structure of falsifiable tests. In the final

analysis, however, almost all conclusions were based on intuition alone since the statistical tests provided little decision power after all. Establishing a test structure in terms of confidence limits is a necessary but insufficient condition for providing a meaningful yes-or-no decision. The sufficient condition is that the test be designed to give such binary decision power within the range of conventional confidence limits.

## ACKNOWLEDGMENT

This paper is one of a series of reports of the Chicago Area Expressway Surveillance Project sponsored by the Illinois Division of Highways in cooperation with the Bureau of Public Roads, Cook County, and the city of Chicago. The cooperation received from Northwestern University in carrying out this study is greatly appreciated.

## REFERENCES

1. May, A. D., Jr. Discussion of "Freeway Level of Service as Influenced by Volume and Capacity Characteristics," by D. R. Drew and C. J. Keese. Highway Research Record 99, pp. 39-43, 1965.
2. Wardrop, Charles. Some Theoretical Aspects of Road Traffic Research. Proc. Inst. of Civil Eng. (London), Vol. 1, pp. 325-362, 1952.
3. Gazis, D. C., Herman, R., and Potts, R. Car-Following Theory of Steady-State Traffic Flow. Operations Research, Vol. 7, pp. 499-595, 1959.
4. Gazis, D. C., Herman, R., and Rothery, R. W. Nonlinear Follow-the-Leader Models of Traffic Flow. Operations Research, Vol. 9, pp. 545-567, 1961.
5. Greenshields, B. D. A Study in Highway Capacity. HRB Proc., Vol. 14, p. 468, 1935.
6. Ellis, R. H. Analysis of Linear Relationships in Speed-Density and Speed-Occupancy Curves. (unpublished report) Northwestern University, Dec. 1964.
7. Greenberg, H. An Analysis of Traffic Flow. Operations Research, Vol. 7, pp. 78-85, 1959.
8. Underwood, R. T. Speed Volume, and Density Relationships: Quality and Theory of Traffic Flow. Yale Bureau of Highway Traffic, pp. 141-188, 1961.
9. Edie, L. C. Car Following and Steady-State Theory for Non-Congested Traffic. Operations Research, Vol. 9, pp. 66-76, 1961.
10. Quandt, R. E. The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. Journal of the American Statistical Association, Vol. 53, pp. 873-880, 1958.
11. Quandt, R. E. Tests of the Hypothesis That a Linear Regression System Obeys Two Separate Regimes. Journal of the American Statistical Association, Vol. 25, pp. 324-330, 1960.
12. Drake, J. S. Test for a Hierarchy in Southwest Iowa: A Methodological Exercise. (unpublished report) Northwestern University, March 1965.

# *Appendix*

## GRAPHS OF HYPOTHESES

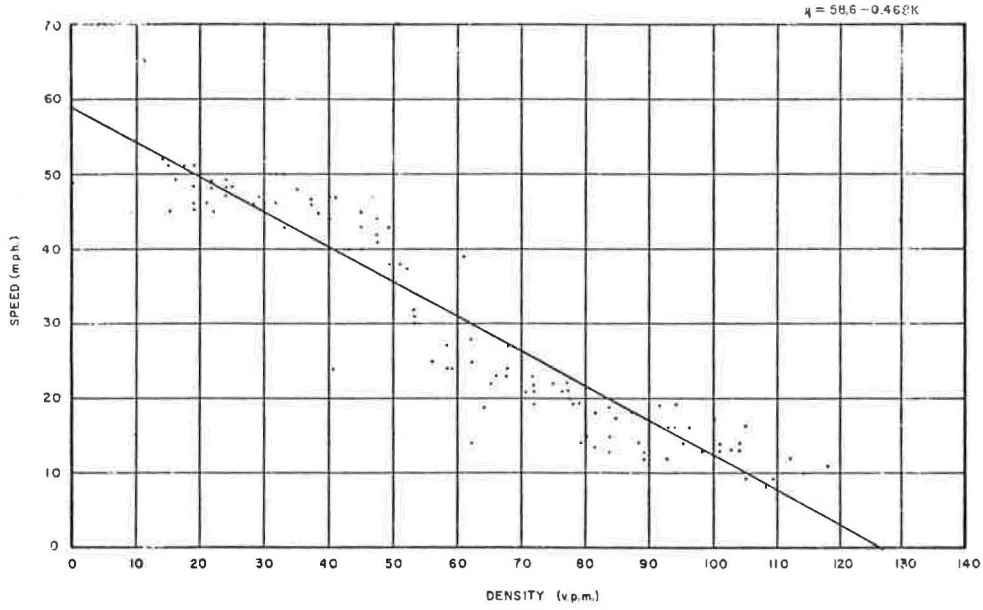



Figure 11. Speed-density relationship—Greenshields hypothesis.

