TRANSPORTATION RESEARCH RECORD 772

Traffic Flow Theory, Characteristics, and Capacity

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1980

Transportation Research Record 772
Price $7.20
Edited for TRB by Mary McLaughlin

mode
1 highway transportation

subject area
55 traffic flow, capacity, and measurements

Library of Congress Cataloging in Publication Data
Traffic flow theory, characteristics, and capacity.

(Transportation research record; 772)
Reports prepared for the 59th annual meeting of the Transportation Research Board.
1. Highway capacity—Addresses, essays, lectures. 2. Traffic flow—Addresses, essays, lectures. 3. Traffic engineering—Addresses, essays, lectures. 4. Transportation planning—Addresses, essays, lectures. I. National Research Council (U.S.). Transportation Research Board. II. Series.
TE7.H5 no. 772 [HE336.H48] 380.5s 81-4563
ISBN 0-309-03116-8 [388.3'1] ISSN 0361-1981 AACR2

Sponsorship of the Papers in This Transportation Research Record

GROUP 3—OPERATION AND MAINTENANCE OF TRANSPORTATION FACILITIES
Adolf D. May, University of California, Berkeley, chairman

Committee on Highway Capacity and Quality of Service
James H. Kell, JHK & Associates, chairman
William R. McShane, Polytechnic Institute of New York, secretary
Brian L. Allen, Robert C. Blumenthal, Arthur A. Carter, Jr., Joseph W. Hess, Thomas D. Jordan, Frank J. Koepke, Jerry Kraft, Walter H. Kraft, Joel P. Leisch, Herbert S. Levinson, Edward B. Lieberman, Frederick D. Rooney, John L. Schlaefli, Gerald W. Skiles, Alexander Werner, Jeffrey M. Zupan

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

David K. Witheford, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each report. The organizational units, officers, and members are as of December 31, 1979.

Contents

Authors of the Papers in This Record

Andrews, B. J., KLD Associates, Inc., 300 Broadway, Huntington Station, NY 11746
Courington, William, Crain and Associates, 120 Santa Margarita Avenue, Menlo Park, CA 94025
Davila, M. C., KLD Associates, Inc., 300 Broadway, Huntington Station, NY 11746
Easa, Said M., Institute of Transportation Studies, University of California, Berkeley, 109 McLaughlin Hall, Berkeley, CA 94720
Glauz, William D., Midwest Research Institute, 425 Volker Boulevard, Kansas City, MO 64110
Glazer, Lawrence Jesse, Crain and Associates, 120 Santa Margarita Avenue, Menlo Park, CA 94025
Habib, Philip A., Polytechnic Institute of New York, 333 Jay Street, Brooklyn, NY 11201
Harwood, Douglas W., Midwest Research Institute, 425 Volker Boulevard, Kansas City, MO 64110
Heimbach, C. L., Department of Civil Engineering, North Carolina State University, Raleigh, NC 27607
Hoban, C. J., Monash University, Clayton, Victoria 3168, Australia
Khasnabis, S., Department of Civil Engineering, Wayne State University, Detroit, MI 48202
Lieberman, E. B., KLD Associates, Inc., 300 Broadway, Huntington Station, NY 11746
May, Adolf D., Institute of Transportation Studies, University of California, Berkeley, 109 McLaughlin Hall, Berkeley, CA 94720
McShane, William R., Polytechnic Institute of New York, 333 Jay Street, Brooklyn, NY 11201
Ogden, K. W., Department of Civil Engineering, Monash University, Clayton, Victoria 3168, Australia
Pignataro, Louis J., Polytechnic Institute of New York, 333 Jay Street, Brooklyn, NY 11201
Roess, Roger P., Polytechnic Institute of New York, 333 Jay Street, Brooklyn, NY 11201
St. John, Andrew D., Midwest Research Institute, 425 Volker Boulevard, Kansas City, MO 64110
Yedlin, M., KLD Associates, Inc., 300 Broadway, Huntington Station, NY 11746

Determining the Lateral Deployment of Traffic on an Approach to an Intersection

E. B. LIEBERMAN

An analytic model that predicts the lateral (i.e., lane-specific) deployment of traffic on an approach to an intersection is described. The basis for this model is a variation of Wardrop's first principle: that every motorist will select a lane on an approach consistent with his or her intended turn maneuver and with any specified lane channelization so as to minimize his or her perceived travel time. The specified conditions include the approach geometry; lane channelization, if any; specification of control policy at the intersection; service rate for each turn-movement component of the traffic stream; and turn percentages of the traffic stream discharging from the approach. The model yields the proportion of the total traffic volume entering the approach that is deployed in each lane, stratified by turn movement. The model is one component of a larger analysis that predicts the capacity of an approach and, by aggregation, intersection capacity. A brief outline of this capacity model is also included.

Intersection capacity, as defined in the 1965 Highway Capacity Manual (HCM), "actually represents individual approach capacity", whereas intersection service volume "usually refers to service volume on a particular approach" (1, p. 111).

Many investigators have developed analytic models to estimate approach capacity, while others have developed computational procedures for the practicing engineer. An excellent bibliography of this past work is available elsewhere (2). There still remained, in my view, a need to develop an analysis that includes an explicit description of all of the important factors that influence approach capacity. These factors are the following: geometrics, lane channelization, service rate for through vehicles, service rate for unimpeded right turners, service rate for unimpeded left turners, gap acceptance for left turners, percentage of left turners, percentage of right turners, oncoming approach geometrics, opposing volume, opposing service rate for through vehicles, cycle length, duration of green phase, signal phasing, and right turn on red.

In previous work, some of these factors have often been disregarded, or simplifying assumptions have been applied to reduce the complexity of the real-world process. Examples of such simplifications include the assignment of

1. "Passenger-car-equivalence" (PCE) factors to account for the different service rates associated with different turn movements,
2. "Left-turn factors" based on opposing traffic volumes to account for the dependence of left-turn service rates on oncoming traffic volume, and
3. Traffic volume on an approach to specific lanes in a somewhat arbitrary manner.

The assertion of simplifying assumptions in order to reduce a complex process to a form that is analytically tractable is, of course, a necessary and acceptable practice. But the impact of these assumptions on the efficacy of the results obtained must always be a potential source of concern, since it is generally very difficult--perhaps impossible--to assess the quantitative impact of such assumptions.

BACKGROUND

The model discussed here is an outgrowth of a project sponsored by the Federal Highway Administration (FHWA) to develop a macroscopic simulation program (TRAFLO). Since the integrity of such a model is strongly dependent on the accuracy of estimates of approach service rates, a survey of existing models was undertaken to identify one that would be acceptable for this application. No treatment that possessed the required generalism and precision could be located. Consequently, work was begun on developing a satisfactory capacity model for TRAFLO. This work resulted in the identification of the interdependence of the determination of approach service rates and the determination of the lateral deployment of traffic, by lane, on an approach.

Stated formally, the approach capacity ($\bar{s}$) can be expressed as follows:

$$\bar{s} = \sum_{\text{over all lanes } i} \bar{s}_i \qquad (1)$$

$$\bar{s}_i = \sum_{\text{over all movements } x} \bar{s}_i^x$$

$$\bar{s}_i^x = f_i(P_i^x; V_j) \qquad (2)$$

and

$$P_i^x = g_i(\bar{s}_i; V_k) \qquad (3)$$

where

$\bar{s}_i$ = service rate for all vehicles discharging lane i,
$\bar{s}_i^x$ = service rate for vehicles discharging lane i by executing maneuver x (x = L,T,R),
P_i^x = proportion of vehicles discharging lane i that are executing maneuver x (x = L,T,R),
L,T,R = left-turn, through, and right-turn maneuvers, respectively, and
V_j, V_k = sets of variables that may be treated as parameters, i.e., independent variables.

The functional relations $f_i(\ )$ and $g_i(\ )$, in aggregate, form a system. Because these functions are interdependent and highly nonlinear and are not all expressible explicitly in an algebraic format, it is necessary to solve the system in an iterative manner.

The system consists of an assemblage of interrelated models:

1. One model calculates the lateral deployment of vehicles on an approach. The lane-specific service rates $\bar{s}_i$ for all lanes are presumed to be known. Other required information includes approach-specific turn fractions, approach geometry, and lane channelization. This model yields the values of P_i^x.
2. One model calculates the service rate for traffic discharging as a mix of left-turning and through vehicles from an approach lane. The proportions of turn-movement-specific vehicles on this lane (P_i^x) are presumed to be known. Other required information includes mean acceptable gap for left turners, unimpeded discharge headways for through and left-turn vehicles, opposing traffic

volume and service rate, and signal-control timing. This model yields the values of $\bar{s}_i^x$ and $\bar{s}_i$ for this lane.

3. One model calculates the service rate for traffic discharging as a mix of right-turning and through vehicles from an approach lane. The proportions of turn-movement-specific vehicles on this lane (P_i^x) are presumed to be known. Other required information includes pedestrian intensity, unimpeded discharge headways for through and right-turn vehicles, and signal-control timing. This model yields the values of $\bar{s}_i^x$ and $\bar{s}_i$ for this lane.

The procedure for applying these models as a system is outlined below:

1. Estimate approach service rates for each lane ($\bar{s}_i$);
2. Apply the lateral deployment model to calculate the turn-maneuver proportions (P_i^x) for each lane;
3. Apply the left-turn model, given P_i^x for the inside lane, to calculate the values of $\bar{s}_i^x$ and $\bar{s}_i$ for that lane;
4. Apply the right-turn model, given P_i^x for the outside lane, to calculate the values of $\bar{s}_i^x$ and $\bar{s}_i$ for that lane;
5. Compare these computed values of $\bar{s}_i$ with those used for the prior iteration and, if any service rate differs significantly from its previous value, continue the iteration by returning to step 2 above; and
6. At convergence, all values of $\bar{s}_i$ and P_i^x are known.

The above models are designed for signal-controlled approaches and for uncontrolled approaches. Additional models were developed for sign-controlled approaches. Documentation of the entire model is given by Lieberman and others (3).

Fortunately, this iterative procedure is stable and rapidly convergent, and the entire process is computationally efficient. It is interesting to note that a procedure similar in concept (although less detailed), and one that is also rapidly convergent, is presented in the Swedish Capacity Manual (4).

This paper describes the first of the three models outlined above.

MODEL FORMULATION

The lateral--i.e., lane-specific--deployment of traffic on an approach is assumed to satisfy the following variation of Wardrop's first principle: that every motorist will select a lane on an approach consistent with his or her intended turn maneuver and with any specified lane channelization so as to mimimize his or her perceived travel time. On this basis, it will be shown that the following factors influence the lateral deployment of traffic on an approach: (a) specified turn proportions (T_L and T_R) for the traffic discharging from the approach; (b) number of lanes (N) on the approach at the stop line; (c) specified lane channelization, if any; and (d) defined service rates for traffic on each lane ($\bar{s}_i$).

The analysis is developed in stages, and the following approach configurations are treated:

1. An approach with a single lane (N = 1);
2. An approach with two lanes (N = 2), neither of which is channelized;
3. An approach with more than two lanes, none of which is channelized; and
4. An approach with at least two lanes, one or more of which are channelized exclusively for turning vehicles (an immediate extension of approach 3 above, not presented in this paper).

Approach with a Single (Unchannelized) Lane

For an approach with a single (unchannelized) lane, since there is only one lane, the solution is trivial:

$$P^L = T_L; P^R = T_R; P^T = 1 - P^L - P^R \tag{4}$$

where P^L, P^R, and P^T are the proportions of vehicles in the subject lane that execute left turns, right turns, and through movements, respectively.

Approach with Two Unchannelized Lanes

For an approach with two unchannelized lanes, the right-turning volume $Q_R = T_R Q$, the left-turning volume $Q_L = T_L Q$, and the through volume $Q_T = (1 - T_L - T_R)Q$, where Q is total entering volume. It follows immediately that

$$Q_1 = Q_R + (1 - p_2)Q_T \tag{5}$$

and

$$Q_2 = Q_L + p_2 Q_T \tag{6}$$

where Q_i = traffic volume discharging from lane i (i = 1,2), where lane 1 is the outside (curb) lane and lane 2 is the inside (median) lane; and p_2 = proportion of through traffic that discharges into (i.e., selects) lane 2 (the inside lane).

The above expressions can be written as follows:

$$Q_1 = Q[T_R + (1 - p_2)(1 - T_L - T_R)] \tag{7}$$

$$Q_2 = Q[T_L + p_2(1 - T_L - T_R)] \tag{8}$$

These equations assume that all right-turning vehicles will select the outside lane and all left-turning vehicles the inside lane. Note that the condition $Q = Q_1 + Q_2$ is satisfied by these equations.

Two conditions must be considered: (a) $1 - T_L - T_R = 0$ and (b) $1 - T_L - T_R > 0$.

The first case asserts that no vehicles execute through movements on leaving the approach (i.e., $Q_T = 0$). In this case, the value of p_2 has no meaning since $Q_1 = QT_R$ and $Q_2 = QT_L$; that is, the inside lane contains only left-turning vehicles (if any), and the outside lane contains only right-turning vehicles (if any). Consequently, $p_2^L = p_1^R = 1$.

The second case is more interesting, since the parameter p_2 exists and must be evaluated. Let $\bar{s}_i$ = mean service rate for traffic discharging from lane i in vehicles per second of green time (i = 1,2). These service rates must be known--i.e., must be determined by other models (see the paper by McShane and Lieberman elsewhere in this Record). Furthermore, let $\bar{h}_i$ = mean discharge headway for traffic discharging from lane i [$\bar{h}_i = (1/\bar{s}_i)$] and t_i = time to discharge the vehicles in lane i in one signal cycle C, measured from the start of the green (i = 1,2). By definition, $t_i = Q_i \bar{h}_i C$.

The cited variation of Wardrop's law is stated formally as

$$\left.\begin{array}{l} t_1 = t_2 \\ \text{or} \\ Q_1\bar{h}_1 = Q_2\bar{h}_2 \end{array}\right\} \tag{9}$$

Substituting Equations 7 and 8 yields

$$[T_R + (1 - p_2)(1 - T_L - T_R)]\,\bar{h}_1 = [T_L + p_2(1 - T_L - T_R)]\,\bar{h}_2 \quad (10)$$

Solving for p_2,

$$p_2 = [(1 - T_L)\bar{h}_1 - T_L\bar{h}_2]/[(1 - T_L - T_R)(\bar{h}_1 + \bar{h}_2)] \quad (11)$$

subject to $0 \leq p_2 \leq 1$.

Now derive the expressions for p_i^L, p_i^R, and p_i^T by using Equations 7 and 8. For the outside lane (i = 1),

$$P_1^L = 0;\ P_1^R = (Q_R/Q_1) = T_R/[T_R + (1 - p_2)(1 - T_R - T_L)] \quad (12)$$

For the inside lane (i = 2),

$$P_2^R = 0;\ P_2^L = (Q_L/Q_2) = T_L/[T_L + p_2(1 - T_R - T_L)] \quad (13)$$

In either lane,

$$P_i^T = 1 - P_i^L - P_i^R \quad (14)$$

Approach with More Than Two Unchannelized Lanes

For an approach with more than two unchannelized lanes (N > 2), as before, assume that all turning vehicles select their respective lanes, either inside (median) or outside (curb). The through vehicles can deploy over all N lanes in a manner to be determined.

In case a $(1 - T_L - T_R = 0)$, following the same rationale as for the two-lane approach,

$$P_N^L = P_1^R = 1 \quad (15)$$

and there is zero flow in the middle lane(s); m = 2,..., N - 1. In case b $(1 - T_L - T_R > 0)$, let p_i equal the proportion of through vehicles on the approach that select border lanes i = 1 and i = N. The remaining through vehicles are assumed to deploy uniformly over the middle lanes m. Then

$$Q_1 = Q_R + p_1 Q_T \quad (16)$$

$$Q_N = Q_L + p_N Q_T \quad (17)$$

$$Q_m = (1 - p_1 - p_N)Q_T/(N - 2) \qquad m = 2, 3, \ldots, N - 1 \quad (18)$$

which can be written as

$$Q_1 = Q[T_R + p_1(1 - T_R - T_L)] \quad (19)$$

$$Q_N = Q[T_L + p_N(1 - T_R - T_L)] \quad (20)$$

$$Q_m = Q[(1 - p_1 - p_N)(1 - T_R - T_L)]/(N - 2) \qquad m = 2,3, \ldots, N - 1 \quad (21)$$

Proceeding as before,

$$t_1 = t_m = t_N \quad (22)$$

where $t_i = Q_i\bar{h}_i C$.

Asserting that $t_1 = t_m$ and substituting Equations 19 and 21, then asserting that $t_N = t_m$ and substituting Equations 20 and 21, yield

$$[T_R + p_1(1 - T_R - T_L)]\,\bar{h}_1 = [(1 - p_1 - p_N)(1 - T_R - T_L)/(N - 2)]\,\bar{h}_m \quad (23)$$

$$[T_L + p_N(1 - T_R - T_L)]\,\bar{h}_N = [(1 - p_1 - p_N)(1 - T_R - T_L)/(N - 2)]\,\bar{h}_m \quad (24)$$

Solving these two equations simultaneously and recognizing that $\bar{s} = 1/\bar{h}$ yield

$$p_1 = p_N(\bar{s}_1/\bar{s}_N) + [T_L/(1 - T_R - T_L)](\bar{s}_1/\bar{s}_N) - [T_R/(1 - T_R - T_L)] \quad (25)$$

and

$$p_N = [1/(1 - T_R - T_L)]\,\{[1/(N - 2)(\bar{s}_m/\bar{s}_N) + 1 + (\bar{s}_1/\bar{s}_N)] - T_L\} \quad (26)$$

Then, with Equation 25,

$$p_1 = [1/(1 - T_R - T_L)]\,\{[p_N(1 - T_R - T_L) + T_L](\bar{s}_1/\bar{s}_N) - T_R\} \quad (27)$$

Finally,

$$p_m = (1 - p_1 - p_N)/(N - 2) \quad (28)$$

All values of p_i are subject to the condition $0 \leq p_i \leq 1$.

It is instructive to reexamine Equation 11, which can be written as follows:

$$p_2 = [1/(1 - T_R - T_L)]\,\{[1/1 + (\bar{s}_1/\bar{s}_2)] - T_L\} \quad (29)$$

It is seen that this expression is identical to Equation 26 for N = 2. It can also be shown that Equation 27 is valid for N = 2. Consequently, Equations 26-28 and the condition for p_i $(0 \leq p_i \leq 1)$ are applicable for all approaches where the number of unchannelized lanes is two or more.

To calculate the lane-specific mix of traffic, proceed as follows:

$$P_N^L = T_L/[T_L + p_N(1 - T_R - T_L)] \qquad P_N^R = 0 \quad (30)$$

$$P_1^R = T_R/[T_R + p_1(1 - T_R - T_L)] \qquad P_1^L = 0 \quad (31)$$

$$P_m^L = P_m^R = 0 \qquad m = 2,3, \ldots, N - 1 \quad (32)$$

For any lane i,

$$P_i^T = 1 - P_i^L - P_i^R \quad (33)$$

Note that $P_i^x = 0$ whenever $T_x = 0$ (x = L,R).

When $p_N = 0$, i.e., when only left turners occupy lane N, it implies that $t_N \geq t_m$ and $t_N \geq t_1$. To calculate p_1 when $P_n = 0$, only the condition $t_1 = t_m$ can be applied. Equation 23 then becomes

$$[T_R + p_1(1 - T_R - T_L)]\,\bar{h}_1 = [(1 - p_1)(1 - T_R - T_L)/(N - 2)]\,\bar{h}_m \quad (34)$$

Then,

$$p_1 = 1/[(N - 2)(\bar{s}_m/\bar{s}_1) + 1]\,\{1 - [T_R(N - 2)(\bar{s}_m/\bar{s}_1)/(1 - T_R - T_L)]\} \quad (35)$$

When $p_1 = 0$, i.e., when only right turners occupy lane 1, it implies that $t_1 \geq t_m$ and $t_1 \geq t_N$. To calculate p_N when $p_1 = 0$, only the condition $t_N = t_m$ can be applied. Equation 24 then becomes

$$[T_L + p_N(1 - T_R - T_L)]\,\bar{h}_N = [(1 - p_N)(1 - T_R - T_L)/(N - 2)]\,\bar{h}_m \quad (36)$$

Then

$$p_N = 1/[(N - 2)(\bar{s}_m/\bar{s}_N) + 1]\,\{1 - [T_L(N - 2)(\bar{s}_m/\bar{s}_N)/(1 - T_L - T_R)]\} \quad (37)$$

When $T_L = 0$, it can be shown that

$$p_1 = 1/[(N - 1)(\bar{s}_m/\bar{s}_1) + 1]\,\{1 - [(N - 1)T_R(\bar{s}_m/\bar{s}_1)/(1 - T_R)]\} \quad (38)$$

Then Equation 31 yields P_1^R and $P_m^L = P_m^R = 0$, where m = 2,3,...,N. When $T_R = 0$, it can be shown that

$$p_N = 1/[(N - 1)(\bar{s}_m/\bar{s}_N) + 1]\,\{1 - [T_L(N - 1)(\bar{s}_m/\bar{s}_N)/(1 - T_L)]\} \quad (39)$$

Then Equation 30 yields P_N^L and $P_m^L = P_m^R = 0$, where m = 1,2,..., N - 1.

This analysis also applies, with minimal

modification, to approaches with one or more lanes channelized exclusively for turning vehicles and/or for buses. Hence, the model is completely general for approaches of any geometric configuration and traffic demand of any intensity. Note that traffic volume is not a parameter--i.e., it does not appear in the equations yielding p_i and $P_i{}^x$.

The foregoing analysis has led to the development of formulas that permit the calculation of the lateral (lane-specific) deployment of traffic on an approach. The proportion of through and turning vehicles in each lane is shown to be dependent on known, or estimated, values of (a) number of lanes (N), (b) approach-specific turn-movement percentages (T_L and T_R for left- and right-turning vehicles, respectively), and (c) lane-specific service rates [$\bar{s}_N$, $\bar{s}_m$, and $\bar{s}_1$ for the inside (median), middle (if any), and outside (curb) lanes, respectively].

The values of n, T_L, and T_R are exogenous variables that must be known or estimated by the analyst. The mean value of service rates for through vehicles (s_T) must also be provided by the analyst. In addition, the service rates s_L and s_R, for unimpeded left- and right-turning vehicles, respectively, must be provided by the analyst. All of these service rates should be estimated by direct observation, preferably for the approach under consideration or for a similar approach configuration.

Since the middle lanes service only through vehicles, $\bar{s}_m = s_T$. The inside lane generally services a mix of left-turn and through vehicles. Its service rate ($\bar{s}_N$) is dependent on known, or estimated, values of (a) unimpeded left-turn service rate (s_L), (b) oncoming traffic volume (Q_{op}), (c) opposing approach geometry (number of through lanes, N_{op}), and (d) proportion of left-turning vehicles ($P_N{}^L$) in the lane.

It was shown that the value $P_N{}^L$ is a function of, among other parameters, the mean service rate $\bar{s}_N$ of traffic on the inside lane. Yet this service rate is not known a priori, since it, in turn, is a function of $P_N{}^L$. This condition requires the application of an iterative procedure that simultaneously produces the values of $P_N{}^L$ and $\bar{s}_N$. Similar comments apply to the variables $P_1{}^R$ and $\bar{s}_1$ (see Equations 1-3).

This iterative procedure is outlined below for $N \geq 2$:

1. Assert that $P_N{}^L = 1.0$ and that $P_1{}^R = 1.0$ and calculate $\bar{s}_i^{(0)}$ on this basis (the superscript attached to $\bar{s}_i$ denotes the iteration n.) Set n = 1.

2. Calculate p_N by using Equation 26. If $p_N \leq 0$, continue with step 3. Otherwise, continue with step 4.

3. Calculate p_1 by using Equation 35 and continue with step 6.

4. Calculate P_1 by using Equation 27. If $P_1 \leq 0$, continue with step 5. Otherwise continue with step 6.

5. Calculate p_N by using Equation 37 and continue with step 6.

6. Calculate p_m by using Equation 28 if $N > 2$, and $P_N{}^L$, $P_1{}^R$, and $P_1{}^T$ with Equations 30-33.

7. Calculate $\bar{s}_N^{(n)}$ and $\bar{s}_1^{(n)}$ and compare with $\bar{s}_N^{(n-1)}$ and $\bar{s}_1^{(n-1)}$, respectively. If $\bar{s}_N^{(n)} \approx \bar{s}_N^{(n-1)}$ and $\bar{s}_1^{(n)} \approx \bar{s}_1^{(n-1)}$ within an acceptable tolerance, then the procedure is complete. Otherwise, set n = n + 1 and continue iterating at step 2 by using the most recent values of $\bar{s}_i^{(n-1)}$.

Since the values of $\bar{s}_N^{(n)}$ and $\bar{s}_1^{(n)}$ each increase monotonically as $P_N{}^L$ and $P_1{}^R$, respectively, decrease, the iteration will converge.

In view of the conditions implied by the cited variant of the first principle (i.e., $t_1 = t_2 = \ldots = t_N$), it suffices to require that t_i = minimum. Then, by using Equations 19-22, we may express the above requirement as follows:

$$\tau_N = [T_L + p_N(1 - T_L - T_R)]/\bar{s}_N = \text{minimum} \quad (40)$$

where $\bar{s}_N = (\bar{h}_N)^{-1}$. Similarly,

$$\tau_1 = [T_R + p_1(1 - T_L - T_R)]/\bar{s}_1 \quad (41)$$

and

$$\tau_m = (1 - p_1 - p_2)(1 - T_L - T_R)/(N - 2)\bar{s}_m \quad (42)$$

Note that the first principle is only satisfied for those (unchannelized) lanes that service some portion of the through component of traffic on an approach. A violation of this principle implies that traffic volume for one or more turning movements exceeds approach capacity for that movement. In this case, the value of t for that lane or those lanes will exceed the values of t for the lanes servicing the other turn-movement components of traffic, which implies over-saturated conditions.

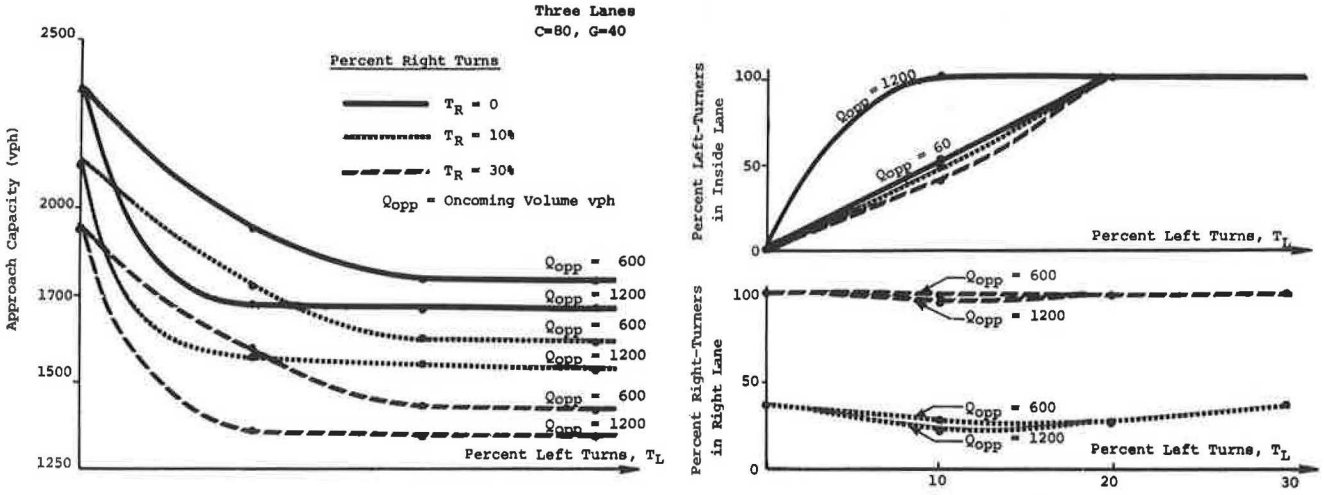
REPRESENTATIVE RESULTS

Consider a three-lane unchannelized approach to an intersection. The two-phase signal control operates on an 80-s cycle with a split of 50 percent. Left turners from the subject approach must contend with oncoming traffic and right turners with pedestrian interference. The model examines the effects on service volumes of varying turn percentages and of opposing volume on the oncoming approach (i.e., approach capacity).

Consider Figure 1 for the case of 10 percent right turners and Q_{opp} = 600 vehicles/h. At zero left-turn percentage, both the inside and middle lanes can provide saturation service volume for the through vehicles. The outside lane contains 38.6 percent right turners, and the remainder are through vehicles. The outside lane is less attractive to the through vehicles because the right turners discharge at a lower service volume (it was necessary to implement the entire analytical system of equations, including the service-rate models, to generate these results). Hence, only 15 percent (approximately) of the total through vehicles entering the approach select the outside lane with 85 percent split between the other two lanes.

At 10 percent left turners, the inside lane contains 47.1 percent left turners and 52.9 percent through vehicles when the opposing volume is 600 vehicles/h and 100 percent left turners when the opposing volume is 1200 vehicles/h. Since left turners encounter more impedance than right turners, the through vehicles are less attracted to the inside lane than to the outside lane even when both turn percentages are the same. In this case (opposing volume = 600 vehicles/h), 11 percent of all through vehicles select the inside lane while nearly 25 percent select the outside lane; the

Figure 1. Approach capacity and lane deployment as a function of several factors.



remaining 64 percent select the middle lane. Note that the percentage of through vehicles in the outside lane increased when the incidence of left turners rose to 10 percent.

At 17 percent left turners and an opposing volume of 600 vehicles/h, the inside lane contains only left turners and this movement experiences saturation conditions. That is, the service volume for left turners is at a maximum. Increasing the percentage of left turners entering the approach will create an unstable queue on the inside lane but will not change capacity.

It has been shown that lane deployment and approach capacity (i.e., service rates) are strongly influenced by many factors. The heightened level of detail embodied in this model--and in the associated service-rate models--provides results that cannot be extracted from models that depend on more simplifying assumptions.

A computer program exists that can produce the necessary curves and/or tabulations that will permit the development of a manual procedure to estimate approach, and intersection, capacity, based on this system of analytic models. Results produced by this system have been compared, in a limited study, with data reduced from 16-mm film, and the results are favorable. A more extensive validation effort would provide the necessary evidence for using this system as a reliable tool for estimating approach capacity.

ACKNOWLEDGMENT

The effort that produced this system of models was supported by an FHWA contract. Guido Radelat, who was the FHWA contract manager for this project, was supportive of this effort. Thanks are also due to William McShane of the Polytechnic Institute of New York and Charles Berger of Sperry Systems Management, whose development of the service-rate models made this work possible.

REFERENCES

1. Highway Capacity Manual. HRB, Special Rept. 87, 1965.
2. Development of an Improved Highway Capacity Manual: Phase 1 of NCHRP Project 3-28. JHK and Associates, San Francisco, Dec. 1978.
3. E. Lieberman and others. Macroscopic Simulation for Urban Traffic Management: The TRAFLO Model. Federal Highway Administration, U.S. Department of Transportation, Vols. 1-7, 1980.
4. Swedish Capacity Manual. National Swedish Road Administration, Stockholm, 1977.

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

Abridgment

Development of a TRANSYT-Based Traffic Simulation Model

M. YEDLIN AND E. B. LIEBERMAN

The Level II urban traffic simulation model contained within the TRAFLO traffic simulation package is described. Level II is a tool by which to evaluate transportation system management strategies and is an adaptation of the traffic flow model embedded in the TRANSYT signal optimization program. The Level II model describes traffic flow patterns in the form of statistical histograms. These histograms express flow rate as a function of time on each network link, stratified by turning movement; buses are treated in somewhat more detail. Platoon dispersion is treated explicitly, and service rates at an intersection are related to turning movements and to the signal control. The flow model used in TRANSYT is discussed along with the modifications and extensions that were incorporated in Level II. Data requirements and the measures of effectiveness generated by Level II are presented. Model validation results and program efficiency are also discussed.

This paper describes the Level II urban traffic simulation model contained within the TRAFLO traffic simulation package (1). Level II is a tool for evaluating transportation system management (TSM) strategies and is an extension of the traffic flow model embedded in the TRANSYT signal optimization program (2). The TRANSYT program models traffic flow on a network represented by "nodes" (intersections) that are connected by unidirectional "links" (one-way portion of a roadway).

The average pattern of traffic flow past a point on a network link is represented by a statistical histogram that relates flow rate to time (see Figure 1). All calculations are carried out by the manipulation of such histograms; no representation of individual vehicles is made.

A major conceptual difference between Level II and TRANSYT is the treatment of the independent variable, time. Whereas TRANSYT is a quasi-steady-state model, Level II is a dynamic simulation model that accommodates time-varying traffic demands and turn percentages and fixed-time signals of arbitrary cycle lengths. Traffic conditions generally vary from one "time interval" of specified duration to the next.

DESIGN OF LEVEL II SIMULATION MODEL

The key to the Level II model is the histogram representation of traffic flow at points along a link (see Figure 2).

The ENTRY flow histogram describes the entering traffic flow as seen by an observer at the entry point (upstream end) of a link. An INPUT flow histogram represents the time-varying pattern of arriving traffic as seen by an observer just upstream of the stop line. An OUTPUT histogram indicates the pattern of discharging traffic as seen by an observer just downstream of the stop line.

In addition to these flow histograms, the Level II model constructs two other types of histograms: a SERVICE histogram that represents the time history of service rates provided by the control device at the downstream intersection and a QUEUE histogram that describes the time history of the number of vehicles queued at the stop line.

Each histogram is constructed by the model for each turn movement component of traffic over one time interval; a simulation run extends over a sequence of many time intervals. During each time interval, the model computes the various histograms successively for each link in the network. Continuity of flow from one link to the next is modeled by aggregating the turn-movement-specific OUTPUT histograms of all feeder links to build the ENTRY histogram of the receiving link.

In this scheme, a link cannot have a complete ENTRY histogram unless all feeders are processed first. In a grid network, it is not possible to satisfy this requirement. So a link-sequencing algorithm was developed that guarantees that at least one feeder of each link will be processed first and ensures that most links will have all of their feeders processed first. This eliminates the need to specify "dummy" links, as was done in the TRANSYT program. Since the Level II simulation is designed to describe a dynamic traffic environment, the time lag (i.e., travel time) experienced by traffic as it moves from the upstream node of a link to the stop line must be explicitly represented. The concept of an influence zone represents this time lag explicitly.

The influence zone ABCD shown in Figure 3 contains all traffic arriving at the stop line during the current time interval. The length of this zone is the length of the street, and its width reflects the duration of the time interval. The parallel sides, AD and BC, represent free-flow vehicle trajectories whose slopes are equal to the free-flow speed $\bar{u}$. The pattern of traffic arriving at the stop line during the current time interval reflects only the pattern of traffic that enters the street within the influence zone (between time points A and B) and that portion from the prior influence zone that was not discharged.

It is necessary to process the known ENTRY histograms in order to construct the required histogram that represents traffic entering the influence zone along line AB. Figure 3 shows that the ENTRY histogram can be subdivided into several "flow regimes". These are described below:

1. Regime F_1 is defined by the aggregation of all the known turn-movement-specific OUTPUT histograms representing the traffic discharged from all feeder links during the current time interval.
2. Regime F_2 of the ENTRY histogram was "left over" from the ENTRY histogram aggregated during the prior time interval.
3. Regime F_3 describes the traffic flow pattern that lies within the influence zone at the upstream end of the subject link. Note that this histogram is formed by concatenating the F_2 regime with the early portion of the F_1 regime. This traffic will disperse as it travels along the street and will arrive at the stop line to form the INPUT histogram within the current time interval.

The influence-zone concept is applicable to any traffic environment and control policy and is therefore suitable for a general-purpose traffic simulation model.

The processing of traffic on a link involves several activities:

Figure 1. Representation of platoon structure in TRANSYT.

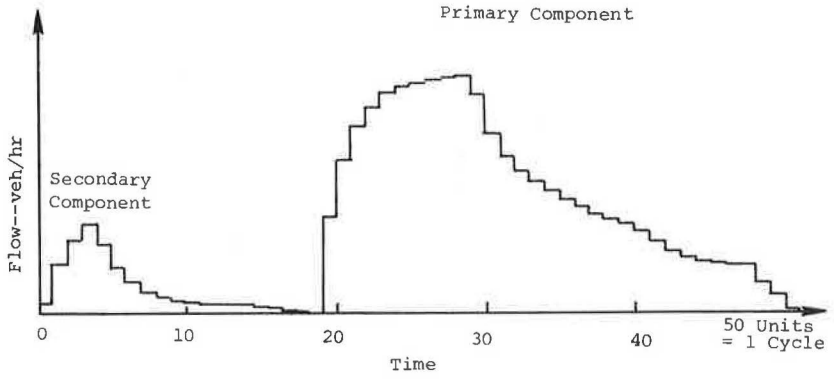
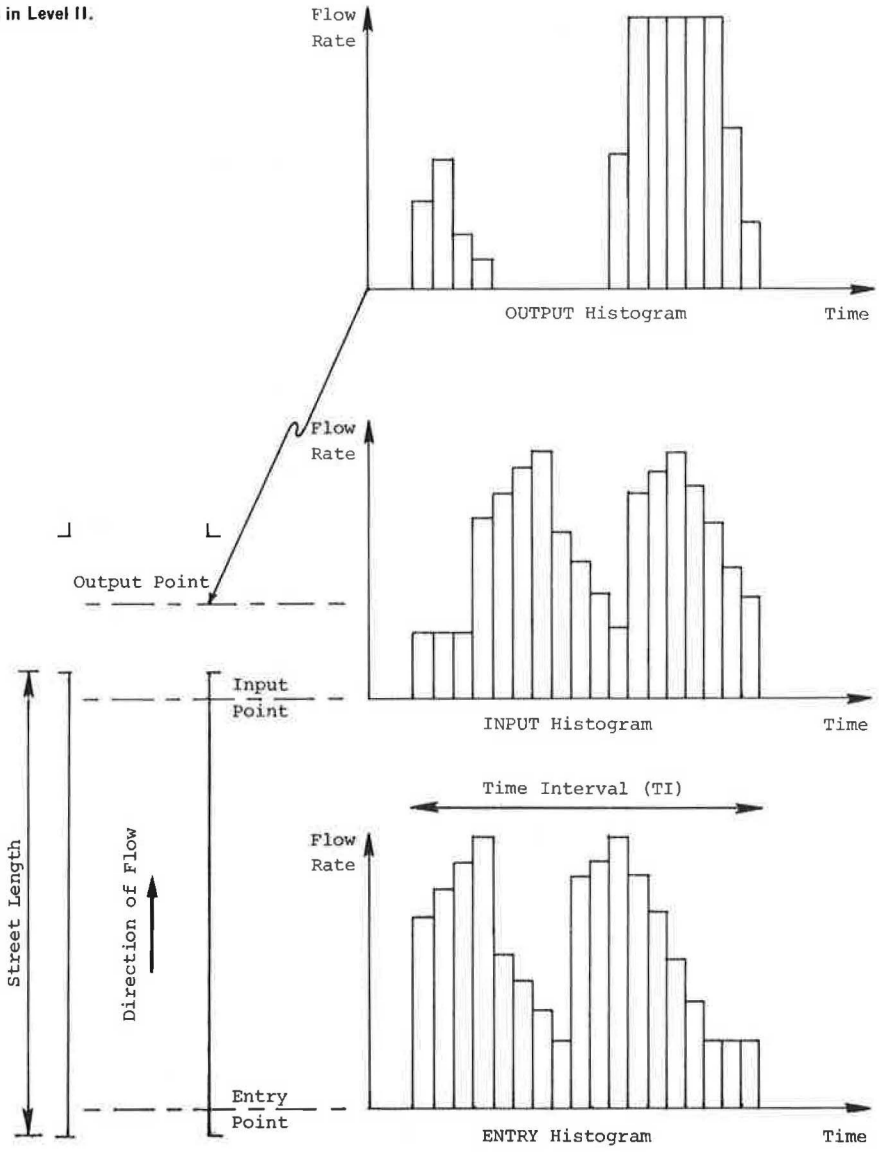


Figure 2. Flow histograms in Level II.



1. Form the ENTRY histogram for traffic entering the influence zone (regime F_3), as described above.
2. Disaggregate this histogram (regime F_3) into turn-movement-specific components.
3. Store that portion of the original ENTRY histogram (regime F_1) that lies beyond the influence zone for subsequent use in the next time interval (this portion becomes regime F_2 for the next time interval).
4. Calculate the respective INPUT histograms (Figure 2) by applying the platoon-dispersion relation developed by Robertson and used in TRANSYT (2).
5. Generate a SERVICE histogram for each

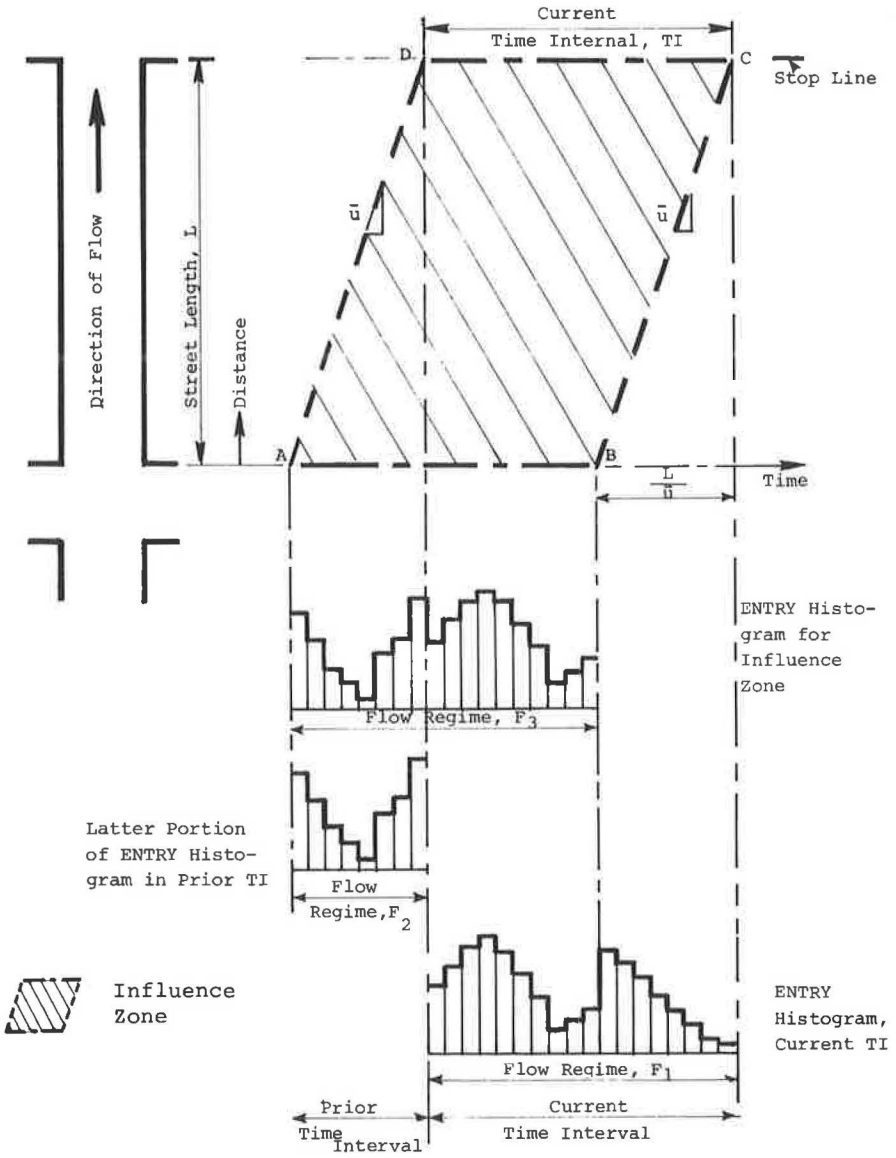


Figure 3. Space-time representation of influence-zone concept.

movement. These histograms represent the service rates permitted by the control. These service rates are computed by a new capacity model (described in papers by Lieberman and by McShane and Lieberman elsewhere in this Record).

6. Transform each turn-movement-specific INPUT histogram into its counterpart OUTPUT histogram. This transformation reflects the interaction of each INPUT histogram with its respective SERVICE histogram. In the process, a time history of vehicle queues (QUEUE histogram) is produced to facilitate the computation of vehicle stops and delays.

Vehicle delays, vehicle stops, and the random component of delay are computed by using the TRANSYT concepts described in detail by Robertson (2). Continuity of flow is carefully preserved between adjoining time intervals.

Input

The Level II model requires a detailed network description to accurately represent traffic flow. Data requirements include geometrics, traffic operating characteristics, traffic control specifications, traffic demand volumes, turning movements on each link, and mass transit data.

The user may specify certain time-varying data: lane channelization, entering traffic volumes, intralink source-sink volumes (from-to parking facilities), link-specific turn percentages, and bus headways and station dwell times. To improve the ability of the model to accept these time-varying data, the concept of a user-specified time period was developed. A time period consists of a sequence of time intervals. During a time period, all input data remain fixed. At the end of each period, the user may specify any changes in the above parameters.

Output

The Level II model offers a wide range of statistics. The output generated by the model is the "measures of effectiveness" used by traffic engineers to evaluate management strategies.

Cumulative output of both (a) networkwide link-specific and (b) bus-route-specific measures is provided at the end of each time period. More frequent and more detailed output can be requested by the user.

REPRESENTATIVE RESULTS

The TRAFLO Level II model was validated on a network in downtown Washington, D.C., that consisted of 96 links and 51 nodes and represented a wide range of geometrics. Validation runs were made for morning peak and off-peak periods, and a wide range of turn movements and traffic volumes was reflected. Each run was executed for 32 min as a sequence of eight 4-min time periods. Sperry Systems Management validated the model, reporting results on a link-by-link basis for each of the eight time periods (1). The field measurement of networkwide average speed over the 32-min morning peak period was 9.71 miles/h compared with a model estimate of 10.29 miles/h. For the off-peak period, the model estimated an 8.79-mile/h average speed, which compared very favorably with an observed speed of 8.73 miles/h.

PROGRAM EFFICIENCY

The Level II model was executed on a CDC 7600 computer at the Brookhaven National Laboratory in Upton, New York. Computer time for the model depends strongly on the size of the network. Runs of the validation network of 96 links indicate a ratio of simulated time to computer time of approximately 160:1 and a cost of less than $8 for a 32-min simulation and "fill" time of 6 min. The computer memory requirement is reasonable. For IBM computers, less than 250 K bytes are required; on CDC machines, less than 40 K words are needed.

ACKNOWLEDGMENT

The development of a model such as that described in this paper represents the contributions of many people. In particular, we want to acknowledge the contributions of Guido Radelat and George Tiller of the Traffic Systems Division of the Federal Highway Administration, Barbara Andrews and Manfredo Davila of KLD Associates, Inc., William McShane of the Polytechnic Institute of New York, and Fred Wagner of Wagner/McGee Associates. This work was performed under a contract with the U.S. Department of Transportation.

REFERENCES

1. E.B. Lieberman and others. Macroscopic Simulation for Urban Traffic Management: The TRAFLO Model. Federal Highway Administration, U.S. Department of Transportation, Vols. 1-7, 1980.
2. D.I. Robertson. 'TRANSYT': A Traffic Network Study Tool. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Rept. LR 253, 1969.

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

TRAFLO: A New Tool to Evaluate Transportation System Management Strategies

E. B. LIEBERMAN AND B. J. ANDREWS

The TRAFLO model, which combines the attributes of traffic simulation with traffic assignment, is described. TRAFLO was developed as a tool for use in transportation planning and traffic engineering to test transportation management strategies. It is a software system, programmed in FORTRAN, that consists of five component models that interface with one another to form an integrated system. Four of the models simulate traffic operations, and the fifth is an equilibrium traffic assignment model. The operating characteristics of the component simulation models are described. These models are capable of simulating traffic on one or more of the following networks: freeways, corridors that include the freeway/ramp/service-road complex, urban and suburban arterials, and grid networks representing the central business districts of urban centers. Also described is the traffic assignment component, which can be used in conjunction with the simulation components to determine the response of a traffic system to a transportation management strategy.

In recent years, events have shifted attention to the need for providing safe, efficient, and economical movement of people and goods on existing highway facilities. Furthermore, there is a growing awareness that factors such as air and noise pollution and the conservation of energy must be weighted heavily in any decision process involving the nation's transportation system.

These considerations have led to the emergence of transportation management as the basis for improving the mobility of people and goods. The application of the transportation management process requires the ability to quantitatively assess alternative transportation management strategies to identify those that best satisfy the stated objectives.

