TRANSPORTATION RESEARCH RECORD 971

Traffic Capacity and Characteristics

TRB
TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1984

TRANSPORTATION RESEARCH BOARD
National Research Council

ERRATA 1982-1985

Special Report 201

page 17, column 2, second paragraph, should read
"Tools also need to change as the nature of options changes significantly. Emerging policy options are not largely focused on network-expansion investments, whereas traditional models were developed long ago to deal with such options."

Special Report 200

page 3, column 1
Change the caption for the bottom figure to
"A new AM General trolley bus starts down the 18 percent grade on Queen Anne Avenue North in Seattle in October 1979 (photograph by J. P. Aurelius)".

Transportation Research Record 1040

page ii
Under "Library of Congress Cataloging-in-Publication Data," delete "Meeting (64th: 1985: Washington, D.C.)" and "ISBN 0361-1981"

Transportation Research Record 1020

page 7, Figure 1
The histogram should reflect that the rail mode is represented by the black bar and that the highway mode is represented by the white bar.

Transportation Research Record 1017

page 19, column 1, 7 lines above Table 1
Change "ranged from 1 $in.^2$ to nearly 30 $in.^2$ of runoff" to "ranged from 1 area inch to nearly 30 area inches of runoff"

page 22, column 1, last line
Change "1 to nearly 30 $in.^2$" to "1 to nearly 30 area inches"

page 22, column 2, first line
Change "13 $in.^2$" to "13 area inches"

Transportation Research Record 1011

page 12, Figure 4
Figure does not show right-of-way structure for O-Bahn. See discussion on page 11, column 1, paragraph 3.

Transportation Research Record 996

page 49
Insert the following note to Figure 2:
"The contour lines connect points of equal candlepower."

page 49
Insert the following note to Figure 3:

"The candlepower contours are superimposed on a 'headlight's-eye-view' of a road scene. The candlepower directed at any point in the scene is given by the particular candlepower contour light that overlays that point. For example, 1400 candlepower is directed at points on the pedestrian's upper torso. For points between contour lines, it is necessary to interpolate."

page 50
Insert the following note to Figure 3:

"Where

ρ = the azimuth angle from the driver's eye to a point P on the pavement;
θ = the elevation angle from the driver's eye to a point P on the pavement;
EZ = the driver's eye height above the pavement; and
DX, DY, DZ = the longitudinal, horizontal, and vertical distance between the headlamp and eye point.

Then

$EX = EZ/\text{Tan}\,\theta$ $\quad HZ = EX\text{-}DZ$
$H1^2 = EZ^2 + EX^2$ $\quad HX = EX\text{-}DX$
$EY = H1\ \text{Tan}\,\rho$ $\quad HY = EY\text{-}DY$
$H2^2 = H1^2 + EY^2$ $\quad H3^2 = HX^2 + HZ^2$
$\alpha = \text{Tan}^{-1}\,(HZ/HX)$, $\beta = \text{Tan}^{-1}\,(HY/H3)$, $H4^2 = H3^2 + HY^2$"

Transportation Research Record 972

page 30, column 2, 22 lines up from bottom
Reference number (5) should be deleted

page 31, column 2, 5 lines up from bottom
Reference number should be 5, not 4

page 34, column 2, 8 lines above References
Reference number (5) should be deleted

Transportation Research Record 971

page 31, reference 3
Change to read as follows:

Merkblatt für Lichtsignalanlagen an Landstrassen, Ausgabe 1972. Forschungsgesellschaft für das Strassenwesen, Köln, Federal Republic of Germany, 1972.

Transportation Research Record 965

page 34, column 1, Equation 1
Change equation to

$$r_u = \gamma_w \cdot h/\gamma \cdot z$$

where

γ_w = unit weight of water,
γ = moist unit weight of soil,
h = piezometric head, and
z = vertical thickness of slide.