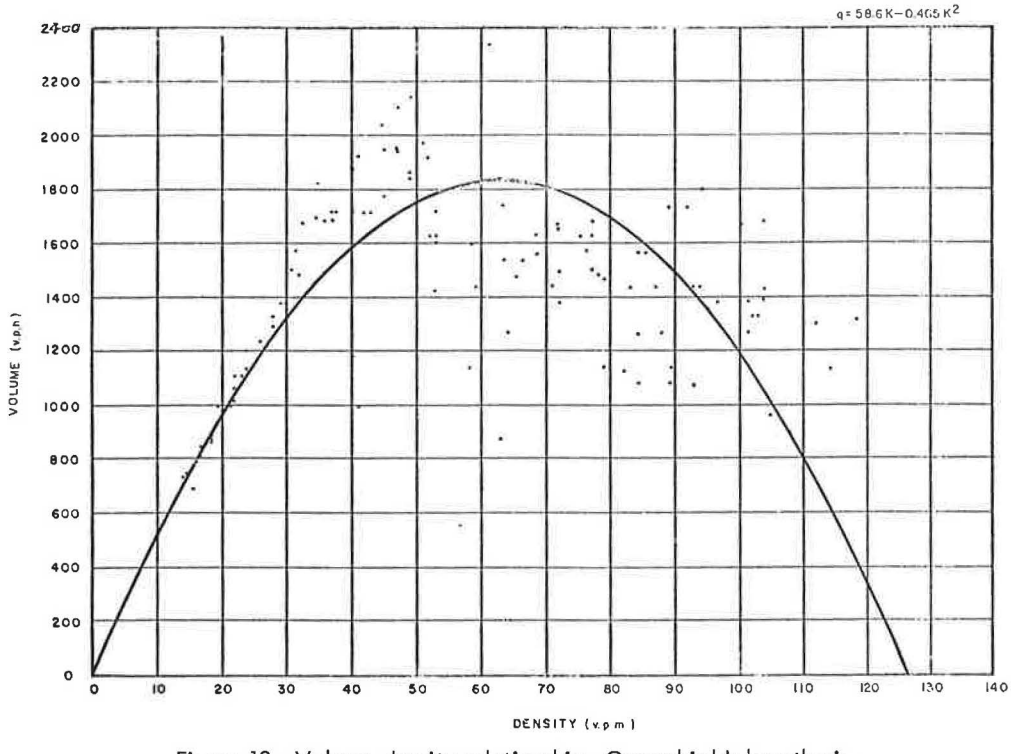



Figure 12. Volume-density relationship—Greenshields hypothesis.

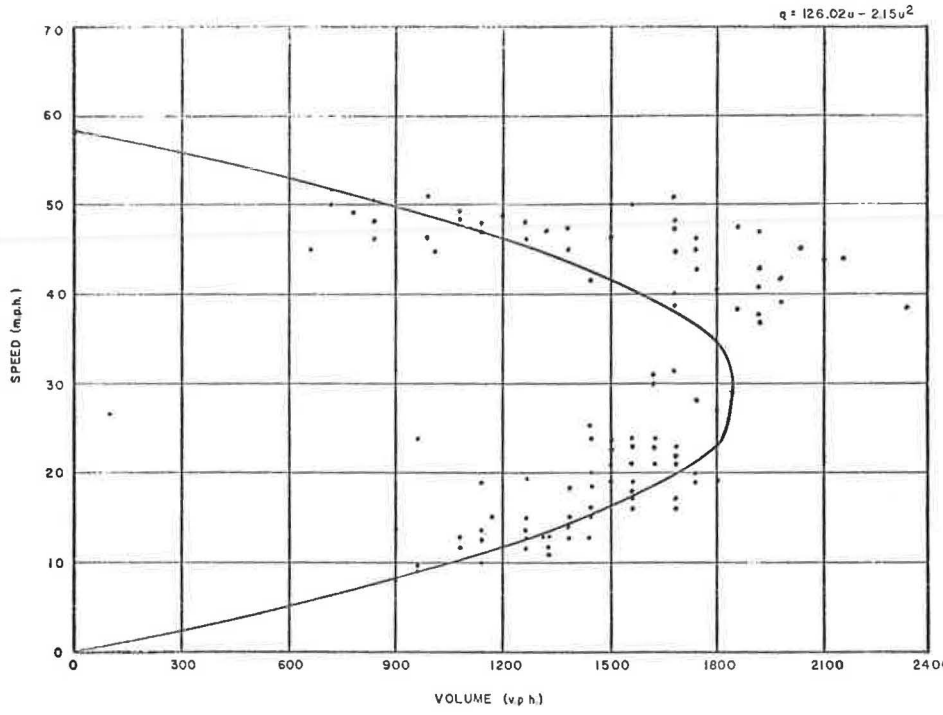



Figure 13. Speed-volume relationship—Greenshields hypothesis.

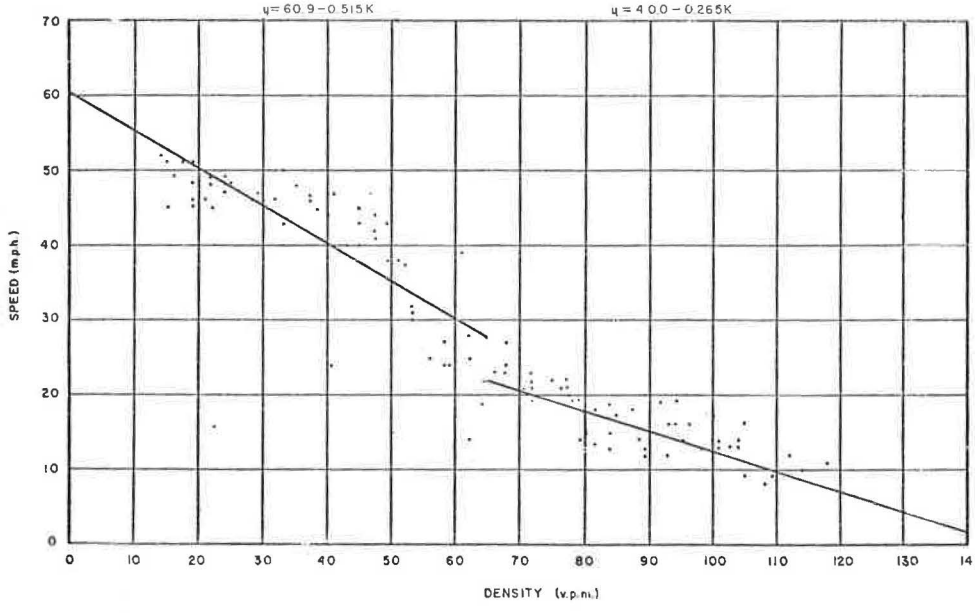



Figure 14. Speed-density relationship—linear, two regime hypothesis.