The scope of the process involves both the transportation planning and traffic engineering disciplines. The involvement of these two disciplines reflects the intrinsic dependence of behavioral responses (trip generation, distribution, and assignment and modal choice) on the performance of the transportation system as expressed in terms of travel time, cost, and accessibility.

It is clear that the need to develop effective transportation management strategies implies a requirement to develop analytic tools for that purpose. Furthermore, these tools must be sufficiently broad in scope to meet the objectives of the transportation management process for both disciplines identified above.

One tool that is particularly effective for evaluating transportation management strategies applied to a dynamic environment is traffic simulation. Simulation models provide the means for evaluating a wide spectrum of traffic management schemes within the framework of a controlled

experiment. This simulation approach is far more appealing and practical than a strictly empirical approach, for the following reasons:

1. It is much less costly.
2. Results are obtained in a fraction of the time required for field experimentation.
3. The data generated by simulation include measures of effectiveness that cannot, in a practical sense, be obtained empirically.
4. Disruption of traffic operations, which often accompanies field experimentation, is completely avoided.
5. Many transportation management strategies involve significant physical changes that are not acceptable for experimental purposes.

Most projects undertaken by transportation planners address a time period that lies in the future and thus require estimates of transportation demand. Based on these estimates, trip tables that delineate traffic demands between origin and destination zones within a region are developed. It is then necessary to identify a transportation management strategy that will satisfy the mobility, environmental, and economic objectives perceived for that future time period. To do this, a traffic assignment model must be applied to estimate the distribution of traffic demand over a regional network, consistent with the projected trip tables.

An assumption implied in every traffic assignment model is that the traffic environment is steady state. That is, it is assumed that the specified trip tables reflect constant traffic volumes on each network link and that any dynamic effects do not influence the assignment process. Furthermore, all estimates of flow impedances included in the assignment models are also based on the assumption of steady-state conditions, and it is assumed that dynamic interactions between traffic on adjoining network links may be disregarded. The efficacy of the results provided by a traffic assignment model depends on the degree of validity of these underlying assumptions and on the accuracy of the estimates of traffic impedance. Finally, the transportation planner has no means for verifying the estimates of travel time on each network link that are provided by the traffic assignment model.

Simulation models, on the other hand, are specifically designed to describe the dynamic effects of traffic flow. Factors that impede traffic are explicitly represented at a high degree of detail. Consequently, simulation tools can provide a detailed description of the dynamic performance of traffic over a network.

The availability of an analytic tool that combines the attributes of traffic simulation with those of traffic assignment will greatly expand the opportunity for the development of new and innovative transportation management concepts and designs. Transportation planners and engineers will no longer be restricted by the lack of a mechanism for fully testing these designs prior to field demonstration.

Such a tool is of value to both transportation planners and traffic engineers. It gives the planner the opportunity to examine the net result of a design based on transportation system management (TSM) principles. These results are expressed as measures of effectiveness (MOEs), which describe traffic operations on each network link. With this detailed information, the planner can reexamine the estimates of travel time and accessibility that were involved earlier in the transportation management process (e.g., when preparing the trip generation data).

This tool provides the traffic engineer with the information needed to explore candidate operational solutions to resolve bottleneck conditions, expedite mass transit operations, or satisfy other TSM objectives. He or she can apply the simulation model repeatedly as an integral part of an iterative design procedure.

Of course, any design improvement implemented by the traffic engineer can "feed back" and influence the results obtained from the planning process. It is this interdependence, noted earlier, that requires a strong interaction between the two disciplines. The TRAFLO model, which is designed to provide the capability described above, can act as a primary mechanism for encouraging this strong interaction and providing the information needed for designing effective transportation management strategies.

The TRAFLO model was conceived as the tool to fill this role by the Traffic Systems Division of the Federal Highway Administration (FHWA), in consultation with that agency's Planning Division and with personnel of the Urban Mass Transportation Administration (UMTA). This paper describes several innovative concepts and design features incorporated in the model, which has been implemented as a computer program.

GENERAL MODEL DESCRIPTION

TRAFLO is a valuable tool in the transportation management process. Its design includes features that permit the analyst to conduct a wide variety of studies on large roadway networks of general configuration. These networks may contain components such as freeways, corridors that include the freeway/ramp/service-road complex, urban and suburban arterials, and grid networks representing the central business district (CBD) of urban centers. The analyst has complete flexibility to configure the network according to his or her needs. The network may consist of any one or more of the components mentioned above.

Since the TRAFLO model is actually a system composed of well-defined component models, the analyst can also select those component models that are most responsive to his or her needs. This flexibility enables the user to apply TRAFLO in the most cost-effective manner.

The simulation models that constitute the TRAFLO program describe traffic flow macroscopically. Past experience with other models has demonstrated that it would be possible to retain sufficient accuracy for evaluating transportation management strategies if the less detailed macroscopic representation were used. It has also been concluded that TRAFLO should provide a hierarchy of macroscopic detail for simulating traffic on urban streets. That is, the user can select among three levels of simulation detail: The more detail, the greater is the accuracy obtained and the higher is the associated computing cost. This hierarchy would permit the user to decide on the optimal trade-off between the accuracy required and the computer resources at his or her disposal. Regardless of this selection process, TRAFLO is far more economical in every respect than any of the existing microscopic models.

The selection of an existing traffic assignment model for inclusion in TRAFLO was based on the idea that the most satisfactory traffic assignment models, from a mathematical viewpoint, are those that use Wardrop's principles of equilibrium (1). These models apply optimization theory to calculate the assignment of traffic over a network, taking into consideration all link impedances. As expected, these models are computationally more

complex, although computational costs are reasonable. The accurate estimation of link impedances is necessary if the potential of these models is to be fully realized.

Since the TRAFLO software is designed to be machine independent, it was necessary to select a model already coded in FORTRAN. Fortunately, such a model does exist (2) and, in fact, has been carefully validated (3). This model is similar to the equilibrium traffic assignment model currently embedded in the UROADS module of the Urban Transportation Planning System (UTPS) package developed by UMTA.

MAJOR PROGRAM FEATURES

The TRAFLO program is actually a software system that consists of five functionally independent models. The logical structure of TRAFLO is designed to permit these independent models to interface with one another so as to form a coherent, integrated system. Four of the models simulate traffic operations over a specified network of roadways; the fifth model is an equilibrium traffic assignment model.

Representing the Traffic Environment

In order to use any of the simulation models in the TRAFLO program, the user must specify the following features of the physical traffic environment:

1. The topology of the roadway system;
2. The geometrics of each roadway component;
3. The channelization of traffic on each roadway component;
4. Motorist behavior, which, in aggregate, determines the operational performance of vehicles in the system;
5. Circulation pattern of traffic on the roadway system;
6. Traffic control devices and their operational characteristics;
7. Volumes of traffic entering and leaving the roadway system;
8. Traffic composition; and
9. The configuration of the mass transit system, i.e., bus routes, bus stations, and frequency of service.

In using the traffic assignment model, the user must also specify the trip table that defines the volume of traffic traveling from each origin to each associated destination.

To provide an efficient framework for defining these specifications, the physical environment is represented as a network. The unidirectional links of the network generally represent roadway components--either urban streets or freeway segments. The nodes of the network generally represent urban intersections or points along the freeway where a geometric property changes (e.g., a lane drop or a change in grade).

Figure 1 shows an example of a network representation. The freeway is defined by the sequence of links (1, 2), (2, 3),...,(5, 6). Links (8000, 1) and (6, 8001) are entry and exit links, respectively. An arterial extends from node 7 to node 15 and is partially subsumed within a grid network.

Each of the four simulation models in TRAFLO describes traffic operations in a subnetwork. That is, the user may partition the analysis network into subnetworks if he or she wishes to apply more than one simulation model concurrently [if the network consists of a freeway and urban streets, it must be partitioned (at least) into freeway and urban subnetworks, each of which may consist of several noncontiguous sections]. The user must also specify "interface nodes" at the juncture of the various subnetworks.

Urban Level I Model

The Urban Level I model is the most detailed of the macroscopic simulation models. Since it treats each vehicle in the traffic stream as a separate, identifiable entity, the representation of traffic can be considered microscopic. The treatment of the traffic stream, however, which is intermittent, or event based, can be considered macroscopic.

By treating each vehicle in the traffic stream individually, it is possible to explicitly distinguish between different types of vehicles (automobiles, trucks, and buses) and to treat each type according to its respective operating characteristics. Hence, the interaction of these vehicle types and the impact of lane channelization of bus-only or truck-only streets, and other detailed traffic management strategies, can be studied in adequate detail. Furthermore, much of the stochastic nature of the traffic-flow process can be explicitly represented.

Each vehicle is processed (i.e., moved) as infrequently as possible. This frequency depends on the conditions encountered by the vehicle immediately downstream. The less impedance a vehicle encounters, in the form of queues and no-go signal indications, the fewer processing steps are required to move a vehicle a given distance.

Associated with each vehicle is an "activation time" (AT), which is expressed in terms of the simulation clock time. When the simulation clock time equals the vehicle's AT, the vehicle will be processed (i.e., moved). The vehicle is generally moved to a point downstream--either on its current link or onto a receiving link--and its new location, speed, and AT are calculated. This vehicle then remains "dormant" until the simulation clock time advances to this new AT, whereupon the vehicle is again processed. [In contrast, a microscopic model such as NETSIM (4) moves all vehicles every time step and generates detailed trajectories.]

When a vehicle is processed, the determination of its new location, speed, and AT (i.e., its status) depends on conditions downstream of its starting point. A small number of scenarios (or "cases") have been identified that, in aggregate, span the entire spectrum of possible conditions. For each such case, explicit analytic expressions have been derived to compute the vehicle's new status. Spillback conditions that arise from inadequate capacity on one or more network links are also properly accounted for.

Urban Level II Model

The Urban Level II model is an extension and refinement of the flow model used in the TRANSYT signal optimization program (5). This flow model in the Level II simulation represents the traffic stream in the form of movement-specific statistical histograms. Figure 2 shows this histogram representation, which preserves the platoon structure of the traffic stream.

The Level II logic constructs a total of five such histograms for each (turn) movement on each network link:

1. The ENTRY histogram describes the platoon flow at the upstream end of the subject link. This histogram is simply an aggregation of the appropriate OUT turn-movement-specific histograms of all feeder links.

Figure 1. Representative analysis network.

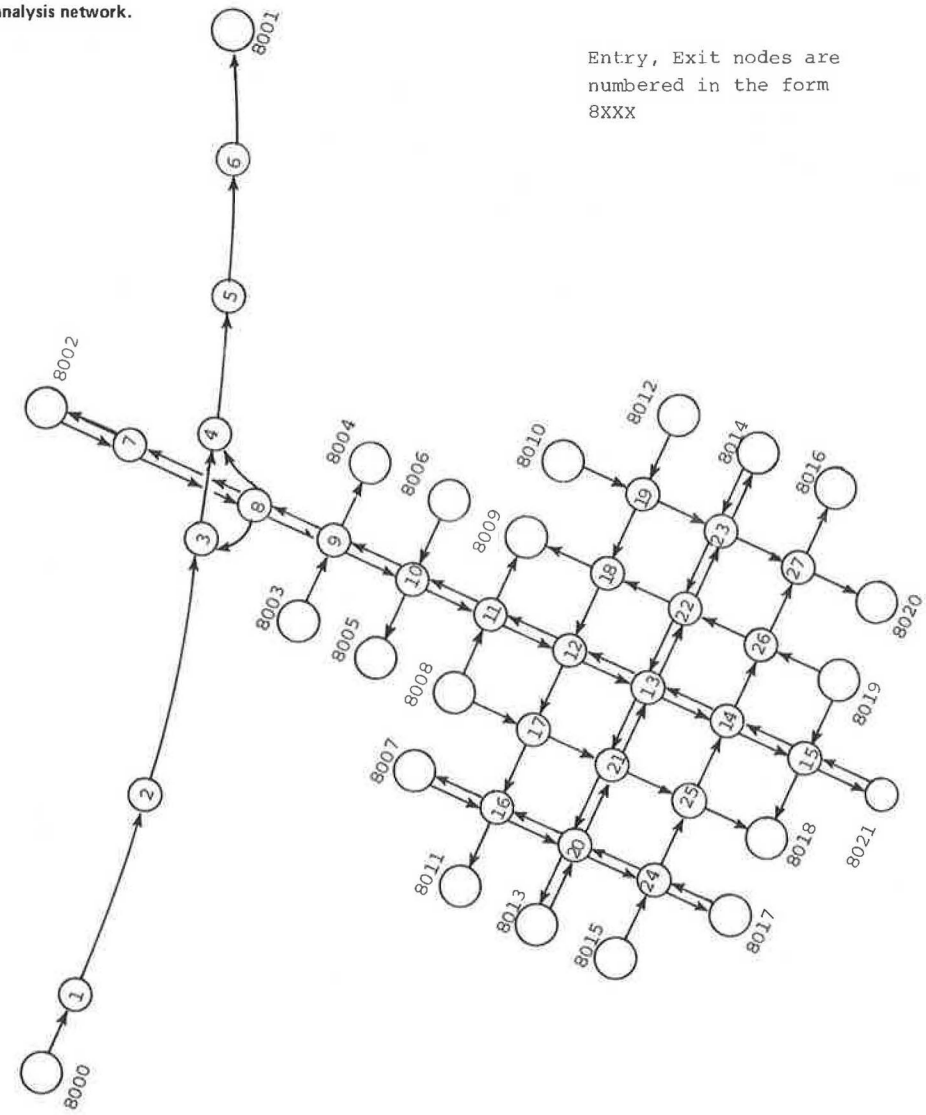
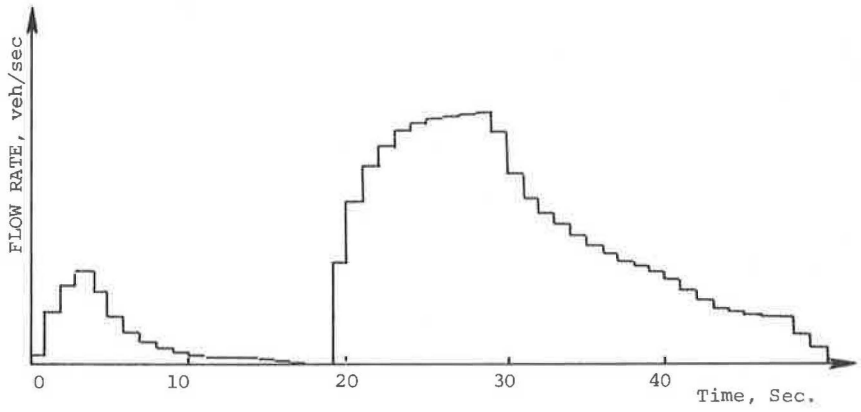


Figure 2. Statistical representation of traffic-stream platoon structure.



2. The IN histograms describe the platoon flow pattern arriving at the stop line. These dispersed histograms are turn-movement-specific and are obtained by disaggregating the ENTRY histogram to reflect the specified turn percentages for the subject link and then dispersing each histogram separately. This histogram representation, of course, describes the end result of the physical dispersion of platoons as they travel along the subject link to the stop line.

3. The SERVICE histograms describe the service volumes for each turn movement. These service volumes reflect the type of control device on this approach; if it is a signal, the histogram reflects the specified movement-specific signal phasing. A separate model was developed to estimate service

volumes for each turn movement, given that the control is "go".

4. The QUEUE histograms describe the time-varying ebb and growth of the queue formation at the stop line. These histograms are derived from the interaction of the respective IN histograms with the SERVICE histograms.

5. The OUT histograms describe the pattern of traffic discharging from the subject link. Each of the IN histograms is transformed into an OUT histogram by the control applied to the subject link. Each of these OUT histograms is added into the (aggregate) ENTRY histogram of its receiving link.

Note that this approach provides the Level II model with the ability to identify the characteristics of each turn-movement-specific component of the traffic stream. Each component is serviced at a different saturation flow rate, as it is in the real world. Furthermore, the Level II logic is able to recognize when one component of the traffic flow is encountering saturation conditions even if the others are not.

Algorithms provide estimates of delay and stops, reflecting the interaction of the IN histograms with the SERVICE histograms. Level II logic also provides for representing bus traffic as separate entities (although at a lower level of detail than Level I) and for properly treating spillback conditions.

Urban Level III Model

Level III logic is designed for major arterials that act as collectors, distributors, circulators, or connectors. As a collector, an arterial would serve to feed traffic from, say, an outlying region (or suburb) to a region of higher traffic density. As a distributor, the arterial would serve a reverse role, servicing a high demand level at one end and distributing this traffic to cross streets throughout its length. An arterial that serves primarily to provide access to adjoining traffic generators can be called a circulator. Finally, a connector arterial links two high-density areas, each of which would be modeled in greater detail (at Level I or Level II).

A user may determine that, although an arterial plays an important functional role, as described above, a detailed anlysis of its traffic operations lies outside his or her realm of interest. For example, a planner may wish to determine travel time along the arterial from various points to a particular location (e.g., a shopping center or a rail station). A traffic engineer may wish to determine whether congested conditions will occur as part of a quick precursor study to find out whether a more detailed analysis is necessary.

To satisfy such needs it is not necessary to explicitly simulate traffic elements either as individual vehicles or as platoons. It is sufficient to calculate the MOEs associated with a traffic environment that is described in relatively gross, aggregate terms.

Many investigators have developed explicit analytic expressions that relate delay to traffic volume, control settings, and saturation flow rate by using various techniques and asserting a variety of assumptions. One formulation that is widely accepted is that developed by Webster (6). The Level III simulation model uses an extension of Webster's formula to calculate vehicle delay.

Although the Level III model is far less detailed than Level II, it is still necessary to properly represent the "time lags" in the system; that is, all the link-segment-specific values of volume are time dependent.

In addition, it is necessary to recognize that the delay experienced by vehicles at intersections is a function of the turn-movement percentages, the presence or absence of oncoming traffic opposing left turners, and the channelization of lanes on the link.

Freeway Model

FREFLO, the freeway traffic simulation model included in TRAFLO, is an extension and refinement of the MACK model developed at the University of Southern California (7). This macroscopic simulation model represents traffic in terms of aggregate measures associated with sections of freeway generally less than 1 mile (1.6 km) in length. The aggregate measures used are flow rate, density, and space mean speed within the section. The formulation is based on a fluid-flow analogy to traffic operations.

The earliest modeling work (7) used a conservation equation and an equilibrium speed-density relation. In FREFLO, the equilibrium speed-density relation is incorporated into a dynamic speed equation. Another extension allows vehicles to be distinguished by type in three categories: (a) automobiles and trucks, (b) buses, and (c) carpools.

Most of the capabilities of FREFLO are shown in Figure 3. For each freeway section, there is a variable for entry flow rate, exit flow rate, density, and space mean speed. These variables are distinguished by vehicle type according to the three categories given above.

Vehicles enter the freeway subnetwork either at the upstream end of a freeway segment or by way of on ramps. In the on-ramp case, it should be noted that FREFLO represents only the movement on the freeway main line so that vehicles are introduced at the ramp gore and immediately merged. Vehicles exit the freeway subnetwork at the downstream end of a freeway segment or by way of off ramps. In the off-ramp case, FREFLO represents movement only up to the off-ramp gore so that movement down the ramp is represented within the adjoining subnetwork.

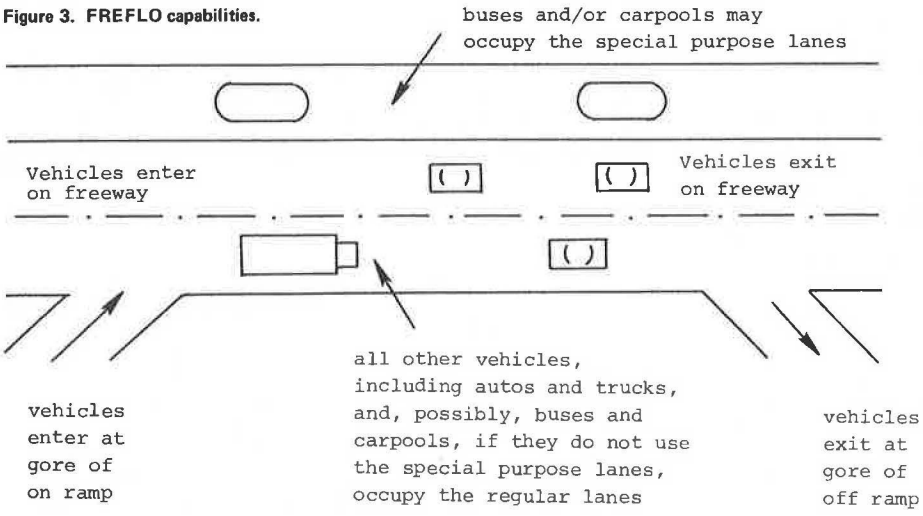
Traffic is associated with two types of lanes: (a) special-purpose lanes that can be designated to allow use by buses and/or carpools only and (b) regular lanes that can accommodate all other traffic and all vehicle types, including buses and carpools. Traffic is not associated with a particular lane but is considered to be uniformly distributed over the special-purpose and regular lanes, separately. The number of lanes of each type is arbitrary.

The network that can be represented is quite general. Disjoint segments or more general disjoint pieces are accommodated. Freeway-to-freeway connectors that involve merge and diverge points, as well as several connected freeways, can be accommodated. FREFLO provides for the representation of an incident on the freeway by allowing for the specification of a reduced number of available lanes and a constraint on the flow rate past the incident site.

Traffic Assignment Model

The equilibrium traffic assignment model embedded in TRAFLO interfaces with the simulation models. That is, the traffic assignment model in TRAFLO accepts a specified trip table (matrix of origin-destination demand volumes) and assigns these trips over the specified network. The assigned traffic volumes are then transformed into link-specific turn percentages, as required by the simulation models.

Figure 3. FREFLO capabilities.



The software then executes the simulation model(s) requested by the user to generate statistical measures that quantify the performance of traffic operations over the analysis network. This entire process is automatic, requiring no manual intervention beyond the initial preparation of the input data.

It is the inclusion of a traffic assignment model that makes TRAFLO a tool for the transportation planner as well as the traffic engineer. TRAFLO provides the planner with a description of the dynamic response of a transportation system to an applied transportation management strategy for the specified current, or projected, travel demand pattern (trip table). The information provided by TRAFLO offers far more detail and accuracy than are currently available to the planner.

The TRAFFIC model (2) uses the U.S. Bureau of Public Roads formulation to relate link travel time to volume. It then calculates the link-specific volumes that minimize an objective function representing Wardrop's first principle (i.e., user optimization). These data are subsequently transformed into link-specific turn percentages, as required by the simulation models, and the simulation is then implemented.

OPERATING CHARACTERISTICS

The TRAFLO software was developed by rigorously applying structured design and programming methodologies to reduce subsequent maintenance costs and to provide an efficient FORTRAN code.

The computing time for the Level I model depends almost linearly on the number of vehicles that occupy the analysis network. The relevant statistic for this model is expressed in terms of the ratio of the number of vehicle seconds of travel time to computer time. Based on results obtained on the CDC 7600 computer, Level I provides a ratio of 20 000:1. For the validation network of 95 links, and an average content of 375 vehicles, the total execution time for a simulation of 32 min plus 5 min of fill time was 26 s, which corresponds to a cost of less than $15.

The computing times consumed by the Level II and Level III models, as well as the FREFLO model, depend strongly on the size of the network rather than on the traffic volume. For a network of approximately 95 links, the ratio of simulated time to computing time for the Level II model is 160:1 on the CDC 7600 computer. For the validation case noted above, the Level II program consumed 10 s of computer time at a cost of $5. The computing cost for the Level III and FREFLO models is insignificant (on the order of less than 5 s) regardless of network size or volume.

Table 1. TRAFLO card types.

Set	Group	Card Type
Network independent	Run specification	00-05
Subnetwork specific	Urban link characteristics	11
	Freeway link characteristics	15
	Turning movements	21
	Freeway turning movements	26
	Freeway incident specifications	27
	Link specifications: Level III	30-31
	Freeway parameters	34
	Fixed-time signal control	35-36
	Actuated signal control	39-41
	Traffic volume	50-52
	Subnetwork delimiter	170
Global network	Traffic assignment	175-177
	Bus transit	185-189
	Time period delimiter	210

INPUT REQUIREMENTS

The input stream for TRAFLO is partitioned into sets of cards. Each set consists of one or more card groups, and each group contains one or more card types (see Table 1). Only those card types that are required for a particular application need be specified.

Although a substantial data base is required to adequately define the traffic environment under study, care has been taken to minimize user effort. For example, default values are available wherever possible, and all input data items are integers. Exhaustive diagnostic tests protect the user against improper inputs to the extent possible. The user also has the option to review his or her specified inputs prior to the execution phase of TRAFLO to further reduce the prospects of incorrect results.

VALIDATION

All three urban simulation models (Level I, Level II, and Level III) have been carefully validated by comparing model results with field data on a statistical basis. Details are provided elsewhere (8).

CURRENT STATUS OF TRAFLO

The TRAFLO program is now complete and is currently undergoing in-house testing by FHWA personnel. More detailed descriptions of TRAFLO appear elsewhere (8).

ACKNOWLEDGMENT

The development of a model such as that described in this paper is the result of the contributions of many people. In particular, we want to acknowledge Guido Radelat and George Tiller of the Traffic Systems Division of FHWA; Mark Yedlin and Manfredo Davila of KLD Associates, Inc.; William McShane of the Polytechnic Institute of New York; and Fred Wagner of Wagner/McGee Associates.

This work was performed under a U.S. Department of Transportation contract. A portion of the first part of the paper was extracted from the Request for Proposal statement of work.

REFERENCES

1. J.G. Wardrop. Some Theoretical Aspects of Road Traffic Research. Proc., Institute of Civil Engineers, Part 2, Vol. 1, 1952, pp. 325-378.
2. S. Nguyen. An Algorithm for the Traffic Assignment Problem. Transportation Science, Vol. 8, No. 3, Aug. 1974.
3. M. Florian and S. Nguyen. Recent Experience with Equilibrium Methods for the Study of a Congested Urban Area. Proc., International Symposium on Traffic Equilibrium Methods, Univ. of Montreal, Nov. 1974.
4. Network Flow Simulation for Urban Traffic Control System: Phase II (NETSIM). Peat, Marwick, Mitchell, and Co., Washington, DC; and KLD Associates, Inc., Huntington Station, NY, 1973. NTIS: PB 230 760-4.
5. D.I. Robertson. TRANSYT: A Traffic Network Study Tool. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Rept. LR 253, 1969.
6. F.V. Webster and B.M. Cobbe. Traffic Signals. Her Majesty's Stationery Office, London, Road Research Tech. Paper 56, 1966.
7. H.J. Payne, W.A. Thompson, and L. Isaksen. Design of a Traffic-Responsive Control System for a Los Angeles Freeway. Trans., Institute of Electrical and Electronics Engineers, Systems Science Cybernetics, Vol. SSC-3, May 1973, pp. 213-224.
8. E.B. Lieberman and others. Macroscopic Simulation for Urban Traffic Management: The TRAFLO Model. Federal Highway Administration, U.S. Department of Transportation, Vols. 1-7, 1980.

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

Abridgment

Hybrid Macroscopic-Microscopic Traffic Simulation Model

M. C. DAVILA AND E. B. LIEBERMAN

The Level I model, a component of the TRAFLO macroscopic traffic simulation program designed to evaluate transportation system management strategies, is described. The Level I model is designed to explicitly treat traffic control devices, include all channelization options, and describe traffic operations at grade intersections in considerable detail. Other features include actuated signal logic, right turn on red, pedestrian interference, and source-sink flow. Automobiles, buses, carpools, and trucks are explicitly treated as individual entities. The simulation processing uses "event-based" logic, which moves these vehicles intermittently, as required, rather than at every time step (interval scanning). Thus, this model is hybrid in the sense that the entities are microscopic but the processing is macroscopic in treatment. An overview of the Level I model logic is presented, the input requirements and measures of effectiveness provided by the model are indicated, and program efficiency and validation results are discussed.

This paper briefly describes the Level I model, a component of the TRAFLO macroscopic traffic simulation program (1), which has been designed to evaluate transportation system management (TSM) strategies. Level I is the most detailed simulation model within TRAFLO. It provides a microscopic description of the traffic stream and a macroscopic description of each vehicle movement. This approach is designed to provide a reasonably high resolution of detail as well as economy of operation.

Ideas embedded in several existing traffic simulation models have been selected, synthesized, refined, and expanded to form the Level I logic. These include the System Development Corporation macroscopic model (2), the TRANS model (3), the NETSIM (formerly UTCS-1) model (4), and the SCOT-Q model (which is not documented). The basic concept of processing each vehicle only when it is time to do so is called (in GPSS terminology) "event-based transactions". The intrinsic benefit of this concept is that it greatly reduces computing time, particularly when each event is widely spaced in time.

A careful analysis of these existing models revealed that it would be feasible and desirable to use an event-based approach in processing all vehicles; that is, even when a vehicle is in queue state, it could be "jumped" to the stop line and yet the mechanism of the queue discharge expansion wave could be preserved. By treating each vehicle in the traffic stream individually, the model is able to

explicitly distinguish between different vehicle types (automobiles, carpools, trucks, and buses) and to treat each type according to its respective operating characteristics. Furthermore, much of the stochastic nature of the traffic flow process can be explicitly represented.

MODEL DESIGN

For the purpose of simulation, in the Level I model an urban street network is decomposed into a set of unidirectional links (streets) and nodes (intersections). Each link may contain as many as six moving lanes. The control at a node may take the form of a traffic signal, two-way stop signs, and yield sign control. To accurately model delay at an intersection, each vehicle at the head of a queue, prior to the onset of the green signal phase, is stochastically assigned a start-up delay that must be exhausted after the green phase is activated and before the vehicle is discharged. When a vehicle is discharged, the remaining members of the queue move up in response to a "green wave" propagating upstream at a speed of 1 vehicle/s (≈20 ft/s). Thus, the fifth vehicle in a queue when the green phase is activated remains motionless for 4 s after the first vehicle discharges. As each following vehicle comes to the head of the queue, it is stochastically assigned a discharge headway that is related to (a) the specified mean headway value, (b) the statistical distribution about the mean, (c) the vehicle's original position in queue, (d) its vehicle type, (e) the type of the vehicle previously discharged, and (f) any additional surcharge to account for pedestrian interference.

Bus traffic is explicitly and realistically treated by the model. Buses have their own vehicle length and operational characteristics and traverse prescribed paths (routes) through the network, servicing those stations assigned to that route. The probability of stopping and the duration of dwell time are assigned to each vehicle stochastically. Impedance from other traffic, queuing that prevents buses from accessing their stations, the possibility of a saturation of station capacity, and many similar factors are rigorously modeled.

Scheduling of Vehicle Movement

Since the Level I model uses an event-based simulation logic, it is necessary to create a schedule of events internally. Two vehicle-scheduling arrays are used. In one array for near-term events, vehicles with an activation time (AT) within the current scheduling period (i.e., an AT lower than some clock time into the future) are stored. All other vehicles fall by default under the "imaginary" long-term-event array and are not stored. When the simulation clock reaches the end of a scheduling period or the near-term scheduling array is empty, a new scheduling period is defined. At this point, all vehicles in the network are scanned and those whose AT is within the new scheduling period are stored in the near-term scheduling array.

Vehicle Processing

A vehicle is only processed (i.e., moved) when its AT is equal to the current simulation clock time. Each vehicle is then generally moved to a point downstream, either on its current link or onto a receiving link, subject to the constraints imposed by other vehicles and by the signal control. The vehicle's new location, speed, and AT are computed in the process. This vehicle then remains "dormant" until the simulation clock time advances to the new AT, whereupon the vehicle is again processed.

A small number of cases that, in aggregate, span the entire spectrum of possible conditions were identified (see Figure 1). For each such case, explicit analytic expressions have been derived.

The number of times that a vehicle is processed on any link varies according to the following conditions:

1. Whether the vehicle will encounter a queue,
2. Whether the vehicle will be stopped by the control at the intersection,
3. The vehicle's turn movement,
4. Whether the vehicle will experience cycle failure as a result of congestion,
5. Whether intersection blockage prevails, and
6. Whether the control is actuated.

It appears that a vehicle will be processed once if it is unconstrained and encounters a go indication. If it is constrained (by control or by joining a queue), and if the control is fixed time and there is no intersection blockage, it will be processed twice. In the presence of intersection blockage or cycle failure, a vehicle may be processed three or more times. If the signal control is actuated and its indication is no-go, it is not possible for the software logic to predict when the go phase will become active. The program is then forced to revert to a time-scanning approach for the lead queued vehicle until the indication switches to go.

Car-Following Model

To ensure that the kinematic relation between the subject vehicle and its leader is realistic, a car-following relation used in the Integrated Traffic Simulation (INTRAS) model (5) is applied to cases 3, 6, and 9 in Figure 1.

The car-following relation applies only when the presence of the lead vehicle actually affects the trajectory of the subject vehicle. This depends on the separation of the two vehicles in time and space and on their respective speeds at the activation time of the subject vehicle.

Intersection Capacity

An analytical model of approach capacity that is applicable to all geometric and control configurations was developed as part of this project and is discussed in a paper by McShane and Lieberman and a paper by Lieberman elsewhere in this Record. This model provides relevant lane capacities (service rates), given the geometrics, control policy, conflicting flows, and lane-specific mix of through and turning traffic volume. It also finds the lane-specific deployment of traffic, stratified by turn movement, given the geometrics, traffic volume, control policy, and turn-movement-specific service rates.

Unlike some models that use a constant, reconstrained speed (2,3), the Level I model treats the acceleration and deceleration of all vehicle trajectories explicitly. In addition, the calculation of vehicle trajectories is based on the known distance L (of the time of vehicle movement) to the rear of a queue or to the stop line. This refinement affects the results obtained.

Input

In general, the inputs of the Level I model include geometric characteristics, traffic volumes, traffic control specifications, and driver and traffic

Figure 1. Flow chart of decision process used to obtain vehicle-processing cases.

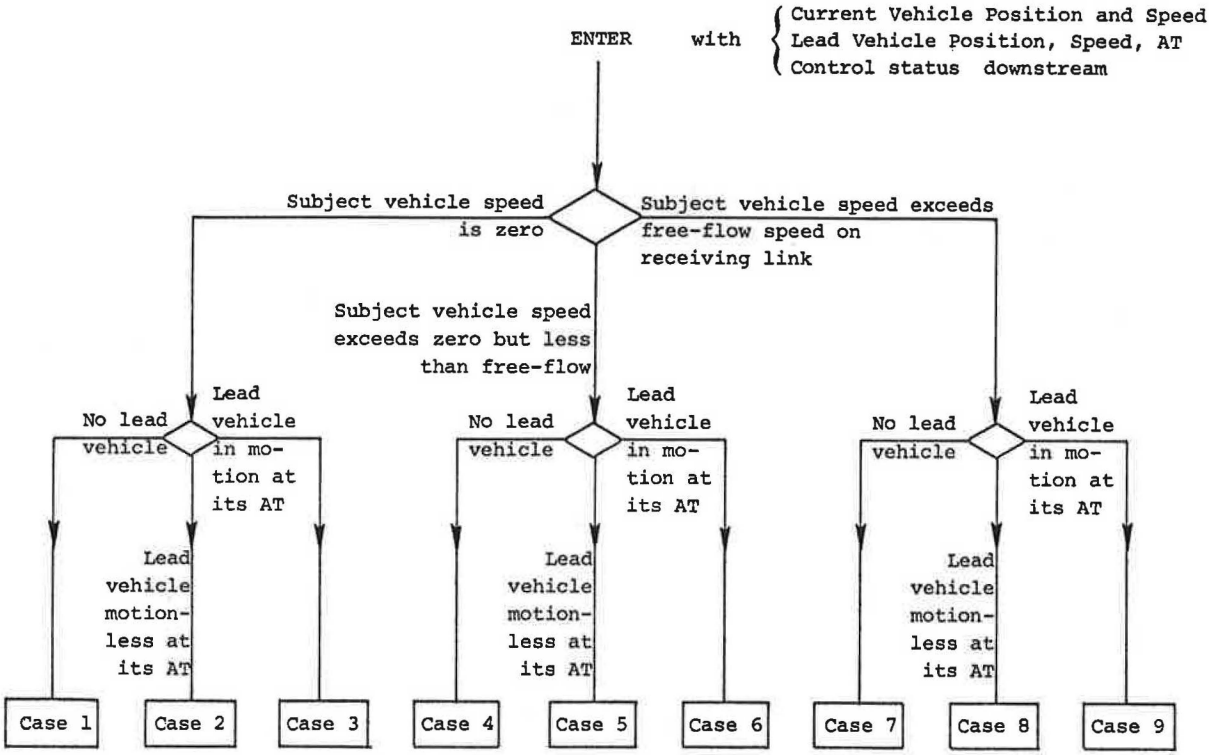


Table 1. Level I validation: networkwide comparison of measures of effectiveness.

		Measure							
		Travel Time (min)		Vehicle Miles		Total Delay (min)		Mean Speed (miles/h)	
Case	Run	Model	Field	Model	Field	Model	Field	Model	Field
Peak	1	11 015	10 505	1780	1701	6911	6479	9.69	9.71
	2	10 919	10 505	1776	1701	6822	6479	9.76	9.71
	3	11 024	10 505	1780	1701	6918	6479	9.69	9.71
Off-peak	4	8 522	8 841	1330	1286	5479	5814	9.36	8.73
	5	8 621	8 841	1335	1286	5564	5814	9.29	8.73
	6	8 349	8 841	1321	1286	5325	5814	9.50	8.73

operational characteristics. Considerable effort was made to design the input requirements so as to minimize data collection and preparation and the computer storage required to accommodate the data.

Output

A comprehensive set of traffic performance measures is generated periodically, either as cumulative output or as more detailed intermediate output, at the option of the user. The cumulative output is presented both in link-specific form and for the network as a whole. Intermediate outputs provide "snapshots" of system status and additional information for individual links. Separate statistics are provided for bus traffic along the lines of the NETSIM model output. Statistical estimates of fuel consumption and vehicle emissions for each vehicle type are also provided (6).

VALIDATION RESULTS

The Level I model was successfully validated by using a network in the central business district of Washington, D.C. This network exhibited many short links, significant bus traffic, turn pockets, and other features that rigorously tested the model.

Three model runs (replications) were executed on morning peak and off-peak traffic situations so that the influence of stochastic variations on the test results could be determined. Each simulation run was extended over a (simulated) time of 32 min, broken down into a sequence of eight 4-min time periods. Input data, such as traffic volumes and turning movements, varied for each time period.

Table 1 gives some networkwide comparisons of measures of effectiveness in the peak and off-peak cases for model results and results obtained in the field. The results show that the Level I model performs with a high degree of accuracy.

OPERATING EFFICIENCY

Computer running time is affected by the duration of the simulation, the number of vehicles, and the specification of user options. For the Level I model, the most meaningful operational statistic is the ratio of vehicle seconds of travel time to seconds of computer processing time. This ratio is 20 000:1 for the CDC 7600, which is approximately

five times faster than the ratio for the NETSIM model (4).

ACKNOWLEDGMENT

The development of a model such as that described in this paper represents the contributions of many people. In particular, we wish to extend our thanks to Guido Radelat and George Tiller of the Traffic Systems Division of the Federal Highway Administration; Barbara Andrews and Mark Yedlin of KLD Associates, Inc.; William McShane of the Polytechnic Institute of New York; and Fred Wagner of Wagner/McGee Associates. This work was performed under contract with the U.S. Department of Transportation.

REFERENCES

1. E.B. Lieberman and others. Macroscopic Simulation for Urban Traffic Management: The TRAFLO Model. Federal Highway Administration, U.S. Department of Transportation, Vols. 1-7, 1980.
2. Program Manual for Macroscopic Simulation Model of Diamond Interchange Traffic Operations. System Development Corp., Los Angeles, 1972. NTIS: PB 216 990.
3. Refinement and Testing of Urban Arterial and Network Simulation (TRANS). Planning Research Corp., Los Angeles, 1967. NTIS: PB 177 605.
4. Network Flow Simulation for Urban Traffic Control System: Phase II (NETSIM). Peat, Marwick, Mitchell, and Co., Washington, DC; and KLD Associates, Inc., Huntington Station, NY, 1973. NTIS: PB 230 760-4.
5. KLD Associates, Inc. Integrated Traffic Simulation (INTRAS): Program Design, Parameter Calibration, and Freeway Dynamics Component Development. Federal Highway Administration, U.S. Department of Transportation, Rept. DOT-FH-11-8502, 1977.
6. S.L. Cohen and G. Euler. Signal Cycle Length and Fuel Consumption and Emissions. TRB, Transportation Research Record 667, 1978, pp. 41-48.

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

Service Rates of Mixed Traffic on the Far Left Lane of an Approach

WILLIAM R. McSHANE AND EDWARD B. LIEBERMAN

The effect of left-turning vehicles on approach capacity has been observed and studied, but a complete model has been lacking. As part of the development of the TRAFLO simulation model for the Federal Highway Administration, the capacity of a signalized intersection approach was modeled. The most complex component of this capacity model, that for left-turn lanes, is discussed. Three basic types of intervals for the vehicle discharge process are identified, and the probabilities of each are found. Expressions for expected vehicle discharge per cycle and for saturation flow rates per hour of green are reported. Reasonable results have been obtained.

The problem of a left-turn lane that serves both turning and through vehicles has been addressed in various ways, as has the problem of lanes that serve turners only (1-3). The 1965 Highway Capacity Manual (4) provides some empiric assessment of left-turn capacity without special phasing; other treatments include equivalencies for left turners facing opposing traffic (5). But none of the existing treatments provides the level of information and detail necessary for inclusion in the TRAFLO model currently being developed for the Federal Highway Administration (FHWA).

The TRAFLO model includes a determination of service rates on approaches to a signalized intersection. (It also includes appropriate treatments of unsignalized intersections, but these are not treated in this presentation.) The discharge service rates of vehicles on the individual lanes are strongly linked to the mechanism by which the through vehicles distribute themselves among the lanes in a self-optimizing fashion. Computationally, the service rates of the mixed traffic lanes (i.e., those that contain turning and through vehicles) are functions of the percentage of turning vehicles in the lanes. This interdependence of service rates and the lateral distribution of vehicles on an approach creates a feedback that can be addressed by the following iterative computations:

1. Assuming that all vehicles in the outermost lanes turn, the service rates of these lanes can be computed from appropriate models to begin the procedure.
2. Based on these service rates, the principle of drivers optimizing their individual travel time is applied in order to compute the implied percentage of turns in each lane that serves mixed (i.e., through and turning) traffic.
3. If the lane percentages differ from the assumed values, the initial computation is repeated but the latest lane percentage is used.

The overall procedure implied in these steps is discussed and justified in a paper by Lieberman elsewhere in this Record. It has been shown that convergence is obtained and that few iterations are required.

This paper is restricted to the model of the lane that contains through and left-turning vehicles, as required in the statement "computed from appropriate models" in step 1 above. The complexity of the process makes a closed-form solution infeasible, but the formulation obtained is both complete and tractable. It is certainly well suited to the original purpose--inclusion in the TRAFLO model--and could potentially be a component in a general procedure for estimating intersection capacity.

LEFT-TURN-LANE MODEL

The far left lane at an intersection generally contains mixed traffic, of which a fraction λ turns left and a fraction $(1 - \lambda)$ goes through. The activity during a particular green phase is handled by three submodels:

1. An A model, which handles the discharge of the subject lane while an opposing queue is discharging or a coherent oncoming platoon is discharging;
2. A B model, which handles the discharge of vehicles on the subject lane during unstructured opposing arrivals, usually after the A model; and
3. A C model, which handles the discharge from the subject lane by a number of left-turn laggers at the end of the green phase and a "jumper" at the beginning of the phase.

Each vehicle in the subject lane is assigned a probability λ of being a left turner, except as explicitly noted. Gaps in the opposing flow are assumed to be exponential during the B model. This assumption is reasonable and has two important advantages: (a) If each gap in each opposing lane is exponential, it can be shown by order statistics that the resultant gaps are also exponential and (b) in the B model, truncated gaps and conditioned gaps play an important role. Note that the distribution of the conditioned gaps of an exponentially distributed sequence of gaps is also exponential.

For the purposes of the work presented in this paper, we assume that discharges from both the subject lane and the opposing queue (if any) are deterministic at specified rates. An endless supply of vehicles is assumed in the subject lane, for it is the capacity value that is of interest in this work. Clearly, if the necessary demand is lacking, the actual output can be less than the levels that result from the computations.

It must be recognized that the model presented here was developed to be both logical and complete and to yield known sensitivities that are not represented in previous formulations and at the same time to remain tractable and computationally feasible. Some refinements are possible but were judged unnecessary for the precision with which the model would be used. For instance, duration for the A model is actually a random variable, but a simple deterministic computation is done to estimate its duration. Likewise, a weighted average headway is used in the B model to simplify the formulation and the computations.

A Model

For simplicity, completely random (i.e., exponentially distributed) opposing arrivals are assumed in the A model. As Figure 1 shows, a queue forms on the opposing approach. The queue discharges during the "A period", effectively blocking any discharge of left turners on the subject approach during this time.

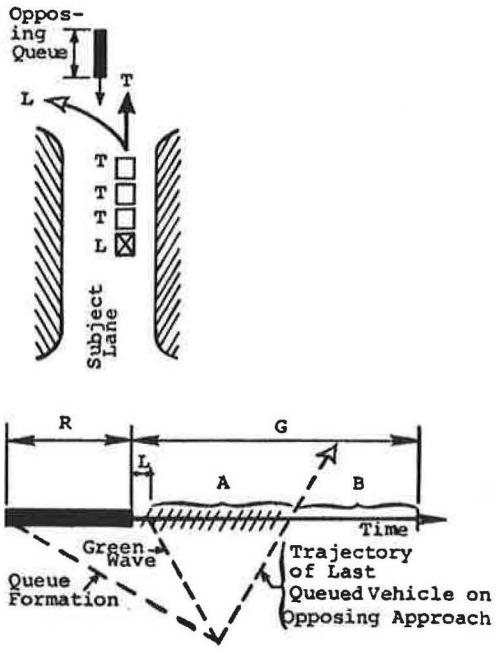
The A duration is given by

$$A = Q_{opp}(R + L)/(S - Q_{opp}) \quad (1)$$

where

Q_{opp} = total opposing flow (vehicles/h),
R = red-phase duration (s),
L = total start-up lost time for discharging queue (s), and
S = total opposing saturation flow (vehicles/h).

Figure 1. The A model.



How many vehicles can be discharged from the subject lane during the A duration? Only leading through vehicles can discharge, since the first left turner must wait until the opposing queue is discharged. The maximum number of through vehicles that can discharge (M*) is

$$M^* \doteq (A/t_0) \quad (2)$$

where t_0 is the mean discharge headway for a through vehicle in the lane of interest (s/vehicle).

Consider the situation of X through vehicles preceding the first turner. The probability of there being x through vehicles preceding the first left turner is

$$P_r(X = x) = \begin{cases} (1 - \lambda)^{x\lambda} & x < M^* \\ (1 - \lambda)^{M^*} & x = M^* \end{cases} \quad (3)$$

and the expected value of x is

$$N_A = E[X] = \sum_{x=0}^{x=M^*} x \cdot P_r(X = x) \quad (4)$$

or

$$N_A = [(1 - \lambda)/\lambda] - [(1 - \lambda)^{M^*+1}/\lambda] \quad (5)$$

For small λ, the asymptotic form $N_A \simeq M^*[1 - (M^* + 1)\lambda/2]$ is used to avoid numeric problems with the division by λ as it approaches zero.

It is necessary to compute the gap-distribution parameter α for the B model. This value is based on the total opposing flow minus that portion that is discharged during the A duration.

Let N_B equal the number of vehicles discharged from the opposing approach during the B duration on one cycle:

$$N_B = (Q_{opp}C/3600) \cdot \{1 - [(R + A)/C]\} = Q_{opp}B/3600 \quad (6)$$

where C is the cycle length in seconds. The rate of flow during the B duration is therefore

$$N_B/B = Q_{opp}/3600 \quad (7)$$

and, consequently, $\alpha = Q_{opp}/3600$ vehicles/s.

B Model

The B model, which is relatively complex, is presented in two parts: the "start" operation and the "normal" operation.

Start Operation

At the beginning of the B interval, there is a probability P_R that the first vehicle is a left turner [this is the "left turner first" (LTF) situation]:

$$P_R = \sum_{x=1}^{x=M^*+1} (1-\lambda)^{x-1}\lambda = 1 - (1-\lambda)^{M^*+1} \quad (8)$$

This left turner must wait through N gaps in the opposing flow that are too small, followed by a gap that is at least the minimum size. The probability that a gap is too small is

$$P_1 = (1 - e^{-\alpha T}) \quad (9)$$

where α is the parameter for the opposing-flow gap distribution (1/s) and T is the gap required by the first left turner in a given gap (s). The expected value of the waiting time through the succession of such gaps is

$$E(W) = \sum_{i=0}^{\infty} i\mu_1 P_1^{\,i}(1 - P_1) = \mu_1 P_1/(1 - P_1) \quad (10)$$

where μ_1 is the mean of the rejected gap, as shown in Figure 2a:

$$\mu_1 = \int_0^T t f_1(t)dt = (1/\alpha) - T(1 - P_1/P_1) \quad (11)$$

After this, there is a gap that is big enough. The mean of this acceptable gap, denoted as μ_2, is given by

$$\mu_2 = [T + (1/\alpha)] \quad (12)$$

which reflects the properties of the exponential distribution (see Figure 2b).

This same gap may also service other vehicles. If the following vehicle is another turner, it takes H_o s to discharge. If it is a through vehicle, it takes t_o s to discharge. To make the modeling tractable, it is assumed that both take the same time H, a weighted average:

$$H = \lambda H_o + (1 - \lambda)t_o \quad (13)$$

The number of vehicles that can be accommodated in this gap is shown in Figure 3. It can be shown that the expected numbers of left-turn and through vehicles in this gap are as follows:

$$E(LT) = [1 - (1 - \lambda)U]/(1 - U) \quad (14)$$

$$E(THRU) = (1 - \lambda)U/(1 - U) \quad (15)$$

$$E(TOTAL) = 1/(1 - U) \quad (16)$$

where $U = e^{-\alpha H}$. Note that these are the expected number of vehicles to be discharged in the time D:

$$D = E(W) + \mu_2 \quad (17)$$

defined by the LTF situation that begins the B interval.

Figure 2. Two types of gaps in the start of the B interval.

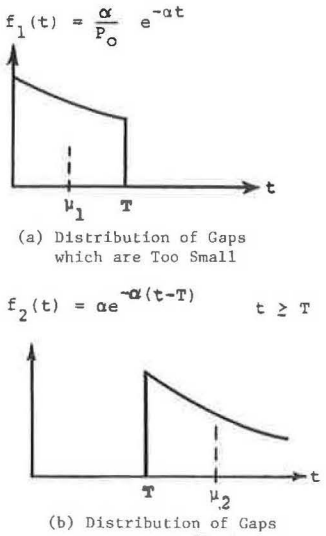


Figure 3. Number of vehicles accommodated in a gap.

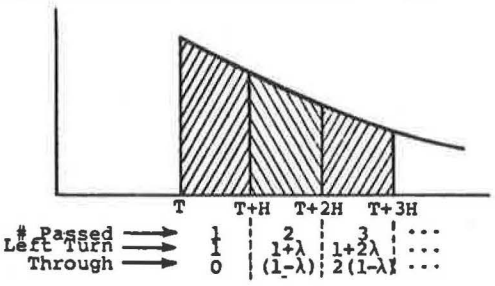
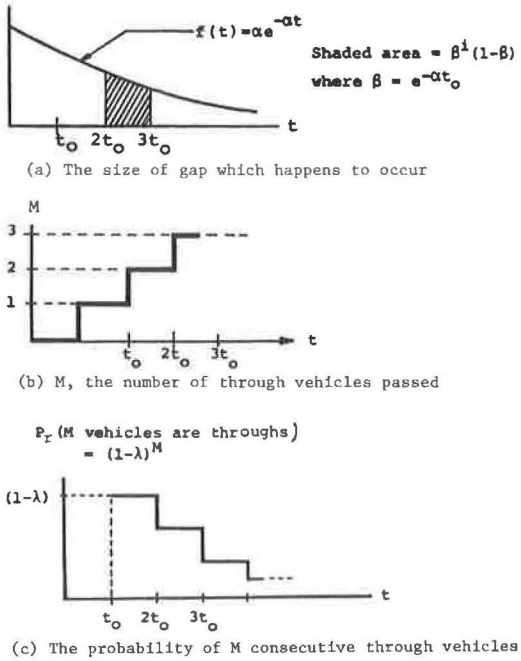


Figure 4. All through vehicles discharged during a gap in the opposing stream.



A special case arises when the B duration is less than the acceptable gap (i.e., D > B).

Normal Operation

After the time described above, other gaps occur. These can be referred to as normal gaps (i.e., not special, as the start situation was). Two distinct possibilities occur: (a) All vehicles processed are through vehicles, or (b) only the first M vehicles are through vehicles. These two possibilities will be considered here individually.

In the first case, note that the opposing gap is "used" by through vehicles only in the sense that they occupy the same time and preclude its use by others (i.e., left turners). Let P_2 be the probability of the gap "passing" only through vehicles and N_2 be the expected number of vehicles passed (given that all are through vehicles). It should be noted that $(1 - \lambda)^M\lambda$ is the probability string of exactly M through vehicles followed by cutoff and β, the probability that a gap will "see" only through vehicles, is equal to $e^{-\alpha t_o}$.

Figure 4 summarizes the sequence for an arbitrarily selected gap size: If the gap falls between $2t_o$ and $3t_o$, it can pass M = 2 through vehicles; the probability that there are two consecutive through vehicles is $(1 - \lambda)^2$. Summing over all possibilities, P_2 = the probability of M consecutive through vehicles times the probability that the gap is smaller. Thus,

Figure 5. Summary of activities in each phase of the model.

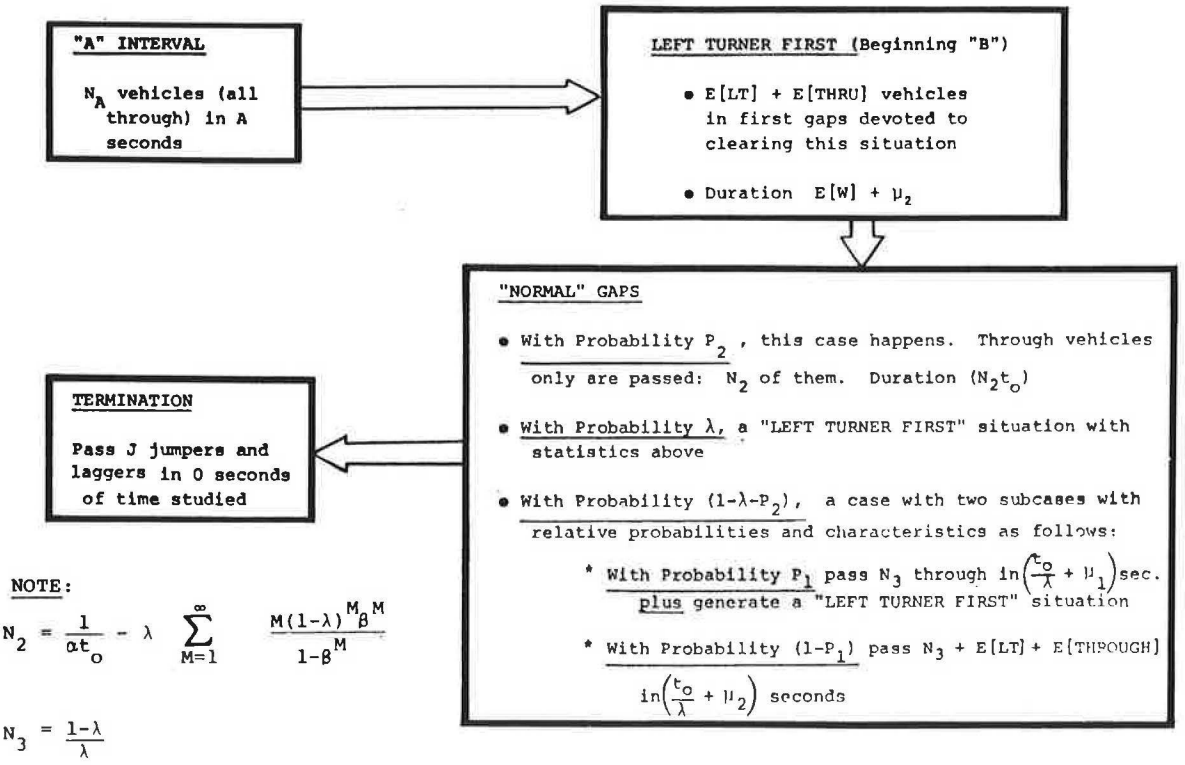
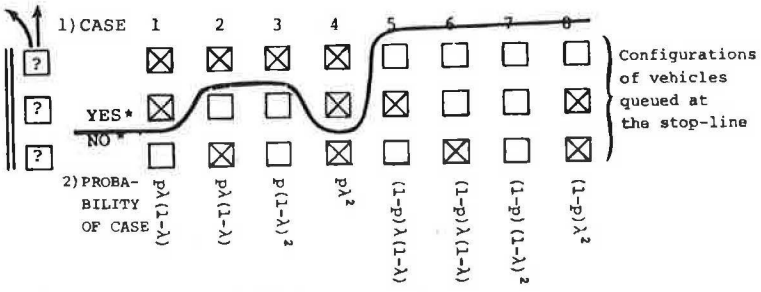


Figure 6. Vehicles at the end of the phase: the C interval.



⊠ Denotes a left-turn vehicle

□ Denotes a "through" vehicle

YES* indicates those who are judged to "push" through intersection

NO* indicates those who will not "push" through intersection

$$P_2 = \sum_{M=1}^{\infty} (1-\lambda)^{M\lambda}(1-\beta^M) = (1-\lambda) - \{\lambda\beta(1-\lambda)/[1-\beta(1-\lambda)]\} \quad (18)$$

One can similarly develop an expression for N_2.

For the second possibility--that only the first M vehicles are through vehicles--there are three subcases: (a) M = 0, (b) M > 0 and the first left turner cannot be passed, or (c) M > 0 and the first left turner can be passed. In the interests of brevity, a detailed development is not presented here.

Figure 5 shows a summary of these and other subcases. The relative frequency of the different subcases, and how one begets the other, are summarized later in this paper. Clearly, if M = 0, the first vehicle is a turner and the situation is identical to all LTF situations, as in the case of the B model.

C Model

Let there be J left-turn "jumpers" or "laggers" per cycle. One must find an expression for J.

At the end of the cycle, several configurations of left turners may "push" themselves through, as shown in Figure 6. It is plausible to assume that no more than 2 left-turn laggers/cycle will push across the stop line. From Figure 6, it can be estimated that there are $(p + p\lambda)$ expected laggers at the end of the cycle, where p is the probability that the vehicle or vehicles indicated in Figure 6 are left turners.

Note that the probability p is not the typical probability of a left turner, λ. It depends on the fact that the configurations in Figure 6 are created by the B interval activity. The value of p is discussed in the summary of model and time allocation.

The "NO" in Figure 6 also gives insight into the probability of a left turner leading the A interval: $p\lambda^2$. Based on observations made in Washington, D.C., the probability of that vehicle jumping is taken as 1/3. Thus, a reasonable estimate of the number of jumpers and laggers per cycle is

$$J = p[1 + \lambda + (\lambda^2/3)] \quad (19)$$

SUMMARY OF MODEL AND TIME ALLOCATION

Figure 5 summarizes the activity in each of the key intervals of the phase. However, this does not provide an estimate of how many events happen in the phase, from which the productivity can be computed.

Assume that there are η events in the B interval, not including the initial LTF situation. Figure 7 shows the sequence in which these events occur.

There are $[1 + \lambda\eta + (1 - \lambda - P_2)P_1\eta]$ LTF situations in a green phase, which leaves the following time (in seconds) for other situations:

$$B - [P_R + \lambda\eta + (1 - \lambda - P_2)P_1\eta][E(W) + \mu_2] \quad (20)$$

There are $(1 - \lambda)\eta$ other situations from the illustration above, each of which averages f seconds. From Figure 5, f can be computed as follows:

$$f = \{P_2(N_2 t_0) + (1 - \lambda - P_2)P_1[(t_0/\lambda) + \mu_1] + (1 - \lambda - P_2)(1 - P_1)[(t_0/\lambda) + \mu_2]\} / [P_2 + (1 - \lambda - P_2)P_1 + (1 - \lambda - P_2)(1 - P_1)] \quad (21)$$

Further simplification is possible, but not necessary, given the anticipated numeric solution. Thus, the situations use $(1 - \lambda)f\eta$ seconds.

The following equality,

$$(1 - \lambda)f\eta = B - [P_R + \lambda N + (1 - \lambda - P_2)P_1\eta][E(W) + \mu_2] \quad (22)$$

leads to

$$\eta = \{B - [E(W) + \mu_2]P_R\} / \{(1 - \lambda)[\lambda + (1 - \lambda - P_2)P_1][E(W) + \mu_2]\} \quad (23)$$

Knowing η, one can obtain expressions for the expected number of through vehicles (THRU) and left-turn vehicles (LEFT) in a phase:

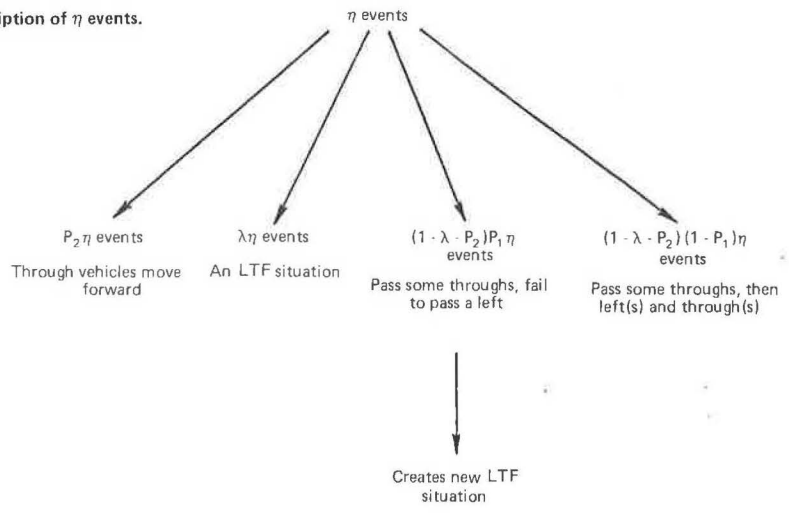
$$THRU = N_A + P_2N_2\eta + E(THRU)[P_R + (1 - P_2)\eta] + N_3\eta(1 - \lambda - P_2) \quad (24)$$

$$LEFT = J + E(LT)[P_R + (1 - P_2)\eta] \quad (25)$$

Given these, simple multiplication by (3600/green time) will yield the components of the saturation flow for the lane, in vehicles per hour of green.

Figure 5 and the separation of cases illustrated

Figure 7. Description of η events.



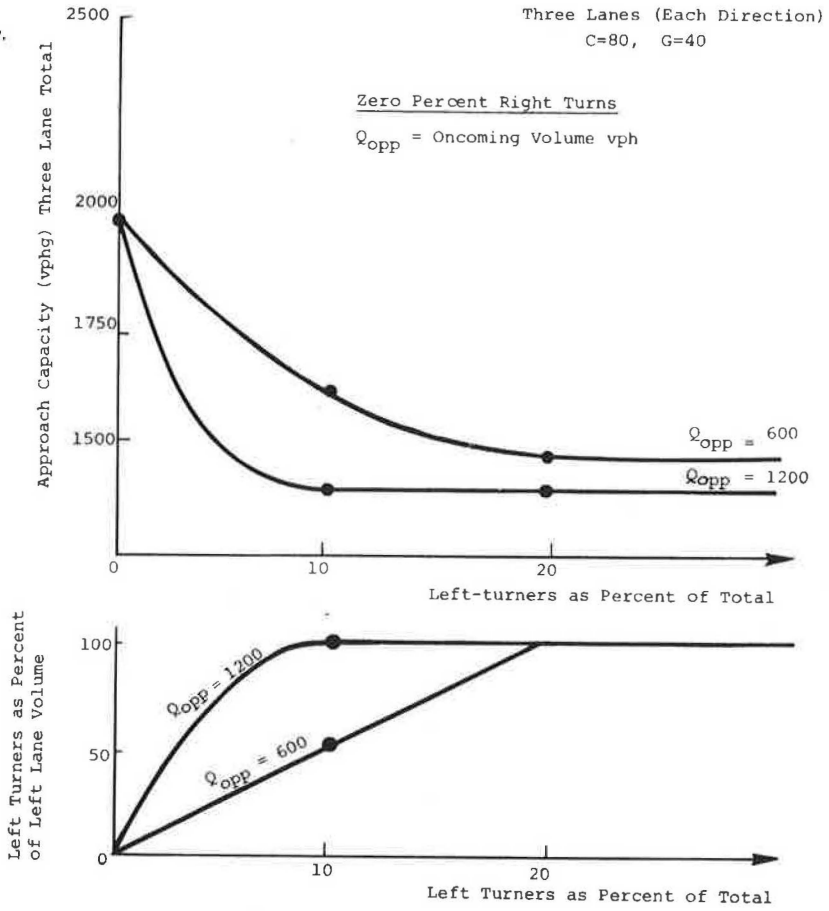


Figure 8. Approach capacity and lane distribution as a function of percentage of left turners and opposing flow.

for the η events can also be used as the basis for estimating the p of the previous section:

$$p = P_R + \lambda\eta + (1 - \lambda - P_2)P_1\eta/P_R + \eta \quad (26)$$

USE AND EFFECTIVENESS OF THE MODEL

The detailed model formulated here is well suited to a computational model such as TRAFLO. It exposes the basic elements of left-lane use, identifies the probability of each case, and is extremely useful in considering sensitivities and interactions. For saturation conditions in the far left lane, our computations have resulted in flow rates that are within 5 percent of values observed in a data base collected by other researchers for FHWA.

Figure 8 shows the effect of the subject model. A six-lane arterial is considered, with a cycle length of 80 s and a green phase of 40 s facing the approach. The fairly rapid fall-off of the discharge capacity with increasing turns replicates patterns observed in actual data better than alternative models. The sensitivity to the opposite-direction flow is interesting. Of particular interest is the fact that through vehicles quickly avoid the far left lane as the opposing flow or the number of turners increases.

For simplicity, Figure 8 does not show the interaction effect of right turners on the far left lane and vice versa. The complete TRAFLO approach model does have a right-turn component model.

ACKNOWLEDGMENT

The work reported in this paper was conducted by KLD Associates, Inc., under contract to FHWA. However, we, the authors, are solely responsible for the accuracy of the model reported.

The right-turn model cited in the text was created by Charles Berger of Sperry Systems Management, who also participated in the TRAFLO model work. Guido Radelat is the FHWA project manager.

REFERENCES

1. D. B. Fambro, C. J. Messer, and D. A. Andersen. Estimation of Unprotected Left-Turn Capacity at Signalized Intersections. TRB, Transportation Research Record 644, 1977, pp. 113-119.
2. E. Fellinghauer and D. S. Berry. Effects of Opposing Flow on Left-Turn Reduction Factors at Two-Phase Signalized Intersections. TRB, Transportation Research Record 489, 1974, pp. 13-18.
3. A. J. Miller. Signalized Intersections Capacity Guide. Australian Road Research Board, Vermont South, Victoria, Rept. 79, April 1978, 26 pp.
4. Highway Capacity Manual. HRB, Special Rept. 87, 1965.
5. R. L. Carstens. Some Traffic Parameters at Signalized Intersections. Traffic Engineering, Aug. 1971, pp. 33-36.

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

Generalized Procedure for Estimating Single- and Two-Regime Traffic-Flow Models

SAID M. EASA AND ADOLF D. MAY

Macroscopic traffic-flow models play an important role in the planning, design, and operation of transportation facilities. Evaluation of these models is often required to select the appropriate model that best represents prevailing operating conditions. For this purpose, a technique is needed that will enable the analyst to easily and quickly estimate model parameters. The technique should be easy to understand and use and inexpensive to apply and should generate results that reasonably represent actual traffic behavior. The development of such a technique is described. The proposed estimation procedure is based principally on the theoretical relations between model parameters and traffic-flow criteria. Such relations were developed for both single- and two-regime approaches. To facilitate use of the procedure, generalized nomographs were developed to model the complexity of the theoretical aspects involved. These nomographs are capable of directly providing the values of model parameters that satisfy specified evaluation criteria. This procedure significantly reduces the need for regression analysis in estimating model parameters and thus appears to be of particular use in a wide range of transportation applications.

Considerable research has been undertaken to model the interrelationships among traffic-flow variables, and researchers have developed several models that describe the behavior of traffic flow on highways. In general, traffic-flow models can be classified into two major classes: microscopic and macroscopic. Microscopic models consider the spacing and speed of individual vehicles as model elements. Macroscopic models, on the other hand, describe the operations of traffic flow in terms of the speed, flow, and density of the traffic stream.

The macroscopic models are generally adequate for most practical purposes and have been widely used in the planning, design, and operation of transportation facilities. Before any particular macroscopic model can be used, however, the analyst should estimate model parameters that best represent prevailing traffic characteristics.

There is a need for a technique that will enable the analyst to directly estimate model parameters. It is obviously desirable that such a technique exhibit several important features, including efficiency, flexibility, accuracy, and generality: It should be easy to understand and use and be inexpensive to apply, it should allow flexible treatment of the various variables and parameters involved and generate results that reasonably represent actual traffic behavior, and, most important, it should be general in nature and allow a wide range of transportation applications. With these features in mind, a generalized procedure for estimating single- and two-regime models has been developed.

This paper presents a background of microscopic and macroscopic modeling theories and briefly discusses the concept of the proposed procedure. A detailed description of the evaluation procedure for single-regime models is given, and the evaluation procedure for the two-regime models is described.

BACKGROUND

The general macroscopic theory of traffic flow is based principally on the microscopic (car-following) theory. These two classes of theories are described briefly below.

Microscopic Theory

The microscopic description of vehicular traffic flow was first formulated by Reuschel (1) and Pipes (2). They formulated the phenomena of pairs of vehicles following each other:

$$X_n - X_{n+1} = L + S(\dot{X}_{n+1}) \quad (1)$$

In this formulation, it is assumed that driver (n + 1) maintains a separation distance from driver n proportional to the speed of his or her vehicle ($\dot{X}_{n+1}$) plus a distance L. The factor L is the distance headway at standstill ($\dot{X}_n = \dot{X}_{n+L} = 0$). The constant S has the dimension of time, and the differentiation of Equation 1 gives

$$\ddot{X}_{n+1} = (1/S)(\dot{X}_n - \dot{X}_{n+1}) \quad (2)$$

where $\ddot{X}_{n+1}$ = the acceleration (or deceleration) rate.

This differential equation is generally referred to as the basic equation of the car-following theory. This basic stimulus-response relation was investigated further by Chandler, Herman, and Montroll (3), who formulated a linear mathematical model that took the following form:

$$\ddot{X}_{n+1}(t+T) = \lambda[\dot{X}_n(t) - \dot{X}_{n+1}(t)] \quad (3)$$

where T = the time lag of response to the stimulus and λ = the sensitivity factor.

This formulation was refined by Gazis and others (4,5), and a more general expression of the sensitivity factor was proposed:

$$\lambda = \alpha\left\{\dot{X}^m_{n+1}(t+T)/[X_n(t) - X_{n+1}(t)]^{\ell}\right\} \quad (4)$$

where α is the constant of proportionality.

The general expression for the microscopic theory thus becomes

$$\ddot{X}_{n+1}(t+T) = \alpha\left\{\dot{X}^m_{n+1}(t+T)/[X_n(t) - X_{n+1}(t)]^{\ell}\right\}[\dot{X}_n(t) - \dot{X}_{n+1}(t)] \quad (5)$$

Macroscopic Theory

Gazis, Herman, and Rothery (5) have shown that, by integrating the generalized microscopic equation (Equation 5), the following expression is obtained:

$$f_m(u) = c' + cf_{\ell}(s) \quad (6)$$

where

u = steady-state speed of the traffic stream,
s = constant average spacing, and
c and c' = some appropriate constants consistent with physical restrictions.

The integration constant c' is related to free-flow speed u_f or jam spacing s_j, depending on the values of ℓ and m. The jam spacing s_j can be transformed to jam density k_j by $s_j = 1/k_j$.

By using this general solution of Gazis and others, May and Keller (6) developed a matrix of the steady-state flow equations for different ℓ and m values. This matrix was modified by Ceder (7) and has been further refined here to properly establish some regions of the matrix. The final version is shown in Figure 1.

The matrix shows the speed-density relations for different combinations of ℓ and m parameters in four "regions". In region 1 ($\ell < 1$ and $m > 1$), the boundary conditions are not satisfied. Models in region 2 ($\ell < 1$ and $m < 1$) have no intercept with the speed axis, $u_f \to \infty$. Models in region 3 ($\ell > 1$ and $m > 1$) have no intercept with the density axis, $k_j \to \infty$. Region 4 ($\ell > 1$ and $m < 1$) contains models that have intercepts with both axes.

It should be noted that this paper is concerned only with the three specially delineated parts of the matrix shown in Figure 1. These include region 4 and the two portions of regions 2 and 3 that correspond to $\ell = 1$ and $m = 1$, respectively. For consistency and ease of reference, models in region 4 will be referred to throughout as single-regime models, those in region 2 as congested-flow models, and those in region 3 as non-congested-flow models.

Single- and Two-Regime Approaches

There are generally two approaches to representing traffic-flow relations: single-regime and two-regime. In the single-regime approach, the entire range of operations is represented by a single model (normally from region 4) as shown in Figure 2(a), but one could represent the regimes of noncongested and congested flow by separate models, as shown in Figure 2(b). This two-regime representation, first proposed by Edie (8), provides a theory that accounts for the discontinuity often observed in traffic-flow data. As defined in this paper, in the two-regime approach the non-congested-flow regime can be represented by a model from region 3 or region 4, and the congested-flow regime can be represented by a model from region 2 or region 4.

CONCEPT OF THE PROPOSED PROCEDURE

An illustration of the proposed estimation concept is shown in Figure 3. The procedure is based principally on the theoretical relations between traffic-flow criteria and model parameters. The traffic-flow criteria include free-flow speed (u_f), optimum speed (u_o), jam density (k_j),

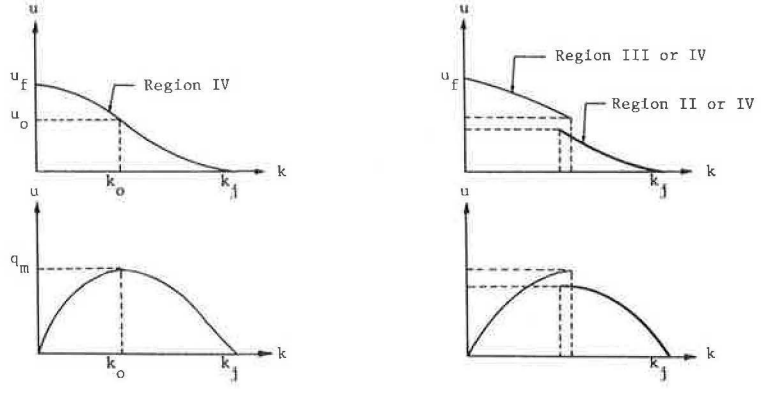
Figure 1. Matrix of steady-state flow equations for different values of ℓ and m.

ℓ \ m	$m < 1$	$m = 1$	$m > 1$
$\ell < 1$	$u^{1-m} = ck^{\ell-1} - ck_j^{\ell-1}$ where $c = \alpha \frac{1-m}{1-\ell}$ and $\ell > m$	$\ell n\, u = \alpha \frac{k^{\ell-1}}{1-\ell} + d_4$	$u^{1-m} = \alpha \frac{1-m}{1-\ell} k^{\ell-1} + d_3$
$\ell = 1$	$u^{1-m} = (1-m)\alpha\, \ell n(k_j/k)$	$\ell n\, u = \alpha \ell n\, (d_1/k)$	$u^{1-m} = (1-m)\, \alpha \ell n\, (1/k) + d_2$
$\ell > 1$	$u^{1-m} = u_f^{1-m}\, 1-(k/k_j)^{\ell-1}$	$\ell n\, u = \ell n\, u_f + \frac{\alpha}{1-\ell} k^{\ell-1}$	$u^{1-m} = u_f^{1-m} + ck^{\ell-1}$ where $c = \alpha \frac{1-m}{1-\ell}$ and $\ell > m$

(boundary conditions not satisfied)

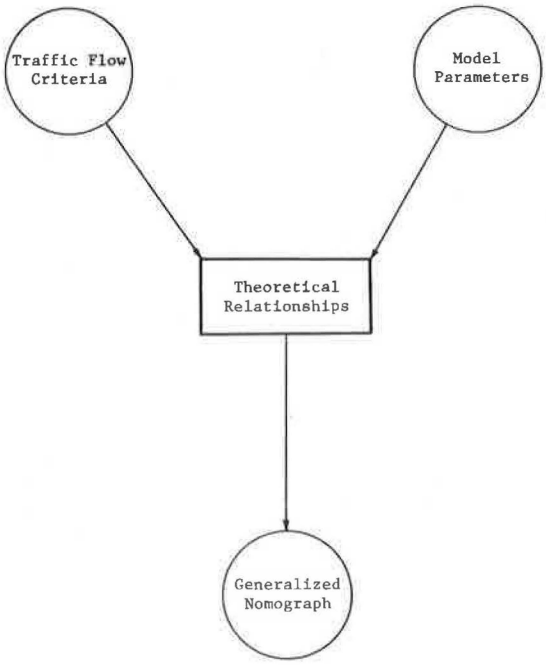
II I IV III

Figure 2. Single-regime and two-regime approaches.



(a) Single Regime (b) Two Regime

Figure 3. Concept of proposed estimation procedure.



optimum density (k_o), and maximum flow (q_m). These five criteria, shown in Figure 2(a), represent the critical points of the traffic-flow relations. In addition, model parameters include α, m, and ℓ, which are contained in the matrix of the general macroscopic models. To make the procedure easier and more flexible to use, the theoretical relations are translated into a generalized nomograph that can be used to directly determine (or output) model parameters that satisfy specified (or input) traffic-flow criteria.

This concept has been proposed and applied to the estimation of single-regime models by Easa (9) and is further extended in this paper to the estimation of two-regime models. The basic principles of the single-regime approach will be repeated here for purposes of integrity and because the single-regime approach is complementary to the two-regime approach, as will be discussed later in this paper.

SINGLE-REGIME APPROACH

As mentioned previously, the single-regime approach, as defined in this paper, is limited to models in region 4 of Figure 1. The generalized procedure for the estimation of models in this region is described here in four parts:

1. Establishment of the theoretical relations between model parameters and traffic-flow criteria,
2. Development of the nomograph,
3. Description of the procedure for establishing the feasible region of model parameters, and
4. A sensitivity analysis of various aspects involved in the procedure.

Theoretical Development

The steady-state flow equation for region 4, shown in the matrix in Figure 1, is as follows:

$$u^{1-m} = u_f^{1-m}\,[1 - (k/k_j)^{\ell-1}] \tag{7}$$

where u and u_f = steady-state and free-flow speeds, respectively; and k and k_j = density and jam density, respectively. This equation represents a single-regime model that has an x intercept (jam density) and a y intercept (free-flow speed) and corresponds to combinations of m and ℓ values so that $m < 1$ amd $\ell > 1$.

Ceder and May (10) have shown that a relation between m and ℓ parameters and traffic-flow characteristics can be obtained. Such a relation includes k_j and u_f and optimum variables u_o and k_o of speed and density, respectively. From Equation 7, one can obtain the following relations, in which q is expressed as a function of k and u, respectively:

$$q = u_f\, k[1 - (k/k_j)^{\ell-1}]^{1/(1-m)} \tag{8}$$

$$q = u\, k_j[1 - (u/u_f)^{1-m}]^{1/(\ell-1)} \tag{9}$$

At maximum flow, $dq/dk = 0$. Therefore, by differentiating Equation 8 with respect to k and equating the derivative to zero, one obtains:

$$(k_o/k_j)^{\ell-1} = (1 - m)/(\ell - m) \tag{10}$$

In addition, at maximum flow, $dq/du = 0$. Therefore, by differentiating Equation 9 with respect to u and equating the derivative to zero, one obtains:

$$(u_o/u_f)^{1-m} = (\ell - 1)/(\ell - m) \tag{11}$$

Equations 10 and 11 are related as follows:

$$(u_o/u_f)^{1-m} = 1 - (k_o/k_j)^{\ell-1} \tag{12}$$

Rearranging to obtain m as a function of ℓ, (u_o/u_f), and (k_o/k_j) gives

$$m = 1 - \ln[1 - (k_o/k_j)^{\ell-1}]/\ln(u_o/u_f) \tag{13}$$

Substituting Equation 13 into Equation 10 gives

$$\ln(u_o/u_f) = [1/(\ell - 1)]\left\{[1/(k_o/k_j)^{\ell-1}]^{-1}\right\}\ln[1 - (k_o/k_j)^{\ell-1}] \tag{14}$$

The reader can see that Equations 13 and 14 relate model parameters ℓ and m to traffic-flow criteria k_j, u_f, k_o, and u_o. Furthermore, the maximum flow q_m is related to k_o and u_o, as follows:

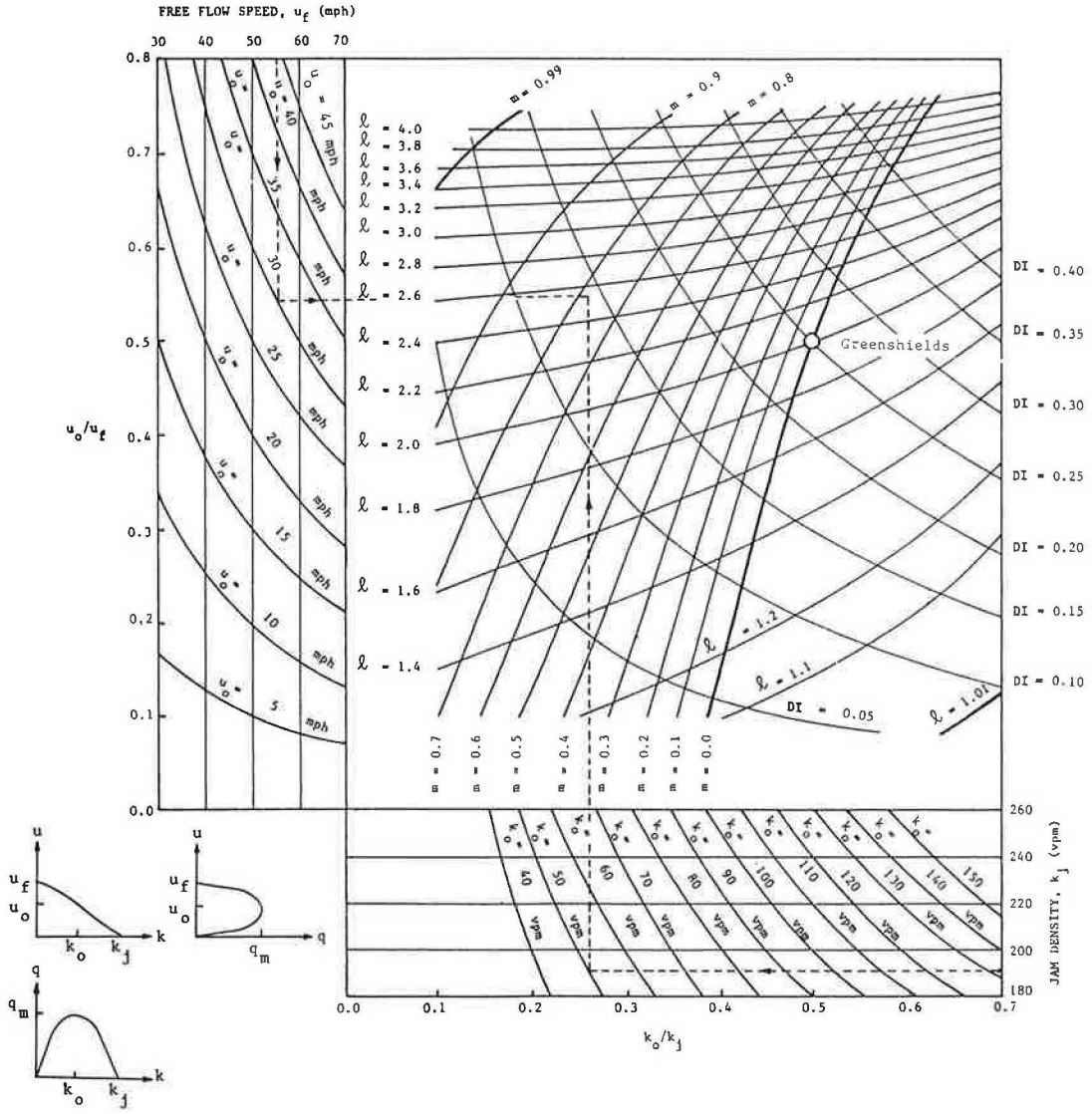
$$q_m = k_o\, u_o \tag{15}$$

Equations 13-15 now relate model parameters ℓ and m to the five traffic-flow criteria, k_j, k_o, u_f, u_o, and q_m. By establishing these criteria from traffic-flow data, one can use these equations to determine the corresponding parameter values. A generalized nomograph developed to simplify this process is described below.

Nomograph Development and Use

The mathematical relations of Equations 13-15 were represented by the nomograph shown in Figure 4 (9), which incorporates contour lines for the five traffic-flow criteria and for model parameters ℓ and m. In the lower right-hand portion of the nomograph, values for jam density (k_j), ranging from 180 to 260 vehicles/mile, are provided. In addition, contour lines for optimum density (k_o) are provided for values ranging from 40 to 150 vehicles/mile. In the upper left-hand portion, values for free-flow speed (u_f), ranging from 30 to 70 miles/h, are given, and contour lines for optimum speed (u_o) are estblished for values ranging from 5 to 45 miles/h.

Figure 4. Generalized nomograph for single-regime models ($\ell > 1$, $m < 1$).



Contour lines for ℓ and m parameters were established by using these values of the traffic-flow criteria. Values of ℓ ranging from 1.1 to 4.0 are included. In addition, the ℓ contour corresponding to a value of 1.01 is provided and represents the limit after which the models would have no intercept with the y axis ($u_f \to \infty$) and would belong to region 2 of Figure 1. The m values range from 0.0 to 0.9, and a value of m = 0.99 is included to represent the limit after which models would have no intercept with the x axis ($k_j \to \infty$) and would belong to region 3. The thick line shown in the middle of the nomograph corresponds to a value of m = 0 and represents a lower limit of the m values. The negative values of m were considered undesirable; because such values have the effect of shifting the speed variable in the sensitivity term of Equation 4 from the numerator to the denominator, they were not included in the nomograph.

The final set of contours provided in Figure 4 is the set related to the maximum flow (q_m). Clearly, contours for q_m cannot be established in Figure 4, since this variable depends on k_o and u_o, which are provided as contours. To solve this problem, a variable DI was introduced. DI is defined as follows:

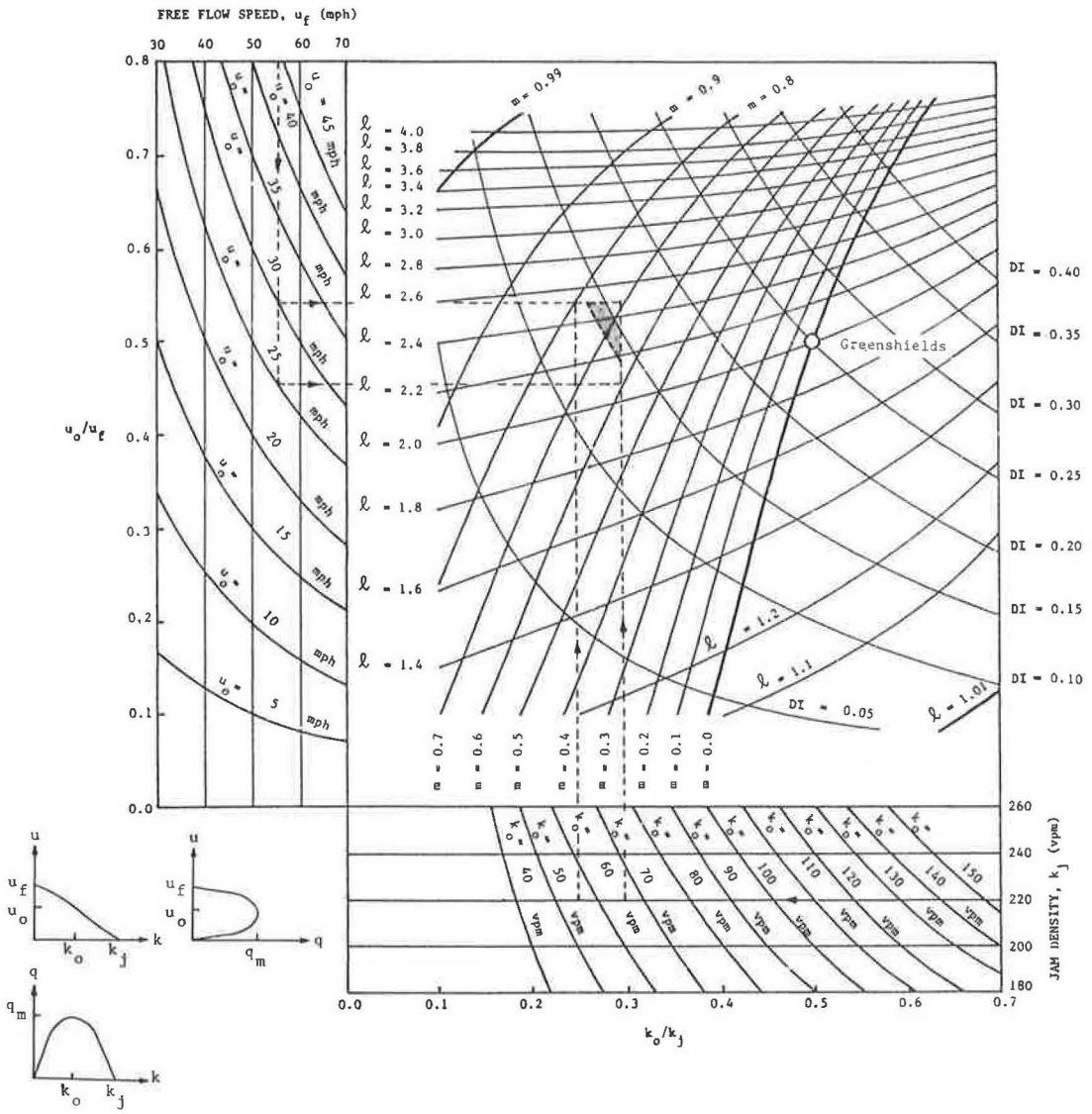
$$DI = (q_m/k_j u_f) = (k_o/k_j)(u_o/u_f) \quad (16)$$

By using this definition, contour lines for DI were constructed in Figure 4 for values ranging from 0.05 to 0.40. These contours are used to establish the maximum flow criteria, which will be described later. It should be noted that the Greenshields model (11) is a special case of single-regime models and corresponds to ℓ = 2.0 and m = 0.

To illustrate the use of the nomograph, let us consider an example. Suppose that the traffic-flow criteria are established as k_j = 190 vehicles/mile, u_f = 55 miles/h, k_o = 50 vehicles/mile, and u_o = 30 miles/h. The corresponding values of model parameters ℓ and m must be determined. To do this, the following steps are performed (Figure 4):

1. Enter at k_j = 190 vehicles/mile and draw a horizontal line that intersects with the contour corresponding to k_o = 50 vehicles/mile.
2. From that point draw a vertical line.
3. Enter at u_f = 55 miles/h and draw a vertical line that intersects with the contour corresponding to u_o = 30 miles/h.
4. From that point draw a horizontal line.
5. The intersection point of the vertical and

Figure 5. Establishing the feasible region for the single-regime approach ($\ell > 1$, $m < 1$).



horizontal lines established in steps 2 and 4, respectively, defines the required values of model parameters. It can be seen that this point corresponds to $\ell = 2.55$ and $m = 0.78$.

6. Check the maximum flow value by reading the value of DI at the intersection point and multiplying it by k_j and u_f to determine q_m. From the diagram, DI = 0.14 and, therefore, q_m = 1463 vehicles/h (55 x 190 x 0.14). Obviously, in this simple example, q_m can be directly calculated from input values of k_o and u_o = 1500 vehicles/h.

With the values of $\ell = 2.55$ and $m = 0.78$, the steady-state flow equations can now be defined. For example, the speed-density relation (Equation 7) can be described as follows:

$$u = 55\,[1 - (k/190)^{1.55}]^{4.55} \quad (17)$$

Establishing the Feasible Region

It should be noted that the procedure described above is intended for use when the traffic-flow criteria are specified as single values. In many situations, however, the analyst might be interested in information on the feasible range of model parameters that satisfy specified ranges (rather than single values) of the traffic-flow criteria. The estimation procedure for such cases can be described as follows, by using a hypothetical example.

Suppose that, based on a given set of traffic-flow data, one has established the following values of the traffic-flow criteria: k_j = 220 vehicles/mile, u_f = 55 miles/h, k_o = 55-65 vehicles/mile, u_o = 25-30 miles/h, and q_m = 1700-1800 vehicles/h. One must determine the feasible region of ℓ and m values that satisfy the above evaluation criteria. The estimation procedure, shown in Figure 5, consists of the following four basic steps:

1. Draw a horizontal line corresponding to the k_j value and two contour lines corresponding to the limits of the k_o range. The intersection of these lines defines two points.
2. Draw a vertical line corresponding to the u_f value and two contour lines corresponding to the limits of the u_o range. The intersection of these lines defines two points.
3. From the two points defined in step 1, draw two vertical lines and, similarly, draw two horizontal lines from the two points defined in step

2. The rectangular region defined by these lines includes the values of model parameters that satisfy the traffic-flow criteria k_j, k_o, u_f, and u_o.

4. Finally, establish the appropriate DI contours that correspond to the criteria range of k_j, u_f, and q_m. Since k_j and u_f can be expressed as ranges, Equation 16 is used with special considerations. Specifically, the lower and upper limits of DI are generally calculated as follows:

$$DI\ (lower) = q_m(lower)/u_f(upper)\ k_j(upper) \quad (18)$$

$$DI\ (upper) = q_m(upper)/u_f(lower)\ k_j(lower) \quad (19)$$

where lower and upper on the right-hand side refer to the limits of the established ranges of the traffic-flow criteria k_j, u_f, and q_m. (Note that, in this example, u_f and k_j are established as single values; the upper and lower limits are equal.)

By using Equations 18 and 19, it can be found that DI (lower) = 0.14 and DI (upper) = 0.15. The DI contours that correspond to these values are shown in Figure 5. The area between these contours that overlaps with the previously determined rectangular region defines the feasible region of model parameters. Clearly, values of ℓ and m within that region satisfy all of the traffic-flow criteria. It should be noted that the area below the DI (lower) contour includes all points with q_m < 1700 vehicles/h. Similarly, the area above the DI (upper) contour includes all points with q_m > 1800 vehicles/h.

The feasible region of model parameters determined above provides a range of models that satisfy specified ranges of the traffic-flow criteria and, consequently, confine the traffic-flow data from which these criteria are established. For this reason, the feasible region would be useful for a variety of transportation applications in which sensitivity to changes in the traffic-flow relations is of particular concern. It is also important to note that, for any model in the feasible region, the associated traffic-flow criteria are immediately defined. For example, for the model that corresponds to ℓ = 2.3 and m = 0.7, it can be determined that k_o = 61 vehicles/mile and u_o = 28 miles/h. Noting that u_f = 55 miles/h and k_j = 220 vehicles/mile, the speed-density relation, for instance, can be described as follows:

$$u = 55\ [1 - (k/220)^{1.3}]^{3.33} \quad (20)$$

It is important to note that, in establishing the feasible region, both u_f and k_j were specified as single values. However, these two criteria can also be established as ranges, in which case additional computations are needed to establish the traffic-flow criteria associated with any selected model (9).

Sensitivity Analysis

In the estimation procedure previously described, a feasible region of model parameters was defined that includes all models that satisfy the established traffic-flow criteria. To investigate the likely variations among models of the feasible region, boundary models were investigated. Figure 6 shows three models in the previously defined feasible region that approximately bound other models in the region. Obviously, other models in the feasible region would lie somewhere within the band of models shown in Figure 6. It is interesting that, if the regression analysis technique (6) had been used to determine model parameters, the selected model with that technique would have lain within that band. This is essentially true, since such a model should fulfill the specified traffic-flow criteria. It is noted in Figure 6 that the expected maximum variations among models are relatively small and do not generally exceed the length of the speed criteria range (5 miles/h in this example).

Another important point related to the sensitivity of model parameters is worthy of note. The shape of model-parameter contours shown previously in Figure 4 clearly exhibits the sensitivity of ℓ and m values to variations in the traffic-flow criteria. To further illustrate the sensitivity of model parameters to these criteria, contours for k_o/k_j and u_o/u_f were established on an ℓ versus m diagram for the region of greater interest (ℓ = 1.8 to 2.8, m = 0.0 to 0.9), as shown in Figure 7. It is noted that the ℓ parameter is considerably more sensitive to u_o/u_f than to k_o/k_j, and this parameter tends to be almost insensitive to k_o/k_j at higher values of ℓ. On the other hand, the m parameter is slightly more sensitive to k_o/k_j than to u_o/u_f. These characteristics appear to be useful as guidelines for the relative effort to be expended in establishing the traffic-flow criteria. Figure 7 also shows the relative locations of high- and low-design facilities and the region of models that correspond to actual traffic-flow data analyzed in previous research work (12).