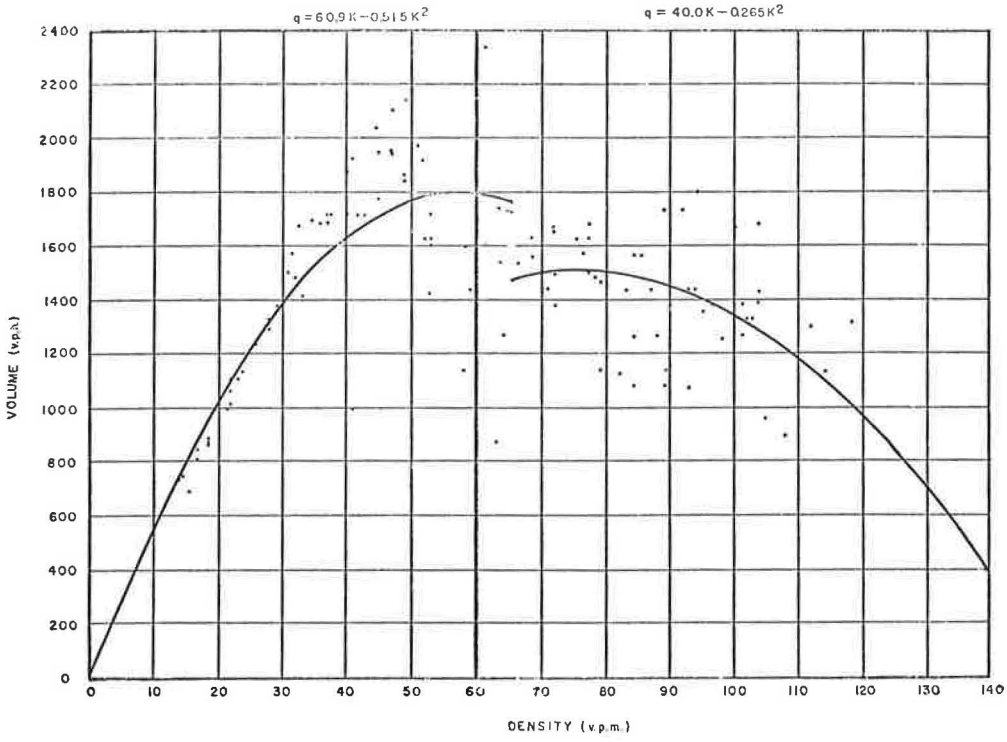



Figure 15. Volume-density relationship—linear, two regime hypothesis.

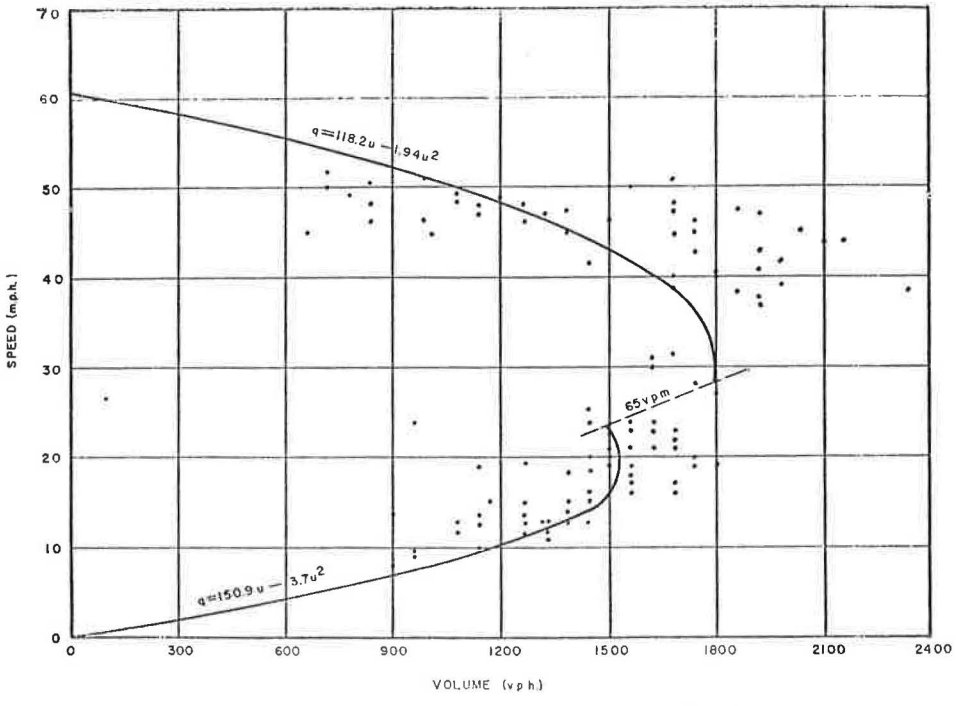



Figure 16. Speed-volume relationship—linear, two regime hypothesis.

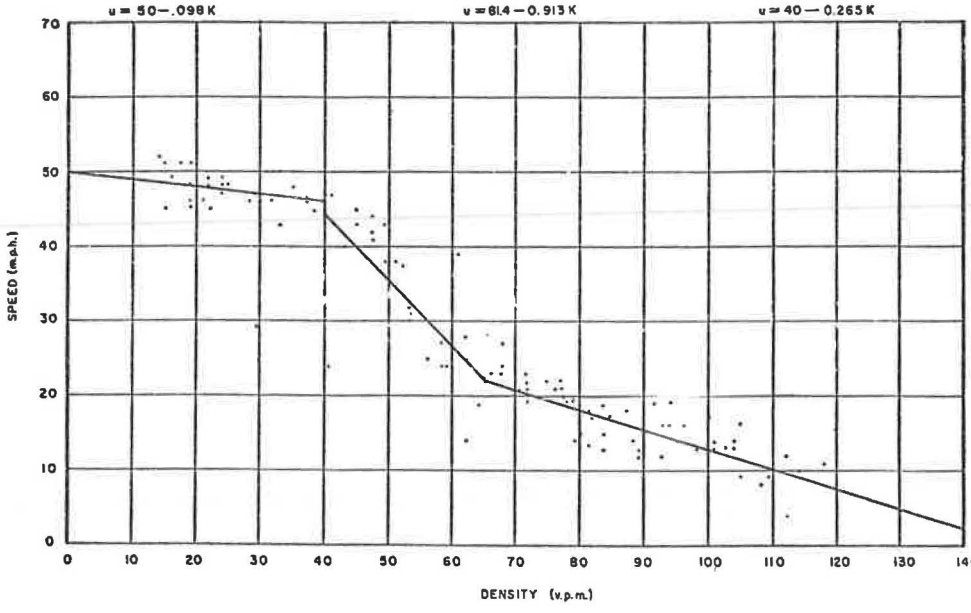



Figure 17. Speed-density relationship—linear, three regime hypothesis.

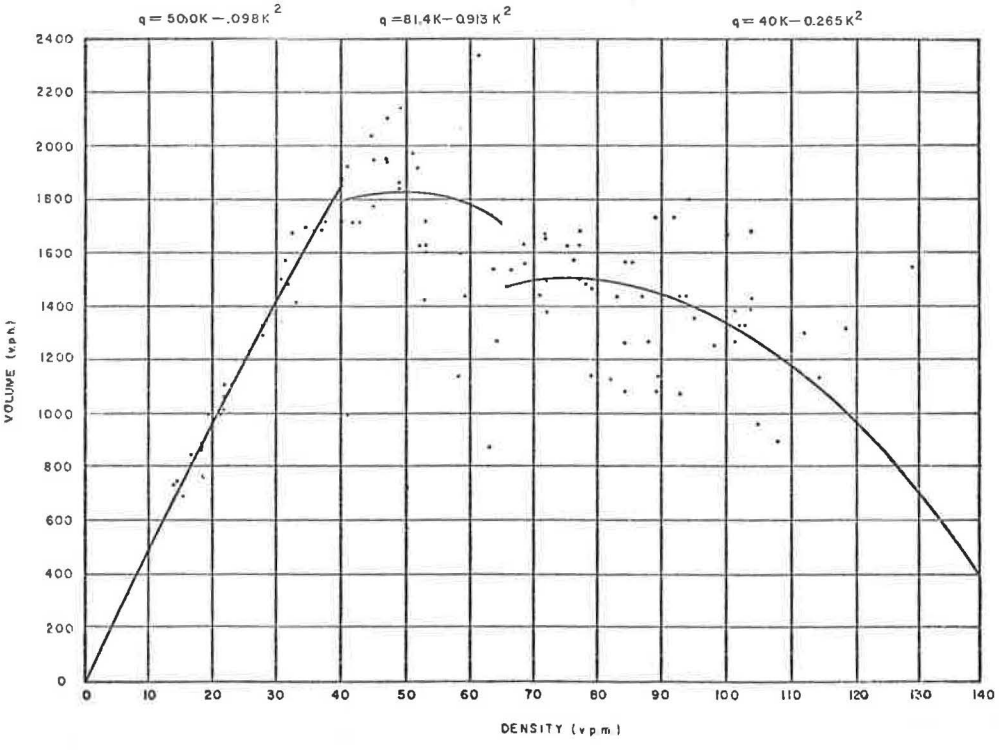



Figure 18. Volume-density relationship—linear, three regime hypothesis.

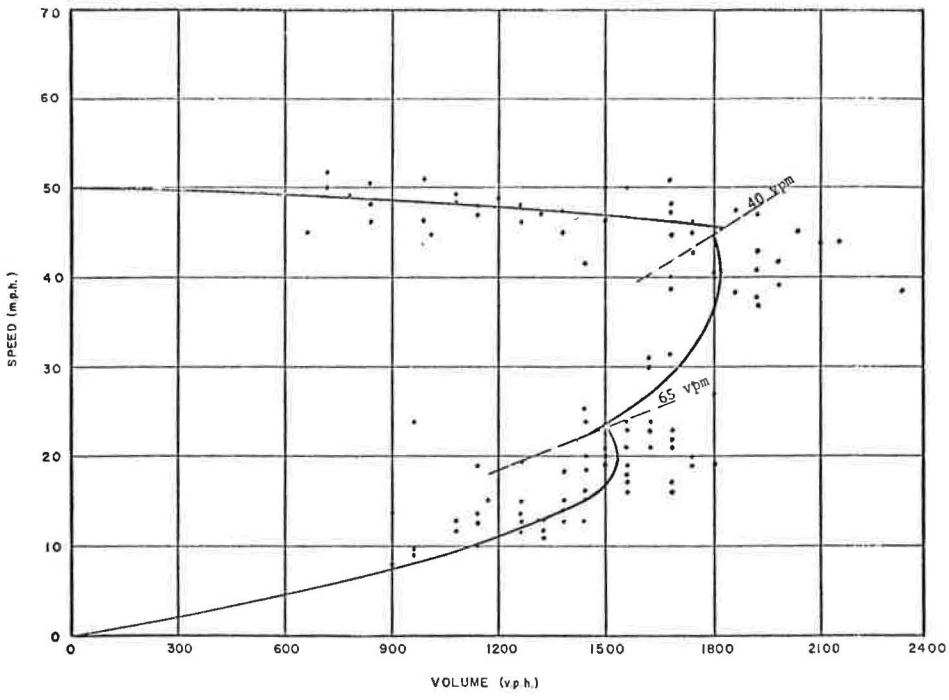



Figure 19. Speed-volume relationship—linear, three regime hypothesis.

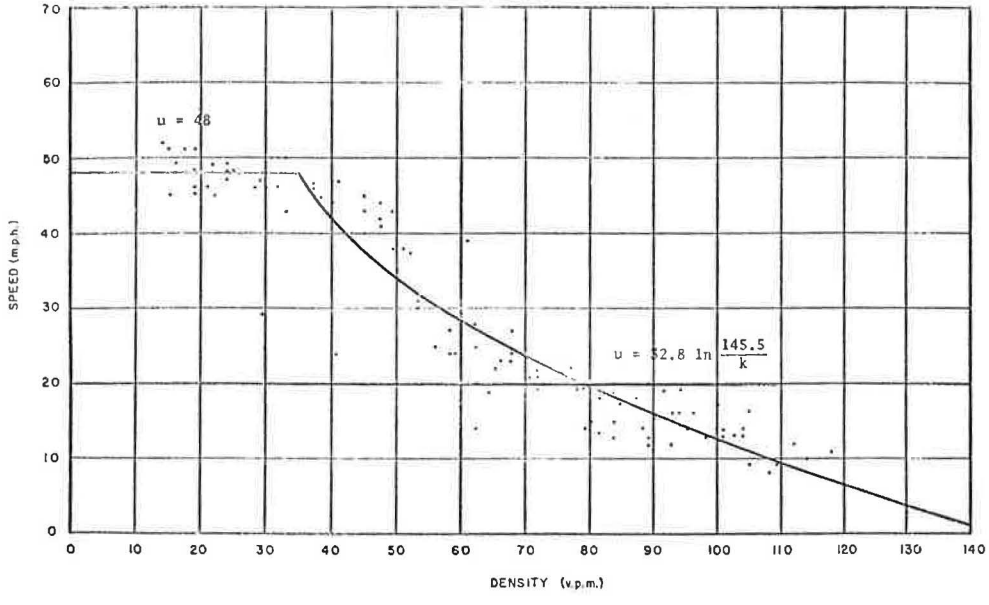



Figure 20. Speed-density relationship—modified Greenberg hypothesis.