TWO-REGIME APPROACH

It should be remembered that models in region 4, (the region for which the generalized nomograph described above was developed) have intercepts with both speed and density axes and are designated as single-regime models. In the single-regime approach, the model determined by the nomograph is used to represent the entire range of operations. In addition to its use for the single-regime approach, the nomograph presented above can be used for the two-regime representation in which two models would be established--one for the non-congested-flow regime and the other for the congested-flow regime. In this case, the nomograph is used twice by using the traffic-flow criteria that correspond to each of the two regimes. Such a process is a straightforward application of the procedure described previously and will not be elaborated on further here.

In the two-regime approach, one can represent the non-congested-flow regime by a model from the region (ℓ > 1, m = 1) and the congested-flow regime by a model from the region (ℓ = 1, m < 1). The purpose of this section is to describe the generalized estimation procedure for the two-regime approach by using models from these two regions. This description is presented in three parts:

1. Development of the theoretical relations between model parameters and the traffic-flow criteria for the non-congested-flow and congested-flow regimes,
2. Presentation of nomographs for both regimes as well as a brief description of their intended use, and
3. Description of the procedure for establishing the feasible region of model parameters.

Theoretical Development

For the non-congested-flow region (ℓ > 1, m = 1), the steady-state flow equation, shown in Figure 1, is as follows:

$$\ln u = \ln u_f + [\alpha/1 - \ell)]\ k^{\ell-1} \quad (21)$$

where

k = density,
α = constant of proportionality, and
ℓ = model parameter.

Equation 21 represents a model that has no intercept with the density axis ($k_{j\to\infty}$). This relation can be used to establish the relations between traffic-flow criteria and model parameters. Such relations include free-flow speed u_f, optimum speed u_o, and optimum density k_o, as well as model parameters ℓ and α (note that α is referred to as model parameter).

From Equation 21, one can obtain the following relations, in which q is expressed as a function of k and as a function of u, respectively:

$$q = ke^{\ln u_f + [\alpha/(1-\ell)]\, k^{\ell-1}} \quad (22)$$

$$q^{\ell-1} = [(1-\ell)/\alpha]\; u^{\ell-1} \ln (u/u_f) \quad (23)$$

where e is the base of the natural logarithm.

At maximum flow, $dq/dk = 0$. Therefore, by differentiating Equation 22 with respect to k and equating the derivative to zero, one obtains

$$\alpha = 1/k_o^{\ell-1} \quad (24)$$

In addition, at maximum flow, $dq/d_u = 0$. Therefore, by differentiating Equation 23 with respect to u and equating the derivative to zero, one obtains

$$u_o/u_f = e^{-[1/(\ell-1)]} \quad (25)$$

It is interesting to note that Equation 25 can also be obtained from Equation 11 for the

Figure 6. Variations of models within the feasible region.

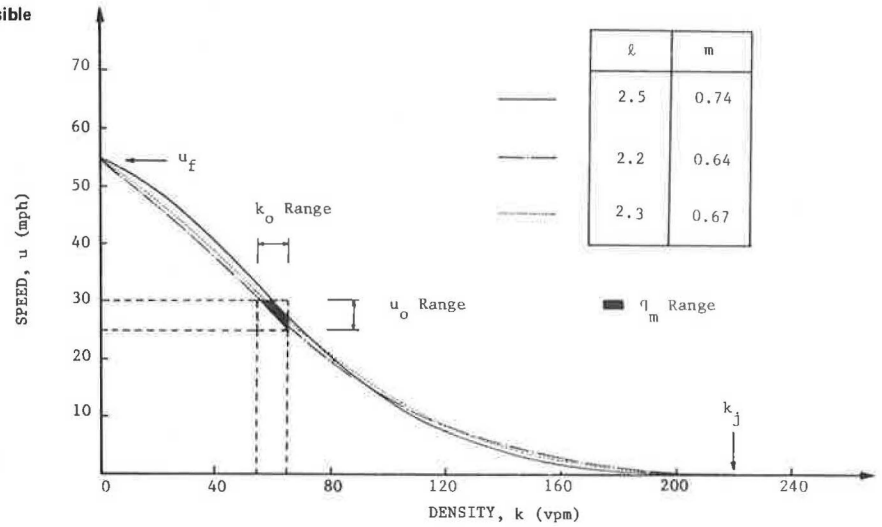
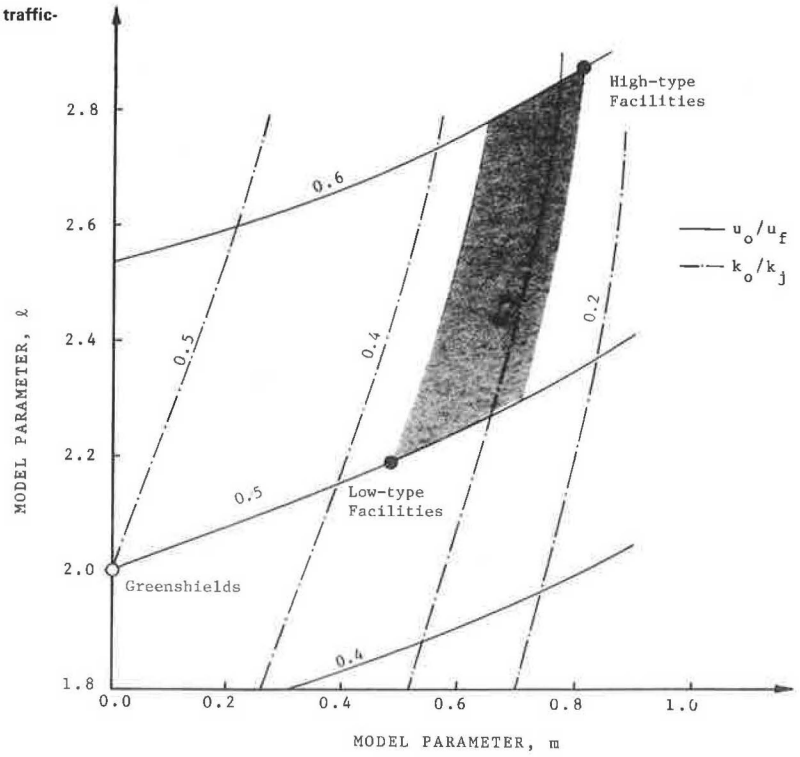


Figure 7. Sensitivity of model parameters to traffic-flow criteria.



single-regime approach. When m approaches 1 in Equation 11, u_o/u_f becomes

$$\lim_{m\to 1}(u_0/u_f) = \lim_{m\to 1}[(\ell-1)/\ell-m)]^{1/(1-m)} = \lim_{m\to 1}[(1+\{1/[(\ell-m)/(m-1)]\})^{(\ell-m/m-1)}]^{-[1/(\ell-m)]} = e^{-[1/(\ell-1)]} \quad (26)$$

which is the same as Equation 25. This feature indicates the continuity of the u_o/u_f ratio between the single-regime models and the non-congested-flow models. Now, rearranging Equation 25 to obtain ℓ as a function of u_o/u_f, one obtains

$$\ell = 1 - [1/\ln(u_0/u_f)] \quad (27)$$

Furthermore, the maximum flow $q_m (= u_o k_o)$ can be expressed as follows:

$$q_{1m} = k_0 u_f e^{-[1/(\ell-1)]} \quad (28)$$

It can be seen that Equations 24, 27, and 28 relate model parameters ℓ and α to the traffic-flow criteria u_f, u_o, k_o, and q_m (note that k_j does not exist).

For the congested-flow region ($\ell = 1$, $m < 1$), the steady-state flow equation, shown previously in Figure 1, is as follows:

$$u^{1-m} = \alpha(1-m)\ln(k_j/k) \quad (29)$$

where k_j = jam density and m = model parameter.

Equation 29 represents a model that has no intercept with the speed axis ($u_f \to \infty$). By following a procedure similar to that described for the non-congested-flow regime, the relations between the traffic-flow criteria k_j, k_o, and u_o and model parameters m and α can be established. From Equation 29, one can express q as a function of k and as a function of u, as follows:

$$q = k[\alpha(1-m)\ln(k_j/k)]^{1/(1-m)} \quad (30)$$

$$q = uk_j \exp\left\{-[u^{1-m}/\alpha(1-m)]\right\} \quad (31)$$

In addition, α and k_o/k_j can be expressed as follows:

$$\alpha = u_0^{1-m} \quad (32)$$

$$k_0/k_j = e^{-[1/(1-m)]} \quad (33)$$

It is worthy of note that, by taking the limit of k_o/k_j in Equation 10 (presented for the single-regime approach) when ℓ approaches 1, one would obtain Equation 33. This can be proved in a way similar to that described earlier for the non-congested-flow models. Similarly, this feature indicates the continuity of the k_o/k_j ratio between the single-regime models and the congested-flow models. Now, rearranging Equation 33 to obtain m as a function of k_o/k_j, one obtains:

$$m = 1 + [1/\ln(k_0/k_j)] \quad (34)$$

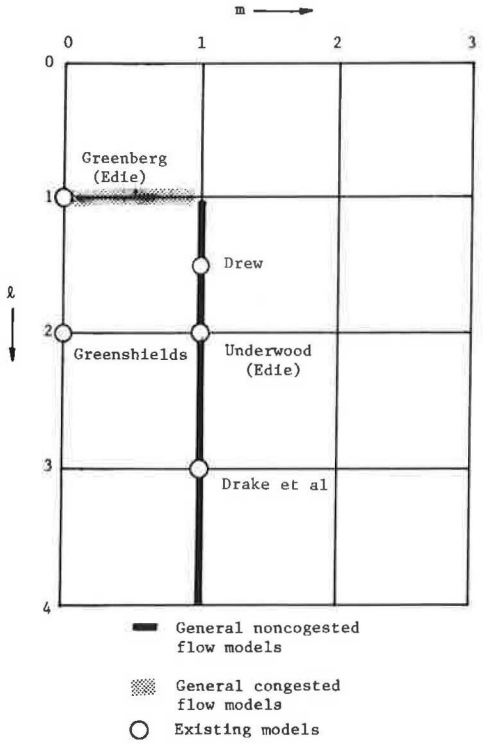
The maximum flow $q_m (= u_o k_o)$ is expressed as follows:

$$q_m = u_0 k_j e^{-[1/(1-m)]} \quad (35)$$

It can be seen that Equations 32, 34, and 35 relate model parameters m and α to the traffic-flow criteria k_j, u_o, k_o, and q_m (note that u_f does not exist).

It is important to note that existing macroscopic models are special cases of the formulations generalized above. Figure 8 shows the locations of the generalized and existing two-regime models. As seen, the three models developed by Drew (13), Underwood (14), and Drake and others (15) correspond to $\ell = 1.5$, $\ell = 2$, and $\ell = 3$, respectively, on the general non-congested-flow models. In addition, the model developed by Greenberg (16) is a special case of the general congested-flow models when m = 0. The key elements of both the general and existing two-regime models are summarized in Figure 9. Clearly, the generalized formulations presented in Figure 9 provide a wider range of models and effect a more flexible treatment in the estimation process for two-regime models.

Figure 8. Locations of generalized and existing two-regime models on ℓ-versus-m matrix.



Development and Use of Nomographs

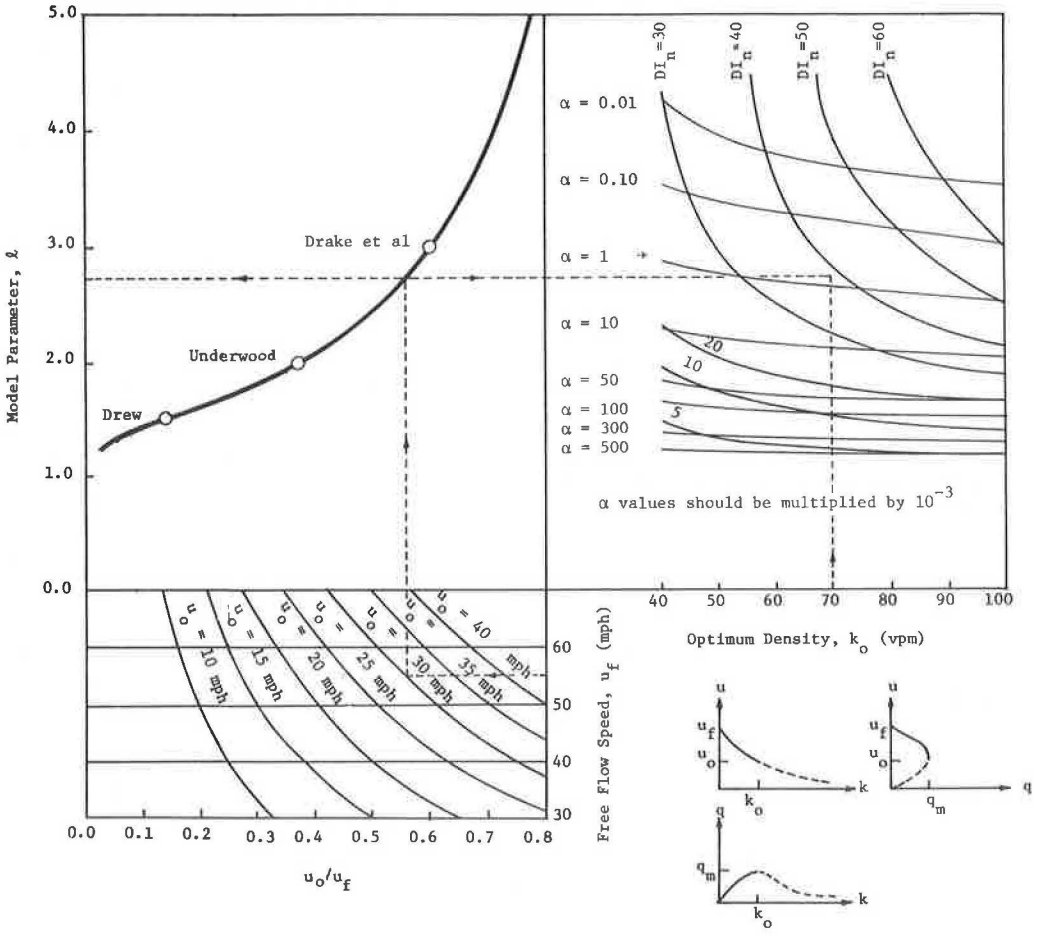
Once the theoretical relations between model parameters and traffic-flow criteria were established, as described above, these relations were translated into a practical tool, and generalized nomographs for non-congested-flow and congested-flow regimes were developed to graphically represent the theoretical aspects involved.

For the non-congested-flow regime ($\ell > 1$, $m = 1$), Figure 10 shows a generalized nomograph that relates model parameters ℓ and α to the traffic-flow criteria u_f, u_o, k_o, and q_m. The nomograph encompasses the basic relation between ℓ and u_o/u_f (thick curve) and three sets of contours for u_o, α, and q_m. The basic relation (Equation 27) was established for values of u_o/u_f ranging from 0.02 to 0.78, which correspond to ℓ values of approximately 1.2-5. The locations of existing non-congested-flow models are shown on the curve. Note also that the curve has an inflection point at $u_o/u_f = 1/e^2$, which corresponds to $\ell = 1.5$ [Drew model (13)]. The contour lines for optimum speed u_o were established for values ranging from 10 to 40

Figure 9. Characteristics of generalized and existing two-regime models.

ELEMENT	NONCONGESTED FLOW REGIME ($\ell > 1$, $m=1$)				CONGESTED FLOW REGIME ($\ell=1$, $m < 1$)	
	GENERAL	DREW	UNDERWOOD	DRAKE ET AL	GENERAL	GREENBERG
Equation of State	$u = u_f e^{\frac{\alpha}{1-\ell} k^{\ell-1}}$	$u = u_f e^{-2(\frac{k}{k_o})^{0.5}}$	$\mu = \mu_f e^{-(\frac{k}{k_o})}$	$\mu = \mu_f e^{-\frac{1}{2}(\frac{k}{k_o})^2}$	$\mu^{1-m} = \alpha (1-m) \ln(\frac{k_j}{k})$	$\mu = \mu_o \ln(\frac{k_j}{k})$
Constant of proportionality	$\alpha = \frac{1}{k_o^{\ell-1}}$	$\frac{1}{\sqrt{k_o}}$	$\frac{1}{k_o}$	$\frac{1}{k_o^2}$	μ_o^{1-m}	μ_o
Parameter ℓ	$\ell = 1 - \frac{1}{\ln(\frac{\mu_o}{\mu_f})}$	1.5	2	3	1	1
Parameter m	$m = 1$	1	1	1	$1 + \frac{1}{\ln(\frac{k_o}{k_j})}$	0
Optimum speed	$\mu_o = \mu_f e^{-\frac{1}{\ell-1}}$	$\frac{u_f}{e^2}$	$\frac{\mu_f}{e}$	$\frac{\mu_f}{\sqrt{e}}$	$\alpha^{\frac{1}{1-m}}$	α
Optimum density	$k_o = \alpha^{-\frac{1}{\ell-1}}$	$\frac{1}{\alpha^2}$	$\frac{1}{\alpha}$	$\frac{1}{\sqrt{\alpha}}$	$k_j e^{-\frac{1}{1-m}}$	$\frac{k_j}{e}$
Capacity	$q_m = k_o \mu_f e^{-\frac{1}{\ell-1}}$	$\frac{k_o u_f}{e^2}$	$\frac{k_o \mu_f}{e}$	$\frac{k_o \mu_f}{\sqrt{e}}$	$\mu_o k_j e^{-\frac{1}{1-m}}$	$\frac{\mu_o k_j}{e}$

Figure 10. Generalized nomograph for non-congested-flow regime ($\ell > 1$, $m = 1$).



miles/h. As shown, the free-flow speed u_f associated with these contours ranges from 30 to 70 miles/h.

Contours for the constant of proportionality α were established based on Equation 24, which relates α to both ℓ and k_o (k_o ranges from 40 to 100 vehicles/mile). As noted, values of α contours range from 10^{-5} to 0.5. The final set of contours are those related to q_m. To establish contours representing q_m, a variable DI_n was introduced (based on Equation 28). This variable is defined as follows:

$$DI_n = q_m/u_f = k_0 e^{-[1/(\ell-1)]} \quad (36)$$

From Equation 36, it should be noted that DI_n is a function of ℓ and k_o. By using this equation, contours for DI_n were established for values ranging from 5 to 60. As will be described later, these contours are used to ensure that the criteria range for q_m is satisfied.

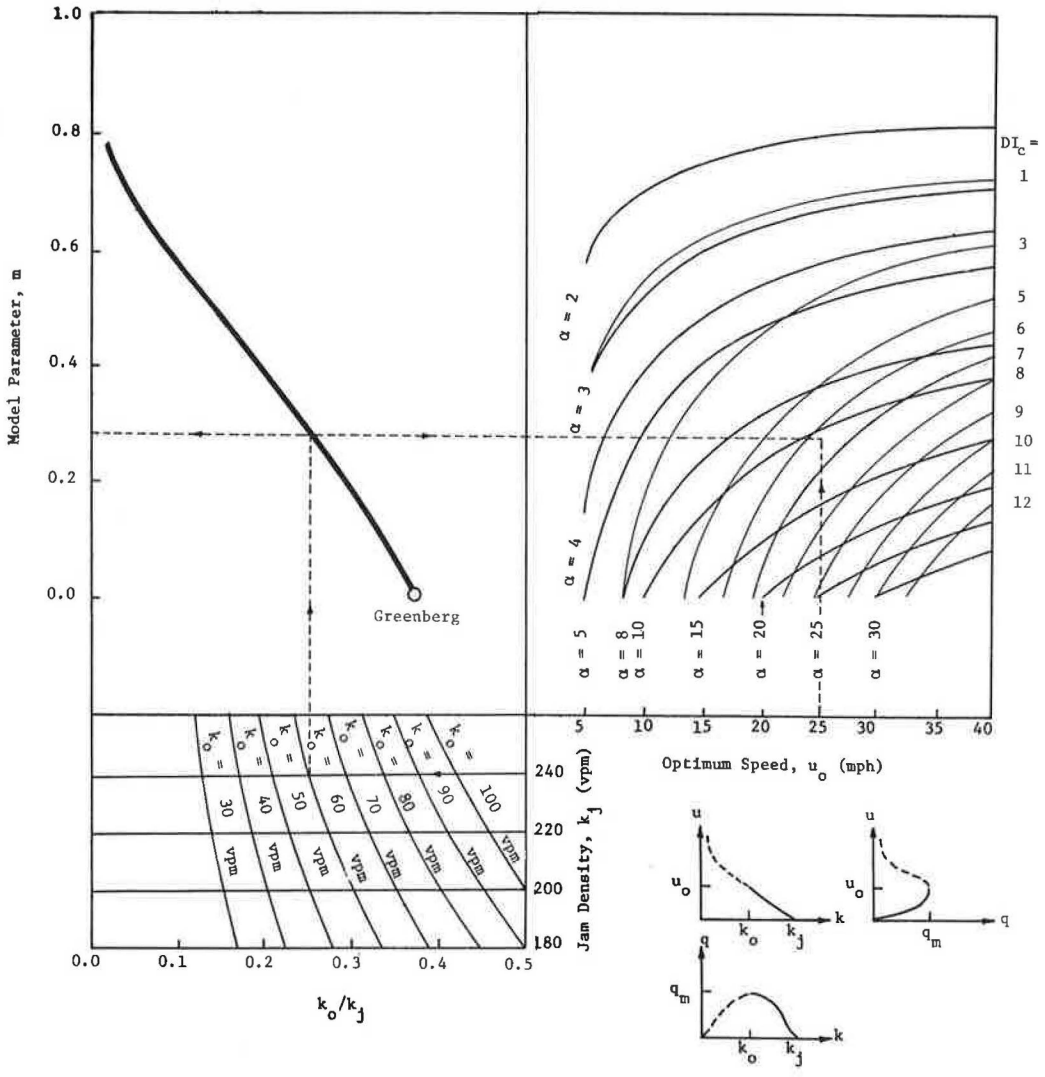
To illustrate the use of the nomograph, let us consider an example. Given that u_f = 55 miles/h, u_o = 30 miles/h, and k_o = 70 vehicles/mile, one must determine the corresponding model parameter ℓ and constant of proportionality α by performing the following steps (Figure 9):

1. Enter at u_f = 55 miles/h and draw a horizontal line that intersects with the contour corresponding to u_o = 30 miles/h.
2. From that point, draw a vertical line that intersects with the basic (thick) curve.
3. At the intersection point, draw a horizontal line (to the left) and read the value of ℓ = 2.6. Extend this horizontal line to the right.
4. Enter at k_o = 70 vehicles/mile and draw a vertical line that intersects with the horizontal line in step 3 above. At the intersection point, read the value of $\alpha = 0.9 \times 10^{-3}$.
5. Check the maximum flow value by reading the value of DI_n at the intersection point and multiplying it by u_f to determine q_m. From the diagram, DI_n = 39; therefore, q_m = 2145 vehicles/h (39 x 55). Obviously, in this simple example q_m can be directly calculated from input values of k_o and u_o as $k_o u_o$ = 2100 vehicles/h.

Having determined the values of ℓ = 2.6 and $\alpha = 0.9 \times 10^{-3}$, one can now define the steady-state flow equations. For example, the speed-density relation (Equation 21) can be described as follows:

$$u = 55 \exp(-0.56 \times 10^{-3} k^{1.6}) \quad (37)$$

Figure 11. Generalized nomograph for congested-flow regime (ℓ = 1, m < 1).



A similar nomograph was developed for the congested-flow regime ($\ell = 1$, $m < 1$) and is shown in Figure 11. The nomograph represents the relations between model parameters m and α and traffic-flow criteria k_j, u_o, k_o, and q_m and encompasses the basic relation of m and k_o/k_j and three sets of contours for k_o, α, and q_m. The basic relation (thick curve) is based on Equation 34. The curve has an inflection point at $k_o/k_j = 1/e^2$, which corresponds to $m = 0.5$. In addition, the Greenberg model is located at a point corresponding to $m = 0$. It should be noted that the curve is not extended beyond the Greenberg model for values of $m < 0$. This was used because such negative values have the effect of shifting the speed variable in the sensitivity term of Equation 4 from the numerator to the denominator, which is considered undesirable.

Contours for k_o and α for the congested-flow regime were established in a way similar to that for the non-congested-flow regime. In addition, to establish contours representing q_m, a variable DI_c was introduced (based on Equation 35). This variable is defined as follows:

$$DI_c = q_m/k_j = u_o e^{-[1/(1-m)]} \quad (38)$$

DI_c, which is similar to DI_n for the non-congested-flow regime, is used to ensure that the criteria range for q_m is satisfied. The use of the nomograph is similar to that for the non-congested-flow regime. Figure 11 further illustrates the use of the nomograph. In this example, the input values of the traffic-flow criteria are $k_j = 240$ vehicles/mile, $k_o = 60$ vehicles/mile, and $u_o = 25$ miles/h. By using these values, the reader can ascertain the value of $m = 0.28$ and $\alpha = 10.5$. Substituting these values into the speed-density equation (Equation 29), for example, yields

$$u = 16.60\,[\ln(240/k)]^{1.39} \quad (39)$$

Establishing the Feasible Region

To this point, the procedures outlined above are intended for use in cases where the traffic-flow criteria are established as single values. Knowledge of the feasible region of model parameters is important for two-regime, as for single-regime, models. Such a feasible region is obtained when the traffic-flow criteria are established as ranges. The procedure for such cases can be described as follows.

The use of the various steps involved in establishing the feasible regions for the non-congested-flow and congested-flow regimes can be illustrated by means of an example. Suppose that, based on a given set of traffic-flow data, one has established the ranges of the traffic-flow criteria for the non-congested-flow and congested-flow regimes as follows (k_j and u_f are established as single values):

Criterion	Non-Congested-Flow Regime	Congested-Flow Regime
k_j (vehicles/mile)	-	250
u_f (miles/h)	46	-
k_o (vehicles/mile)	80-90	70-80
u_o (miles/h)	15-25	15-20
q_m (vehicles/h)	1450-1550	1300-1400

One must now determine model parameters that satisfy these criteria.

For the non-congested-flow regime, the feasible region of model parameters ℓ and α that satisfy

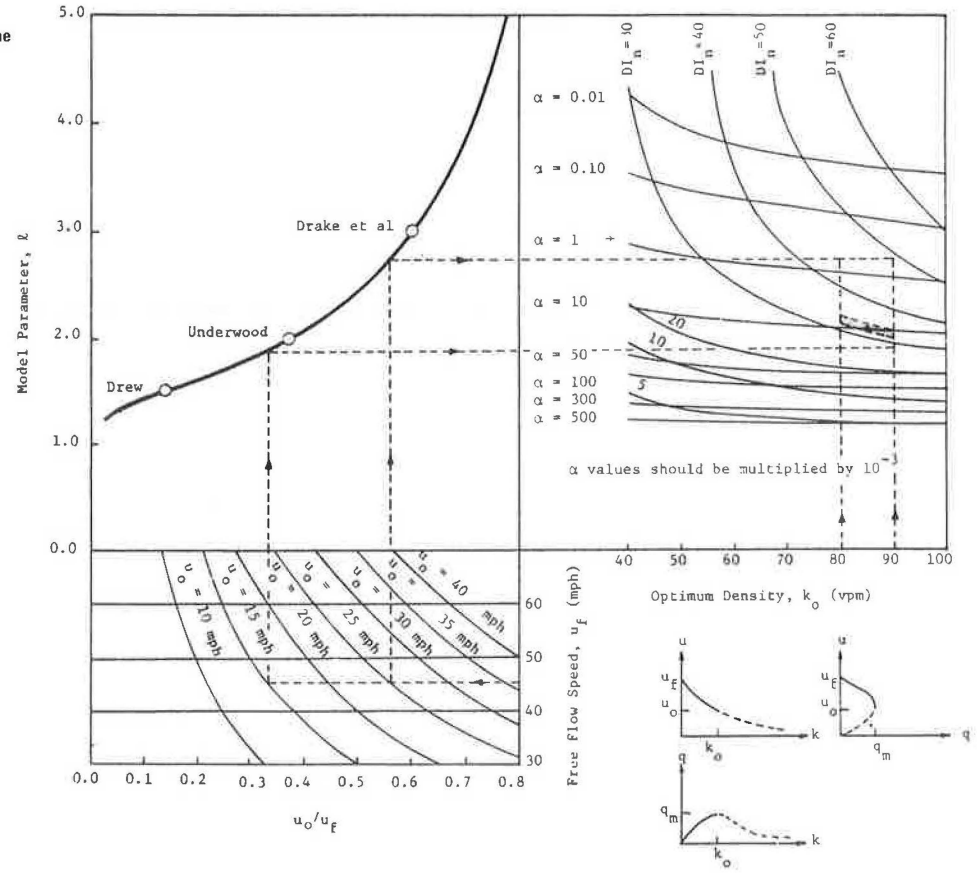


Figure 12. Establishing the feasible region for the non-congested-flow regime ($\ell > 1$, $m = 1$).

given traffic-flow criteria can be established as shown in Figure 12. The procedure consists of the following basic steps:

1. Draw a horizontal line corresponding to u_f = 46 vehicles/mile. This line intersects with the two contour lines that correspond to the range of u_o (15 and 25 miles/h).
2. From the intersection points, draw two vertical lines that intersect with the basic (thick solid) curve at two points.
3. From these points, draw two horizontal lines to the right.
4. At the limits of the k_o range (80 and 90 vehicles/mile), draw two vertical lines that intersect with the two horizontal lines drawn in step 3 above. The rectangular region defined by these lines contains model parameters that satisfy traffic-flow criteria u_f, u_o, and k_o.
5. Finally, establish the appropriate DI_n contours that correspond to the criteria range of u_f and q_m. In general, the lower and upper limits of DI_n can be calculated (based on Equation 36) as follows:

$$DI_n(\text{lower}) = q_m(\text{lower})/u_f(\text{upper}) \quad (40)$$

$$DI_n(\text{upper}) = q_m(\text{upper})/u_f(\text{lower}) \quad (41)$$

where "lower" and "upper" on the right-hand side refer to the limits of established ranges of the traffic-flow criteria q_m and u_f.

By using Equations 40 and 41 and noting that u_f (lower) = u_f (upper) = 46 miles/h, it can be found that DI_n (lower) = 32 and DI_n (upper) = 34. The DI_n contours corresponding to these values are drawn in Figure 11. The area between these contours that overlaps with the previously determined rectangular region defines the feasible region of model parameters. Clearly, values of ℓ and α within that region satisfy all four traffic-flow criteria: u_f, u_o, k_o, and q_m. It should be noted that the area below the DI_n (lower) contour contains models with q_m < 1450 vehicles/h, whereas the area above the DI_n (upper) contour contains models with q_m > 1550 vehicles/h. In addition, for any model in the feasible region, one can immediately define the associated traffic-flow criteria. As an example, for the model corresponding to ℓ = 2.05 and $\alpha = 10^{-2}$, the associated criteria of k_o and u_o are determined as 85 vehicles/mile and 18 miles/h, respectively. For these values, the speed-density relation (Equation 21), for example, can be described as follows:

$$u = 46 \exp(-9 \times 10^{-3} k^{1.05}) \quad (42)$$

For the congested-flow regime, the feasible region is established in a way similar to that for the non-congested-flow regime. The feasible region is established as shown in Figure 13 by using the traffic-flow criteria given in the in-text table above. It should be noted that contours for DI_c

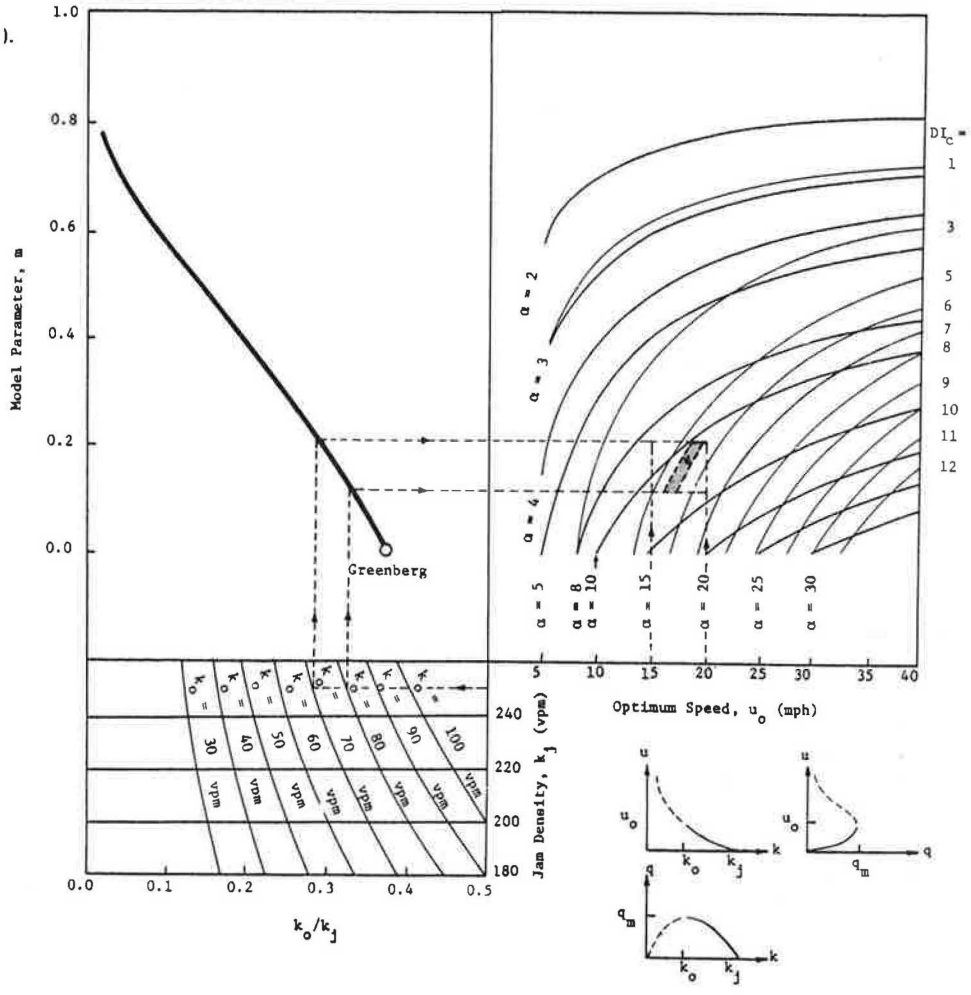


Figure 13. Establishing the feasible region for the congested-flow regime (ℓ = 1, m < 1).

are generally established (based on Equation 39) as follows:

$$DI_c(\text{lower}) = q_m(\text{lower})/k_j(\text{upper}) \quad (43)$$

$$DI_c(\text{upper}) = q_m(\text{upper})/k_j(\text{lower}) \quad (44)$$

By using these equations and noting that k_j (lower) = k_j (upper) = 250 vehicles/mile, one can find that DI_c (lower) = 5.2 and DI_c (upper) = 5.6. It should be noted that, for any model in the feasible region, the associated traffic-flow criteria can be immediately defined. For example, for the model corresponding to m = 0.19 and α = 10.5, the associated traffic-flow criteria are u_o = 18 miles/h, k_o = 75 vehicles/mile, and k_j = 250 vehicles/mile. For these values, the speed-density relation (Equation 29), for example, can be described as follows:

$$u = 14.03\,[\ln(250/k)]^{1.24} \quad (45)$$

It should be remembered that the feasible regions of model parameters established for the non-congested-flow and congested-flow regimes are intended for use in cases in which a range of models representing the traffic-flow data is of interest. In addition, u_f and k_j were specified as single values in an attempt to simplify the procedure. However, these criteria can be specified as ranges, in which case the estimation procedure would involve some additional computations to determine the associated traffic-flow criteria.

CONCLUSIONS

This paper presents a generalized procedure for estimating single- and two-regime traffic-flow models. The procedure is based principally on the theoretical relations between model parameters and traffic-flow criteria. Emphasis has been given to translating the theoretical aspects into practical analysis tools. Generalized nomographs developed for both modeling approaches are capable of directly providing the user with the values, or the feasible region, of model parameters that satisfy specified traffic-flow criteria.

Based on this research work, a few important observations can be made:

1. The input to the nomograph procedure is rather simple. It includes the traffic-flow criteria, which can be established from traffic-flow data for a particular facility. The output of the nomograph includes model parameters that satisfy these criteria. Clearly, if the traffic-flow criteria are carefully selected, the resulting model is likely to provide a reasonable representation of the data. Similarly, a good selection of the criteria ranges would result in a feasible region of model parameters that is more representative of the data characteristics.
2. The nomograph procedure should be viewed as complementary to rather than as a substitute for the existing regression analysis procedure for model estimation. The nomograph procedure is intended for use in situations in which a reasonable estimation of model parameters would suffice. When a relatively high degree of accuracy is required, the regression analysis procedure should be used.
3. The nomograph procedure appears to represent a powerful and flexible estimation tool that enables the analyst to adjust the evaluation criteria and to directly determine their effect on the evaluation results.
4. It should be emphasized that the nomograph procedure is based solely on theoretical aspects and so is general in nature; it is not a site-specific procedure. As a consequence, it appears to be of particular value for a wide range of transportation applications.

Future research work should be devoted to the following areas:

1. The nomograph procedure appears to provide a basis for the development of a facility design index for various highway types. Such an index would characterize highway facilities by specific combinations of model parameters. To this end, the variables DI, DI_n, and DI_c used in this paper could be further investigated by using real traffic-flow data.
2. Guidelines for the selection of single- or two-regime approaches under varying operating conditions should be developed. Future work to investigate alternative methods of modeling the two-regime approach is required.
3. A similar nomograph procedure for the remaining portions of the matrix of macroscopic models should be developed.

REFERENCES

1. A Reuschel. Fahrzeugbewegungen in der Kolonne. Oesterreichischen Ingenieurarchiv, Vol. 4, 1950, pp. 193-215.
2. L.A. Pipes. An Operational Analysis of Traffic Dynamics. Journal of Applied Physics, Vol. 24, 1953, pp. 274-281.
3. R.E. Chandler, R. Herman, and E. Montroll. Traffic Dynamics: Studies in Car-Following. Operations Research, Vol. 7, 1958, pp. 165-184.
4. D.C. Gazis, R. Herman, and R. Potts. Car-Following Theory of Steady-State Traffic Flow. Operations Research, Vol. 7, 1959, pp. 499-595.
5. D.C. Gazis, R. Herman, and R.W. Rothery. Nonlinear Follow-the-Leader Models of Traffic Flow. Operations Research, Vol. 9, 1961, pp. 545-567.
6. A.D. May and H.E. Keller. Non-Integer Car-Following Models. HRB, Highway Research Record 199, 1967, pp. 19-32.
7. A. Ceder. Investigation of Two-Regime Traffic Flow Models at the Micro- and Macroscopic Levels. Univ. of California, Berkeley, Ph.D. dissertation, Jan. 1975.
8. L.C. Edie. Car Following and Steady-State Theory for Noncongested Traffic. Operations Research, Vol. 9, 1961, pp. 66-76.
9. S.M. Easa. Generalized Procedure for Estimation of Single-Regime Traffic Flow Models. Institute of Transportation Studies, Univ. of California, Berkeley, 1979.
10. A. Ceder and A.D. May. Further Evaluation of Single- and Two-Regime Traffic Flow Models. TRB, Transportation Research Record 567, 1976, pp. 1-15.
11. B.D. Greenshields. A Study in Highway Capacity. HRB, Proc., Vol. 14, 1934, p. 468.
12. A.D. May and H.E. Keller. Evaluation of Single- and Two-Regime Traffic Flow Models. Proc., 4th International Symposium on the Theory of Traffic Flow, Karlsruhe, Federal Republic of West Germany, 1968.
13. D.R. Drew. Deterministic Aspects of Freeway Operations and Control. HRB, Highway Research Record 99, 1965, pp. 48-58.
14. R.T. Underwood. Speed, Volume, and Density Relationships: Quality and Theory of Traffic Flow. Yale Bureau of Highway Traffic, Yale Univ., New Haven, CT, 1961, pp. 141-188.

15. J.S. Drake, J.L. Schofer, and A.D. May, Jr. A Statistical Analysis of Speed-Density Hypotheses. Expressway Surveillance Project, Chicago, Rept. 16, May 1965.
16. H. Greenberg. An Analysis of Traffic Flow. Operations Research, Vol. 7, 1959, pp. 79-85.

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

Projected Vehicle Characteristics Through 1995

WILLIAM D. GLAUZ, DOUGLAS W. HARWOOD, AND ANDREW D. ST. JOHN

The U.S. Department of Transportation has established fuel-consumption standards for passenger vehicles and light trucks that will result in increasingly fuel-efficient vehicles in the future. Projections of characteristics of the mix of vehicles on the road that can be expected to change as a result of industry compliance with the standards are presented through 1995, based on a variety of government and industry publications. The average mass (weight), power, and engine size of passenger vehicles—including light trucks with a maximum gross vehicle weight of 3860 kg (8500 lb)—will obviously decrease during this period. Fuel economy will continue to improve steadily, and average acceleration performance will not decline appreciably after the 1983-1985 period. The characteristics of recreational vehicles will change, mostly in the next few years. All of these changes in on-the-highway averages will be brought about through "replacement" of heavy and high-performance vehicles by others of more modest weights and powers rather than through the introduction of very small or low-performance vehicles, which will lead to a more homogeneous vehicle population.

The fuel embargo of 1973 and 1974 and the spot fuel shortages of the summer of 1979 have aroused wide public reaction and contributed to a change in consumer buying habits. Vehicle purchasers are, on the average, seeking more fuel-conserving cars. In response to this demand and to U.S. Department of Transportation (DOT) mandates, the automobile industry is gradually changing its fleet mix to produce vehicles that generally have better fuel-consumption characteristics. This is being accomplished primarily through size and weight reductions as well as a shift to smaller engines (with accompanying performance impacts). To improve overall efficiency, other changes in vehicle technology are also being introduced.

It is of interest to project the long-range impact of these changes on vehicle operations, traffic safety, and overall fuel consumption. To do this, it is first necessary to predict the distributions of the characteristics of vehicles that will be on the road in future years. This prediction process and the results obtained are the subject of this paper. The process assumes an orderly progression of changes based on present rule making and associated projections. It does not consider possible catastrophic events, such as curtailment of automobile production, cessation of fuel imports, or imposition of fuel rationing. The projected characteristics can then be used in analyses or models to estimate impacts of interest.

This paper deals with two basic vehicle categories: passenger and recreational vehicles. The first category includes American and imported automobiles as well as all light trucks (e.g., pickups and vans) with a gross vehicle weight (GVW) of as much as 3860 kg (8500 lb). Recreational vehicles include motor homes, pickup campers, and passenger-vehicle/trailer combinations.

POPULATIONS OF PASSENGER VEHICLES

The aim of this study was to estimate the average characteristics of vehicles that will be on the road in future years as well as the distributions around the averages. The estimation process required, first, breaking down each year's sales into identifiable vehicle categories, each described in terms of such factors as weight, engine size and power, and production. Then all of the sales over the 15-year period prior to the year of interest were accumulated. This accumulation process accounted for the scrappage rates of the vehicles as well as the decreasing annual mileage with age. Finally, averages and other quantities were determined on a travel-weighted basis (that is, vehicles driven more kilometers in the year of interest counted more heavily in the averaging process). Thus, the averages and distributions should be representative of what one would find by measuring all vehicles passing a given location. The assembly process, which involved summing over 3000-4000 identifiable vehicle categories, was made feasible by using specially written computer programs.

For convenience, passenger vehicles were generally divided into three groups: American cars, foreign cars, and light trucks. Then detailed vehicle characteristics were assembled only for selected model years (because of the rather painstaking process required). The characteristics for intervening years were estimated by the computer program, by use of interpolation. The following subsections provide more detail about the assembly process.

Data on Vehicle Characteristics

The most important determinant of acceleration performance is the ratio of a vehicle's net engine power to its mass [commonly, but imprecisely (from a technical viewpoint), called its weight]. Other characteristics, such as transmission and axle ratios, frontal areas, and aerodynamic drag coefficient, also have an effect. Unfortunately, these latter characteristics are not generally available other than on a special-case basis. Therefore, performance capability was estimated solely on the basis of power-to-mass ratio. More specifically, for each vehicle model identified, the maximum net power of the engine and an appropriate vehicle mass (weight) were recorded. For automobiles, this was taken as the curb weight (empty vehicle weight plus fuel and coolant) plus a

driver and a passenger--e.g., curb weight plus 140 kg (300 lb). It was assumed that light trucks are usually more heavily loaded than automobiles; a load of approximately 230 kg (500 lb) was therefore assigned for these vehicles. Projections for future American automobiles are available from the National Highway Traffic Safety Administration (NHTSA) (1). For these vehicles, the "inertial weights", which are roughly equivalent to curb weights plus 140 kg, were given, and the vehicles were placed in inertial weight-class intervals of 114 kg (250 lb). Thus, less precision is available regarding these vehicles.

Estimates of fuel consumption are more complex than estimates of vehicle performance. Some studies (2) use the concept of an engine map (a representation of specific fuel consumption as a function of engine speed and engine power output) to estimate total fuel consumption. Unfortunately, engine maps are available for only relatively few current engines. In this study, only those vehicle characteristics that have the strongest effect on fuel consumption and are readily available for nearly all vehicles were considered. These characteristics are loaded weight and engine displacement. Other vehicle parameters that have an effect on fuel consumption and are being improved from year to year (such as reduced frictional losses and other means of increasing efficiencies) could be incorporated in an approximate fashion based on model year.

Assembly of Vehicle Data by Model Year

Detailed data on vehicle characteristics were assembled for the model years 1967, 1971, 1974, 1977, and 1981-1985. Vehicle data for the 1967 model year were taken from a previous research report by St. John and Glauz (3), which analyzed available data for that year and presented the resulting vehicle distributions. That work provided 19 vehicle categories, each with a different power-to-mass ratio and corresponding sales volume. A typical vehicle weight for each class was taken from Table E-1 of that study. All powers quoted at that time were gross powers. A regression analysis performed in another study (4) indicated that the net engine power is approximately 0.746 times the gross power. This correction was made to all 1967 values.

Data for the 1971 model year had already been analyzed in a previous project (4). The results of that analysis yielded a distribution of performance characteristics represented by 31 vehicle categories. This distribution also required that gross powers be converted to net powers by using the multiplicative factor 0.746.

Data for the 1974 and 1977 model years were assembled by using a similar process. Data for American automobiles, foreign automobiles, and light trucks were compiled separately. The 1977 American automobile population, for example, was determined largely from the April 1977 Automotive Industries annual report, which gives, for each model of each manufacturer (e.g., American Motors Pacer, Chrysler LaBaron, and Ford LTD-II), the number of each of the optional engine sizes installed. The report also gives engine displacement, net power, and other details for each engine. Vehicle masses were obtained from NHTSA documentation (5). Some judgment and approximation were required in some instances. For example, the same basic engine size [e.g., 5752 cm^3 (351 in^3)] was often available at more than one power rating as a result of such factors as different carburetors. Production data for subdivisions within models (e.g., station wagons) are not quoted.

American automobile data for 1974 were assembled in essentially the same way as the 1977 data except that engine sales were not given at as fine a level of detail. Where this detail was lacking, the 1977 figures were used as a guide to apportion the optionally available engines to each vehicle model. The April 1975 Automotive Industries report (and, thus, the 1975 engine data) was used to determine the net powers because the figures quoted in 1974 were gross powers.

The foreign-automobile distributions were also determined by incorporating data from the 1974, 1975, and 1977 Automotive Industries reports. Automotive Industries, however, provides sales information for only the 10 leading foreign manufacturers, which in 1977 accounted for only 83 percent of U.S. sales of foreign vehicles. Therefore, NHTSA documentation (6) was used to supplement this information. Vehicles for which sales were extremely small, such as Rolls Royce, Lotus, and Ferrari, were not included in the final tabulation.

NHTSA documentation includes data on the sales of 1976-1977 light trucks by manufacturer and by type of light truck (7). Light trucks include pickups, vans and panel trucks, general utility trucks (such as jeeps), station wagons built on a truck chassis (such as the American Motors Cherokee), and other types (such as light platform or stake trucks). The April 1977 Automotive Industries report provides some engine-sales distributions for each manufacturer's light trucks by model. Additional data were obtained from U.S. Department of Energy (DOE) documentation of data obtained from NHTSA (8). Where sales data were not explicitly provided, it was necessary to augment these data by judgment or assumption to split sales figures among options. Loaded-vehicle weights were obtained from a recent issue of the Official Used-Car Guide of the National Automobile Dealers Association (9). Light-truck data for 1974 were assembled by following a similar process.