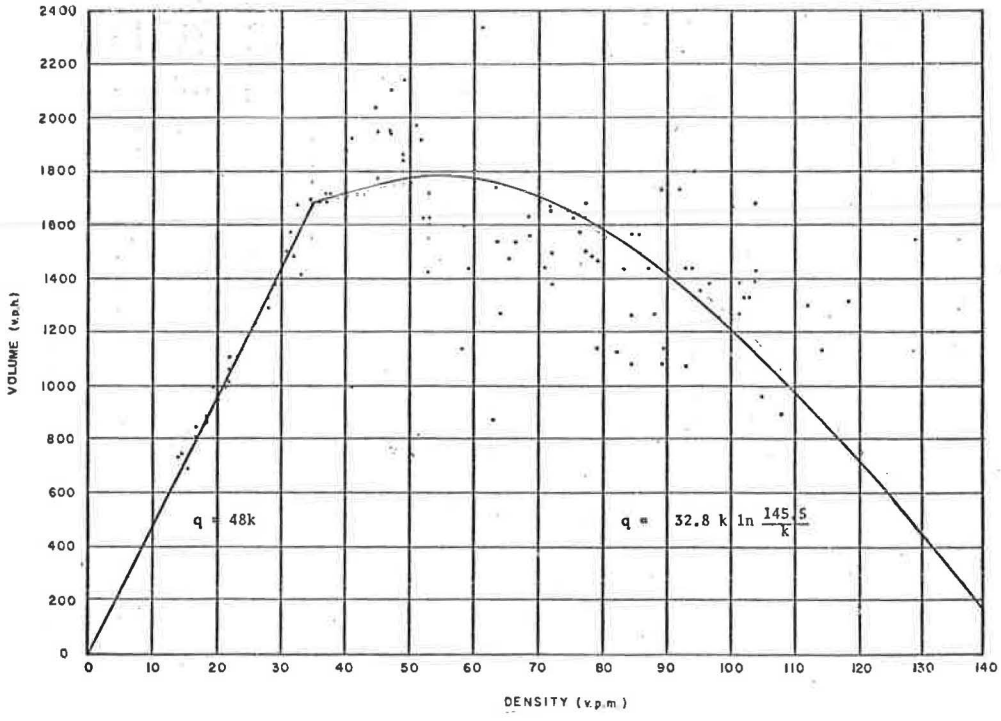



Figure 21. Volume-density relationship—modified Greenberg hypothesis.

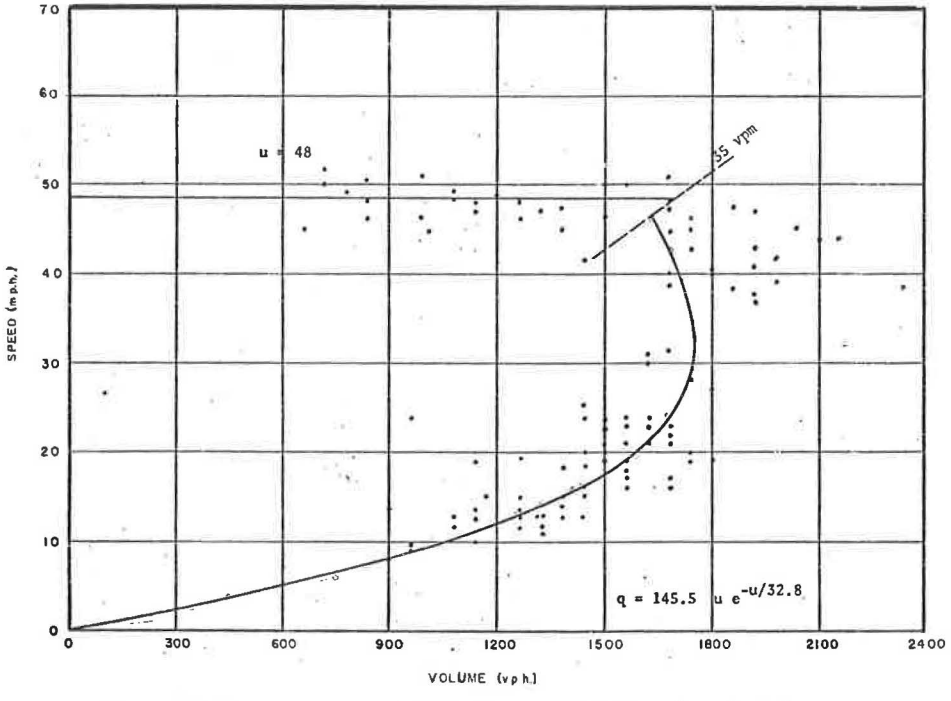



Figure 22. Speed-volume relationship—modified Greenberg hypothesis.

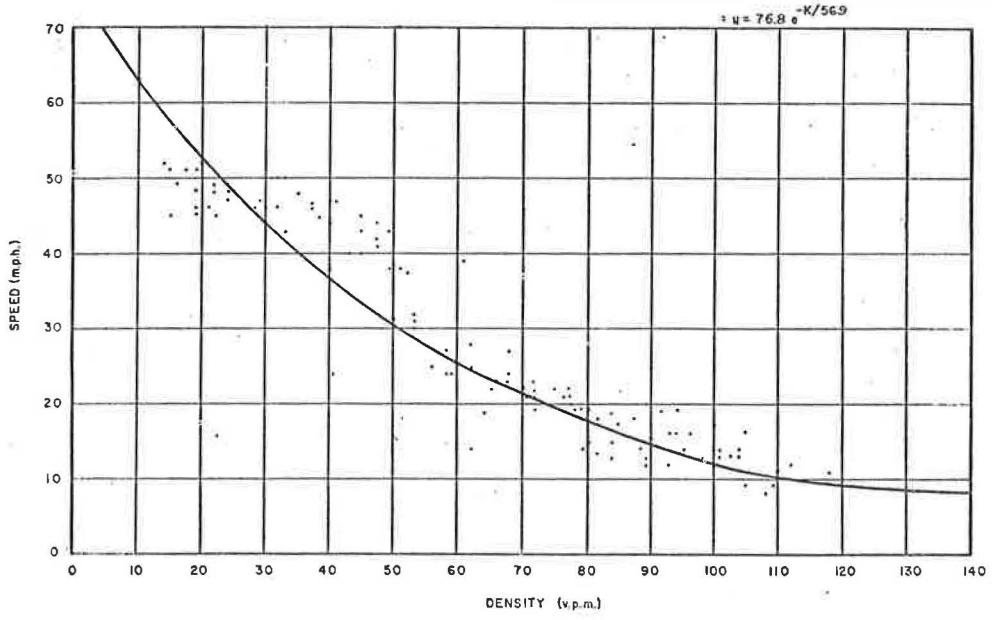



Figure 23. Speed-density relationship—Underwood hypothesis.