Rather complete sales projections and characteristics of 1981-1985 American automobiles have been assembled by NHTSA (1). NHTSA had analyzed four alternative projections. In agreement with their final decision regarding rule making for vehicle fuel-consumption requirements, the projection for "alternative 2" was used. This alternative includes predicted weight reductions resulting from body redesign and substitution of materials, a selection of engines with efficient spark ignition, technological improvements to spark-ignition engines, the addition of a torque lockup clutch and fourth gear to automatic transmissions, and improvements in other areas such as lubricants and accessories. Alternative 2, however, does not include any significant market penetration of diesel engines or additional penalties attributable to emission controls. The figures given were all based on the 1971-1975 average sales and were adjusted to agree with the total annual sales projected by NHTSA for each of the years from 1981 through 1985 (6).

Only limited information is available on 1981-1985 projections for sales of foreign cars (6) and light trucks (7,10). Procedures were devised to quantify expected trends in the weights, powers, and engine sizes of these vehicles (11).

In agreement with the NHTSA rule-making assumptions, we projected that there would be no further changes in model-year vehicle characteristics or sales mixes after 1985. Thus, the detailed 1985 data were simply adjusted to obtain future total vehicle sales as estimated by NHTSA (sales were assumed to increase at a basic rate of 2 percent/year).

Combining the Assembled Data

After assembling the detailed data for certain model years, a computer program interpolated for the intervening model years and extrapolated for model years prior to 1967 and after 1985. Then the annual vehicle travel distance for each vehicle category (i.e., make, model, year, engine size, etc.) was calculated based on the number of vehicles sold, the fraction not yet scrapped by the year of interest, and the annual distance traveled by all vehicles of a given vintage. The program then computed the weighted characteristics previously described, ordered them in terms of increasing power-to-mass ratio, and printed the calculated values together with cumulative sums. From these data, averages and distributions could readily be determined.

Resulting Passenger-Vehicle Distributions

Figure 1 shows the expected average trends in passenger-vehicle characteristics through 1995. The average vehicle mass is projected to decline by more than 230 kg (500 lb), or 14 percent, during this time interval. Most of this decline will occur between 1981 and 1985, and very little change is projected after 1990. The average engine size is expected to decrease rather steadily between now and 1985--by more than 23 percent--and little thereafter. Likewise, average engine power should decrease almost 24 percent. Performance, as reflected in power-to-mass ratio, will only decrease about 15-16 percent, and most of this change will have occurred by 1983. Thus, although average weights and engine sizes will continue to decrease appreciably for about 10 or 12 years (leading to continuously improving fuel economy), the decrease in average performance will occur much quicker, and the most noticeable effects should be felt within the next 5 or 6 years.

Traffic operations are expected to be affected not just by average vehicle characteristics but also, and more importantly, by differing characteristics among road users. Thus, the distribution of characteristics is of extreme interest. The distributions of performance capabilities, as indicated by power-to-mass ratio, are shown in Figure 2 for the years 1978, 1981, 1985, and 1995. Clearly, the major change during this period will be a decrease in the fraction of moderate and high-performance vehicles; relatively little increase will be observed in the percentage of vehicles that have very poor acceleration capabilities. In other words, the spread in vehicle performance should lessen in future years, and this should lead to a more homogeneous mix of passenger vehicles.

For purposes of comparison, Figure 2 also shows the distribution of power-to-mass ratio for the 1967 model year. This distribution is much broader than current or projected distributions. It includes a substantial fraction of vehicles that perform very poorly (e.g., the 1967 Volkswagen Beetle and Microbus, which had relatively small engines) as well as a large fraction of high-performance vehicles [those with 6600-cm^3 (400-in^3) and larger displacement engines]. This comparison suggests that the most pronounced changes in the distribution of the performance characteristics of passenger vehicles may, in fact, have occurred prior to 1978.

It is often convenient, in making projections or performing analyses, to deal with several discrete vehicle "types"--each of which has a prescribed set of characteristics--rather than with the total distribution. As an example, assume that the total distribution is represented by five vehicles, of which the first type represents the 10 percent of vehicles on the road that have the poorest acceleration capabilities and the others simulate the next 15, 20, 25, and 30 percent, respectively, according to performance. The power-to-mass ratios and other characteristics for these vehicle types are given in Table 1 for selected years between 1978 and 1995. Although performance capabilities of each type decrease with time, the most pronounced changes are for the higher-performance types of vehicles (31 and 4 percent decreases for the highest- and lowest-performance types, respectively).

The average mass (weight) for the lowest-performance category remains rather stable, but the mass of the next category increases. This category contains a substantial fraction of pickup and panel trucks, the sales of which are projected to increase (and have increased) more rapidly than those of other passenger vehicles. The masses (weights) of the remaining vehicle types will all decrease, especially between 1978 and 1985. These types are predominantly American automobiles.

The data given in Table 1 also show the decrease in engine displacements. Again, the effect of light trucks on the second performance category is obvious. A relative stability after 1985 is also evident.

The previous projections are, in a sense, somewhat conservative regarding vehicle acceleration performance. They assume, in agreement with NHTSA presumptions, no significant market impact for any but gasoline-powered engines. Yet some scientists anticipate a marked change in power plants, including diesels, electric vehicles, and hybrid vehicles (which have electric and internal-combustion engines). Indeed, diesel-powered automobiles are already being extensively promoted by one domestic manufacturer (General Motors), as well as by at least three foreign firms (Mercedes-Benz, Volkswagen, and Peugeot). An alternative (more extreme) estimate of vehicle performance characteristics might be made by projecting extensive market-share impacts by these other types of vehicles.

A recent report (12) contains results of rather extensive modeling of the markets for postulated future vehicles. In addition to gasoline-powered vehicles, several other alternatives are considered. The "most likely" results, adapted from that report, are given in Table 2.

Two important conclusions can be drawn from these projections:

1. Diesel engines will become a substantial fraction (28 percent) of total passenger vehicles, which definitely contradicts the NHTSA assumption of no significant market impact by diesels.
2. The dominant electric vehicles that are projected will have much greater performance capabilities than current electric vehicles.

The projections imply that there will be substantial fractions of vehicles with new power plants, whose average power-to-mass ratios will approximate the calculated 1995 average for (predominantly) gasoline-powered vehicles. Thus, these new projections (12) would not greatly change the performance characteristics given earlier.

It is informative to examine traffic characteristics under hypothesized "worst-case" performance distributions. The report by Train (12) was critically examined for this purpose. Train's projected power-to-mass average of 49-W/kg (0.03-hp/lb) powered vehicles seems rather high compared with that for current vehicles. The three foreign imports range from 30 to 35 W/kg (0.018-0.021 hp/lb); the ratio for the Oldsmobile diesel is about 43 W/kg

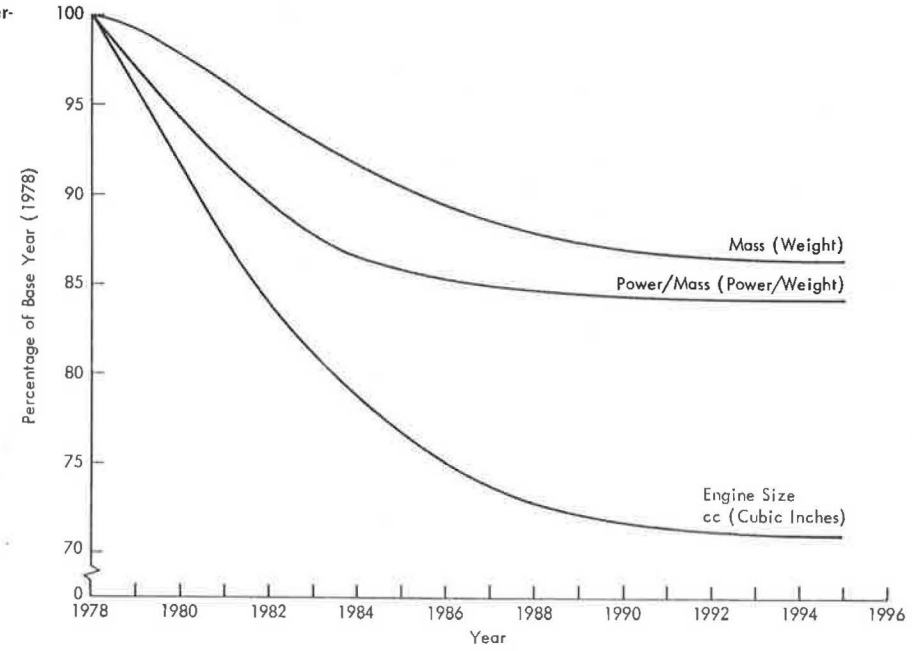


Figure 1. On-highway average passenger-vehicle characteristics.

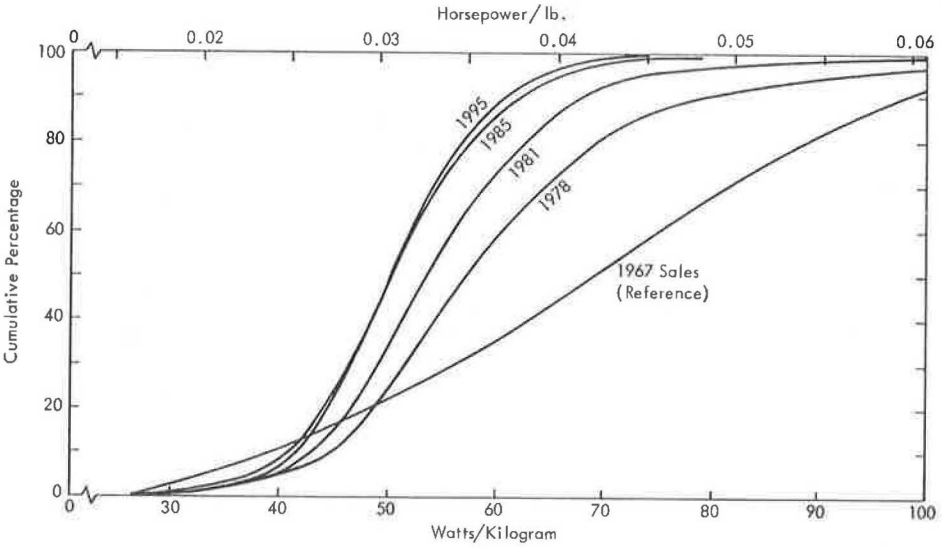


Figure 2. Distribution of power-to-mass ratios of passenger-vehicle populations, accumulated by vehicle kilometers of travel.

Table 1. Characteristics of representative vehicles.

Year	Performance Category[a] (%)	Power-to-Mass (W/kg)	Mass (kg)	Engine Displacement (cm^3)
1978	0-10	39.0	1511	2606
	10-25	47.8	1638	3557
	25-45	52.9	1754	4278
	45-70	59.5	1785	4851
	70-100	77.8	1888	6704
1985	0-10	37.8	1499	2606
	10-25	43.9	1761	4032
	25-45	47.8	1547	3458
	45-70	52.0	1494	3328
	70-100	61.2	1651	4474
1995	0-10	37.3	1500	2622
	10-25	43.9	1741	3852
	25-45	47.5	1445	3032
	45-70	51.6	1439	3229
	70-100	59.2	1536	4016

Note: 1 W/kg = 0.0006 hp/lb; 1 kg = 2.205 lb; 1 cm^3 = 0.061 in^3.

[a]Ranked according to increasing power-to-mass ratio.

Table 2. Alternative vehicle characteristics for 1995.

Power Plant	Mass (kg)	Power-to-Mass (W/kg)	Range (km)	Percentage Owned
Gasoline				
1	1360	50	—[a]	29.0
2	900	60	—[a]	41.0
Diesel	1500	50	—[a]	28.0
NiZn battery	420	16	80	0.01
Hi-temp battery				
1	1300	33	160	0.03
2	1900	50	240	2.0

Note: 1 km = 0.62 mile; 1 kg = 2.205 lb; 1 W/kg = 0.0006 hp/lb.

[a]Unlimited in the sense that the vehicles can be rapidly refueled as needed.

(0.026 hp/lb). Thus, it is not clear why such an increase is hypothesized. Furthermore, in addition to the many flaws that Train pointed out, his model did not impose any new constraints, economic or otherwise, on gasoline-powered vehicles; that is, it did not assume any increase in gasoline prices or any type of rationing (which many believe to be possible). The model therefore reflected little of an incentive to own electric vehicles that would balance the negative factors of limited range and hypothesized higher initial and operating costs.

In this study, to examine a more nearly worst-case situation, it was assumed that either 5 or 20 percent of the 1995 vehicles were new, low-powered automobiles with a pessimistic power-to-mass ratio of 33 W/kg (0.02 hp/lb). If these vehicles replace a uniform, random selection of the other vehicles, and not just a low-performance or high-performance segment of them, the performance capabilities of the representative 1995 vehicles would be as follows (1 W/kg = 0.0006 hp/lb):

	Power/Mass (W/kg)		
Performance Category (%)	Baseline	5 Percent Low Powered	20 Percent Low Powered
0-10	37.3	33.9	32.1
10-25	43.9	42.6	36.5
25-45	47.5	46.2	43.7
45-70	51.6	51.1	49.8
70-100	59.2	59.0	57.9

Based on these assumptions, the average power-to-mass ratio would decline modestly, from 50.5 W/kg (0.03 hp/lb) to 49.6 or 47.0 W/kg (0.029 or 0.028 hp/lb), but the performance range would increase.

Method of Modeling Vehicle Performance

The assembly process aggregated passenger vehicles into populations on the road by performance class. Conceptually, this process is complicated by the difficulty of defining the term "performance".

Performance is defined here as the acceleration capability (a) of a vehicle as a function of speed (v) on level pavement. One may model this function as a straight line:

$$a = a_0(1 - v/v_m) \quad (1)$$

where a_0 is the extrapolated acceleration at zero speed and v_m is the extrapolated speed at zero acceleration (conceptually, a "maximum" speed). However, this model is known to be inaccurate at low speeds; initial acceleration is substantially less than a_0 at v = 0 and then increases rapidly at low speeds to exceed the modeled values. In addition, at high speeds this model tends to slightly underestimate acceleration capability and maximum speed. After several alternative approaches were examined, this model was used, but it was fitted to best represent midrange speeds--say, 50-100 km/h (30-60 miles/h), the speed at which most maneuvers of interest occur.

A major problem is the lack of uniformly good source data. A number of organizations routinely test and report on vehicle performance. Unfortunately, their data vary greatly in quantity as well as in comparability. Even for vehicles that are nominally identical, widely disparate results are reported by different organizations, and these differences are usually consistent; some organizations' results will generally be higher than others. Data from various issues of Consumer Reports and Consumer's Research were emphasized. These data tend to be more conservative than some, but they are probably relatively typical of data on average vehicles driven (hard) by persons other than professional race drivers.

Not all current and past model-engine combinations have been tested, and obviously no performance data exist for future vehicles. Therefore, one must determine the a-versus-v curve for each combination by using available or projected vehicle characteristics, such as weight and power.

Research to date (3-5) suggests that the slope a_0/v_m is about the same for most vehicles--between 0.08 and 0.10/s. Moreover, little explanation can be found for variations between vehicles; the differences appear to be more dependent on test or analysis procedures than on readily identifiable vehicle parameters. A constant value (for all vehicles) of $a_0/v_m = 0.085/s$ is suggested, since this value is about at the midpoint for current vehicles, after one accounts for possible exaggeration in the test procedures used by some organizations.

The next step involves estimating a_0 as a function of vehicle characteristics. Preliminary evidence suggests that the best single parameter is the mass-to-power ratio, although better predictions could clearly be made by using additional quantities such as gear-train characteristics and aerodynamic drag. Unfortunately, since these data are not usually available, they cannot routinely be used. The most obvious model would be of the following form:

$$a_0 = \alpha + \beta(W/kg) \quad (2)$$

In fact, this is what was used in the NHTSA study (3).

A more general approach was used here. Since the quantity usually measured directly is time, not acceleration, a relation of the following form was postulated:

$$t = K(W/kg)^{\beta} \quad (3)$$

where t is the time required to accelerate from one speed to another [48.3-96.6 km/h (30-60 miles/h)]. This relation was made linear by using a logarithmic transformation:

$$\ln t = \ln K + \beta \ln(W/kg) \quad (4)$$

Linear regression techniques were used to evaluate K and β. The data base included 186 vehicles (1977 and 1978) reported on in various issues of Consumer Reports, Consumer's Research, Motor Trend, and Car and Driver. These data covered accelerations over different (but similar) speeds in the midspeed ranges of the vehicles. The results showed that the exponent β did not vary appreciably by publication or by speed (in this midspeed range). Moreover, β was not significantly different from the simple value -1. Thus, the following simplified expression is sufficiently accurate:

$$t = K(kg/W) \quad (5)$$

The linear regression analysis yielded K = 500.6 and a standard deviation of 6.1, which indicates a very good fit (simple Pearson correlation coefficient r = 0.83).

If the basic equation (Equation 1) is integrated over the speed range of 48.3-96.6 km/h (30-60 miles/h) and the above values of a_0/v_m and K are inserted, one obtains

$$a_0 = 1.14[(2 - e^{-42.55R})/(1 - e^{-42.55R})] \quad (6)$$

in meters per second squared and

$$v_m = a_0/0.085 \quad (7)$$

in meters per second, where R is the mass-to-power ratio. The resulting acceleration performance curves are shown in Figure 3.

POPULATIONS OF RECREATIONAL VEHICLES

The populations of recreational vehicles (RVs) were assembled by using a procedure similar to that used for passenger vehicles. The computer program described earlier was modified to match the characteristics of RVs to the characteristics of potential towing vehicles--the future passenger-car and light-truck populations.

The characteristics of RVs that are representative of the current population--including weight, engine power, frontal area, and aerodynamic drag coefficient--are given in Table 3. These data were based on surveys of RV weights recently conducted for NHTSA (13) and on published and unpublished information assembled by the Midwest Research Institute (3,4). The basic types of RVs considered were motor homes, slide-in camper boxes, camping trailers, and travel trailers. Two types each of motor homes and travel trailers were considered because the range of weights for these vehicles is much greater than that for camper boxes and camping trailers.

There are currently two opposing trends that will affect the size and weight of future RVs:

1. Owners are demanding more fully equipped vehicles, and the added features tend to increase the vehicle size and, especially, the total vehicle weight.
2. Faced with rising material costs and the public's interest in greater fuel economy, manufacturers are attempting to use lighter materials and improved designs to reduce vehicle weights.

Because the available data are insufficient to forecast either trend accurately, we have assumed that the trends will offset one another. Thus, Table 3 represents the characteristics of future as well as current RVs and RV components.

The performance of RVs, unlike that of passenger cars, must be based on two independent parameters: zero-speed acceleration (a_o) and "maximum" speed (v_m). These parameters can be determined from the combined characteristics of the RV (and its towing vehicle, if any). The zero-speed acceleration for RVs was determined from the power-to-mass ratio by using the relation previously developed for passenger cars. Greater dispersion about this relation was found for RVs than for passenger cars, but no justification for modifying the relation was found.

Maximum speed (in meters per second) was determined (4) as

$$v_m = 3.133 + 0.0977[1/(\alpha R)]^{1/3} \quad (8)$$

where

$\alpha = (\rho C_D A/2W) + C_{RV}$ ($s^2 m^2$),
ρ = atmospheric mass density = 1.226 kg/m^3,
C_D = aerodynamic drag coefficient (Table 3),
A = projected frontal area of vehicle (m^2) (Table 3),
W = combined weight of RV and towing vehicle (N),
C_{RV} = coefficient for rolling resistance = 6.7×10^{-6} for speed (m/s), and
R = mass-to-power ratio (the reciprocal of power-to-mass was used for purposes of simplification).

Certain simplifying assumptions were made in assembling the population of RVs on the road for any given year:

1. For camping trailers and travel trailers, only towing-vehicle models that represent more than 0.1 percent of total vehicle miles of travel in the year of interest were considered.
2. For slide-in camper boxes, only pickup-truck models that represent more than 0.01 percent of total vehicle miles of travel in the year of interest were considered.
3. A minimum power-to-mass ratio of 19.7 W/kg (0.012 hp/lb) was required for the combination of RV and towing vehicle, as a conservative estimate of the lower bound of the current RV population. Thus, it was assumed that future RV owners will not select a grossly underpowered towing vehicle.
4. A minimum "maximum" speed (v_m) of 72 km/h (45 miles/h) was required of the combination of RV and towing vehicle. This cutoff value was selected on the assumption that owners would not choose a

Figure 3. Acceleration performance curves for passenger vehicles.

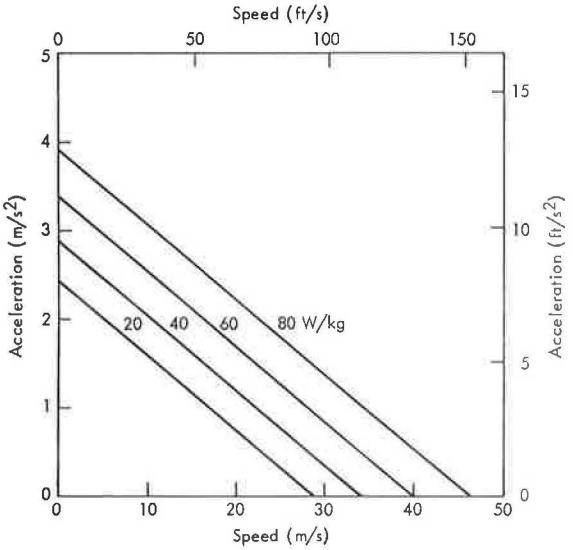


Table 3. Characteristics of representative RVs.

Type of Vehicle	Mass[a] (kg)	Power (kW)	Frontal Area (m^2)	Aerodynamic Drag Coefficient (C_D)[b]
Motor home				
1	4100	139	5.76	0.6
2	5700	167	5.76	0.6
Slide-in camper box	1400	—[c]	5.39	0.59
Camping trailer	1100	—[d]	2.60	1.1[e]-1.24[f]
Travel trailer				
1	1700	—[d]	5.95	0.59[e]
2	3000	—[d]	5.95	0.63[f]

Note: 1 kg = 2.205 lb; 1 kW = 1.34 hp; 1 m^2 = 10.76 ft^2.

[a] Including normal owner's payload.
[b] Based on the frontal area of the vehicle.
[c] Determined by the power of the light-truck population in the year of interest.
[d] Determined by the power of the passenger-car and light-truck population in the year of interest.
[e] When towed by a passenger car.
[f] When towed by a pickup.

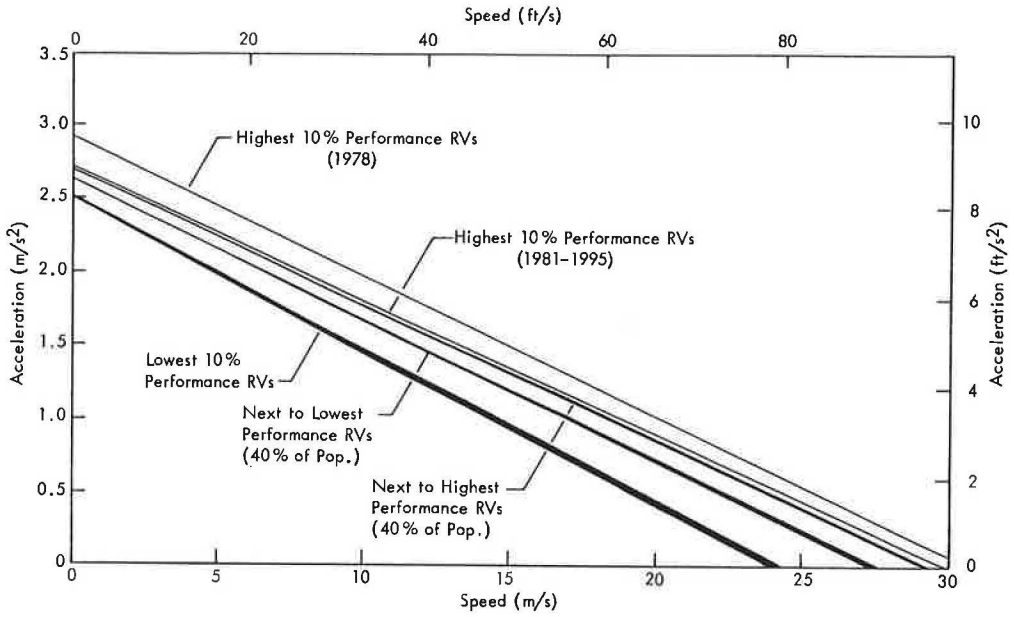


Figure 4. Acceleration performance curves for RVs.

towing vehicle that, on level terrain, could not maintain 72 km/h--currently a typical minimum speed limit on freeways. This limitation and limitation 3 above simply imply that owners will tend to make rational decisions (as they do now) from among available options.

Vehicle miles of travel for each RV combination in the population were estimated based on past sales history, future sales estimates, and survivability for the appropriate RV type and towing-vehicle model, together with average annual vehicle miles per vehicle for each RV type and towing-vehicle model. The distribution of vehicle performances for RVs as a function of the parameters a_o and v_m was then cross tabulated (11).

It was found that the characteristics of most RVs placed them roughly along a diagonal of the tabulation, from low a_o and v_m to high a_o and v_m. This feature of the performance distribution was used to select typical RVs that represent four performance ranges: the lowest-performance 10 percent of the population, the next-to-lowest 40 percent, the next-to-highest 40 percent, and the highest-performance 10 percent of the population. Weighted averages for a_o and v_m were then computed within each RV performance range to represent typical vehicle performance in that range. This was done for five years of interest: 1978, 1981, 1985, 1990, and 1995. The resulting performance capabilities of typical RVs in those years are summarized in Figure 4.

It is apparent in Figure 4 that there is no substantial impact on RV performance characteristics over time except for the highest-performance 10 percent of the RV population, the performance of which drops markedly between 1978 and 1981 and then does not change significantly. The performances of the vehicles that represent the other percentile ranges are hardly distinguishable from year to year. This finding results in part from our assumptions about vehicles selected by RV owners; it is also consistent with the findings about future passenger-car and pickup-truck characteristics, which indicate that the performance of the highest-performance cars will diminish but that very little change will occur in the low-performance end of the mix. Essentially, cars and trucks that offer moderate performance will remain available and will be selected by RV drivers, as they are now, to avoid the limitations of vehicles that offer extremely poor performance.

CONCLUSIONS

The central finding of this study of future vehicle populations is that, based on current NHTSA presumptions, expected changes in vehicle characteristics will not be as drastic as popularly supposed.

The major changes in the distribution of the characteristics of passenger vehicles on the road, including trucks with a maximum GVW of 3860 kg (8500 lb), have already occurred or are projected to occur by 1985. The average weight of such vehicles on the highway will be 10 percent less in 1985 than in 1978 and 14 percent less in 1995. Similarly, average engine size and power will decline 25 percent by 1985 and slightly more by 1995. The power-to-mass ratio, however, will not change appreciably after 1985. This implies that, although fuel-consumption characteristics will continue to improve through 1995, little change in vehicle performance after 1985 is expected.

The reduction in average vehicle weight and power will come about through the replacement of heavy, high-powered vehicles by vehicles of more modest size and performance capabilities and not through the introduction of very small, low-powered vehicles. Thus, the future vehicle mix will be more homogeneous than it is now or has been in the past.

The findings presented in this paper are, for the most part, predicated on projected extensions of today's technology. It appears likely, however, that by 1995 substantial numbers of electric, hybrid, or deisel-powered passenger vehicles may be on the roads. This would probably further reduce average performance and fuel consumption. Again, higher-performance vehicles would be replaced by ones that provide poorer performance (but probably not poorer than that of many vehicles now in use).

Since RVs are towed primarily by passenger cars and pickup trucks, RV projections are based in large measure on passenger-car projections. The projections indicate that the highest-performance RVs will disappear by 1981 but that otherwise the performance distribution for RVs will change very little. In particular, we do not expect the

appearance of RVs that provide extremely poor performance, since passenger vehicles capable of providing current minimum performance levels will remain available to RV owners.

ACKNOWLEDGMENT

Much of the analysis in this paper was performed by Midwest Research Institute as part of a U.S. Department of Transportation contract. We acknowledge this sponsorship but also state that the findings and conclusions of this paper are our own and do not necessarily represent the views of the U.S. Department of Transportation.

REFERENCES

1. Data and Analysis for 1981-1984 Passenger Automobile Fuel Economy Standards: Document 2--Automotive Design and Technology, Volume 2. Office of Automotive Fuel Economy, National Highway Traffic Safety Administration, Feb. 1977.
2. A.C. Malliaris, E. Withjack, and H. Gould. Simulated Sensitivities of Auto Fuel Economy, Performance, and Emissions. Society of Automotive Engineers, Inc., Warrendale, PA, Rept. 760157, Feb. 1976.
3. A.D. St. John and W.D. Glauz. Vehicle Handling, Acceleration, and Speed Maintenance. Federal Highway Administration, U.S. Department of Transportation, 1969.
4. A.D. St. John and D.R. Kobett. Grade Effects on Traffic Flow Stability and Capacity. NCHRP, Rept. 185, 1978.
5. Data and Analysis for 1981-1984 Passenger Automobile Fuel Economy Standards: Document 2--Automotive Design and Technology, Volume 1. Office of Automotive Fuel Economy, National Highway Traffic Safety Administration, Feb. 1977.
6. Rulemaking Support Paper Concerning the 1981-1984 Passenger Auto Average Fuel Economy Standards. National Highway Traffic Safety Administration, July 1977.
7. Rulemaking Support Paper for the 1980-1981 Non-Passenger Automobile Fuel Economy Standards. National Highway Traffic Safety Administration, Dec. 1977.
8. Light Duty Vehicle Fuel Consumption Model: 1975-1986. U.S. Department of Energy, April 1978.
9. NADA Official Used Car Guide. National Automobile Dealers Used Car Guide Co., McLean, VA, 1977.
10. Rulemaking Support Paper: Supplement for the Light Truck and Van Fuel Economy Standards for Model Years 1980 and 1981. National Highway Traffic Safety Administration, May 1978.
11. W.D. Glauz and A.D. St. John. Implications of Light-Weight, Low-Powered Future Vehicles in the Traffic Stream. Midwest Research Institute, Kansas City, MO, Tech. Memorandum, Feb. 1979.
12. K. Train. The Potential Market for Non-Gasoline-Powered Automobiles. Transportation Research, Vol. 14A, Nos. 5 and 6, Oct. and Dec. 1980.
13. N. Ludtke. Survey of Suspension Systems. National Highway Traffic Safety Administration, various vols., 1976-1977.

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

Headway-Distribution Models for Two-Lane Rural Highways

S. KHASNABIS AND C. L. HEIMBACH

The distribution of vehicle headways on two-lane, two-way roadways has been the subject of continuing research for a number of years. The growing interest in headway-generation models is related to the increased application of simulation techniques to describe traffic-flow patterns through the use of digital computers. A headway-distribution model developed for varying traffic-volume conditions (80-630 vehicles/h/lane) is described. The model was developed as part of a research project on the feasibility of using simulation techniques for depicting traffic flow on two-lane highways. A total of 18 sets of headway data (2 sets for each site) were collected from nine sites in North Carolina. The process of model development consisted of testing the field data by using a number of existing simple models and progressing with increasing degrees of complexity until an acceptable match between the field data and the model output was obtained. The study showed that none of the existing models (the Negative Exponential, Pearson Type III, and Schuhl models) provided satisfactory results for the wide range of traffic volumes tested. A modified form of the Schuhl model, incorporating parameters developed from the North Carolina data, provided the most reasonable approximation of the arrival patterns noted in the field. Parameters developed in the study are presented, along with a nomograph that can be used by traffic researchers to describe the time spacing between successive arrivals of vehicles on two-lane highways.

The distribution of vehicle headways, or the time spacing between successive arrivals of vehicles on two-lane roadways, has been a subject of continuing research for a number of years. Several past studies have attempted to describe mathematically the distribution of vehicle headways in two-lane traffic streams. The growing interest in headway-generation models is related to the increased application of simulation techniques to describe traffic-flow patterns through the use of digital computers. The development of a headway-prediction model as an appropriate descriptor of the input traffic stream is considered a mandatory requirement of any such simulation model. The importance of the headway generator, as a part of the simulation program, derives from the fact that the distribution of vehicle headways constitutes the single most important characteristic of traffic-flow patterns on two-lane roadways. The ability to accurately predict the arrival patterns of vehicle traffic by use of a headway-distribution model is thus the primary prerequisite for such a simulation model.

Drew (1), in his book on traffic-flow theory and control, discusses the theoretical concepts and practical implications of the mathematical models

developed by various researchers to predict vehicle headways. Although the development of most of these models dates back to original work by Schuhl (2) and Adams (3), these concepts have been successfully applied to actual traffic data on two-lane and multilane facilities by a number of researchers in the United States (4-8). In their 1976 Transportation Research Board monograph on traffic-flow theory, Gerlough and Huber (9) provide an extensive discussion of various headway models developed by different researchers.

As part of a larger research effort conducted at North Carolina State University from 1969 to 1973, a headway-distribution model was developed and calibrated by using data collected from two-lane roadways in North Carolina. A number of original headway models were reviewed and tested by using field data. The result was the development of a headway model, based on North Carolina data, that was then incorporated as part of a program designed to predict an input queue of vehicles for a larger simulation program.

BACKGROUND OF THE RESEARCH

The basic purpose of the research study from which this paper originates was to test the feasibility of using simulation techniques to evaluate the effects on traffic of selective and systematic removal of no-passing barriers from two-lane rural roadway sections under varying geometric and traffic conditions (10). The study was conducted as part of a cooperative highway research program and was sponsored by the North Carolina Department of Transportation and Highway Safety in cooperation with the U.S. Department of Transportation. The simulation model developed in this study is capable of duplicating traffic-flow characteristics, including passing maneuvers on two-lane rural roadways. A major component of this simulation model is a "speed-headway" program that was developed for the purpose of predicting individual vehicle speeds and headways to be used as input in the simulated roadway section. This program, as developed, is capable of generating headways on a lane-by-lane basis according to the traffic volume and directional distribution specified. The input queue generated by the speed-headway program provides an ordered list of vehicles to be simulated along with assigned speeds and headways in accordance with the specified input parameters. The specification of these parameters was a part of the overall process of model calibration.

The development and application of the overall simulation model have been reported elsewhere (11, 12). These publications did not, however, clarify the development of the headway-distribution model, which is an integral component of the speed-headway program. The purpose of this paper is to explain the development of the headway-distribution model and its importance in the complete model. The necessary data base was formed by collecting headway data from nine sites in North Carolina for different volume and traffic-mix conditions. These data were then used to test the ability of some of the existing headway models to adequately describe the observed arrival patterns. Initial efforts were directed toward the use of existing headway models to fit the observed data. Later, it was evident that, in order to reasonably predict the arrival of vehicles on two-lane sections, it would be necessary to develop an appropriate model with North Carolina data. The results of tests of the field data with the existing headway models and the development of the North Carolina model are described in this paper.

SIMPLE VERSUS COMPLEX MODELS

It has been pointed out in the literature that the selection of a suitable headway model represents a compromise between economic considerations and the faithfulness of the model (9). Both simple and complex models were considered in this study. Simple models, by definition, are computationally straightforward and require the development of a minimum number of parameters and a limited data base. Complex models, on the other hand, require a number of intricate mathematical manipulations. This necessitates a large data base because of the number of parameters that must be developed. Finally, the designation of a model as simple or complex must be somewhat arbitrary, since a fine line of demarcation between the two types does not exist.

The theory of error propagation in models suggests that there are essentially two types of error in model development--namely, measurement errors and specification errors (13). Measurement errors arise from inaccuracies in assessing magnitude--in this case, inaccuracies relative to the collection, recording, and transferral of the field data. Specification errors, on the other hand, are the result of misunderstanding or purposeful simplification of the relation between the variables in the model. In the present context, the description of an exponential relation by a simple linear formulation would constitute a typical specification error. Most models are characterized by both types of errors.

Although simple models can be criticized as being too simplistic in nature to consider the intricate relations between variables, they are preferred by researchers when the data base is poor. The reason cited is that the reduction in specification error resulting from the introduction of additional complexities is likely to be eroded by the effect of significant measurement errors (for a poor data base). Complex models are preferred when the data base is highly reliable. Coupled with increasing model complexity is an increase in the model's ability to explain the correlation between the dependent variables and the independent variables (13). In such cases (assuming that the data base is good), the reduction in specification errors with increasing complexity is likely to outweigh the increase in measurement errors.

In this study, the process of model development consisted of testing the data by starting with simple models and progressing to increasing degrees of complexity. Sufficient care was exercised during the collection and reduction of field data to minimize the effect of measurement errors so as to allow the testing of complex models. It was necessary to use complex models because the success of the overall two-lane simulation program, which involves passing maneuvers, depended largely on the ability of the headway model to realistically predict the successive arrival of vehicles at a specified point on the roadway. In this context, the following comment should be noted (9, p. 31):

> As in many engineering selection processes, selection of a suitable headway distribution represents a compromise between economic considerations and faithfulness of the model. Greater faithfulness is often obtained by using a model with a greater number of parameters; such a model, on the other hand, results in a more complex computational procedure. In some cases the intended use of the model can help in the selection procedure . . . . if the objective is simply the computation of delays, the simplest

(i.e., the negative exponential) distribution should be used. If, however, the objective is the determination of gaps for, say, crossing purposes, a more faithful distribution may be needed.

FIELD DATA COLLECTION

Nine two-lane primary rural highway sections in North Carolina were studied, and traffic, geometric, and operational data were collected. The critical data given in Table 1 indicate that the range of lane volumes, covered in a total of 18 sets of field data (nine sites, each with two directions), is between 80 and 632 vehicles/h/lane and that approximately 50 percent of the data sets (8 out of 18) lie within a range of 100-150 vehicles/h/lane. The directional distribution of volume in most cases was generally balanced, lying between 50:50 and 40:60 (except for site 7, where lane 1 had a considerably higher volume than the other lane). The unbalanced distribution of traffic at this site was the result of a number of traffic generators (industrial developments) immediately upstream of the site, and the fact that the data collection period coincided with the period of peak traffic outflow from these generators. The skewed nature of the distribution of the traffic volume and the resulting large variance (exhibited by the "outliers", such as 632 vehicles/h/lane) were considered to provide a wide spectrum of traffic flow ranging from "random" (light-volume) to "nonrandom" (medium to medium-heavy) conditions. The need to develop a single headway model to appropriately describe traffic flow covering this wide volume range presented a special challenge to this research. As later discussion will show, most existing headway models provided a decent fit to the field data under random conditions, but incorporating a mix of random and nonrandom flow characteristics into a single model proved to be a particularly difficult task.

DEVELOPMENT OF A HEADWAY MODEL

A total of six headway-distribution models were tested by using the field data noted above. Constants (parameters) were estimated for each model by using the lane input field data. Then the model was used to generate a headway distribution that was statistically compared with the actual input headway observed and recorded in the field. This procedure was carried out for each model on all of the test sites until some conclusions could be made about each model's accuracy in generating headway distributions comparable to those that had been measured in the field. A form of the Schuhl model (2) was finally selected as providing the best fit over the range of volumes studied. The results of tests of the field data with a total of six headway-distribution models are presented below.

Testing with Individual Models

During the earlier phase of this research, an effort was made to fit the Erlang distribution to the first four sets of data collected. This model was dropped from consideration later during the research because initial testing did not provide an acceptable statistical match between the field data and the model output. Extensive testing was then conducted with three other models--namely, the Negative Exponential, Pearson Type III, and Schuhl models (see Figure 1). Of these three, the first two are considered simple models in that they require the use of one or two parameters only. The Schuhl model, on the other hand, is considered complex in that it requires the use of at least five parameters. The parameters developed by Grecco and Sword (7) with data collected from a two-lane section of US-52 in Lafayette, Indiana, during 1968 were used for the Schuhl model.

The choice of these three models for testing purposes was based on the premise that the model to be developed in this study should be capable of representing both random and nonrandom traffic flow, in light of the high variance associated with the traffic volume from the nine sites. It has been shown by a number of researchers that the distribution of the Negative Exponential model generally provides a good fit for traffic data under

Table 1. Critical field data collected for nine North Carolina study sites.

Site No.	Location	Equivalent Flow Rate (vehicles/h) Lane 1	Lane 2	Trucks in Traffic Stream (%)	Posted Speed Limit (miles/h)
1	US-1; north station is just south of interchange with NC-55	90	117	13	60
2	US-1; north station is just south of site 1	80	98	19	60
3	US-64; east station is 13.7 miles west of junction with US-1	111	115	15	60
4	US-15, 501; south station is 3 miles north of Creedmoor	305	381	17	55
5	US-15, 501; south station is 3 miles north of Pittsboro	125	165	11	55
6	NC-54; east station is 1 mile west of Morrisville	143	122	8	55
7	NC-54; east station is 1 mile west of Morrisville	632	129	2	55
8	US-64; west station is 1.43 miles west of I-40 interchange	271	235	21	55
9	US-301; north station is 1.95 miles south of end of I-75	300	341	15	55

Figure 1. Headway-distribution models tested.

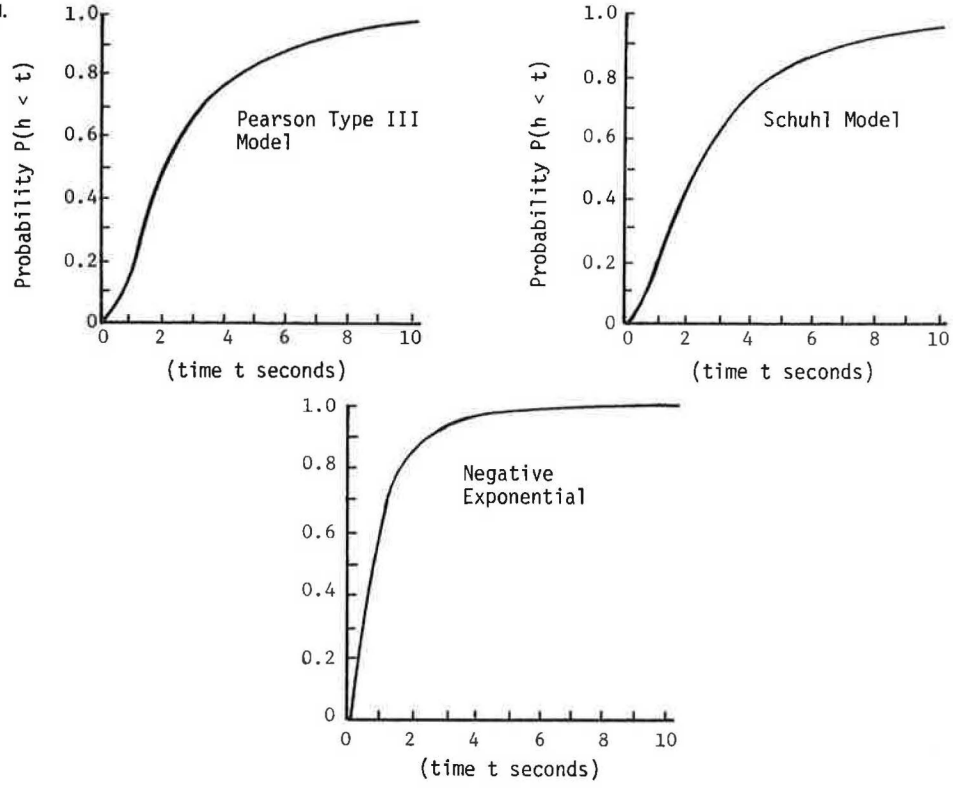


Table 2. Probability distribution models used to generate simulation headways for comparison with headways measured in the field.

Model Category	Goodness of Fit over Range of Lane Volumes Tested[a] (s)			Lane Volume That Gives Best Single Fit for Each Model			
	Underprediction	Good Fit	Overprediction	Vehicles per Hour	χ^2	df	χ^2 0.05
Original							
Negative Exponential	5-30	1-5 and 30-50	>50	143	13.02	11	19.70
Pearson Type III	5-50	1-5 and 50-80	>80	143	14.31	11	19.70
Schuhl (Grecco parameter)	5-30	1-5 and 50-120	30-50	166	6.59	12	21.00
Combined							
Schuhl and Pearson Type III	5-30	1-5 and 30-50	>50	122	9.87	9	16.90
Schuhl and Negative Exponential	5-50	1-5 and 50-80	>80	166	9.20	13	22.40
Modified individual Schuhl (North Carolina parameter)		1-150		166	9.84	14	23.70

[a]80-632 vehicles/h.

light-volume conditions (random). The Pearson Type III and Schuhl models, on the other hand, are both capable of incorporating a combination of random and nonrandom flow through the use of appropriate parameters.

Columns 2-4 in Table 2 give generalized observations concerning the fit of specific models to the 18 different headway distributions. The well-known nonparametric test called the "chi-square" test was used to check the goodness of fit of the model data with the field data. The null hypothesis tested during this analysis was that there is no significant difference in the observed and predicted headway distributions. The number of observations falling into prespecified headway groups was compared between the two sources (field and model) by using the chi-square computational procedure. Where the number of observations in a given cell was less than 5, these were combined with the observations in the next cell.

It is apparent from the results presented in Table 2 that the individual models used underpredict short headways and overpredict longer headways. Columns 5-8 indicate the traffic-lane volume that gives the best single fit for that model and the associated chi-square value. These last four columns clearly show that in all of the models listed the best fit was obtained for low-volume conditions where arrival patterns were generally random in nature.

Testing with Combined Models

An effort was made to combine two headway models

into a composite and a more complex one, assuming that the total area under the probability distribution curve was unity. A computer program was developed to carry out the necessary mathematical manipulation. Headways up to 5 s were described by the Schuhl model, and those greater than 5 s were described by one of the other two models. The Schuhl model was selected to estimate shorter headways within the combined model because our experience indicated that this model was capable of predicting such short headways. Columns 5-8 of Table 2 give the results of fitting the two combined models, Schuhl and Pearson Type III and Schuhl and Negative Exponential, on the field data. The results indicate that this effort was not very successful. Column 2 of Table 2 indicates that the underprediction of short headways is a problem with the combined models as well.

Calibration of Schuhl Model with North Carolina Data

A decision was made to develop parameters for the Schuhl model by using North Carolina data. The decision to adopt the Schuhl model for this study was based on the finding that, as an individual model, the Schuhl model provided the best fit for North Carolina field data, although the question of underprediction or overprediction remained. The choice of this complex model (which requires the use of at least five parameters) was considerably affected by the fact that the simple models tested earlier were incapable of incorporating the mixed random and nonrandom aspects of traffic flow. The principle of least squares was applied in developing the revised model parameters; the field data were tested by conducting an orderly and successive revision of the model parameters until the squared differences between the observed and expected frequencies were minimized.

The Schuhl model (here called the modified Schuhl model) was ultimately selected as the model that provided the best overall fit for the headway distributions surveyed. Table 3 gives the results of the chi-square test in which the headway distributions generated by the Schuhl model were compared with those recorded in the field; as indicated earlier, the hypothesis tested was that there is no difference between the field and the model-generated headway distributions. Table 3 indicates that this hypothesis was rejected in three cases (two lanes at site 9) for the 1 percent level of significance. Each of these three cases had an inordinate number of short headways in the 1- to 5-s range. The acceptance of the null hypothesis (implying "no difference" between the model output and the observed data) in a total of 15 out of 18 cases clearly demonstrated the capability of the model to reasonably duplicate headway patterns observed in the field. It was evident from the testing that underprediction of short headways is a problem with most of the models and that neither the combination of the two models nor the development of new parameters for the Schuhl model resolved the problem satisfactorily. The modified Schuhl model was used to generate a headway distribution for providing a fixed input queue of vehicles to the simulation model.

Table 3. Field versus model-generated headway distributions.

Site No.	$\chi^2_{calculated}$ Value for Comparing Field and Simulated Headway Distributions: Lane 1	Lane 2	χ^2 0.01 (df = 14)	Outcome of Testing[a]
1	26.10	28.42	29.10	Accept H_0
2	26.95	14.51	29.10	Accept H_0
3	25.91	8.60	29.10	Accept H_0
4	22.49	22.82	29.10	Accept H_0
5	16.95	9.84	29.10	Accept H_0
6	32.06	20.40	29.10	Reject H_0 (lane 1) Accept H_0 (lane 2)
7	10.30	24.45	29.10	Accept H_0
8	17.65	22.93	29.10	Accept H_0
9	42.68	39.88	29.10	Reject H_0

[a] H_0 = no difference between field data and simulation data.

The general form of the Schuhl model is as follows:

$$P(h \geq t) = \gamma e^{-(t-\epsilon)/t_1} + (1-\gamma)\, e^{-t/t_2} \quad (1)$$

where

- $P(h \geq t)$ = probability of a headway equal to or greater than time t;
- γ = ratio of vehicles in the restrained group to all vehicles = C x (lane volume/100), where C is a constant;
- ϵ = minimum headway for vehicles in the restrained group (s);
- t_1 = parameter that is a function of the average headway of the restrained group (s);
- t_2 = parameter that is a function of the average headway of the unrestrained group (s) = a - b x (lane volume/100), where a and b are constants; and
- e = base of Naperian logarithms.

The Schuhl model was calibrated by using the following values, based on North Carolina field data: $t_1 = 1.996$, $t_2 = 37.78 - (4.544V/100)$, and $\gamma = 0.2693 + (0.056\ 16V/100)$, where V = lane volume per hour and $\epsilon = 1$ s.

Efforts to fit the Schuhl model for single-lane traffic flow were originally reported by Grecco and Sword in 1968 on their study of US-52 in Lafayette, Indiana, where data were collected for each lane of the two-lane portion of that facility (7). Grecco and Sword recognized that Schuhl had attempted to divide the set of all vehicle spacings into two subsets, that of restrained and unrestrained groups (2, p. 61):

> Before proceeding further it must be observed that the first set spacing might apply to retarded vehicles which are prevented from passing by opposing traffic and the second set to free-moving vehicles which are able to pass at will.

According to Grecco and Sword (7, p. 36),

> By definition the restrained group is composed of those drivers who are traveling at or below their desired speed but are resigned to traveling in a platoon. The unrestrained group includes those drivers, not in a platoon, traveling at their desired speed and those drivers traveling below their desired speed in the platoon who are attempting or desiring to pass.

Clearly, both Schuhl and Grecco and Sword envisioned both the restrained and unrestrained groups to be in the same traffic lane (as shown by Schuhl's words "prevented from passing by opposing traffic" and the two-lane experimental site used by Grecco and Sword). During the testing of the North

Figure 2. Probability distribution nomograph based on parameters developed by Sword and Grecco (7).

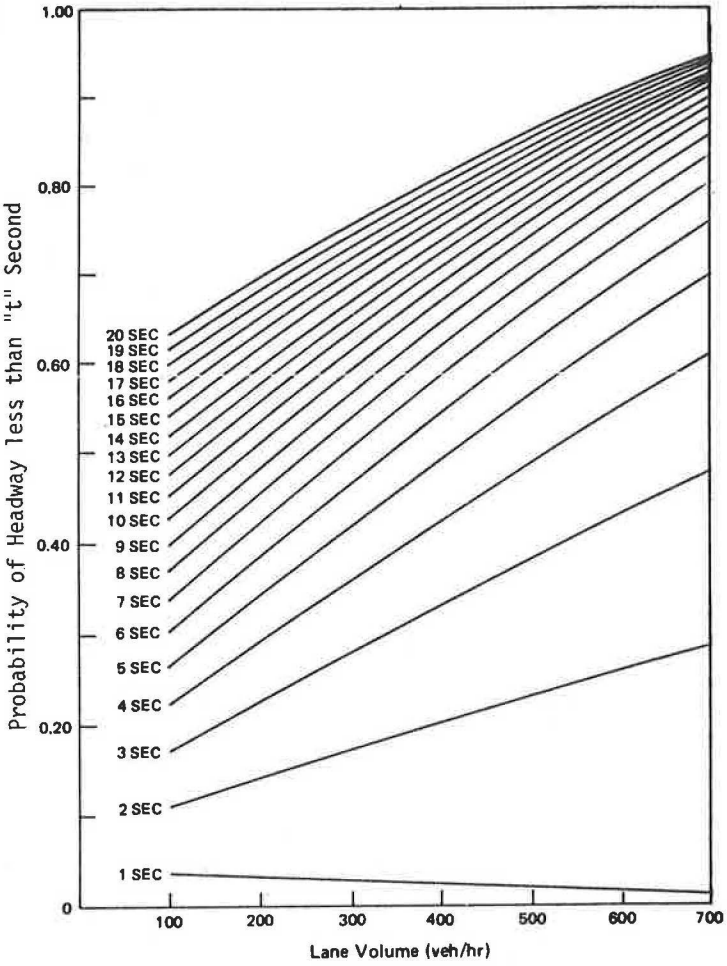
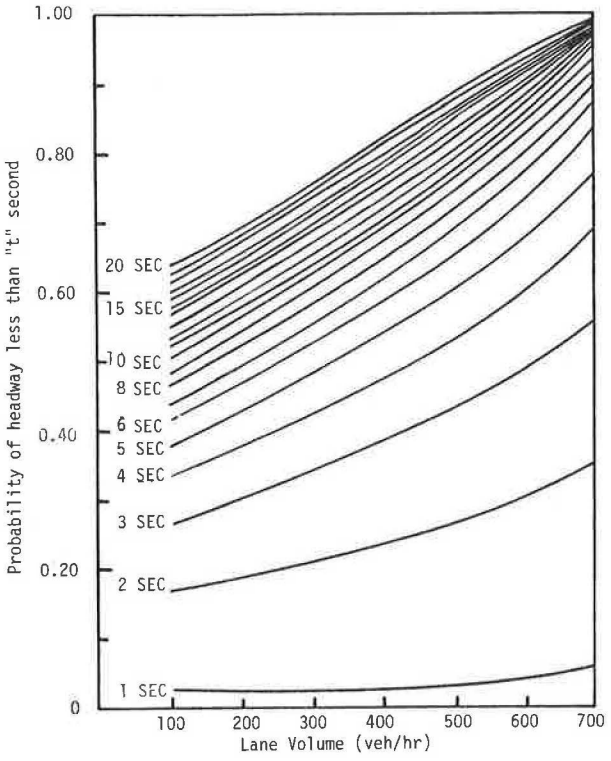


Figure 3. Probability distribution nomograph based on parameters developed in the North Carolina study.



Carolina data with the Schuhl model, we worked under similar postulations--namely, that traffic in a single lane contains two subsets (restrained and unrestrained groups) and that, at higher volume levels, the proportion of the restrained group tends to be higher. Furthermore, the parameters describing the restrained and unrestrained groups (γ, $\bar{t}_1$, and $\bar{t}_2$ in Equation 1) were developed as a part of the model-testing process.

Grecco and Sword (7) developed a nomograph for calculating the probabilities of various headways ("cumulative-percentage-less-than" curves) by using the parameters for the Schuhl model and data collected from Lafayette, Indiana. A similar nomograph was developed in the North Carolina study by using the Schuhl model parameters estimated based on nine sets of field data. These two nomographs are shown in Figures 2 and 3, respectively. The cumulative-percentage-less-than frequencies are also represented in Table 4 for various volume groups for the modified Schuhl model. It should be noted that both of these nomographs are based on the same distribution model--the Schuhl model--although the parameters used in the plots are somewhat different. The differences observed in the two cumulative distribution curves are the result of these different parameters.

DEVELOPMENT OF THE INPUT QUEUE

The preparation of the speed-headway program was carried out as part of the testing and calibration of the overall simulation model (10). A normal distribution function that is completely defined by

Table 4. Probability of headway less than t seconds as predicted by the modified Schuhl model.

Headway (s)	Probability by Lane Volume in Vehicles per Hour						
	100	200	300	400	500	600	700
1	0.0200	0.0212	0.0228	0.0252	0.0289	0.0357	0.0520
2	0.1677	0.1920	0.2172	0.2437	0.2727	0.3071	0.3571
3	0.2642	0.3029	0.3427	0.3844	0.4294	0.4814	0.5525
4	0.3295	0.3772	0.4262	0.4775	0.5327	0.5960	0.6799
5	0.3758	0.4291	0.4839	0.5413	0.6031	0.6735	0.7646
6	0.4103	0.4671	0.5257	0.5870	0.6530	0.7279	0.8223
7	0.4375	0.4966	0.5576	0.6215	0.6901	0.7676	0.8627
8	0.4600	0.5206	0.5832	0.6487	0.7190	0.7978	0.8917
9	0.4796	0.5412	0.6047	0.6713	0.7425	0.8217	0.9132
10	0.4971	0.5594	0.6236	0.6907	0.7623	0.8412	0.9294
11	0.5134	0.5760	0.6406	0.7080	0.7796	0.8576	0.9421
12	0.5286	0.5914	0.6562	0.7236	0.7950	0.8718	0.9521
13	0.5430	0.6060	0.6708	0.7380	0.8080	0.8841	0.9601
14	0.5569	0.6198	0.6845	0.7515	0.8216	0.8951	0.9666
15	0.5702	0.6330	0.6975	0.7641	0.8333	0.9049	0.9720
16	0.5830	0.6457	0.7099	0.7760	0.8442	0.9137	0.9765
17	0.5954	0.6579	0.7218	0.7872	0.8543	0.9216	0.9802
18	0.6075	0.6697	0.7331	0.7979	0.8637	0.9288	0.9833
19	0.6191	0.6810	0.7440	0.8079	0.8725	0.9353	0.9859
20	0.6304	0.6920	0.7544	0.8175	0.8807	0.9412	0.9881

its mean and its standard deviation was used to generate individual (desired) speeds for each input vehicle, given the mean and standard deviation specified by the user. The mean and standard deviation can be calculated by using relations developed as part of the process of model calibration (12).

After individual speeds have been generated, a random list of desired speeds is prepared and assigned to each vehicle to be input to the simulated roadway. The listing of headways generated by the modified Schuhl model is then paired by random assignment with the desired-speed list. The merging of these two arrays provides a sequential list of vehicles ready to be entered into the simulation routine. Each vehicle in the queue is thus assigned a desired speed, headway, and vehicle-type designation. After the vehicle moves onto the simulation roadway, the car-following rule built into the simulation model causes further adjustments to the headways to reflect the effect of the speed differential between the vehicle and the following car. This feature is discussed in more detail elsewhere (11,12).

CONCLUSIONS

The following conclusions can be made as a result of the research reported in this paper:

1. For a wide range of lane volumes--80-632 vehicles/h/lane--no one of the headway models tested in this study (the Negative Exponential, Pearson Type III, and Schuhl models) provided an adequate fit to the field data with an acceptable level of statistical reliability. Simple models particularly were found to be inadequate to describe the arrival patterns for the ranges of traffic volume tested.
2. Combining two headway models into a composite model is not likely to result in any improvement in predictive capability.
3. Underprediction of shorter headways is generally a problem with most headway-distribution models within the volume ranges studied.
4. It is possible to develop specific model parameters for a modified Schuhl model to predict the distribution of headways that incorporate traffic-flow conditions with characteristics ranging from random to nonrandom. The choice of such a complex model over simple ones can be justified by its improved predictive capability.
5. Most of the models studied provided a decent fit for traffic data in light-volume conditions (i.e., approximately 150 vehicles/h/lane).

ACKNOWLEDGMENT

A liaison committee that consisted of the sponsors of the research project (members from the North Carolina Department of Transportation and Highway Safety and the Federal Highway Administration) was extremely helpful and cooperative in furnishing guidance and direction.

The contents of this paper reflect our views, and we are responsible for the facts and accuracy of the data presented. The contents do not necessarily reflect the official views or policies of the North Carolina Department of Transportation and Highway Safety or the Federal Highway Administration.

REFERENCES

1. D.R. Drew. Traffic Flow Theory and Control. McGraw-Hill, New York, 1966.
2. A. Schuhl. The Probability Theory Applied to Distribution of Vehicles on Two-Lane Highways. In Poisson and Traffic, Eno Foundation, Saugatuck, CT, 1955.
3. W.F. Adams. Road Traffic Considered as a Random Series. Journal of Institute of Civil Engineers, 1936.
4. D.L. Gerlough. Traffic Inputs for Simulation on a Digital Computer. Proc., HRB, Vol. 38, 1959, pp. 480-492.
5. F.A. Haight. The Generalized Poisson Distribution. Annals of Institute of Statistical Mathematics, Tokyo, Vol. 11, No. 2, 1959, pp. 101-105.
6. A.D. May and F.A. Wagner, Jr. Headway Characteristics and Interrelationships of Fundamental Characteristics of Traffic Flow. Proc., HRB, Vol. 39, 1960, pp. 524-547.
7. W.L. Grecco and E.C. Sword. Prediction Parameters for Schuhl's Headway Distribution. Traffic Engineering, Feb. 1968.
8. J.E. Tolle. Vehicular Headway Distributions: Testing and Results. TRB, Transportation Research Record 567, 1976, pp. 56-64.
9. D.L. Gerlough and M.J. Huber. Traffic Flow

Theory: A Monograph. TRB, Special Rept. 165, 1976.
10. C.L. Heimbach, J.W. Horn, S. Khasnabis, and G.C. Chao. A Study of No-Passing-Zone Configurations on Rural Two-Lane Highways in North Carolina. North Carolina State Univ., Raleigh, Project ERSD-110-69-3, Final Rept., 1974.
11. C.L. Heimbach, S. Khasnabis, and G.C. Chao. Relating No-Passing-Zone Configurations on Rural Two-Lane Highways to Throughput Traffic. TRB, Transportation Research Record 437, 1973, pp. 9-19.
12. S. Khasnabis and C.L. Heimbach. Traffic Simulation as a Highway Design Tool. Transportation Engineering Journal, ASCE, Vol. 103, May 1977, pp. 369-384.

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

Abridgment

Changes in Traffic Speed and Bunching Near Transition Points Between Two- and Four-Lane Rural Roads

C. J. HOBAN AND K. W. OGDEN

The results of several field experiments conducted to measure changes in traffic performance caused by transitions from two to four lanes on rural highways are reported. Vehicle speed and bunching data were recorded at a number of points upstream and downstream of a transition. A microprocessor-based recorder unit and flat metal-and-rubber detector strips were used. The results show that changes in traffic performance with position occur more rapidly on entering a four-lane road than on merging into two lanes. The effects of a change in road quality at the transition were isolated. The results are applicable to the study of rural overtaking lanes and the validation of simulation models and may also be of interest in the study of rural transition points and temporary detours.

Several field experiments were conducted as part of a simulation study of the performance of rural overtaking lanes. The aims of the experiments were (a) to provide validating data for two- and four-lane simulation models and (b) to investigate directly the changes in traffic parameters that occur in transition between two- and four-lane rural road sections (1).

The experiments were designed to measure traffic mean speed and bunching at a number of points along the road in order to determine equilibrium performance and the rate of change of this performance attributable to the transition in road type. Although the data represent only a limited range of traffic conditions at two sites, the results should be of interest in the study of rural traffic behavior, especially in relation to passing lanes, lane transitions, and temporary detours.

This paper reviews variations in mean speed and bunching with distance downstream of a four- or two-lane merge point or a two- to four-lane demerge point on a rural highway, at various flow rates. The effects of variations in road quality are also briefly discussed.

DATA COLLECTION AND REDUCTION

Field observations were made at two sites near Frankston, Australia, an outer suburban center about 40 km from Melbourne (see Figure 1). Three experiments were conducted between July and October 1978. In all, more than 30 000 vehicle observations were made over 15 h, or 50 traffic-h if each recording station is considered separately. Only Sunday traffic was recorded; this included a significant proportion of recreational traffic but few trucks.

The physical layout of each site and the positions of the stations used for recording are shown in Figure 2. It should be noted that vehicles in Australia travel on the left side of the road. The merge site involved a divided four-lane arterial road merging into a two-lane, two-way carriageway in mildly undulating terrain. The Victoria state speed limit of 100 km/h applied throughout, and curves on the two-lane road section limited overtakings. At the demerge site, a narrow two-lane road with a 90-km/h speed limit joined a newly constructed freeway with a 100-km/h speed limit.

Vehicle speed and headway data were measured by using flat metal-and-rubber detector strips in pairs, coupled to a microprocessor-based recorder unit that stored the information on cassette tape. About 6 km of two-core wire was used to connect recording stations over 2 km of road. Because of wire limitations, some stations were recorded for shorter periods of time than others. Data on opposing traffic were also collected at the merge site.

The field data were later analyzed by using 5-min sample periods to give average values of the following parameters:

V = mean speed (km/h),
F = mean percentage following ("bunching"),
Q = flow rate (vehicles/h),
Q2 = opposing-flow rate at the merge site (vehicles/h), and
X = position downstream of the merge or demerge point (m).

Vehicles were defined as following if their headway from the preceding vehicle was ≤ 3 s. The term "bunching" is used in this paper to refer to the mean percentage of vehicles following in bunches (F). To provide a common basis for comparison, data from two-lane, one-way road sections were artificially merged into a single stream.

DATA ANALYSIS

The aim of this analysis was to establish relations between traffic mean speed and bunching and position

Figure 1. Location of field experiments.

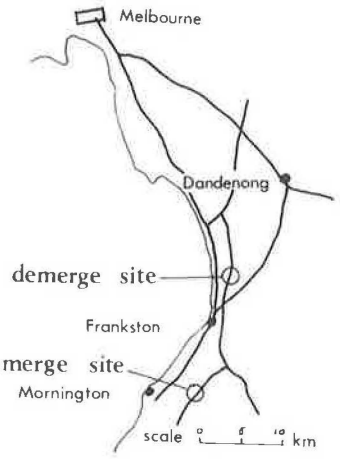
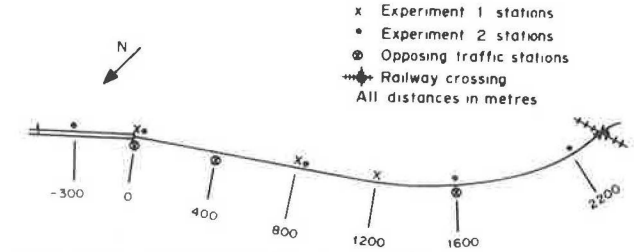
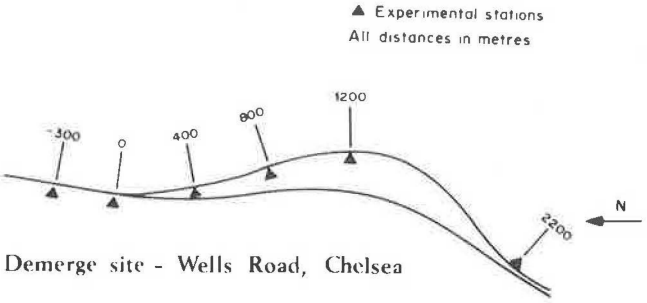


Figure 2. Layout of field experiments.



Merge site - Moorooduc Road, Mt. Eliza



Demerge site - Wells Road, Chelsea

on the road (measured downstream from a merge or demerge point). Intuitively, these relations should be approximately S-shaped, since traffic parameters change from one equilibrium level to another.

V and F, however, are known to vary with Q; in fact, the whole V-X or F-X relations may vary with Q. For the low to medium flow rates considered here, the effect of flow on V has been shown to be linear by Leong (2), Duncan (3), and others. A similar relation between F and Q seems appropriate.

A two-stage analysis procedure was thus adopted. First, for each road type (divided and undivided), at each site, a linear relation of V and F with flow was found. In preparation for the second stage, these relations were used to standardize the field data to the values that would have occurred at fixed flow rates of 400, 800, 1200, and 1600 vehicles/h. The second stage of the analysis was then to establish nonlinear relations for V and F with position, at each flow rate.

Standard regression analysis programs were used to fit relations to the data. Two useful parameters of these analyses are the coefficient of determination (R^2), which gives the proportion of variation in the dependent variable explained by the regression, and the standard error (S_y), which can be used to derive 95 percent confidence bands for the result. Traffic behavior varies widely between individual drivers and vehicles, so that not all of the variation will be attributable to changes in Q and X. It should be noted that R^2 and S_y values are found for mean parameters of 5-min sample periods. All of the relations presented are significant at the 5 percent level.

It was noted earlier that, in addition to the data collected at a succession of stations downstream of the merge point, data were also collected at several stations in the opposing lane. Data for these stations may be taken as representing an equilibrium state for the two-way, two-lane section of the road, since the opposing vehicles had traveled for several kilometers along a relatively uninterrupted two-lane road.

Accordingly, the analyses for the merge situation were repeated, and data from the opposing-flow stations were inserted at a point 5000 m downstream of the merge point; this somewhat arbitrary distance was selected because it was considered that equilibrium conditions would have been reached at that stage. Tests showed that the extended data supported the trend found over the first 2200 m.

RESULTS

Analysis showed that the influence of opposing flow (Q2) on V and F was very small for the two-lane, two-way roads. This is a surprising result that disagrees with the findings of the Highway Capacity Manual (4) and Luk (5). It probably reflects the limited range of flows recorded at each field station and the fact that opposing-flow readings were in some cases taken upstream or downstream of a particular station. Nevertheless, the best relations that could be obtained for this data set were based on one-directional flow only.

Relations for V and F with position are shown in Figures 3-6. Data points are not shown in these figures, but the 95 percent confidence limits indicate the range of data values found. Although the polynomial regression lines are clearly only an approximate indication of the true relations, they are quite useful for comparisons.

Figure 3 shows a quite rapid decrease in bunching with position at the demerge site and a similar trend at all flows. The corresponding sharp increase in mean speed is shown in Figure 4. In both figures, roughly half of the transition occurs within 500 m of the demerge points. Figure 5 shows a more gradual increase in bunching at the merge point, where the transition was half completed only after 2 km downstream. A similar transition in mean speed is shown in Figure 6.

Figures 3 and 5 show that bunching increased considerably with increasing flow. In Figures 4 and 6, mean speeds are shown to increase only slightly with flow. This is because speeds at a given point are drawn from a desired-speed distribution that is independent of flow and an increase in bunching merely gives greater weight to the lower speeds in the distribution. The form of the V-X and F-X relations is very consistent over a wide range of traffic-flow rates.

Whereas bunching is affected only by the transition between two lanes and four lanes, mean

Figure 3. Mean percentage following versus position at demerge site.

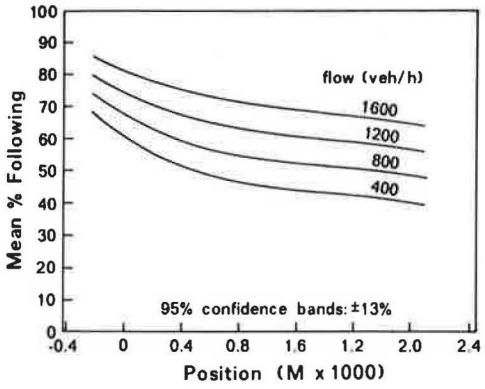


Figure 4. Mean speed versus position at demerge site.

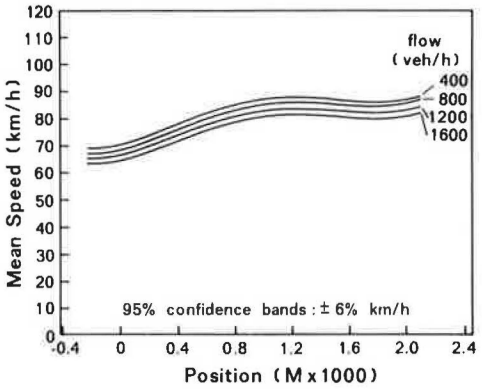


Figure 5. Mean percentage following versus position at merge site.

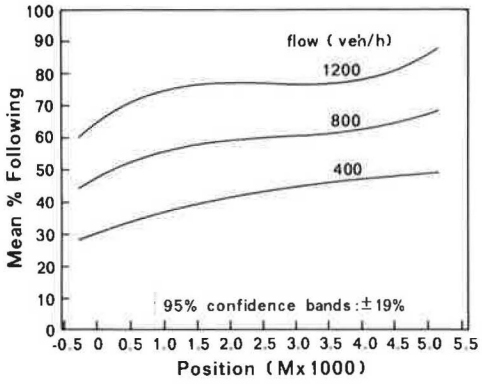


Figure 6. Mean speed versus position at merge site.

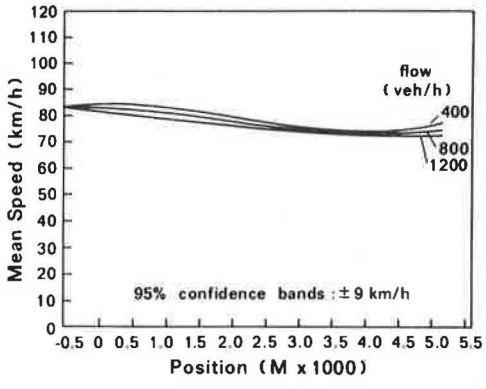


Figure 7. Effects of bunching and desired speed at demerge site.

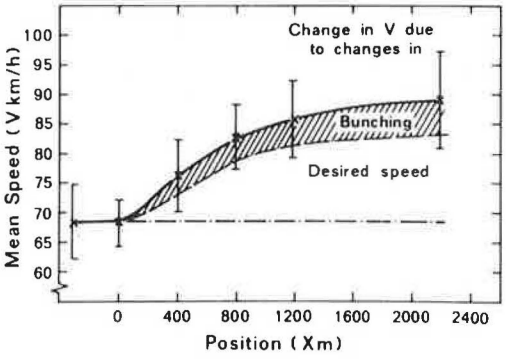
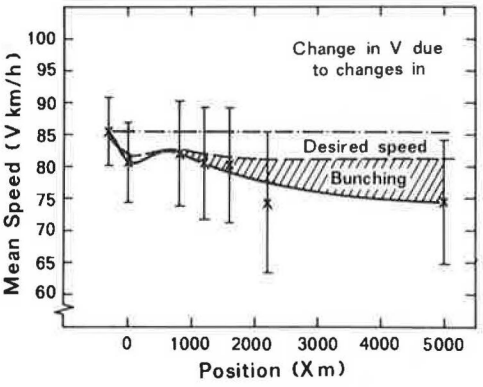


Figure 8. Effects of bunching and desired speed at merge site.



speed is also affected by changes in road quality over the transition. Leong (2) and others have shown that desired traffic speeds vary with such road-quality measures as pavement width, shoulder width, and alignment. Marked differences in the quality of adjacent road sections--e.g., at the demerge site--may have a substantial effect on the transition in mean speeds.

If the desired-speed distribution is known, the mean speed can be found for any degree of bunching. Conversely, the mean desired speed can be estimated for given values of V and F, provided the form of the distribution is known. Hoban (6) describes how this method can be used to isolate the two components of the mean-speed transition.

Figures 7 and 8 show the effect on mean speed of changes in bunching (because of a change in the number of lanes) and desired speeds (because of differences in road quality). The isolation of desired-speed effects, which vary from site to site, is important in examining variations in traffic that are directly attributable to a change in the number of lanes available to traffic.

The effect of changes in desired speed at the demerge site is shown in Figure 7 to account for more than two-thirds of the transition in mean speeds. Figure 8 shows a much smaller change in desired speeds at the merge site, which reflects the

greater similarity in quality of the two road sections. In both figures, the transition in desired speeds took place over approximately 1 km. The data points in Figures 7 and 8 are corrected to 800 vehicles/h, and 95 percent confidence limits are shown.

CONCLUSIONS

The results of the work reported in this paper show clearly that the transition in speed and bunching at a demerge site is considerably more rapid than that at a merge site. Traffic moving from a two-lane road onto a multilane road experiences an improvement in performance that is almost complete within 1 km downstream of the demerge. In the reverse situation, where traffic merges into two lanes, the transition is a gradual one that takes place over several kilometers.

Bunching was found to increase substantially with flow, especially on the two-lane road sections. This led to moderate decreases in mean speed as flow increased. Surprisingly, the form of the V-X and F-X relations was almost unchanged over a wide range of flow rates.

The transitions shown for bunching were directly caused by the change from two lanes to four, or vice versa. The transitions in speed, however, were also affected by changes in road quality. This effect was shown to be quite substantial at the demerge site where the two- and four-lane road sections were of very different quality. Once the effects of road quality are removed from the V-X relation, the remaining transition is entirely attributable to the change in the number of lanes.

ACKNOWLEDGMENT

We wish to acknowledge the cooperation of the Victoria Country Roads Board in the data collection and the helpful comments of A.J. Richardson on earlier drafts of the paper.

REFERENCES

1. C.J. Hoban. Overtaking Lanes on Two-Lane Rural Highways. Monash Univ., Melbourne, Australia, Ph.D. thesis (in preparation).
2. H.J.W. Leong. The Distribution and Trend of Free Speeds on Two-Lane Two-Way Rural Highways in N.S.W. Proc., Australian Road Research Board, Vol. 4, No. 1, 1968, pp. 781-814.
3. N.C. Duncan. Rural Speed/Flow Relations. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, Rept. LR 651, 1977.
4. Highway Capacity Manual. HRB, Special Rept. 87, 1965.
5. J.Y.K. Luk. A Circular Two-Lane Road Model. Australian Road Research Board, Melbourne, ARR Rept. 67, 1976.
6. C.J. Hoban. A Study of Rural Road Transition Points to Predict Passing-Lane Performance. Monash Univ., Melbourne, Australia, Civil Engineering Working Paper 79/12, 1979.

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

Abridgment

Development of a New Traffic-Flow Data-Collection System

LAWRENCE JESSE GLAZER AND WILLIAM COURINGTON

A recently developed hardware-software package for performing "floating-car" or "speed-and-delay" traffic-flow studies is described. The Traffic Recording and Analysis System closely approximates an ideal system. It uses a microcomputerized data recorder that almost totally automates the data-collection process. An enormous amount of data can thus be gathered by one person. No special training or computer background is required. The device uses little power, is truly portable, and is packaged in a rugged metal attaché case. The processing of data from digital cassette tapes is also almost totally automated. An IBM/360 computer analyzes the raw data and produces camera-ready printouts and digital plots. The information presented includes distance, travel time, speed, number of stops, delay time, fuel consumption, air pollution emissions, and vehicle operating costs. Digital plots available are time-speed profiles, speed contour maps, and speed perspectives. Heavy automation of data collection, analysis, and presentation means lower cost per study and thus greater productivity within existing budgets.

Traffic-flow studies have been an important traffic engineering tool for several decades, but currently used techniques are still rather primitive and limiting. Historically, the most common criterion used in measuring the quality of traffic flow has been travel time, which has sometimes included a measure of delay. But circumstances have changed drastically in the past decade, and new demands are now being made on the traffic engineer and the transportation evaluator.

Since automotive air pollution became a major social issue in the late 1960s, traffic engineers have increasingly been required to defend transportation projects with respect to air pollution impacts. Since the 1974 oil embargo, automotive energy consumption and the energy impacts of traffic improvements have also become critical social issues. The current emphasis on cost-effective transportation system management is likewise requiring more detailed project evaluations.

In all of these areas, the tools available to the practicing traffic engineer for measuring the benefits of traffic-flow improvements--including energy savings and reductions in air pollution--have not kept pace with increasing demands. Furthermore, the current economic and political environment is shrinking the budgets that make it possible to do

Figure 1. General report.

PENINSULA TRANSPORTATION DISTR
SPEED AND DELAY STUDY - GENERAL REPORT

MIDPENINSULA REGIONAL TRAFFIC STUDY

```
**********************************************************************************************************
*                                                                                                        *
* STUDY STREET:  MIDDLEFIELD RD       FROM: EMBARCADERO RD       TO: 10TH AVE                            *
* DIRECTION:  EASTBOUND       TIME LIMITS:  4:00PM -  5:36PM      DATE LIMITS:  5/12/78 -  5/18/78       *
*  10 RUNS ANALYZED           DELAY THRESHOLD =   5M.P.H.                                                *
*                                                                                                        *
**********************************************************************************************************
```

LINK #	STREET NAME AT END OF LINK	LINK DISTANCE (FEET) MEAN	STD.DEV.	TRAVEL TIME (SEC.) MEAN	STD.DEV.	LINK SPEED (MPH) MEAN	STD.DEV.	STOP DELAY (SEC)	AVG. # OF STOPS	LEVEL OF SERVICE INDEX #1
	EMBARCADERO RD									
1	CHANNING AVE	2474	19	64.9	17.6	27.7	7.2	8.5	0.4	16
2	HAMILTON AVE	719	15	18.5	7.3	29.1	7.7	1.7	0.2	23
3	UNIVERSITY AVE	896	16	23.5	9.2	28.3	6.8	2.9	0.2	16
4	LYTTON AVE	1507	20	37.1	10.9	29.4	6.8	4.1	0.3	17
5	WILLOW RD	998	17	60.8	26.5	14.7	9.2	34.5	0.7	45
6	RINGWOOD AVE	1130	22	33.5	11.6	25.5	8.1	6.7	0.5	33
7	OAK GROVE AVE	1320	24	28.7	5.0	32.1	4.7	0.6	0.1	12
8	GLENWOOD AVE	1294	26	35.1	7.0	26.1	5.3	3.5	0.4	29
9	ENCINAL RD	431	21	11.0	1.7	27.2	3.7	0.0	0.0	25
10	JAMES AVE	987	15	21.9	2.6	31.1	3.3	0.0	0.0	4
11	MARSH RD	1342	18	28.3	2.5	32.5	2.8	0.0	0.0	4
12	5TH AVE	1505	13	48.7	15.7	23.2	7.6	13.7	0.6	23
13	8TH AVE	1331	15	41.1	5.4	22.5	3.7	6.1	0.8	39
14	10TH AVE	1341	25	55.4	15.0	17.9	6.1	16.4	0.9	42
**	ROUTE DATA:	17272	34	508.5	70.9	23.6	3.3	98.7	5.1	23

good evaluations of traffic-flow improvement projects.

Thus, there is a clear need for better and cheaper techniques for measuring and evaluating traffic flow.

One recent development should provide a partial solution for the beleaguered traffic engineer. The Traffic Recording and Analysis System (TRANS) is a new package of hardware and software for performing traffic-flow studies, sometimes called "speed-and-delay" or "floating-car" studies. The hardware is a data recorder that collects raw data on tape, and the software is a group of computer programs that analyze the raw data to produce a wealth of printed and graphical reports.

DESIGN GOALS

Based on development and extensive use of a first-generation, electronic speed-and-delay device and also on a review of most other available devices of this type, we have developed a new system that overcomes the major problems with past systems. Our major design goals were the following:

1. To improve reliability by reducing machine and operator error,
2. To improve ergonomics (the person-machine interface) for safety and ease of use,
3. To reduce downtime by improving vehicle-to-vehicle portability,
4. To provide a universal recording format plus more analysis programs and thus satisfy a wide variety of local analysis and evaluation needs, and, most important,
5. To reduce the total cost and time required to perform these studies, including data collection, analysis, and report preparation.

The ways in which these design goals were met are described below.

DATA RECORDER

The entire TRANS data recorder is packaged in a rugged, portable, metal attaché case that weighs about 16 lb and so is easily moved from one vehicle to another. In operation, it is strapped to the front seat next to the driver, and connections are made to the vehicle's electricity and speedometer cable. Since the data recorder draws less than 2 A from the vehicle, no special electrical provisions are required. The speedometer cable is fitted with a simple screw-on, electrical-pulse generator.

The data recorder consists of several major components: a "control center", a kneepad keyboard, and a display. The control center includes a read-write digital cassette deck plus a process-control microcomputer that controls all of the functions of the data recorder. The kneepad, a panel of push-buttons that straps to the driver's knee, is used to manually enter data. The display, a long, thin box that lies on the dashboard in front of the driver, supplies feedback from the computer to the driver-operator in clear, English-language messages. The kneepad and display especially demonstrate the careful attention to ergonomics that has often been lacking in previous devices of this type.

The microcomputer accepts header information (date, time, study number, etc.), using the display to "prompt" and the kneepad to read manual data. All header data are checked for error and then recorded on the tape before each run. Thus, every run is uniquely identified to eliminate any possibility of mistaken data. During each run, the microcomputer records the distance traveled each second along with any manually entered remarks (which usually serve to identify the cause of delay). This is the most universal recording format for an instrumented-vehicle data base. The microcomputer also leads the operator through the necessary calibration procedures and performs

Figure 2. Delay analysis.

PENINSULA TRANSPORTATION DISTR
SPEED AND DELAY STUDY - DELAY ANALYSIS

MIDPENINSULA REGIONAL TRAFFIC STUDY

* STUDY STREET: MIDDLEFIELD RD FROM: EMBARCADERO RD TO: 10TH AVE
* DIRECTION: EASTBOUND TIME LIMITS: 4:00PM - 5:36PM DATE LIMITS: 5/12/78 - 5/18/78
* 10 RUNS ANALYZED DELAY THRESHOLD = 5M.P.H.

LINK #	STREET NAME AT END OF LINK	LINK TRAVEL TIME (SEC)	TOTAL STOP DELAY (SEC)	AVG. # OF STOPS	SIGNALS STOP DELAY (SEC)	SIGNALS AVG. # OF STOPS	LEFT TURNS STOP DELAY (SEC)	LEFT TURNS AVG. # OF STOPS	RIGHT TURNS STOP DELAY (SEC)	RIGHT TURNS AVG. # OF STOPS	PEDESTRIANS STOP DELAY (SEC)	PEDESTRIANS AVG. # OF STOPS	OTHER STOP DELAY (SEC)	OTHER AVG. # OF STOPS
	EMBARCADERO RD													
1	CHANNING AVE	64.9	8.5	0.4	8.5	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	HAMILTON AVE	18.5	1.7	0.2	1.7	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	UNIVERSITY AVE	23.5	2.9	0.2	2.9	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	LYTTON AVE	37.1	4.1	0.3	4.1	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	WILLOW RD	60.8	34.5	0.7	34.5	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	RINGWOOD AVE	33.5	6.7	0.5	6.7	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	OAK GROVE AVE	28.7	0.6	0.1	0.6	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	GLENWOOD AVE	35.1	3.5	0.4	3.5	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	ENCINAL RD	11.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	JAMES AVE	21.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
11	MARSH RD	28.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
12	5TH AVE	48.7	13.7	0.6	13.7	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	8TH AVE	41.1	6.1	0.8	5.5	0.7	0.0	0.0	0.0	0.0	0.6	0.1	0.0	0.0
14	10TH AVE	55.4	16.4	0.9	16.1	0.8	0.0	0.0	0.0	0.0	0.3	0.1	0.0	0.0
**	ROUTE DATA:	508.5	98.7	5.1	97.8	4.9	0.0	0.0	0.0	0.0	0.9	0.2	0.0	0.0

Figure 3. Energy/emissions/user-cost report.

PENINSULA TRANSPORTATION DISTRICT
SPEED AND DELAY STUDY - ENERGY & AIR POLLUTION ANALYSIS

MIDPENINSULA REGIONAL TRAFFIC STUDY

* STUDY STREET: MIDDLEFIELD RD FROM: EMBARCADERO RD TO: 10TH AVE
* DIRECTION: EASTBOUND TIME LIMITS: 4:00PM - 5:36PM DATE LIMITS: 5/12/78 - 5/18/78
* 10 RUNS ANALYZED DELAY THRESHOLD = 5 MPH

LINK #	STREET NAME AT END OF LINK	LINK DISTANCE (FEET)	TRAVEL TIME (SEC.)	AVG. SPEED (MPH)	STOP DELAY (SEC)	AVG. # OF STOPS	FUEL USED (GAL/100)	HYDRO-CARBONS (LBS)	CARBON MONOXIDE (LBS)	NITROUS OXIDES (LBS)
	EMBARCADERO RD									
1	CHANNING AVE	2474	64.9	27.7	8.5	0.4	2.87	0.0047	0.0285	0.0057
2	HAMILTON AVE	719	18.5	29.1	1.7	0.2	0.88	0.0014	0.0083	0.0017
3	UNIVERSITY AVE	896	23.5	28.3	2.9	0.2	1.06	0.0017	0.0103	0.0021
4	LYTTON AVE	1507	37.1	29.4	4.1	0.3	1.88	0.0029	0.0174	0.0035
5	WILLOW RD	998	60.8	14.7	34.5	0.7	1.29	0.0019	0.0115	0.0023
6	RINGWOOD AVE	1130	33.5	25.5	6.7	0.5	1.41	0.0022	0.0130	0.0026
7	OAK GROVE AVE	1320	28.7	32.1	0.6	0.1	1.47	0.0025	0.0152	0.0030
8	GLENWOOD AVE	1294	35.1	26.1	3.5	0.4	1.60	0.0025	0.0149	0.0031
9	ENCINAL RD	431	11.0	27.2	0.0	0.0	0.40	0.0008	0.0050	0.0010
10	JAMES AVE	987	21.9	31.1	0.0	0.0	1.09	0.0019	0.0114	0.0023
11	MARSH RD	1342	28.3	32.5	0.0	0.0	1.54	0.0026	0.0155	0.0031
12	5TH AVE	1505	48.7	23.2	13.7	0.6	1.84	0.0029	0.0174	0.0035
13	8TH AVE	1331	41.1	22.5	6.1	0.8	1.65	0.0026	0.0154	0.0031
14	10TH AVE	1341	55.4	17.9	16.4	0.9	1.69	0.0026	0.0154	0.0032
**	ROUTE DATA:	17272	508.5	23.6	98.7	5.1	20.69	0.0332	0.1993	0.0372

numerous error checks during all phases of the operation.

SYSTEM OUTPUTS

A wide variety of outputs can be produced from the raw data on cassettes. A data base maintenance program stores all runs in a single master file and permits retrieval of any combination of runs (e.g., by study number, dates, or times) for subsequent analysis. The analysis software produces both tabular (printed) and graphical (plotted) outputs.

Figure 4. Time-speed profile.

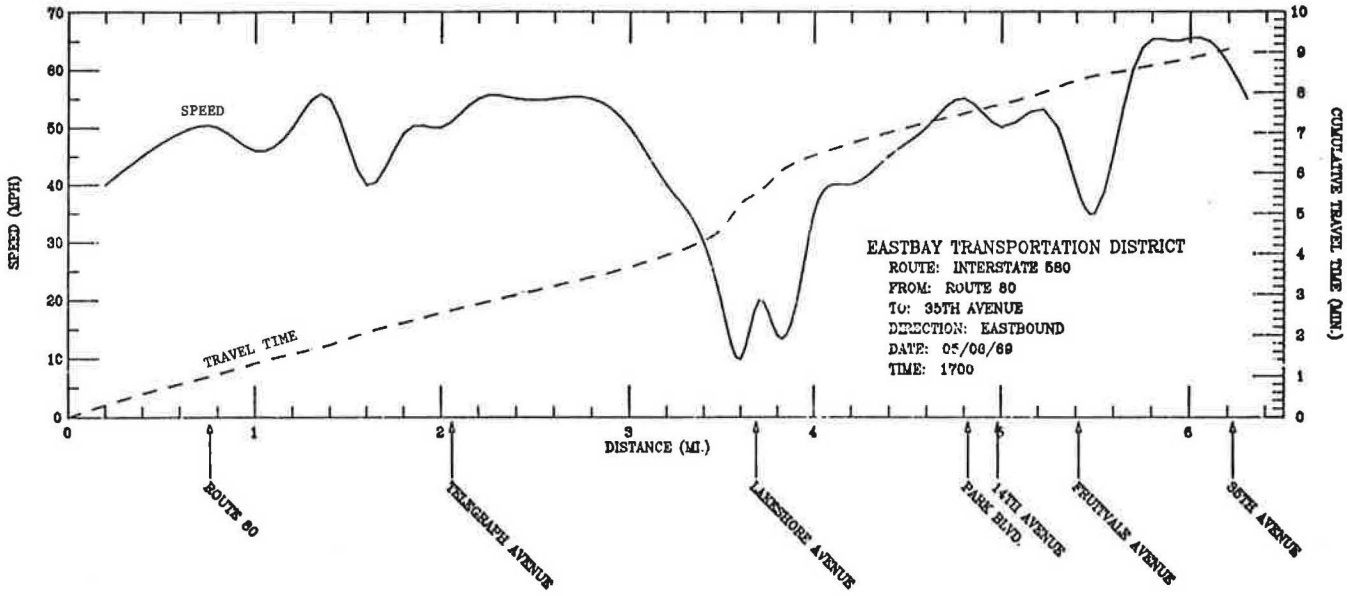
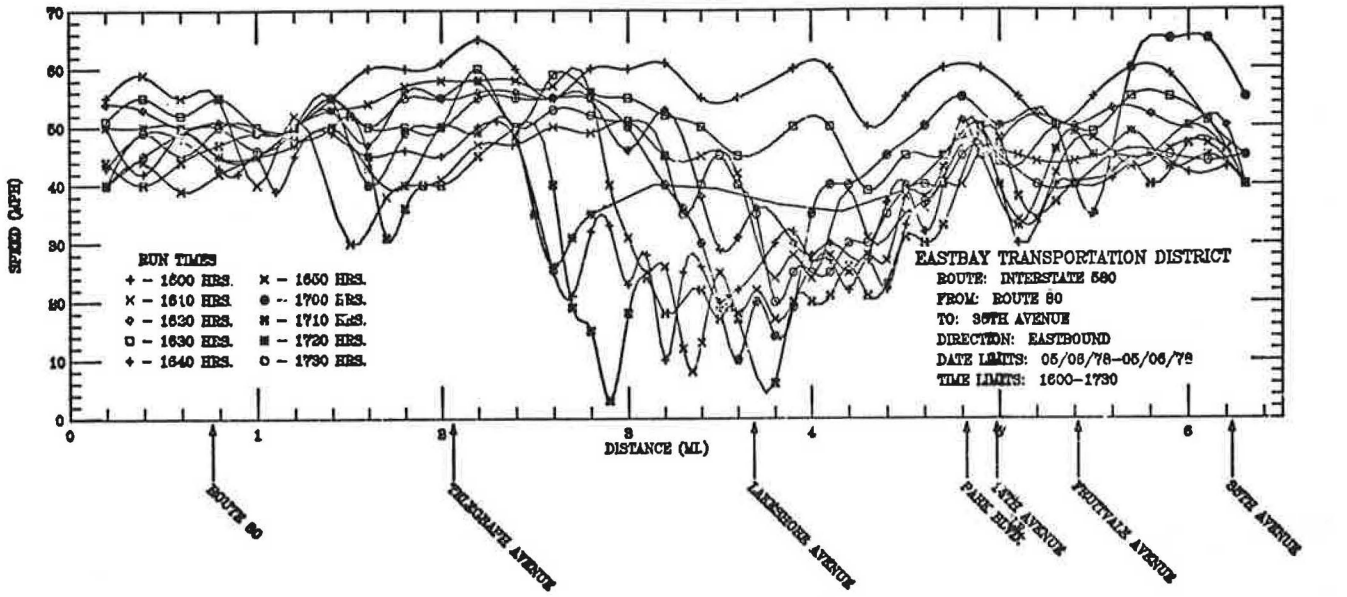


Figure 5. Speed profiles.



With one minor exception, all outputs are of camera-ready quality. Thus, substantial time and cost savings are possible in both the analysis and report-writing phases of traffic-flow studies. For example, the semiautomated RUNCOST procedure described by Parsonson (1), requires 4 h to code and keypunch each hour of tachograph field data. This high degree of automation is probably the major advantage TRANS has over its predecessors, which have tended to concentrate more on hardware than on data analysis and presentation.

Printed Reports

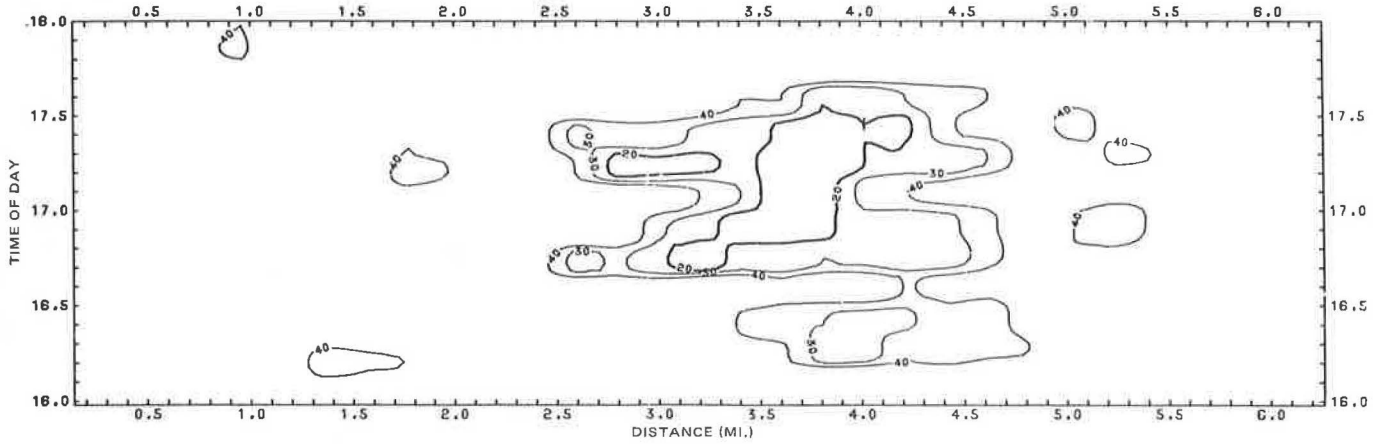
TRANS printed reports include the general report, the delay analysis, and the energy/emissions/user-cost report. Examples of these reports are shown in Figures 1-3.