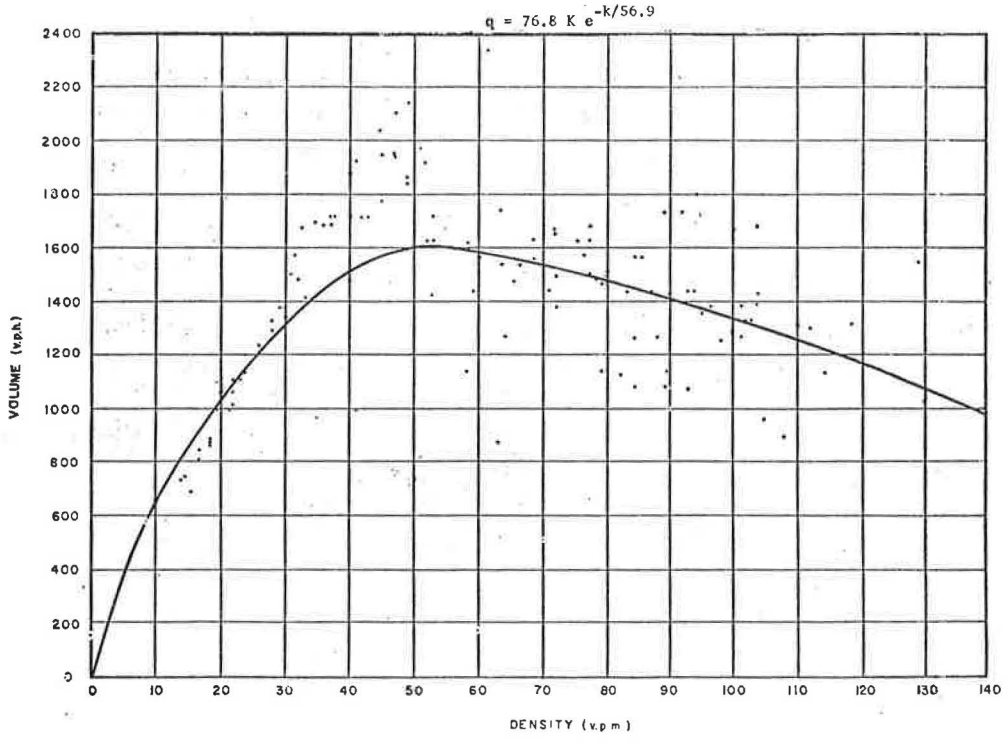



Figure 24. Volume-density relationship—Underwood hypothesis.

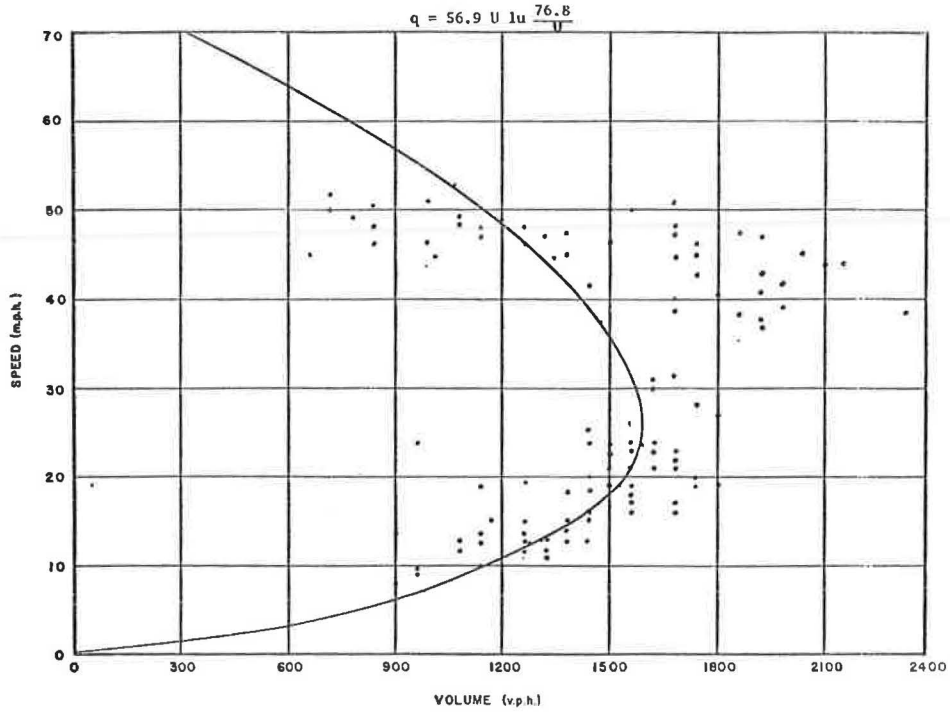



Figure 25. Speed-volume relationship—Underwood hypothesis.

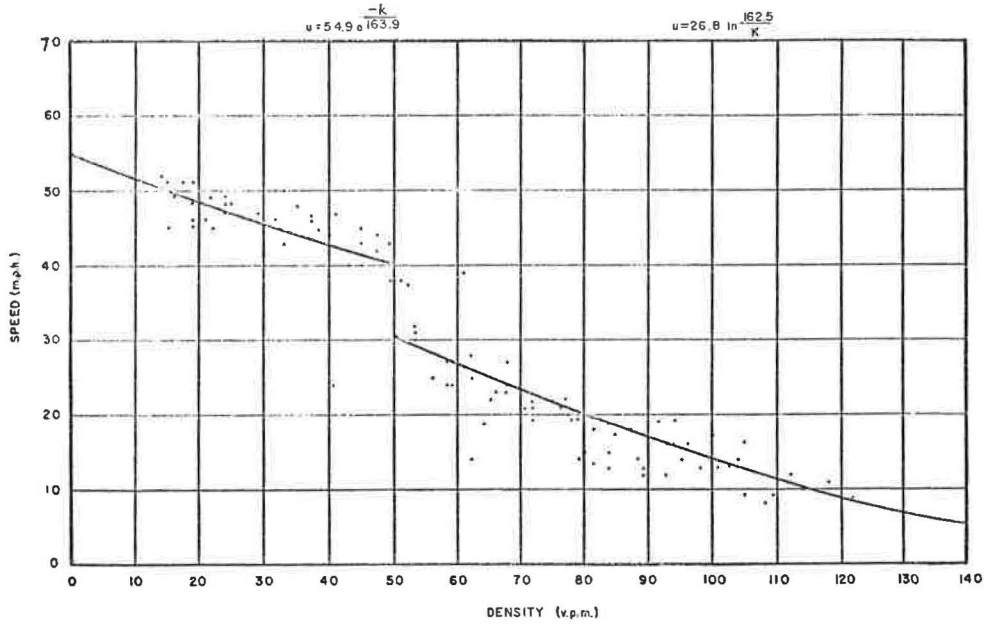



Figure 26. Speed-density relationship—Edie hypothesis.

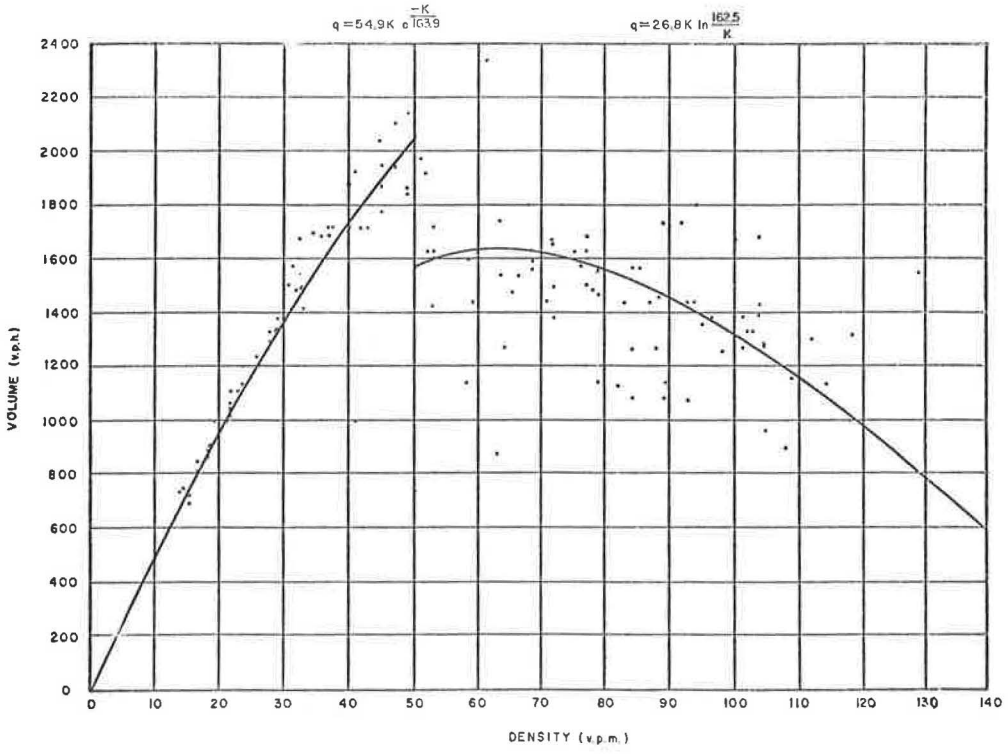



Figure 27. Volume-density relationship—Edie hypothesis.

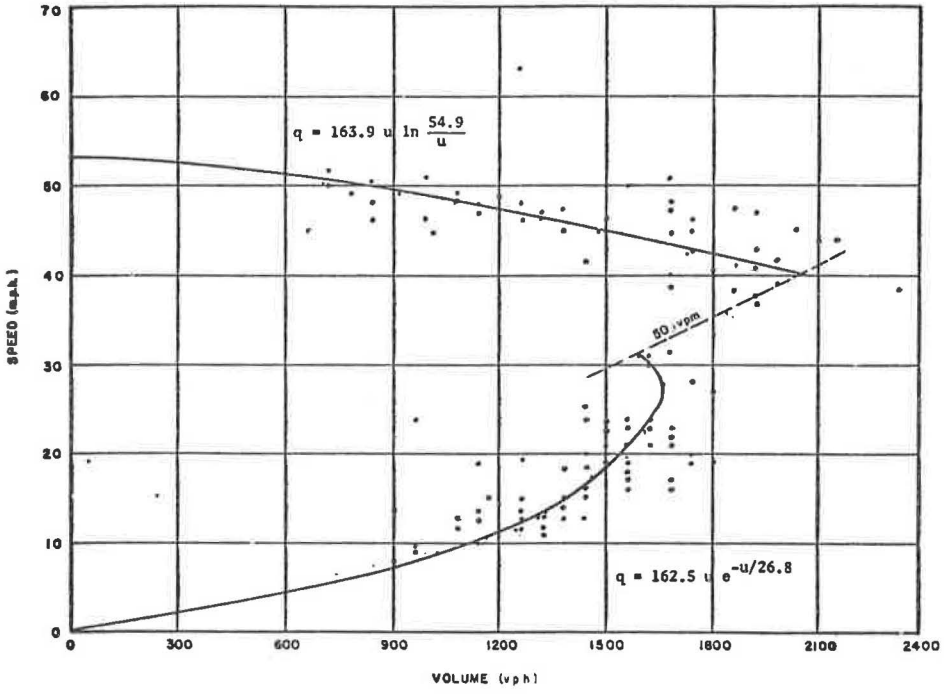



Figure 28. Speed-volume relationship—Edie hypothesis.