The general report (Figure 1) supplies basic operating information: distance, travel time, speeds, delays, stops, and an index of quality of service. The delay analysis (Figure 2) breaks down total delays and number of stops into categories that correspond to the causes of these delays (e.g., signals or pedestrians). The energy/emissions/user-cost report (Figure 3) shows fuel used plus emissions (HC, CO, NO_x). Because the user-cost algorithm is currently under development, these data are not yet shown on the report. Other printouts, such as a run-by-run data matrix, are also available.

Graphical Outputs

The basic plotted output is the time-speed profile for each run (see Figure 4), which shows both cumulative time and speed versus distance. A

Figure 6. Speed contour map.



related plot is the speed profile (see Figure 5), which shows as many as 10 runs on the same plot, for ease of comparison. A different plot is the speed contour map (see Figure 6), which shows speed isopleths (contours) versus distance and time of day. Because it pictorially portrays magnitude, location, and duration of congestion, this is probably the most powerful output. It is meaningful to both the traffic engineer and the layperson.

Any of these outputs are available for any desired combination of runs.

CONCLUSIONS

TRANS uses advanced computer and traffic engineering techniques to achieve dramatic improvements in both the cost and the effectiveness of traffic-flow data collection. Costs are reduced because TRANS automates all three of the steps required for traffic-flow studies: data collection, analysis, and presentation. Effectiveness is greatly improved because TRANS measures all of the traffic-flow parameters that are currently of concern: distance, travel time, speed, number of stops, stop time, flow smoothness, fuel consumption, air pollutant emissions, and user costs. Several valuable digital plots, including the well-recognized speed profiles and the more striking speed contour maps, are also available.

TRANS can be a powerful analysis and evaluation tool for traffic-signal-system projects, freeway ramp-metering projects, roadway channelization and geometric improvement projects, inventories of travel time and operating speed, and energy and air pollution studies.

REFERENCE

1. P.S. Parsonson. Cost-Effectiveness of RUNCOST Evaluation Procedure. TRB, Transportation Research Record 630, 1977, pp. 21-24.

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

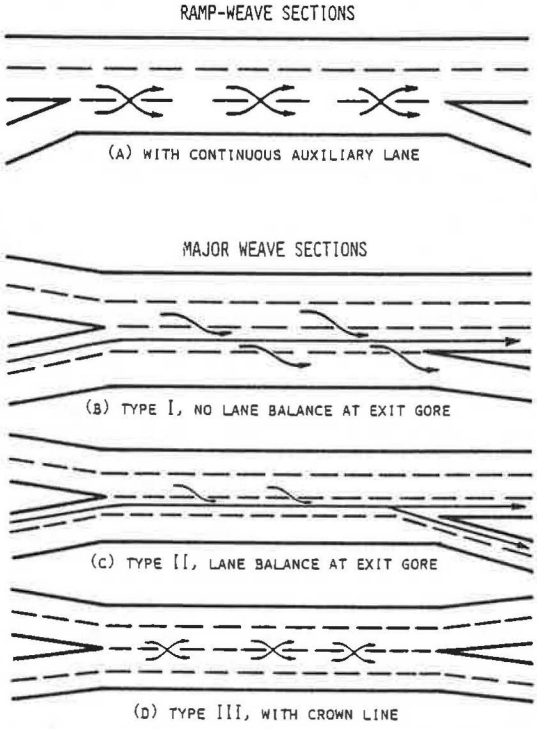
Revision of NCHRP Methodology for Analysis of Weaving-Area Capacity

ROGER P. ROESS, WILLIAM R. McSHANE, AND LOUIS J. PIGNATARO

As part of an effort sponsored by the Federal Highway Administration (FHWA) to revise and update procedures for freeway capacity analysis, the weaving-area methodology developed as a result of National Cooperative Highway Research Program (NCHRP) Project 3-15 was revised with two objectives in mind: (a) to recalibrate the procedure to reflect modified service-volume concepts developed in other parts of the FHWA effort and (b) to simplify the structure of the NCHRP procedure to make it easier to apply and understand while retaining its demonstrated accuracy and sensitivity to lane configuration, a major factor in highway operations. The revised method was developed by using standard multiple regression techniques and a data base consisting of results of the 1963 U.S. Bureau of Public Roads study of weaving-area capacity and the results of extensive data collection on NCHRP Project 3-15. The procedure consists of calibrated relations governing (a) the operation of nonweaving vehicles in weaving areas; (b) the maximum number of lanes that can be occupied by weaving vehicles for various configurations; (c) the "share", or percentage, of weaving-area lanes occupied by weaving vehicles under "balanced" operation; and (d) the relation between the average speed of weaving vehicles and that of nonweaving vehicles. To simplify the application of these relations in design and operational analysis, a series of nomographs has been developed.

In 1973, a major study of weaving-area operations was completed at the Polytechnic Institute of New York for the National Cooperative Highway Research Program (NCHRP) (1,2). The study resulted in the formulation of new procedures and relations for analysis of weaving-area capacity. These were (a) substantially more accurate than the procedures of the 1965 Highway Capacity Manual (HCM) in their representation of field conditions, (b) based on

Figure 1. Lane configurations for weaving areas.



consideration of lane configuration as a principal design and analysis parameter, and (c) able to predict cases in which weaving and nonweaving vehicles were operating "in balance," as well as cases in which they were not.

As part of a recent effort sponsored by the Federal Highway Administration (FHWA) to update and revise freeway capacity-analysis procedures, the 1963 and 1972 data bases were reexamined to determine whether or not the NCHRP procedure could be further improved. This reexamination was motivated by two factors:

1. The NCHRP relations were based on the acceptance of service-volume criteria given in Table 9.1 of the HCM (3). As a result of other findings of the FHWA study, it has been recommended that those service volumes be substantially revised (4).
2. The form of the NCHRP relations was found to be difficult for many people to use, a problem that has reduced the usefulness of the procedure to practitioners.

In the attempt to recalibrate the NCHRP procedure, there were three goals: (a) to improve its accuracy by considering the recalibration of HCM Table 9.1 standards, (b) to simplify the format and use of the procedure, and (c) to retain the advantages of greatly improved accuracy that the NCHRP procedure has over the 1965 HCM procedure.

LANE CONFIGURATION

The NCHRP study found that lane configuration was the principal parameter affecting the operation of weaving areas. Three types of lane configuration were identified as a result of the study; a fourth was added for the purposes of the FHWA effort.

Figure 1 shows the four types of weaving sections. These can be grouped into two broad categories: ramp weaves and major weaves. Ramp-weave sections are formed when an on ramp is followed by an off ramp and the two are joined by a continuous auxiliary lane. Major-weave areas are characterized by at least three of the input and output "legs" having two or more lanes.

The impact of configuration on weaving-area operations is highlighted in the three major-weave sections shown in Figure 1. These weaving areas have the same number of lanes and the same length and yet are substantially different operationally:

1. In the type 1 section, one weaving movement can be made without a lane change. The other weaving movement, however, requires two lane changes. This characteristic hampers the operation of weaving vehicles and limits the total number of lanes that weaving vehicles may occupy.
2. In the type 2 section, one weaving movement can also be made without a lane change. The reverse weaving movement, however, only requires one lane change. Thus, the type 2 section will provide for smoother operation of weaving vehicles and will allow them to occupy a larger number of lanes than a type 1 section. The difference between type 1 and type 2 sections is that the type 2 section provides for lane balance; that is, one lane divides to two at the exit gore so that a vehicle in that lane can travel down either exit leg without making a lane change.
3. The type 3 major-weave section has a weaving crown line--that is, a lane line that divides the section into distinct parts as it starts at the entrance gore point and connects directly to the exit gore point. In such sections, all weaving vehicles must execute one lane change. This somewhat restricts the operations of weaving vehicles; more importantly, it effectively restricts these vehicles to the two lanes adjacent to the crown line.

Ramp-weave sections are similar to type 3 major-weave sections in that they have a crown line. They differ, however, in that one input and one output leg are ramps, often with restrictive geometric features. Major weaves generally involve input and output legs that are high-speed collector-distributor roadways.

The importance of lane configuration in the development of a new or recalibrated procedure on weaving was recognized in the NCHRP study in the following ways:

1. It is not sufficient to define the total number of lanes (N) and the length (L) of a weaving section because operations may differ according to configuration features.
2. Consideration of a total N value is not sufficient to ensure proper design or analysis. The proportion of N used by weaving and nonweaving vehicles must be considered, and this factor is influenced by configuration.
3. Configuration of a given segment of width N (lanes) and length L (feet) is determined by the design and relative placement of entry and exit legs.

These factors were principal considerations in the development of the NCHRP procedure and were considered in the same light in the development of the recalibrated procedure. Configuration is discussed more fully elsewhere (5).

RECALIBRATION OF THE NCHRP PROCEDURE

Data Base

The data base used in the recalibration of the weaving procedure was the same as that used in the NCHRP study. This included data for 38 sites from

the 1963 Urban Area Weaving Capacity Study conducted by the then U.S. Bureau of Public Roads (BPR) and data for 14 sites collected specifically for the NCHRP study. The 1963 BPR data actually included 48 sites, but 10 were arterial cases that were deleted for this recalibration.

The 14 sites for which data were collected for the NCHRP study each contained about 4 h of data, broken into 6-min periods. The BPR data, which generally included fewer data for each site, were also broken into 6-min periods. The 6-min periods were aggregated into 12- and 18-min periods for comparative analysis. The NCHRP study based its calibrations on the 18-min aggregations because it concluded that 6- and 12-min data periods were "statistically unstable" and obscured any inherent relations. The same policy was adopted in this recalibration effort for similar reasons.

In general, the data base contains information on (a) segment geometry (width, length, configuration, and number of lanes), (b) segment volumes (6-min counts by flow components), and (c) segment speeds (6-min average speeds for each flow component). The 14 sites studied for NCHRP also contained additional information on lane changing and lane distribution of weaving and nonweaving vehicles.

Calibration Structure

Critical to any calibration effort that is to explicitly treat configuration as a key parameter is the ability to establish values in the data base for the number of lanes used by weaving vehicles (N_W) and the number of lanes used by nonweaving vehicles (N_{NW}) and to segregate the data base by configuration. Since N_W and N_{NW} were not directly observed, they must be computed or estimated. There are three alternatives:

1. Assume that nonweaving vehicles in a weaving section behave in essentially the same manner as vehicles in a basic freeway section. N_{NW} may then be computed by using the criteria for basic freeway segments, as recalibrated in the FHWA study, and N_W as $N - N_{NW}$. This was the technique used in calibrating the original NCHRP method, which used HCM criteria for basic freeway segments.
2. Assume an arbitrary relation for the behavior of nonweaving vehicles in a weaving section based on general observation of data trends. N_{NW} is then computed by using the assumed relation and N_W as $N - N_{NW}$.
3. Assume that there is a maximum value of N_W that can be achieved for each type of configuration and that cases in which the average speeds of weaving vehicles and nonweaving vehicles differ markedly [by more than 5 miles/h (8 km/h)] have reached this maximum value. Then, for these cases, N_W is set based on assumed values, and N_{NW} is computed as $N - N_W$. A regression relation is developed between N_{NW} and known variables for these cases and applied to compute N_{NW} for all other cases, for which N_W becomes $N - N_{NW}$.

The third alternative involves the concept of "constrained" versus "unconstrained" operation of weaving areas and the identification of such cases in the data base. If there is indeed a maximum practical value of N_W for any given configuration, then weaving vehicles, no matter what their volume, cannot occupy more than that number of lanes. In the normal case, weaving and nonweaving traffic compete for space on the roadway and reach some equilibrium so that both traffic streams experience relatively uniform operating conditions. However, if the configuration and the conditions are such that this equilibrium would occur with a value of N_W that is greater than the maximum value for the configuration, weaving vehicles will be constrained to occupy the maximum value of N_W, which is less than the equilibrium value, and nonweaving vehicles will occupy a value of N_{NW} lanes proportionally larger than the equilibrium value. The result will be that nonweaving vehicles will experience markedly better service than weaving vehicles, as indicated by higher average speeds.

All three techniques outlined above were investigated during this recalibration. Alternative 1 resulted in statistically poor fits to data during later analysis. Alternative 3, which appeared fruitful at first, led to internal inconsistencies in the maximum values of N_W for each configuration. Alternative 2 proved the most successful. Because it requires the postulation of a relation for nonweaving vehicles in weaving sections, numerous trials were required. Each was evaluated for the reasonableness and internal consistency of the results produced and the statistical accuracy of the regression relations generated.

The final output of the recalibration was a series of equations of the following form:

1. Nonweaving vehicles--an equation relating nonweaving volume to the number of nonweaving lanes (N_{NW}) and the average speed of nonweaving vehicles (S_{NW}) (this relation was postulated for each trial calibration);
2. Maximum values of N_W--for each type of configuration, a relation governing the maximum value of N_W, calibrated by using data from cases in which constrained operation was evident;
3. Speeds--for each type of configuration, a relation between the speed of weaving vehicles (S_W) and the speed of nonweaving vehicles (S_{NW}) (this relation was calibrated); and
4. Share of the roadway--for each type of configuration, a relation governing the proportion of total lanes occupied by weaving vehicles (N_W/N) (this relation was calibrated).

Some of the calibrated relations are primary--that is, they are valid for all cases of a particular configuration. Others are secondary, or valid only for unconstrained cases in which the equilibrium value of N_W is less than the maximum value of N_W for the configuration.

A set of equations was calibrated for ramp-weave sections and for type 1 and type 2 major-weave sections. The data base available did not include any cases of type 3 major-weave sections.

Recalibrated Relations

The recalibrated relations are given, in the form previously described, in Table 1, where

V_{NW} = volume of nonweaving vehicles (passenger cars/h),
V_W = volume of weaving vehicles (passenger cars/h),
S_{NW} = average speed of nonweaving vehicles (miles/h),
S_W = average speed of weaving vehicles (miles/h),
N_{NW} = number of lanes occupied by nonweaving vehicles,
N_W = number of lanes occupied by weaving vehicles,
N = total number of lanes in the weaving section = $N_W + N_{NW}$,
$N_W(max)$ = maximum number of lanes that may be occupied by weaving vehicles,

Table 1. Calibrated relations for weaving areas.

Category	Equation No.	Equation	Type of Equation	Regression Coefficient	No. of Samples	F-Test Results
Nonweaving vehicles	1	$V_{NW} = 1500N_{NW} - 50S_{NW} + 1900$	Primary	–	–	–
Maximum value of N_W						
Ramp weaves	2	$N_W(max) = 2.0$	Primary	–	–	–
Type 1 weaves[a]	3	$\log N_W(max) = 0.714 + 0.480 \log R$	Primary	0.788	5	R not significant
Type 2 weaves	4	$\log N_W(max) = 0.896 + 0.186 \log R - 0.402 \log L_H$	Primary	0.655	19	R, L_H significant
Speed						
Ramp weaves	5	$\log S_W = 0.142 + 0.694 \log S_{NW} + 0.315 \log L_H$	Primary	0.883	142	S_{NW}, L_H significant
Type 1 weaves	6	$S_W = 15.031 + 0.819S_{NW} - 24.527$ VR	Secondary	0.982	36	S_{NW}, VR significant
Type 2 weaves	7	$S_W = 2.309 + 0.871S_{NW} + 4.579$ VR	Secondary	0.931	43	S_{NW}, significant VR not significant
Share of the roadway						
Ramp weaves	8	$\log N_W/N = 0.340 + 0.571 \log VR - 0.438 \log S_W + 0.234 \log L_H$	Secondary	0.764	109	VR, S_W, L_H significant
Type 1 weaves	9	$N_W/N = 0.761 - 0.011L_H - 0.005\Delta S + 0.047$ VR	Primary	0.719	41	L_H, significant VR, S not significant
Type 2 weaves	10	$N_W/N = 0.085 + 0.703VR + (234.763/L) - 0.018\ \Delta S$	Primary	0.834	62	VR, L, ΔS significant

[a]Equation valid only for lengths in the range between 400 and 700 ft (122-213 m); outside this range use 85 percent of the value given by Equation 4.

L = length of the weaving section (ft),
L_H = length of the weaving section (ft 00s),
VR = ratio of weaving volume to total volume,
R = ratio of the smaller weaving volume to total weaving volume, and
$\Delta S = S_{NW} - S_W$.

Some key characteristics of these results are discussed below.

Nonweaving Vehicles

Through some of the earlier calibration attempts, two characteristics of nonweaving vehicles in weaving sections had become clear: (a) that nonweaving vehicles behave quite differently in weaving sections than they do on basic freeway sections and (b) that the relation between V_{NW} and S_{NW} appeared to be linear throughout the range of stable flow. Because of the increased level of lane changing and turbulence in the weaving area, speed is sensitive to volume levels not only as volume approaches capacity but throughout the stable flow region, as was found to be the case in basic sections. In general, a nonweaving vehicle traveling at a given speed will occupy more space in a weaving area than on a basic section--a reasonable result considering the additional turbulence caused by weaving.

Ramp Weaves Versus Major Weaves

There is a basic difference between the operational characteristics of a ramp-weave section and a major-weave section. In a ramp weave, ramp vehicles generally enter and exit at significantly reduced speeds, mainly because of restrictive ramp geometry and well-established driving habits. Thus, ramp vehicles are virtually always accelerating or decelerating through the ramp-weave section, and the average speed depends not on the competition for space between weaving and nonweaving vehicles but on the length of the section.

This creates two major difficulties. Since speed depends on length and not the results of the competition for space, it is not valid to identify operating conditions as constrained based on observations of large values of ΔS [>5 miles/h (>8 km/h)]. In shorter sections, large ΔS values will occur whether or not the section is operating in the constrained or unconstrained state.

Values of $N_W(max)$, therefore, could not be calibrated by using constrained-section data. Rather, a relation for $N_W(max)$ was needed to determine whether the segment was constrained or unconstrained. Logically, weaving vehicles in a ramp weave were substantially restricted to the use of the two lanes adjacent to the crown line. A review of the data substantiated this theory; an N_W of 2.0 appeared to be the maximum value achieved. Thus, an $N_W(max)$ of 2.0 was established for ramp-weave cases, and any case in which N_W approached or slightly exceeded this value in the data was categorized as constrained.

Thus, for ramp weaves, S_W is a function of S_{NW} and L_H, where L_H--the length of the section in hundreds of feet--does not depend on the relative presence of nonweaving flows. Furthermore, the relation is primary--that is, it does not depend on whether the section is operating in the constrained or unconstrained state.

In the case of major weaves, all vehicles usually enter and leave the section at normal freeway speeds, and little acceleration or deceleration takes place within the confines of the weaving area. Thus, unless they are prevented from doing so by a configurational constraint, weaving and nonweaving vehicles will compete for space and reach an equilibrium in which the speeds of both are reasonably similar [$\Delta S \leq 5$ miles/h (8 km/h)]. In such cases, large values of ΔS can be used to identify constrained cases.

For major weaves, then, S_W is a function of S_{NW} and VR, where VR is a measure of relative weaving and nonweaving flows. The relation is secondary, however, and holds only for unconstrained cases. Where constraints prevent the balance from being reached, the primary relation is the share-of-the-roadway equation, in which N_W/N is a function of VR, L_H, and ΔS.

$N_W(max)$ Regressions

The regression relations for $N_W(max)$ are the weakest of the set because of the small number of constrained cases available for their calibration. It might reasonably be expected that as length increases $N_W(max)$ does too, since more vehicles have the opportunity to weave via multiple lane changes. L, however, does not even enter the relation for type 1 segments and, for type 2 segments, the opposite trend is exhibited: As length increases, $N_W(max)$ decreases. A review of the data, however, confirms the latter trend. In shorter sections, the weaving turbulence is greater and nonweaving vehicles are more strongly inclined to segregate into outer lanes than they are in

longer sections. In effect, nonweaving vehicles give weaving vehicles "wider berth" in shorter sections to avoid higher turbulence levels.

Significance Levels

The F-test is used to determine whether the coefficient of a particular independent variable is significantly different from zero, in the strict statistical sense. In the development of the recalibrated procedure, four coefficients among the equations developed failed this test. Nevertheless, they are used because (a) their inclusion is necessary in order to produce a procedure capable of considering relevant demand and design variables and (b) the trends displayed in each case are physically meaningful and reasonable.

In each case, the inclusion of the variable did result in a higher multiple correlation coefficient. It is judged that a larger data base would have eliminated these F-test failures and that inclusion of the affected variables as indicated here does not pose a problem.

PROCEDURE FOR USING THE RECALIBRATED EQUATIONS

The methodology adopted for use of these equations is relatively straightforward even though it involves trial-and-error solutions. It is used only in the analysis mode; i.e., given a known situation, compute the expected speeds of weaving and nonweaving vehicles. Design is by trial and error. This is reasonable in that, for a given design, the practical value of N will be limited to two or three feasible integer values and L will be restricted to a range of about ±500 ft (±152 m). Thus, a maximum design within these limits is easily formulated and analyzed.

Before one begins the computations, two preliminary steps must be taken:

1. Convert all flows to passenger cars per hour and peak flow rates:

$$\text{Peak flow rate} = V/(PHF \times Q) \quad (11)$$

where

peak flow rate = passenger cars per hour,
V = volume (vehicles/h),
PHF = peak-hour factor, and
Q = correction factor for the combined effect of trucks, buses, and recreational vehicles on the traffic stream (6).

2. Construct a weaving diagram by using the converted peak flow rates. Compute the required parameters VR and R (as defined earlier).

Steps 1-7 below are iterative. A value of S_{NW} is assumed and then checked through successive computations. When the values agree closely [within ±2 miles/h (±3.2 km/h)], the computations are complete. It is imperative, however, that trials be conducted, starting with the high speeds. For unfamiliar users, computations may start with an assumed value of S_{NW} between 50 and 60 miles/h (80 and 96 km/h).

The steps are as follows:

1. Assume a value of S_{NW}.
2. Compute S_W by using the speed equation in Table 1 for the configuration under consideration.
3. Compute $N_W(max)$ by using the maximum-value-of-N_W equation in Table 1 for the configuration under consideration.
4. Compute N_W/N by using the share-of-the-roadway equation in Table 1 for the configuration under consideration.
5. Compute $N_W = (N_W/N) \times N$. If $N_W > N_W(max)$, the segment is constrained; go to step 6. If $N_W \leq N_W(max)$, the segment is unconstrained; go to step 7.
6. Set $N_W = N_W(max)$ and compute the resulting values of N_{NW} and N_W/N. Compute S_{NW} from Equation 1 in Table 1. Compute S_W from the primary relation for the configuration under consideration. The constrained problem is now complete.
7. Compute $N_{NW} = N - N_W$. Compute S_{NW} from Equation 1 in Table 1. If this S_{NW} is within ±2 miles/h (±3.2 km/h) of the assumed S_{NW}, the problem is complete. If it is not, assume another speed somewhat slower than the computed S_{NW} and repeat the computations.

Nomographs developed for each equation to aid in the computational procedure are shown in Figures 2-5.

SAMPLE PROBLEM

To illustrate the use of the recalibrated methodology, a sample problem is presented. Figure 6 shows a ramp-weave configuration in which the volumes shown on the weaving diagram have already been converted to peak flow rates in passenger cars per hour. The problem is to analyze the operating conditions that are expected to prevail. The procedure includes the following numbered steps:

1. Assume a value of S_{NW} = 60 miles/h (96 km/h).
2. Compute S_W by using Equation 5 in Table 1 (shown in Figure 4a): S_W = 49 miles/h (78.4 km/h).
3. Compute $N_W(max)$ by using Equation 5 in Table 1 (shown in Figure 3): $N_W(max)$ = 2.0 lanes.
4. Compute N_W/N by using Equation 8 in Table 1 (shown in Figure 5a): N_W/N = 0.26.
5. Compute N_W = 0.26 x 4 = 1.04 lanes. Is $N_W \leq N_W(max)$? If yes, the section is unconstrained; go to step 6.
6. Compute N_{NW} = 4 - 1.04 = 2.96 lanes. Compute S_{NW} from Equation 1 in Table 1 (shown in Figure 2): S_{NW} = 45 miles/h (72 km/h).

Since the computed value of S_{NW} (45 miles/h) does not closely agree with the assumed value of 60 miles/h, a second trial is obviously necessary. As indicated in the instructions for step 6, a second iteration will begin with an assumed value of S_{NW} that is somewhat lower than 45 miles/h:

1. Assume a value of S_{NW} = 42 miles/h (67.2 km/h).
2. Compute S_W by using Equation 5 in Table 1 (shown in Figure 4a): S_W = 38 miles/h (60.8 km/h).
3. Compute $N_W(max)$ = 2.0 lanes (as before).
4. Compute N_W/N by using Equation 8 in Table 1 (shown in Figure 5a): N_W/N = 0.29.
5. Compute N_W = 0.29 x 4 = 1.16 lanes. Is $N_W \leq N_W(max)$? If yes, the section is unconstrained; go to step 6.
6. Compute N_{NW} = 4 - 1.16 = 2.04 lanes. Compute S_{NW} from Equation 1 in Table 1 (shown in Figure 2): S_{NW} = 42 miles/h (67.2 km/h).

Since the agreement between the computed and assumed values of S_{NW} is within 2 miles/h (3.2 km/h) (exact in this case), the problem is complete. Operations with an average speed for nonweaving vehicles of 42 miles/h (67.2 km/h) and an average speed for

weaving vehicles of 38 miles/h (60.8 km/h) would be expected.

LEVELS OF SERVICE

Because the procedure developed includes S_W and S_{NW} as explicit parameters in equations and nomographs, the defining of criteria for levels of service becomes primarily an issue of policy. In the FHWA effort, of which the work reported here is a part, levels of service for weaving areas were defined based on the following criteria:

1. As in other parts of the FHWA work, average running speed (also called space mean speed) was used as the defining parameter.
2. Since situations can and do arise in which weaving and nonweaving vehicles experience markedly different operating conditions, levels of service should be separately assigned to describe the operation of weaving and nonweaving flows.
3. It is assumed that for a given level of service weaving drivers will tolerate average running speeds up to 5 miles/h (8 km/h) slower than the speed of nonweaving drivers because of the complexity of the required weaving maneuver. Data analysis bears out the assumption of this maximum 5-mile/h speed differential as the limit of normal unconstrained operation.

Figure 2. Speed-flow relation for nonweaving vehicles in a weaving section.

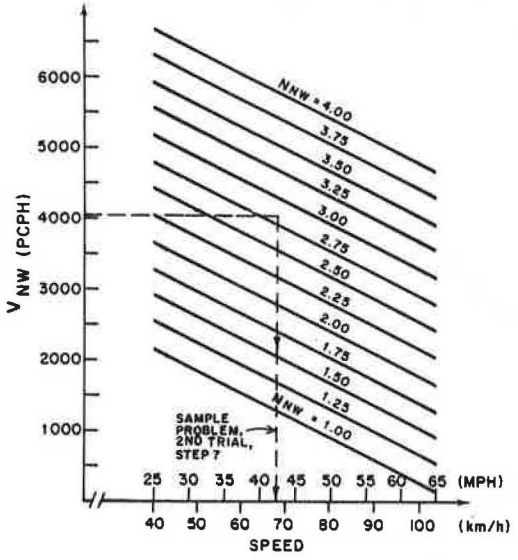


Figure 3. Maximum values of N_W in major-weave sections.

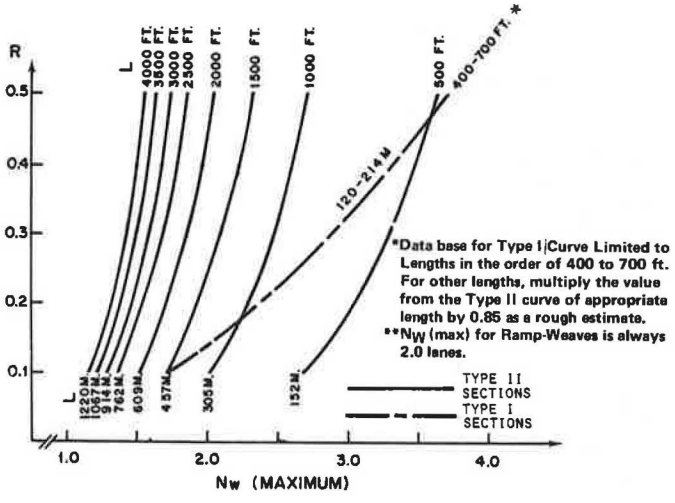
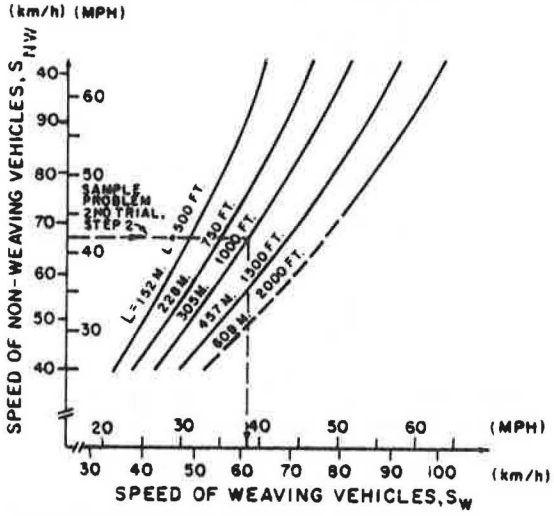
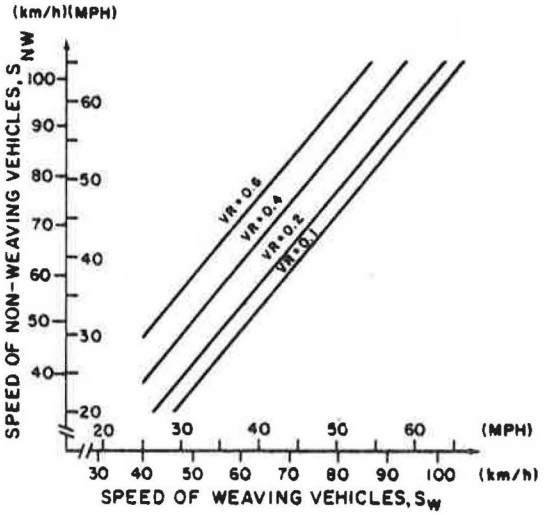


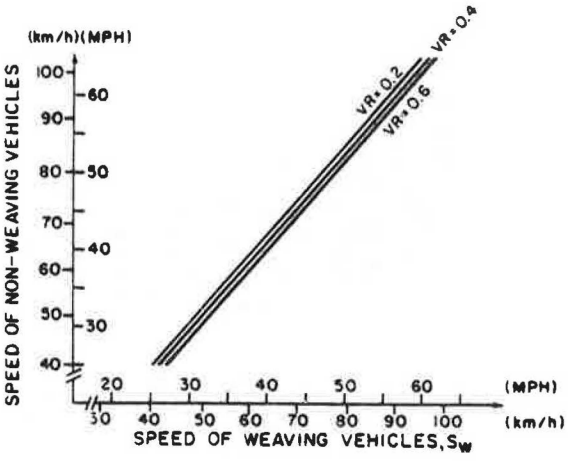
Figure 4. Speed relations for weaving configurations.



(a) RAMP-WEAVES (all Cases)

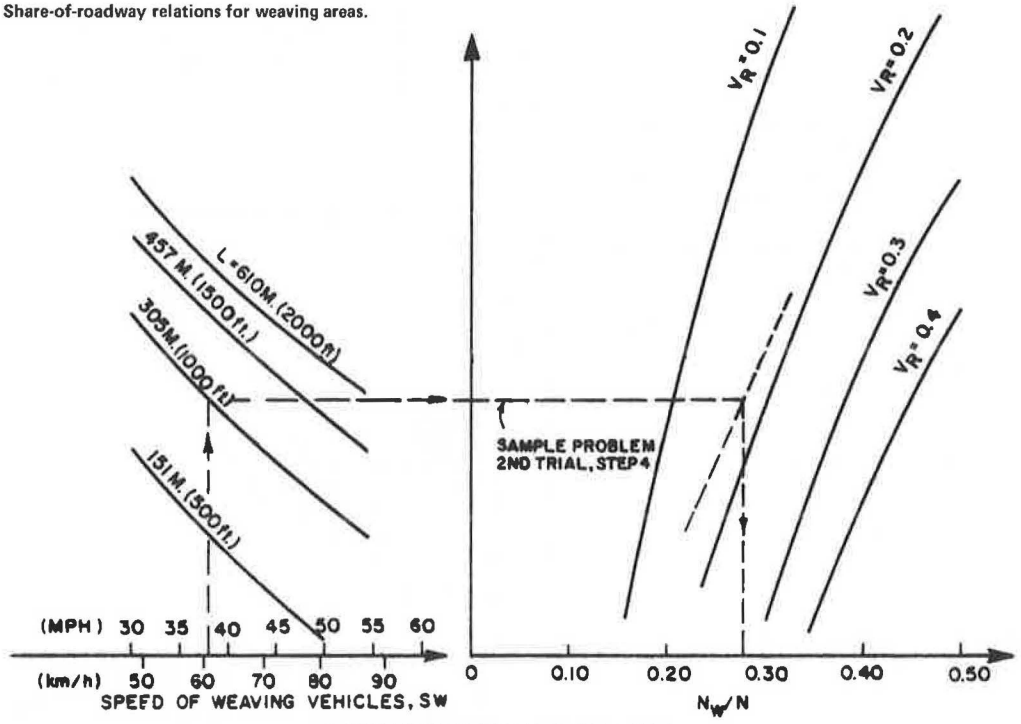


(b) TYPE I MAJOR WEAVES (Unconstrained Cases Only)

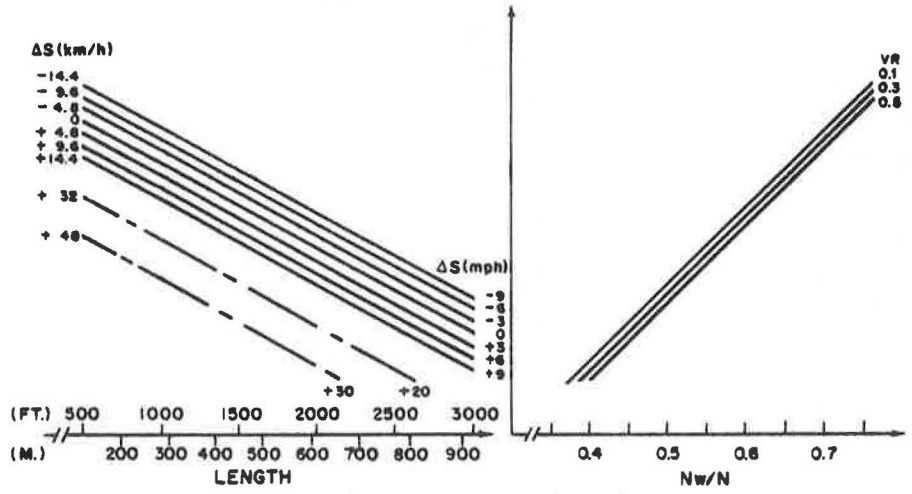


(c) TYPE II MAJOR WEAVES (Unconstrained Cases Only)

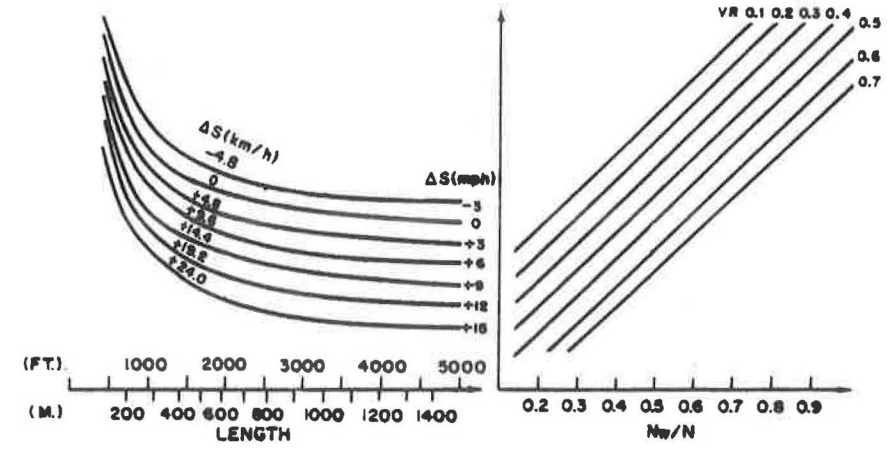
Figure 5. Share-of-roadway relations for weaving areas.



(a) RAMP-WEAVES (Unconstrained Cases Only)

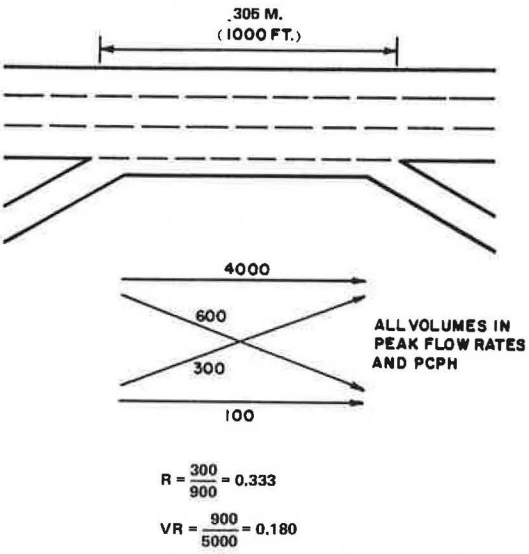


(b) TYPE I MAJOR WEAVES (All Cases)



(c) TYPE II MAJOR WEAVES (All Cases)

Figure 6. Sample problem.



$$R = \frac{300}{900} = 0.333$$

$$VR = \frac{900}{5000} = 0.180$$

4. It is further assumed that nonweaving drivers would expect primarily speeds equal to those experienced on basic freeway segments for a given level of service.

The resulting criteria for levels of service in weaving areas are given below (1 mile/h = 1.6 km/h):

Level of Service	Avg S_{NW} (miles/h)
A	≥50
B	≥45
C	≥40
D	≥35
E	≥30
F	<30

Level of Service for Weaving Vehicles Versus That for Nonweaving Vehicles	ΔS (miles/h)
Same	≤5
One level poorer	≤10
Two levels poorer	≤15
Three levels poorer	≤20
Four levels poorer	≤25

For the sample problem described earlier, the level of service for weaving as well as nonweaving vehicles is C. Had the problem been solved by using the methodology in Chapter 7 of the HCM, the results would be quality of flow III, which predicts operating speeds of about 40 miles/h (64 km/h) for weaving vehicles, and level of service C, which predicts operating speeds for all vehicles of more than 50 miles/h (80 km/h). The solution given here indicates significantly lower speeds, particularly for nonweaving traffic. Furthermore, whereas the HCM speed predictions seem to indicate substantial imbalance between weaving and nonweaving speeds, the solution given here clearly indicates a balanced operation.

VALIDATION

Because of the number of cells into which the data base was divided, there were not sufficient data to withhold cases for validation. The results were, however, checked with those given by the original NCHRP procedure, which had been validated. The results compare favorably: The new procedure is more sensitive to configurational variables than the NCHRP procedure.

An opportunity to apply the procedure to an external case has presented itself through work currently being done by the firm of Howard, Needles, Tammen, and Bergendoff under an FHWA contract to redesign a weaving area on the Shirley Highway (I-95) between the Capital Beltway (I-495) and the Springfield Interchange. The recalibration procedure was applied, and it accurately predicted the existing breakdown conditions in the area (the HCM method indicated level of service D). It was also used to evaluate a series of alternatives that were under consideration and highlighted the importance of configuration to the case.

The complete procedures for basic freeway segments, weaving areas, and ramp junctions developed as a result of the FHWA contract on freeway capacity-analysis procedures have been published elsewhere (7).

ACKNOWLEDGMENT

We wish to express our appreciation to the many individuals who have contributed their advice and judgment to this effort, particularly members of the Transportation Research Board Committee on Highway Capacity and Quality of Service, who reviewed materials and provided a vital discussion forum throughout the project. We also appreciate the cooperation of M. Gerston of Howard, Needles, Tammen, and Bergendoff, who provided us an opportunity to make a validation study by using a project of that firm. We would like to acknowledge the helpful cooperation of Harry Skinner of FHWA, the technical monitor, and the other members of the FHWA advisory panel--A.A. Carter, Jr., W. Prosser, and D. Merritt--throughout the course of this work.

The views and opinions expressed here are ours. The procedures described do not represent standards or methods endorsed by the Federal Highway Administration or any other agency.

REFERENCES

1. L.J. Pignataro and others. Weaving Areas: Design and Analysis. NCHRP, Rept. 159, 1976.
2. L.J. Pignataro and others. Weaving Area Operations Study. NCHRP, Project 3-15, Final Rept., April 1973.
3. Highway Capacity Manual. HRB, Special Rept. 87, 1965.
4. R.P. Roess, W.R. McShane, and L.J. Pignataro. Configuration, Design, and Analysis of Weaving Sections. TRB, Transportation Research Record 489, 1974, pp. 1-12.
5. E.M. Linzer, R.P. Roess, and W.R. McShane. Effect of Trucks, Buses, and Recreational Vehicles on Freeway Capacity and Service Volume. TRB, Transportation Research Record 699, 1979, pp. 17-26.
6. Interim Materials on Highway Capacity. TRB, Transportation Research Circular 212, Jan. 1980.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

Development of Modified Procedures for Analysis of Ramp Capacity

ROGER P. ROESS

As part of an overall effort sponsored by the Federal Highway Administration to update and revise freeway capacity analysis procedures, Highway Capacity Manual procedures for ramps and ramp junctions were revised in order to (a) eliminate the dual procedure for differing levels of service; (b) eliminate cases in which on ramps are followed by off ramps and both are joined by an auxiliary lane, cases that are better treated as weaving sections; (c) adjust criteria to reflect passenger cars per hour rather than a vehicle population with 5 percent trucks; (d) update information on trucks in lane 1 of the freeway; (e) add material on left-handed ramps, ramps on 10-lane freeway segments, and ramps proper; and (f) add illustrative material on the impact of ramp geometry and acceleration-lane design. It is believed that the modifications recommended significantly simplify the use of, and eliminate many potential inconsistencies in, existing Highway Capacity Manual procedures.

There has been little in the way of new, basic research on the subject of ramp capacity by which to update procedures in the 1965 Highway Capacity Manual (1). Indeed, the only significant data base available for study is that collected by the then U.S. Bureau of Public Roads (BPR) and used in calibrating existing Highway Capacity Manual (HCM) procedures.

Nevertheless, in the course of an effort sponsored by the Federal Highway Administration (FHWA) to update and revise freeway-related elements of the HCM, it was considered necessary to modify existing procedures for freeway-ramp junctions to take into account the following important factors:

1. The format of the HCM ramp procedures is somewhat confusing because of the existence of different methods for levels of service A through C and levels of service D and E and because of the use of a large number of nomographs for various geometric configurations.
2. The use in the HCM of 5 percent trucks as a base vehicle population complicates computations and is inconsistent with other freeway-related parts of the manual.
3. HCM ramp procedures are affected by weaving-area procedures adopted elsewhere in the FHWA effort; the weaving procedures recommended are based on the National Cooperative Highway Research Program (NCHRP) method (2), which incorporates several geometric configurations now treated by using ramp techniques in the HCM.
4. Since the development of the 1965 HCM, some new material has been developed that permits treatment of cases and aspects not covered by HCM procedures.

In the light of these factors, HCM ramp procedures were examined for potential format modifications and simplifications and the addition of more recent material where it is available.

BASIC PROCEDURE

In considering a basic format for the presentation of ramp capacity procedures, three alternatives were examined:

1. The HCM procedure specified for levels of service A through C, based on a series of regression relations for various geometric configurations (18 relations, depicted in nomographs, were developed from the BPR data base referred to earlier);
2. The HCM procedure specified for level of service E (also used for level D), based on a limited data base collected in the state of California and often referred to as the California procedure or--in recognition of its developer, the late Karl Moskowitz--the Moskowitz procedure (the data base for this procedure is no longer extant); and
3. The ramp procedure developed by Leisch in 1974 for FHWA (3) (based on the 1965 HCM, this procedure represents a radically different format, using multistep nomographs and a reduced number of different geometric cases).

The first major issue that must be examined is the need and justification for the dual procedures in the 1965 HCM for different levels of service. Examination of the BPR data base used for the HCM regression-based technique reveals that it contains data for all levels of service and is therefore valid throughout the full range of stable traffic flow. Furthermore, in the course of the Weaving Area Operations Study conducted at the Polytechnic Institute of New York for NCHRP, it was found that, throughout the full range of service levels, for the cases examined, the procedure for levels of service A through C was more accurate in the prediction of operating conditions than the procedure for level of service E.

Thus, it appeared that the only reasonable choice was between the HCM regression format and the Leisch format (3), which was based on the same data base. The California procedure was rejected as a basic technique but, because it is not configuration specific and can be applied to all cases, it was retained as a gross estimator for configurations and situations not covered by other methods.

The issue of whether to adopt the HCM format or the Leisch format was a difficult one. The Leisch format significantly reduced the number of different cases to be considered and might be thought to be simpler in application than the HCM format. On the other hand, the accuracy of the HCM relations was verified by the FHWA project team whereas information that would permit similar verification of the Leisch format was not available. In conducting comparative problem solutions, some degree of precision appeared to be lost by using the Leisch format. The FHWA project team therefore recommended the continued use of the HCM format (regression procedure) for ramp analysis, for the following reasons:

1. When problems are solved for level of service by using both the HCM and Leisch methods, results differ by one level in about 35 percent of the cases tried. Although these cases were most often borderline and the percentage difference in a numeric parameter, such as ramp volume (V_R), was generally less than 10 percent, this was considered a significant problem. Since the purpose of such procedures is often to find level of service, the loss of precision in the Leisch format, although not large in percentage terms, does lead to the step-function errors cited above.
2. The Leisch technique treats V_R as the

principal dependent variable. I, and other participants in the FHWA effort, felt that V_R is virtually always a demand value input into an analysis rather than the output of analysis. Philosophically, it was our view that, when a ramp design is considered, other general freeway features, ramp location, and demand volumes are known quantities. The most common use, therefore, is to solve for level of service to see if a given design will work. If not, the ramp location and/or design would have to be reconsidered. Only where ramp controls are being considered would V_R be the most probable output of computations.

3. In conversations with professionals in the field, opinion was mixed as to which format was preferable. Some supported the Leisch technique strongly, whereas others felt that it was somewhat complex. Overall, there was no strong preference among professionals for either technique.

4. The complexity of the HCM format is primarily the result of its presentation, which can be considerably simplified by eliminating ramp-weave cases and the duality of procedure for various service levels, as discussed previously.

5. The HCM methodology is a step-wise procedure that allows, indeed forces, the user to consider intermediate results and values. Members of the FHWA project team felt that this was essential to ensure the reasonableness of the results (it helps catch errors) and to give maximum insight into the analysis of field conditions.

The Leisch format would be useful as a computational aid if the computations could be simplified for the procedure adopted. However, the FHWA project team felt that, for the reasons stated above, it should not itself be adopted as a revised HCM technique.

PROCEDURAL MODIFICATIONS

The procedure for analysis of ramp-freeway junctions involves the determination of volumes in lane 1 of the freeway, just prior to a merge or diverge point. The regression-based procedure of the 1965 HCM contains 18 nomographs or equations to solve for this lane 1 volume (V_1) for various ramp configurations. Of these, 5 can be eliminated, since they represent configurations that should be treated as weaving sections. The recommended procedure, therefore, retains 13 nomographs for the solution of V_1, as well as the California procedure for cases not covered by the nomographs.

The procedural steps and approach of the 1965 HCM (for levels of service A through C) have been retained. That procedure is used in the analysis mode in the following steps:

1. Establish all geometrics and demand volumes for the case to be considered, including upstream and downstream adjacent ramps and volumes.
2. Compute V_1 by using one of 13 nomographs, or estimate it by using the California procedure.
3. Find the percentage of trucks in lane 1 (a nomograph is provided).
4. Convert V_1, V_f, and V_r to peak flow rates (passenger cars per hour) by using appropriate truck factors and the peak-hour factor (PHF).
5. Compute checkpoint volumes: $V_m = V_1 + V_r$ and $V_d = V_1$ (V_f is taken immediately after merge and before diverge).
6. Compare V_f, V_m, and V_d with level-of-service criteria to determine level of service.

The modifications made to the 1965 HCM procedure in an attempt to reduce its complexity are discussed below.

Nomographs

A key element in the ramp-freeway-junction methodology is the determination of which nomograph to use to find V_1. The HCM provides an index to aid in this selection. A similar index was prepared for the revised procedure. This index simplifies the HCM format by (a) eliminating five configurations now treated as weaving sections, (b) including several modifying notes in the HCM text on the index itself, (c) eliminating the ambiguity between the nomograph and California procedures, and (d) organizing the index according to "first ramp" and "second ramp" rather than on ramp and off ramp, as in the HCM, which creates confusion if both ramps are of the same type.