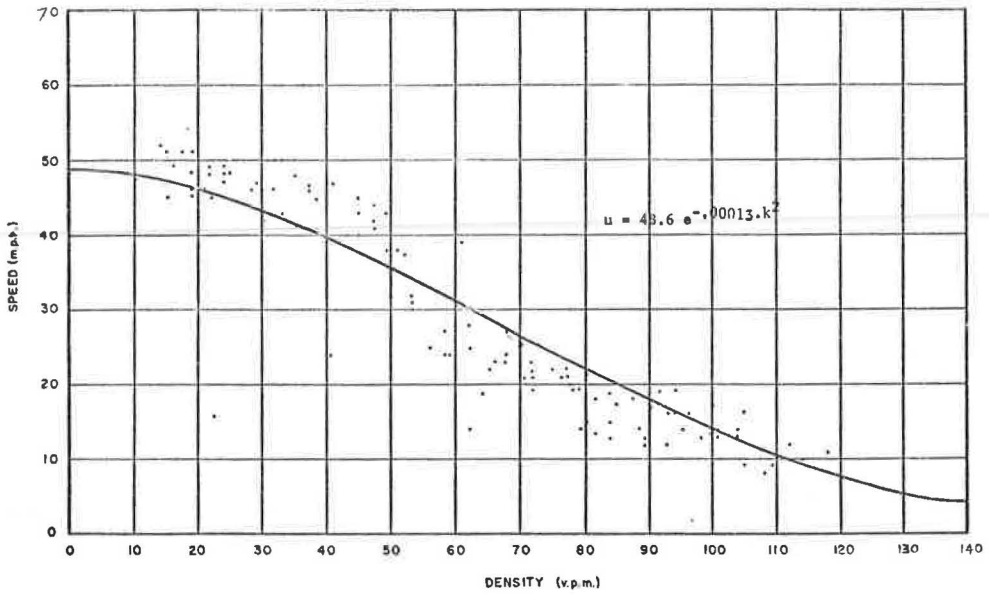



Figure 29. Speed-density relationship—bell curve hypothesis.

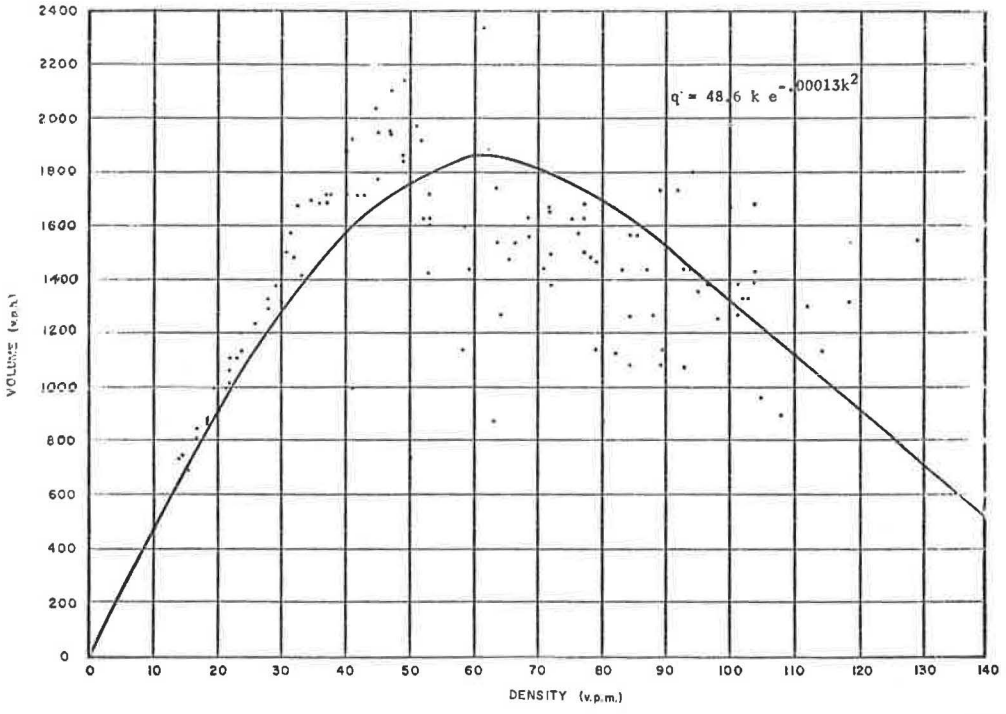



Figure 30. Volume-density relationship—bell curve hypothesis.

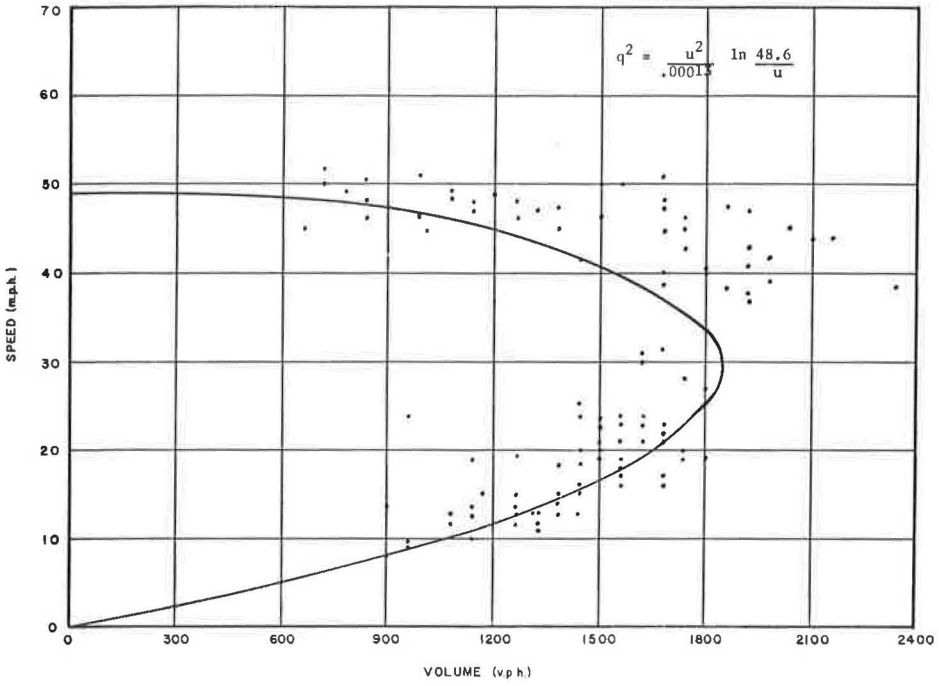



Figure 31. Speed-volume relationship—bell curve hypothesis.