The nomographs themselves have been redrawn in a clarified format, and instructions for their use and special notations on modifications in special circumstances are highlighted. Figure 1 shows the new format developed. The sample nomograph applies to single-lane on ramps on six-lane freeways where there are adjacent upstream on ramps, with or without acceleration lanes. The normal range of the parameters is V_f = 1800-5400 vehicles/h, V_r = 100-1500 vehicles/h, V_u = 100-1400 vehicles/h, and D_u = 500-1500 ft. The steps in the solution are as follows:

1. Draw a line from the V_f value to the V_r value, intersecting turning line 1.
2. Draw a line from the V_u value to the D_u value, intersecting turning line 2.
3. Draw a line from the intersection on turning line 1 to the intersection on turning line 2, and read the solution on the V_1 line.

Equations for each of the nomographs referred to in Table 1 are given in Table 2, where

- V_1 = lane 1 volume immediately upstream of an on ramp or off ramp,
- V_f = total freeway volume immediately upstream of an on ramp or off ramp,
- V_r = ramp volume for the ramp under consideration,
- D_u = distance from ramp under consideration to adjacent upstream on ramp or off ramp (ft),
- V_u = ramp volume for adjacent on ramp or off ramp upstream of ramp under consideration,
- V_d = ramp volume for adjacent on ramp or off ramp downstream of ramp under consideration,
- D_d = distance from ramp under consideration to adjacent downstream on ramp or off ramp (ft), and
- V_c = center-lane volume that divides at a major diverge point.

Trucks in Lane 1

Figure 8.22 of the HCM is used to compute the percentage of trucks in lane 1. Data available through the NCHRP Weaving Area Operations Study allowed for a partial recalibration of this figure, which was shown in that study to be grossly inaccurate for eight-lane freeways. The recalibrated curve is shown in Figure 2.

Left-Hand Ramps and Ramps on 10-Lane Freeway Segments

The 1965 HCM does not contain any material on left-hand ramps or ramps on 10-lane freeway segments. This is a serious lack, for the incidence

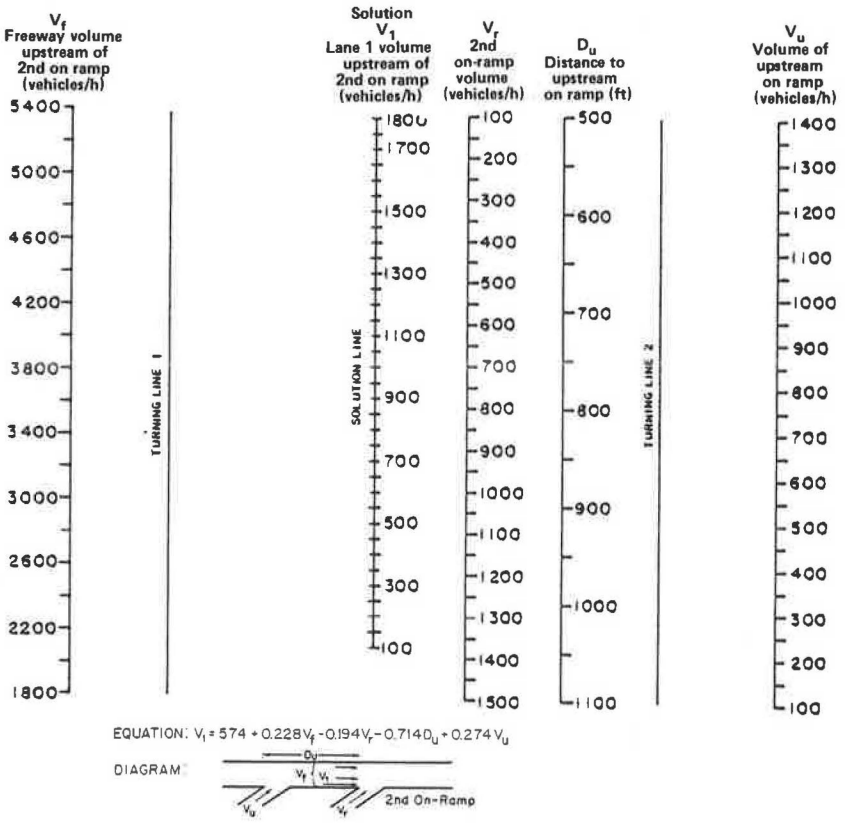


Figure 1. Sample nomograph for determining lane 1 volume upstream of one-lane on ramps on six-lane freeways with upstream on ramps.

of both has increased since 1965. No new data are available for such situations, but Leisch (3) developed approximation procedures for both cases, and these were adopted "as is" for the current work.

Ramp Geometrics

One of the difficulties of the HCM is that it does not account for the effect on performance of such variables as the length of the acceleration or deceleration lane, the angle of ramp convergence with the freeway, and relative grades. Little work has been done in this area, and incorporation of such variables into a working procedure is not yet possible.

Drew (5) has studied the effect on on-ramp merging performance of angle of convergence and acceleration-lane length. By using gap-acceptance models, Drew evaluated the impact of these variables on the percentage of gaps accepted by drivers, using a base, or "ideal", case of a 1200-ft acceleration lane and a 2° angle of convergence. Although interesting for its insights, however, Drew's work cannot be directly incorporated into a capacity analysis methodology because the ideal case adopted by Drew cannot be identified in existing data or procedures and Drew's use of the concept of capacity is not synonymous with the HCM's traditional use and implies, but does not define, a measure of service quality.

Table 3 was developed from Drew's work to show the potential impact of angle of convergence and acceleration-lane length on performance. It is not intended for use as a computational device.

The table does, however, indicate the extreme importance of these factors to ramp operation. The fact that these are not included in current procedures may be a serious deficiency that should be addressed through basic research.

LEVELS OF SERVICE FOR RAMP-FREEWAY JUNCTIONS

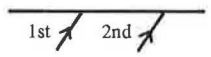
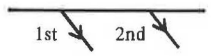
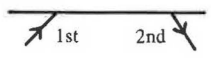
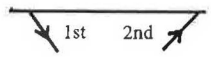
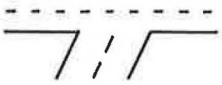
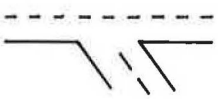
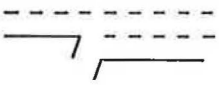
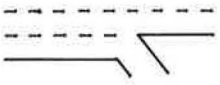
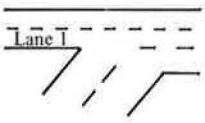
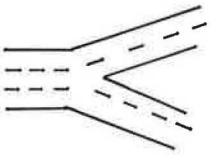
Table 8.1 of the HCM gives criteria for ramp-junction level of service in terms of limiting values of V_m (merge volume), V_d (diverge volume), and V_w (weaving volume per 500 ft of distance). Operating criteria associated with these conditions are vague and are not clearly defined in terms of speeds or other operating parameters. It is implied that these limiting volumes are such that the indicated level of service would prevail on the freeway as a whole, as defined in HCM Table 9.1, for the condition described.

This leads to an immediate problem in the current context. As part of the FHWA effort, Table 9.1 of the HCM has been modified (6). This would suggest that corresponding values of V_m, V_d, and V_w in HCM Table 8.1 should also be recalibrated. Unfortunately, there is no sound basis or data that would make it possible to accomplish this.

Table 4 compares HCM Table 9.1 service volumes for freeways (expressed on a per-lane basis) and HCM Table 8.1 criteria for V_m, V_d, and V_w. This comparison reflects the view of HCM developers that point flows (merge, diverge, and weave) could be higher than single-lane service volumes for a given level of service because of the restrictive area of their influence.

Subsequent research has shown that the extent of the influence of "point" flows can be quite extended, often as much as 5000-6000 ft of highway at poorer levels of service (7). In view of this, it was considered unwise to continue this policy. The criteria recommended for V_d, V_m, and V_w are given in Table 5 and are compared with the criteria recommended for normal freeway conditions as a result of modifying HCM Table 9.1. The selection of these criteria was somewhat judgmental and was based on the following considerations:

Table 1. Index to nomographs and procedures for analysis of ramp terminals.

Configuration	Diagram	Four-Lane Freeway		Six-Lane Freeway		Eight-Lane Freeway	
		First Ramp	Second Ramp	First Ramp	Second Ramp	First Ramp	Second Ramp
Isolated one-lane on ramp		Figure A4.1	–	Figure A4.6	–	Figure A4.9	–
Isolated one-lane off ramp		Figure A4.2	–	Figure A4.7	–	Approximate by using Table 4.3 and Figure 4.3	
Adjacent one-lane on ramps		Figure A4.1	Figure A4.5	Figure A4.6	Figure A4.8	Approximate by using Table 4.3 and Figure 4.3	Approximate by using Table 4.3 and Figure 4.3
Adjacent one-lane off ramps		Figure A4.2[a]	Figure A4.2	Figure A4.7[a]	Figure A4.7	Approximate by using Table 4.3 and Figure 4.3	Approximate by using Table 4.3 and Figure 4.3
On ramp followed by off ramp		Figure A4.1	Figure A4.3	Figure A4.6	Figure A4.7	Figure A4.10	Approximate by using Table 4.3 and Figure 4.3
Off ramp followed by on ramp		Treat as isolated ramps			Figure A4.6	Treat as isolated ramps	
Loop ramps		Figure A4.4	Figure A4.3	Figure A4.6	Figure A4.7	Figure A4.10	Approximate by using Table 4.3 and Figure 4.3
Two-lane on ramps		NA	–	Figure A4.11		NA	–
Two-lane off ramps		–	NA	–	Figure A4.12	–	NA
Addition of lane at on ramp		Merge criteria in Table 4.1 may be applied directly to on-ramp volume as a checkpoint					
Dropping of lane at off ramp		Diverge criteria in Table 4.1 may be applied directly to off-ramp volume as a checkpoint					
Major junctions		Assume that ramp lane B carries an amount of traffic equal to merge checkpoint volume in Table 4.1 for assumed level of service; ramp lane A then carries remaining ramp traffic; compute lane 1 volume by using Figure A4.1 (four-lane freeway), A4.6 (six-lane freeway), or A4.9 (eight-lane freeway), entering with V_r = ramp lane A volume; find checkpoint levels of service; continue computations until assumed level agrees with results					
Major forks (diverges)		NA	–	Figure A4.13	–	NA	–

[a]Use this figure to find V_1 in advance of the first ramp, but enter with a V_r equal to the total volume on both off ramps. This technique is valid where the distance between ramps is less than 800 ft. Where the distance between ramps is between 800 and 4000 ft, use Table 4.3 and Figure 4.3 to approximate the situation. Where the distance between ramps is greater than 4000 ft, the ramps are treated as if they were independent (isolated).

1. It was felt that neither merge-, diverge-, nor weaving-volume criteria for a given level of service should be higher than the average per-lane service volume for that level on basic freeway sections. Given the additional turbulence involved in these maneuvers, and given that the criteria should be such that the freeway as a whole operates at the stated level, V_m, V_d, and V_w criteria should be somewhat lower than corresponding criteria for V_f.

2. In general, a diverge movement is less disruptive to flow than a merge. Therefore, merge-volume criteria should be more restrictive than diverge criteria for any given level of service.

3. The weaving criterion in HCM Table 9.1 appears to be entirely too high. Although this criterion is only used in cases of an on ramp followed by an off ramp without an auxiliary lane, it should be somewhat in line with weaving volumes predicted by the weaving methodology developed for the FHWA effort (4), tempered by the fact that in such sections weaving is really a merge followed by a diverge movement.

4. Criteria should be expressed in passenger cars per hour for consistency with other freeway procedures. This is not a major problem, since the nomographs for solution of V_1 were calibrated and are solved by using mixed vehicles per hour with whatever percentage of trucks exists. Conversions to a base population (5 percent trucks in the HCM) are required only to compare with level-of-service criteria. Thus, the nomograph relations need not be adjusted--only the level-of-service criteria, to reflect passenger cars per hour.

5. For consistency with other material developed for the FHWA, criteria represent peak flow rates (or PHF = 1.00) rather than full-hour volumes.

Clearly, the subject of criteria for ramp level of service is one that should be carefully studied in the future. The recommendations given here are reasonable and consistent. In view of the lack of hard data for analysis, they result from the only

Table 2. Equations for nomographs given in Table 1.

Nomograph	Equation	Notes
A4.1	$V_1 = 136 + 0.345V_f - 0.115V_r$	Not for use if upstream adjacent on ramp exists within 200 ft
A4.2	$V_1 = 165 + 0.345V_f + 0.520V_r$	Not for use if upstream adjacent on ramp exists within 3200 ft
A4.3	$V_1 = 202 + 0.362V_f + 0.496V_r - 0.069D_u + 0.096V_u$	For use only if upstream adjacent on ramp exists within 3200 ft
A4.4	$V_1 = 166 + 0.280V_f$ (for $V_r < 600$ vehicles/h) $V_1 = 128 + 0.482V_f - 0.301V$ (for V_r between 600–1200 vehicles/h)	For use with loop ramps only
A4.5	$V_1 = 123 + 0.376V_f - 0.142V_r$	For use only if upstream adjacent on ramp exists within 2000 ft
A4.6	$V_1 = -121 + 0.244V_f - 0.085V_u + 640\ V_d/D_d$	V_u refers to an upstream off ramp within 2600 ft; if none exists, set $V_u = 50$ V_d/D_d refers to a downstream off ramp within 5700 ft; if none exists, set 640 $V_d/D_d = 5$
A4.7	$V_1 = 94 + 0.231V_f + 0.473V_r + 214\ V_u/D_u$	V_u/D_u refers to an upstream on ramp within 5700 ft; if none exists set 215 $V_u/D_u = 2$
A4.8	$V_1 = 574 + 0.228V_f - 0.194V_r - 0.714D_u + 0.274V_u$	For use only if upstream on ramp exists within 1400 ft
A4.9	$V_1 = -312 + 0.201V_f + 0.117V_r$	For use only if there is no adjacent downstream off ramp within 3000 ft
A4.10	$V_1 = -353 + 0.199V_f - 0.057V_r - 0.486V_d$	For use if downstream adjacent off ramp exists within 1500-3000 ft
A4.11	$V_1 = 54 + 0.070V_f + 0.049V_r$	—
A4.12	—	Special case (4)
A4.13	$V_c = 64 + 0.285V_f + 0.041V_r$	V_c is center-lane volume on six-lane freeway just prior to major diverge into two four-lane freeways

Table 3. Effect of ramp geometrics on gaps accepted by merging vehicles.

Angle of Convergence (°)	Percentage of Ideal Case by Length of Acceleration Lane				
	1200 ft	1000 ft	800 ft	600 ft	400 ft
2	100.0	96.8	90.3	64.5	32.3
4	80.6	77.4	48.4	32.3	17.7
6	45.2	45.2	32.3	24.2	11.3
8	33.8	33.8	25.8	15.5	9.7
10	32.3	32.3	24.5	13.5	8.1

Table 4. Relation between freeway and ramp level of service in 1965 HCM.

Level of Service	Maximum Allowable Value[a]			
	V_f (passenger cars/h/lane)[b,c]	V_m (vehicles/h)[d]	V_d (vehicles/h)[d]	V_w (vehicles/h)[d]
A	800	1000	1100	800
B	1167	1200	1300	1000
C	1600	1700	1800	1450
D	1800	1800	1900	1800
E	2000	2000	2000	2000
F	—	—	—	—

[a]For PHF = 1.00.
[b]For six-lane freeways, 70-mile/h average highway speed.
[c]HCM Table 9.1.
[d]HCM Table 8.1.

Table 5. Relation between freeway and ramp level-of-service criteria recommended in FHWA study.

Level of Service	Maximum Allowable Value[a]			
	V_f (passenger cars/h/lane)[b]	V_m (passenger cars/h)	V_d (passenger cars/h)	V_w (passenger cars/h)
A	800	750	800	500
B	1300	1200	1300	700
C	1700	1550	1650	1300
D	1925	1800	1900	1550
E	2000	2000	2000	2000
F	—	—	—	—

[a]For PHF = 1.00.
[b]For six-lane freeways, 70-mile/h average highway speed.

Figure 2. Percentage of trucks in lane 1.

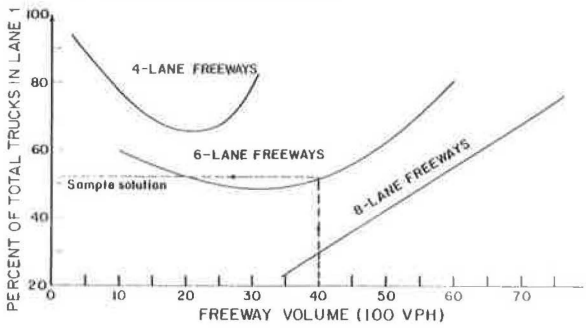
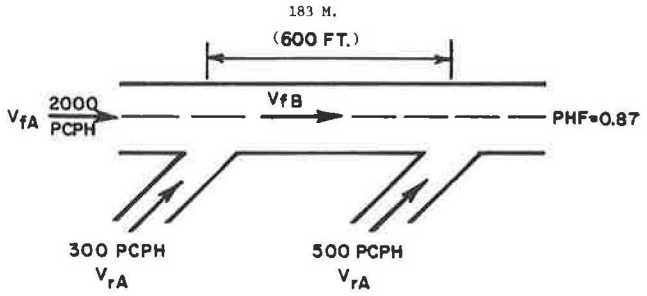


Table 6. Approximate service volumes for single-lane ramps.

Ramp Design Speed (miles/h)	Service Volume by Level of Service (passenger cars/h)				
	A	B	C	D	E
<20	—[a]	—[a]	—[a]	—[a]	1250
20-29	—[a]	—[a]	—[a]	1025	1450
30-39	—[a]	—[a]	1125	1200	1600
40-49	—[a]	1000	1250	1325	1650
≥50	700	1050	1300	1500	1700

Note: Level of service F varies widely.
For two-lane ramps, multiply the given values by 1.7 for ≤ 20 miles/h, by 1.8 for 20-29 miles/h, by 1.9 for 30-49 miles/h, and by 2.0 for ≥50 miles/h.
[a]Level of service not achievable because of restricted design speed.

Figure 3. Illustration for ramp-freeway-terminal problem.



type of approach that can currently be taken.

DEVELOPMENT OF CRITERIA FOR RAMPS PROPER

The 1965 HCM treats only the capacity analysis of ramp-freeway terminals. Material from the American Association of State Highway Officials "Blue Book" (8) was adapted by Leisch (3) to yield the capacity of a ramp based on its design speed. These values were used in this work and were further modified to provide approximate level-of-service guidelines. Level-of-service criteria were developed from capacity figures by (a) assuming that the percentage of capacity for a given ramp level of service should be similar to the percentage of capacity for the same level on basic freeway sections and (b) not allowing for better levels of service on ramps that have a restricted design speed, a safety consideration modeled after similar use in basic freeway criteria, where better levels of service are not allowed for highways with a reduced average highway speed.

The criteria developed are given in Table 6.

RAMP METERING

The subject of ramp metering has become one of increasing importance in recent years, since many urban areas attempt to deal with problems of freeway congestion through the application of systems and methodologies of freeway surveillance and control. Although much study has been devoted to the control aspects of these systems (9), virtually no work has been done on their capacity implications.

Two positions can be taken regarding the capacity implications of ramp metering:

1. Ramp metering has no impact on capacity or service volume. It serves merely to limit demand so that breakdown does not occur and/or certain prescribed levels of service do not deteriorate.

2. Because ramp metering not only limits total demand but also smooths that demand by preventing simultaneous multivehicle entries, it can potentially alter the basic nature and characteristics of merging maneuvers and thus alter basic capacities and/or service volumes.

In the work discussed here, the first position was adopted merely because there are no available data or research on which to base an evaluation of the second position. The FHWA project team believes that some original research will be required to resolve the issue.

Because of the importance of ramp metering, material was developed that treats the subject qualitatively. A special procedure for computing V_r as the dependent variable in an analysis is also given. On the subject of ramp metering, it is reasonable to ask, What limiting value of V_r can be permitted to enter the freeway without causing level of service to be poorer than a given level? The answer can be found by (a) assuming a value for V_r, (b) computing V_1 (by using nomographs or California-procedure approximation), (c) finding the limiting value of V_m for the level of service of interest, and (d) computing $V_r = V_m - V_1$. The procedure is iterated until the V_r assumed is reasonably equivalent to the V_r computed.

In summary, because no substantial new work on the subject of ramps has taken place since the publication of the 1965 HCM, the procedures developed for ramps as part of the FHWA project on freeway capacity-analysis procedures were the results of simplifications and special-case extensions of the HCM. Ramp capacity, therefore, is a subject that will require substantial research if significant improvements on current techniques are to be expected for the mid-1980s edition of the HCM.

SAMPLE PROBLEMS

The simple problems discussed below illustrate the use of some of the revised procedures discussed in this paper.

Analysis of Ramp-Freeway Junctions

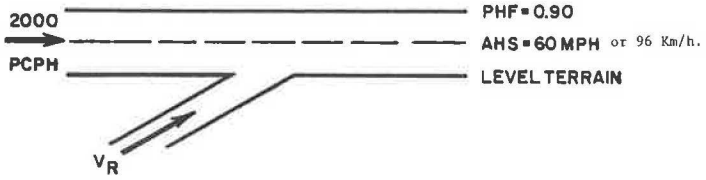
To determine the expected level of service provided at the two ramp-freeway junctions shown in Figure 3, follow the steps given below:

1. Establish all geometrics and demand volumes. This has been done in the problem statement.
2. Compute V_1. Table 1 indicates that Figure A4.1 should be used to compute V_1 for the first on ramp and Figure A4.5 for the second (the equations for these figures are found in Table 2):

$V_{1A} = 136 + 0.345\ V_{fA} - 0.115\ V_{rA}$.
$V_{1A} = 136 + 0.345\ (2000) - 0.115\ (300)$.
$V_{1A} = 136 + 690 - 35 = 791$ passenger cars/h.
$V_{1B} = 123 + 0.376\ V_{fB} - 0.142\ V_{rB}$.
$V_{1B} = 123 + 0.376\ (2300) - 0.142\ (500)$.
$V_{1B} = 123 + 865 - 71 = 917$ passenger cars/h.

3. Find the percentage of trucks in lane 1 (since all volumes are given in passenger cars per hour, this step is not necessary).
4. Convert volumes to peak flow rates in passenger cars per hour. Since volumes are already given in these terms, peak flow rates in passenger cars per hour are computed by dividing by the PHF (0.87 in this case): $V_{fA} = 2000/0.87 = 2299$, $V_{fB} = 2300/0.87 = 2644$, $V_{rA} = 300/0.87 = 345$, $V_{rB} = 500/0.87 = 575$, $V_{1A} = 791/0.87 = 909$, and $V_{1B} = 917/0.87 = 1054$.

Figure 4. Illustration for ramp-metering problem.



5. Compute checkpoint volumes and compare with criteria in Table 5:

Checkpoint Volume	Level of Service
V_{mA} = 345 + 909 = 1254 passenger cars/h	C
V_{mB} = 575 + 1054 = 1629 passenger cars/h	D
V_{fA} = 2299/2 = 1150 passenger cars/h/l	B
V_{fB} = 2644/2 = 1322 passenger cars/h/l	C

Clearly, the second merge is the critical operating element that causes the overall performance level to be level of service D.

Analysis of Ramp Metering

It is desired that V_r be controlled by establishing a maximum flow rate through ramp metering at the location shown in Figure 4. If a fixed-time ramp meter is used, at what rate should ramp vehicles be allowed to enter the traffic stream if the level of service is not to be permitted to be worse than C?

The question asks for a solution of a maximum value of V_r so that the level of service is C. The trial-and-error method described in the previous discussion of ramp metering is used.

From Table 5, the maximum merge service volume for level of service C is 1550 passenger cars/h (peak flow rate). For a PHF of 0.90, this is equivalent to an hourly volume of 1550 x 0.90 = 1395 passenger cars/h. Considering the situation given in the problem, a tabular computation can be set up as follows [V_1 is computed from Figure A4.1 (4), and the formula for the computed V_r is 1395 - V_1]:

Assumed V_θ	V_1	Computed V_θ
200	810	585
400	775	620
600	760	635
650	750	645

A metering rate of 650 passenger cars/h, or one car every 3600/650 = 5.54 s--say, 5.5 s--would be set.

A more precise solution can be found by using the equation for Figure A4.1 directly:

$$V_1 = 136 + 0.345\,V_f - 0.115\,V_r \quad (1)$$

and considering that V_r = 1395 - V_1. Substituting for V_1 in the V_r equation,

$V_r = 1395 - (136 + 0.345\,V_f - 0.115\,V_r)$.
$V_r = 1259 - 0.345\,V_f + 0.115\,V_f$.
$0.885\,V_r = 1259 - 0.345\,V_f$.
$V_r = (1259 - 0.345\,V_f)/0.885$.
$V_r = [1259 - 0.345(2000)]/0.885 = 643$ passenger cars/h.

CONCLUSIONS

It is believed that the modifications to HCM procedures for analysis of ramp capacity reported in this paper both simplify and clarify the application of those procedures. It is hoped, however, that new, basic research in the area of ramp capacity will be possible before the publication of the new HCM in the mid-1980s.

ACKNOWLEDGMENT

I wish to express my thanks to members of the Transportation Research Board Committee on Highway Capacity and Quality of Service, whose members have provided a forum for discussion and have taken the time to review the many documents produced as a result of the FHWA project effort. Their input has meaningfully improved the results of the effort and has been of great value to the project team. I would also like to thank Art Carter, Bill Prosser, and Dave Merritt, members of the FHWA project review panel, and Harry Skinner, the project monitor, for their advice and support throughout the effort.

The views and opinions expressed here are mine. The procedures described do not represent standards or methods endorsed by FHWA or any other agency.

REFERENCES

1. Highway Capacity Manual. HRB, Special Rept. 87, 1965.
2. L. J. Pignataro and others. Weaving Areas: Design and Analysis. NCHRP, Rept. 159, 1976.
3. J. Leisch. Capacity Analysis Techniques for Design and Operation of Freeway Facilities. Federal Highway Administration, U.S. Department of Transportation, 1974.
4. R. P. Roess and others. Freeway Capacity Procedures. In Interim Materials on Highway Capacity, TRB, Transportation Research Circular 212, Jan. 1980.
5. D. R. Drew. Traffic Flow Theory and Control. McGraw-Hill, New York, 1968.
6. R. P. Roess and others. Freeway Level of Service: A Revised Approach. TRB, Transportation Research Record 699, 1979, pp. 7-16.
7. R. D. Worrall, A. G. Bullen, and Y. Gur. An Elementary Stochastic Model of Lane-Changing on a Multilane Highway. HRB, Highway Research Record 308, 1970, pp. 1-12.
8. A Policy on Geometric Design of Rural Highways. AASHO, Washington, DC, 1965.
9. P. Everall. Urban Freeway Surveillance and Control: The State of the Art. Federal Highway Administration, U.S. Department of Transportation, June 1973.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

Effect of On-Street Pickup and Delivery on Level of Service of Arterial Streets

PHILIP A. HABIB

The effect of on-street pickup and delivery (PUD) of freight on arterial streets in central areas is analyzed. The data used were collected throughout the country and included PUD demand, parking patterns, and the impact of lane blockages on traffic. The purpose of the analysis was to develop a method by which on-street PUD operations could be incorporated into level-of-service analysis as well as to provide a mechanism by which to evaluate alternative goods-movement strategies. The findings of the research indicate that lane blockages by PUD vehicles, even at very low demand levels, have a considerable impact on traffic speed and level of service. Tables and charts by which this impact can be determined are presented, and some preliminary strategies that could improve arterial street performance are proposed.

The on-street pickup and delivery (PUD) of freight is the major mode of accomplishing the transfer of materials in central areas. The vehicles used are primarily small and intermediate-sized trucks, which park at the curb or double-park as close to the destination as possible. Several cities, such as Boston and Phoenix, have alleys from which goods can be transferred. In most cities, however, this transfer occurs either at curbside on an artery or off the street in a loading dock or parking lot.

This paper presents an analysis of the effect of curbside and double-parked PUD operations on traffic speed and level of service for the type of streets on which these operations are dominant--arterial streets of two or more lanes per direction. The research on which this paper is based is drawn from a comprehensive study of on-street PUD operations including PUD demand production, parking patterns, and characteristic freight distribution patterns as well as traffic impact. Data for the study were collected in various cities in the United States.

Figure 1. Hypothesized relation between volume and speed reduction.

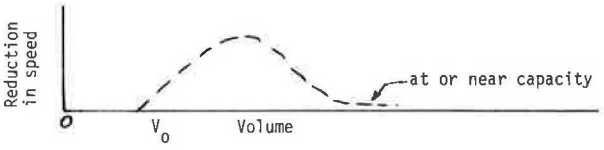


Table 1. NETSIM validation cases.

City	No. of Lanes per Direction	Direction	Vehicles per Hour of Green per Lane[a]
Dallas	3	One-way	1100
San Francisco	2	One-way	1000
	2	One-way	1100
	2	One-way	1000
	2	One-way	800
White Plains, NY	2	Two-way	700
	2	Two-way	650
New York City	3	One-way	800
	2	Two-way	950
	2	Two-way	1000

[a]Represents volume conditions during blockage.

STUDY DESIGN

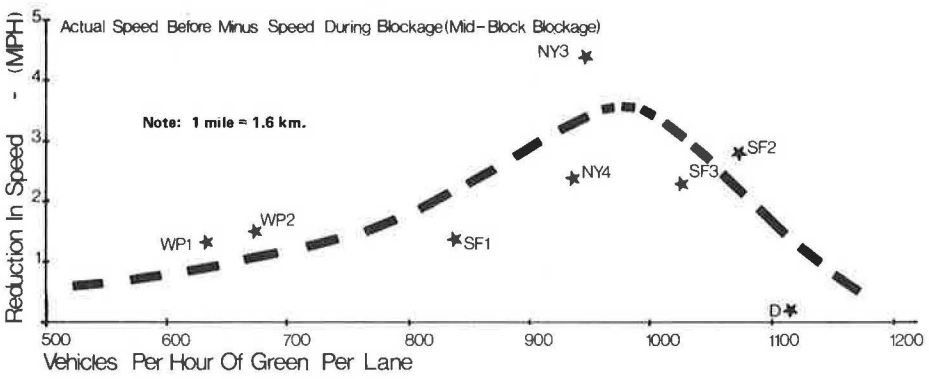
It is hypothesized that, at a certain low volume of traffic, a lane blockage does not affect flow speed. Furthermore, at very high traffic volumes (near capacity), the intersection so controls the flow that a blockage (except one right at the downstream intersection approach) does not affect speed. Figure 1 shows hypothesized volume versus speed reduction for a lane blockage: V_0 would be a threshold volume below which there would be no apparent effect.

The study design for determination of traffic impact was to collect data on traffic blockages by PUD vehicles and use these data to calibrate NETSIM, a widely accepted traffic simulator. Then NETSIM was used to simulate the effect of PUD blockages under various traffic and street conditions. Since the study is restricted to arterial facilities of >2 lanes/direction, the findings would only apply to these larger facilities.

Data Collection

Field data for the study were collected by using 1-s time-lapse photography with a Super 8-mm camera mounted on rooftops, in office buildings, on fire escapes, and in other locations. Approximately 25 blockage cases were filmed throughout the country. Ten cases were selected for use for NETSIM calibration. The selection was based on volume level, duration and type of blockage, number of lanes, signal control, and the degree of influence the blockage had on street performance. Table 1 gives the 10 selected cases.

Figure 2. Speed reduction caused by lane blockage in the various cities studied.



Travel-time data were collected from the films over the affected section (usually one block). Stops were not recorded because of the difficulty of accurately counting locked-wheel stops at a time lapse of 1 s.

Figure 2 shows the impact on speed caused by the lane blockage. It should be noted here that none of the cases shown is an approach-lane blockage at the downstream intersection. Thus, in the cases shown, the intersection characteristics still control the capacity of the block, and it would be expected that speed reduction resulting from the blockage would be directed toward zero as volume approaches capacity (1400-1500 vehicles/h of green/lane).

The principal reason for not considering filmed cases was that the volume past the blockage was too low to provide a measurable impact. The second reason for eliminating certain cases was that other environmental factors--such as cross-street congestion, automobile double parking or other-side-of-street parking, and bus breakdown in the moving lane--controlled arterial flow more than did the lane blockage.

Figure 3. Family of blockage configurations (* denotes location of lane blockage).

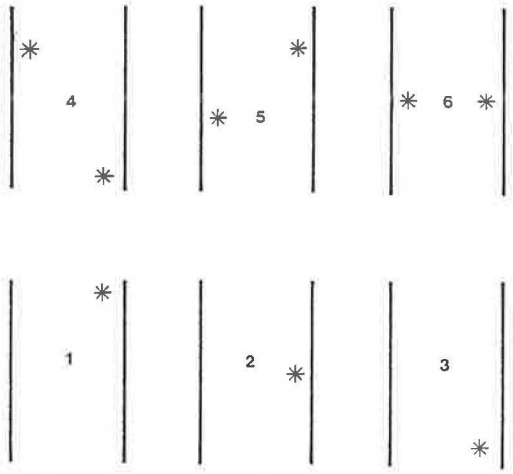
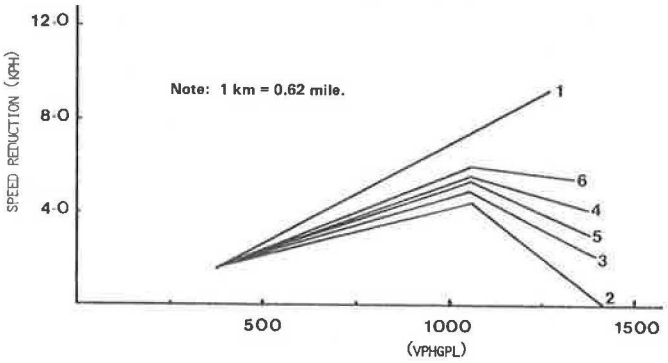


Figure 4. Speed reduction for various blockage configurations.



Analysis Method

NETSIM simulations were designed and run for each case given in Table 1, and relations were developed between simulated conditions and actual field conditions, including signal progression and split, turning volumes, and pedestrian interference levels. Since NETSIM was not created to simulate lane blockages by trucks, the development of relations between simulated and actual conditions was essential in interpreting the further simulations that would be made for the various traffic volumes, street sizes and types, and lane-blockage configurations.

On a typical one- or two-way arterial street, there are a variety of blockage configurations. For instance, on a one-way facility there could be a blockage upstream, midblock, or downstream or a downstream blockage on one side and a midblock blockage on the other side. There are, for all practical purposes, 64 different permutations of blockage configurations on a one-way street and 6 on a two-way street (in one direction), if one considers each block face to be made of three "cells" (upstream, midblock, or downstream) in which a blockage can occur. Each blockage configuration will affect traffic differently.

The analysis process defined six blockage configurations on a one-way street and three

Table 2. Speed reduction for one-way arterials for various levels of PUD double-parking demand on both sides of street.

Volume (vehicles/h of green/lane)	Speed Reduction by No. of PUDs per Hour (km/h)				
	6 PUDs	12 PUDs	24 PUDs	36 PUDs	48 PUDs
500	2.1	2.7	2.9	3.0	2.9
550	2.4	3.0	3.2	3.4	3.3
600	2.6	3.3	3.6	3.7	3.7
650	2.9	3.6	3.9	4.1	4.1
700	3.1	3.9	4.2	4.5	4.5
750	3.4	4.2	4.6	4.9	4.9
800	3.6	4.5	4.9	5.3	5.3
850	3.9	4.8	5.2	5.6	5.8
900	4.1	5.1	5.5	6.0	6.2
950	4.4	5.4	5.9	6.4	6.6
1000	4.7	5.5	5.9	6.5	6.8
1050	4.7	5.5	6.0	6.7	7.1
1100	4.6	5.5	6.1	6.9	7.4
1150	4.5	5.5	6.2	7.2	7.7
1200	4.5	5.5	6.2	7.4	8.1
1250	4.4	5.5	6.3	7.6	8.4
1300	4.3	5.5	6.4	7.8	8.7
1350	4.3	5.5	6.5	8.1	9.0
1400	4.2	5.5	6.6	8.3	9.3

Note: 1 km = 0.62 mile.

Table 3. Speed reduction of two-way arterials for various levels of PUD double-parking demand on one side of street.

Volume (vehicles/h of green/lane)	Speed Reduction by No. of PUDs per Hour (km/h)				
	3 PUDs	6 PUDs	12 PUDs	18 PUDs	24 PUDs
500	1.4	2.2	2.7	2.8	2.8
550	1.6	2.4	3.0	3.2	3.2
600	1.8	2.7	3.4	3.6	3.6
650	1.9	2.9	3.7	3.9	4.0
700	2.1	3.2	4.0	4.3	4.4
750	2.3	3.4	4.3	4.7	4.8
800	2.4	3.6	4.6	5.0	5.2
850	2.6	3.9	5.0	5.4	5.6
900	2.8	4.1	5.3	5.8	6.0
950	2.9	4.4	5.6	6.1	6.4
1000	3.3	4.8	6.0	6.4	6.7
1050	3.2	4.7	6.0	6.6	7.0
1100	3.1	4.6	6.0	6.7	7.3
1150	3.0	4.5	6.0	6.9	7.5
1200	2.9	4.4	6.0	7.0	7.8
1250	2.8	4.3	6.0	7.2	8.0
1300	2.7	4.2	6.0	7.3	8.3
1350	2.6	4.1	6.1	7.4	8.6
1400	2.5	4.0	6.1	7.6	8.8

Note: 1 km = 0.62 mile.

blockage configurations on a two-way street to which all configurations could be reduced for traffic impact purposes. These are shown in Figure 3. As the block face is "divided" into three cells (upstream, midblock, or downstream) a multivehicle blockage in one cell would be represented by one blockage in that cell. The combination of blockages in different cells defines the configurations.

The various configurations shown in Figure 3 were simulated by NETSIM under various traffic volumes, and the results were adjusted by using data from field test cases. A standard block length of 122 m (400 ft) was used, and typical arterial street progression, traffic composition, and turning conditions were assumed. The characteristics of the simulations and the ranges studied are given below:

Item	Range
Number of moving lanes	2, 3, 4
Directions	One-way, two-way
Volume/capacity ratio	0.5, 0.7, 0.8, 0.85, 0.9
Blockage durations (min)	3, 7, 12, 20, 30
Intersection characteristics	10 percent right and left turns; moderate pedestrian volume; green time/cycle length ratio = 0.5
Block length (m)	122
Other considerations	NETSIM does not consider (a) parking versus no parking (only moving lanes) or (b) variations in lane widths; 5 percent trucks

Figure 4 shows the speed-reduction relation developed for the various possible blockage configurations.

Analysis of blockage cases from the field as well as of the NETSIM results showed that there was no consistent representation of the effect of the size of arterial, over the range studied (2-4 lanes/direction), on reduction in speed caused by lane blockage. This insensitivity to number of lanes was also found in a previous study (1). The findings in Figure 4 are therefore presented as a function of vehicles per hour of green per lane.

The probability of an arterial block being in any of the 64 (reduced to 6) configuration states is a function of the demand of PUD vehicles on that block. Therefore, theoretical probability matrices were developed for each configuration state under different PUD demand levels. As demand rises, the probability of the most severe configuration states also rises.

Determining Speed Reduction and Level of Service

Tables 2 and 3 give the expected reduction in speed for a variety of traffic volumes and levels of PUD double-parking demand for one-way and two-way operation on arterial streets. Figures 5 and 6 show the expected resultant level of service on the arterial street under various levels of traffic and PUD demand. These levels of service would be more appropriate descriptors for an arterial segment (several blocks) than for an individual block, since a random PUD arrival pattern based on a uniform distribution along the block was assumed in the analysis process. That is, if a block has a very large downstream generator and little or no generation elsewhere, the tables given would underestimate the impact. On the other hand, should that major generator be midblock, an underestimate would be expected.

The level of service was determined by combining the no-blockage volume-speed curve from the simulation results with a general relation between arterial volume and level of service (2). Figure 7 shows this combination. The method used to find the level of service caused by PUD blockages is as follows: The hypothetical volume determines the no-blockage speed, which in turn is reduced by appropriate values from Table 2 (or Table 3). The resultant location of the point defines the appropriate impact level of service.

Figure 5. Level of service for lane blockages on one-way arterials.

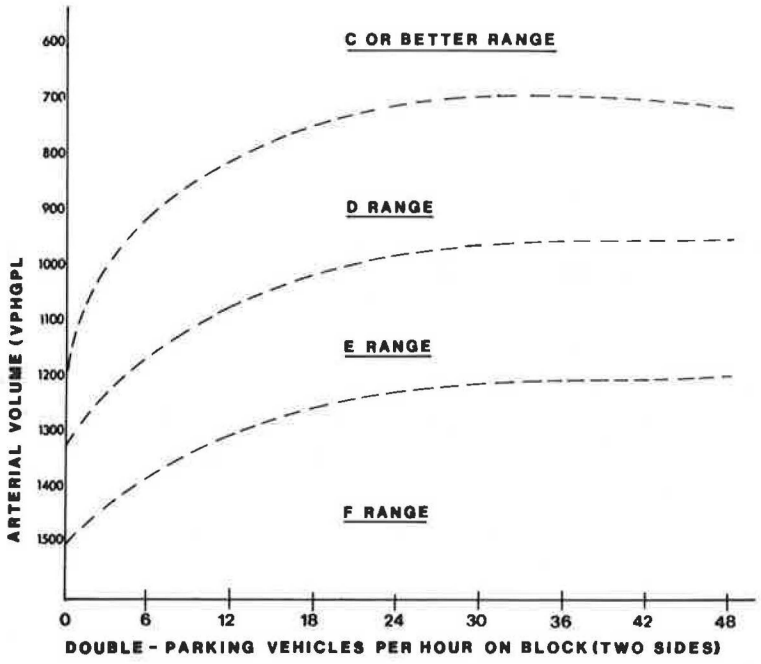


Figure 6. Level of service for lane blockages on two-way arterials.

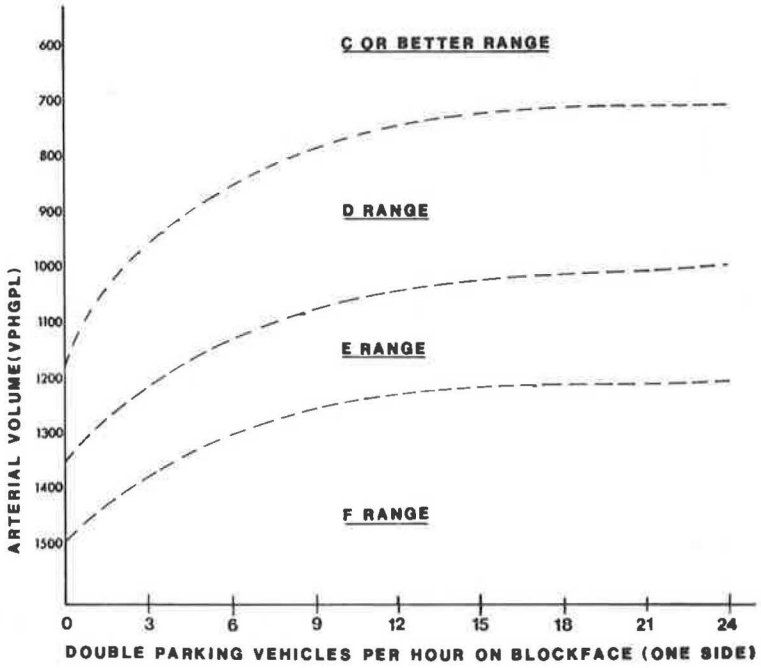
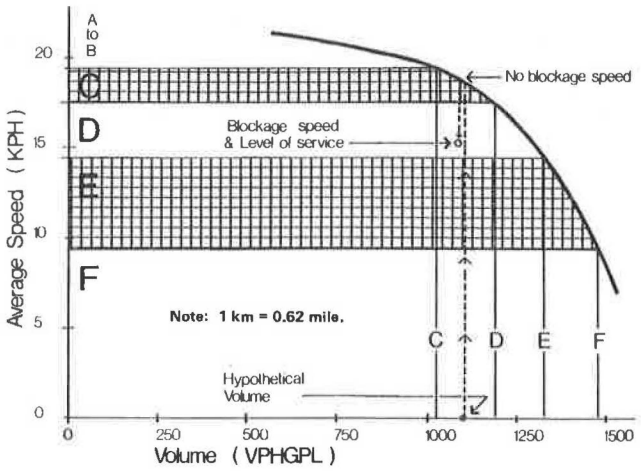


Figure 7. Determination of level of service.



ASSESSMENT OF RESULTS

Figures 5 and 6 show that PUD activities can have a significant detrimental effect on level of service on the arterial street. It should be noted that a single 122-m (400-ft) block face can process 40 PUD operations in a blocked lane. Thus, the PUD demand levels shown are not worst-case conditions of PUD demand. In addition, initial work in this area (2) showed no measurable relation between block length and traffic impact from PUD operations over the range of typical arterial blocks [100-200 m (325-650 ft)]. Therefore, the rates of PUD double-parking demand given in this paper are PUD demand/122 m/h. This implies that the results would be transferable to other block lengths without much loss of accuracy.

Various studies have shown that PUD vehicles will double-park where curbside parking is not available or park in a curbside moving lane when no parking or standing is allowed. The percentage of such vehicles relative to total block PUD generation is 30-40 percent, a sizable amount.

When the amount of PUD activity is very small, the maximum throughput of a downtown arterial street will be reduced from about 1450 to about 1325 vehicles/h of green/lane (more likely about 1200). This reduction of 9-18 percent must be viewed as the minimum effect of PUD operations on system throughput. The range of effects grows larger as PUD demand grows.

Because PUD activity is regular and not a random occurrence, drivers expect certain delays and avoid specific streets as not being on the minimum-impedance path from origin to destination. Therefore, since the volume never reaches the hypothesized level unless this arterial section remains on the minimum path, the hypothetical impact may never occur. However, traffic engineers who calculate downtown street capacity without giving a great deal of attention to the assessment of the generation of on-street PUD operations are avoiding a key determinant of street-segment capacity and quality of flow.

OPTIONS TO IMPROVE LEVEL OF SERVICE

The way to improve traffic performance is to minimize conflict. To do this, the traffic engineer or planner should be fully aware of the elemental characteristics of good traffic movement. A traffic signal cannot be timed adequately unless an elemental characteristic of pedestrians--walking speed--is known. An intersection cannot be analyzed unless various elemental characteristics of the traffic stream--volume, speed, and discharge rates--are known. Knowledge of the elemental characteristics of goods movement--demand, arrival patterns, and parking patterns--seems to me to be the basic element in improving level of service. Acquisition of this knowledge and understanding by traffic engineers and planners will result in the ability to develop realistic strategies by which to lessen conflicts and improve level of service at specific problem locations.

Some of the strategies that I perceive to be

realistic and that will be tested during further phases of the research from which this paper is drawn are

1. Reducing approach-lane blockages by various means, such as short loading zones, relocation of bus stops to the near side, placement of hydrants (if flexibility exists), and provision of loading space at corners on cross streets;
2. Altering the PUD demand pattern (through enforcement, consolidation, and other means) to reduce conflicts in selected time periods;
3. Providing (by purchase or rent) off-street loading space (lots) for PUD vehicles on the most critical arterial sections; and
4. Restriping selected arterials to allow for 4.5- to 5.0-m (15- to 16-ft) double-parking curb lanes where possible [a facility with three 3.7-m (12-ft) travel lanes and two 4-m (13-ft) curb lanes could be restriped to be three 3-m (10-ft) travel lanes and two 5-m (16-ft) curb lanes].

There are more simple and more exotic strategies to be used in addressing the problem of PUD impact on traffic flow. This paper, which is based on a limited amount of data, presents an initial, systematic way of relating the variables of urban goods movement to the variables of arterial traffic flow. In a time when maximum efficiency is being sought for existing traffic facilities, the proper recognition and treatment of on-street goods movement, as presented here and in future research efforts, can go a long way toward achieving that goal.

ACKNOWLEDGMENT

The research reported in this paper is part of the Federal Highway Administration (FHWA) program on metropolitan multimodal traffic management and was performed with the support and under the supervision of the FHWA Office of Research. The conclusions reported are mine and do not necessarily reflect the opinions of FHWA.

REFERENCES

1. K. Crowley and P. Habib. Facilitation of Urban Goods Movement: Year 2 Final Report. Office of University Research, U.S. Department of Transportation, 1975.
2. Interim Materials on Highway Capacity. TRB, Transportation Research Circular 212, Jan. 1980.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.