Transportation Research Record 905

page 60, column 1, 9 lines up from bottom
Change "by Payne (6)" to " by us"

Transportation Research Record 819

page 47, Table 1

Replace with the following table.

Table 1. Summary of interactions between signal-timing parameters and MOEs.

Timing Method	Parameter	Total Delay	Stops	Fuel Consumption	Emissions HC	Emissions CO	Emissions NO_x
Manual	Cycle length	⊕	⊕	⊕	⊕	⊕	⊕
	Speed of progression	+	⊕	+	+	+	+
	Priority policy	+	+	+	+	+	+
	Split method	+					
TRANSYT	Cycle length	⊕	⊕	⊕	⊕	⊕	⊕
	K-factor	+	⊕				
	Priority policy				+		

Note: + = main effect detected from TRANSYT output, and ○ = main effect detected from NETSIM output.

Transportation Research Record 869

page 54, authors' names

The second author's name should read "Edmond Chin-Ping Chang"

Transportation Research Record 847

page 50, Figure 3

Add the following numbers under each block in the last line of the flowchart:

R1, R2, R3, R4, D1, D2, D3, A1, A2

page 50, Figure 4

Make the following changes in the last line of the flowchart.

Change "R4" to "D1" and "Recognition" to "Decision"

Change "R5" to "D2" and "Recognition" to "Decision"

Change "R6" to "D3" and "Recognition" to "Decision"

Change "R7" to "R4"

Change "R8" to "D4" and "Recognition" to "Decision"

Change "R9" to "A1" and "Recognition" to "Action"

Change "R10" to "A2" and "Recognition" to "Action"

Transportation Research Record 840

page 25, column 1, line 5

Change "money" to "model"

Transportation Research Record 831

page ii, column 1

Change ISBN number to "ISBN 0-309-03308-X"

Transportation Research Circular 255

page 6, column 1, third paragraph

Change "Marquette University" to "Northern Michigan University"

NCHRP Synthesis of Highway Practice 87

page ii

Change ISBN number to 0-309-03305-5

NCHRP Synthesis of Highway Practice 84

page ii

Change ISBN number to 0-309-03273-3

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

ADDRESS CORRECTION REQUESTED

Transportation Research Record 971
Price: $18.20
Editor: Elizabeth W. Kaplan
Compositor: Lucinda Reeder

mode
1 highway transportation

subject area
55 traffic flow, capacity, and measurements

Transportation Research Board publications are available by ordering directly from TRB. They may also be obtained on a regular basis through organizational or individual affiliation with TRB; affiliates or library subscribers are eligible for substantial discounts. For further information, write to the Transportation Research Board, National Research Council, 2101 Constitution Avenue, N.W., Washington, D.C. 20418.

Printed in the United States of America

Library of Congress Cataloging in Publication Data
National Research Council. Transportation Research Board.
Traffic capacity and characteristics.

(Transportation research record; 971)
1. Highway capacity–Congresses. 2. Traffic flow–Congresses. I. National Research Council (U.S.). Transportation Research Board. II. Series.
TE7.H5 no. 971 380'.5 s 84-27254 [HE336.H48]
[388.3'14] ISBN 0-309-03753-0 ISSN 0361-1981

Authors of the Papers in This Record

Auslander, David M., Department of Mechanical Engineering, University of California, Berkeley, Calif. 94720
Babcock, Philip S. IV, Department of Mechanical Engineering, University of California, Berkeley, Calif. 94720
Bakare, Adebayo B., Civil Engineering Department, Northwestern University, Evanston, Ill. 60201
Ballard, John L., Industrial and Management Systems Engineering Department, University of Nebraska-Lincoln, Lincoln, Neb. 68588-0531
Goble, John G., Industrial and Management Systems Engineering Department, University of Nebraska-Lincoln, Lincoln, Neb. 68588-0531
Haden, Richard J., Department of Transportation, City of Lincoln, Neb. 68508
Herman, Robert, Department of Civil Engineering and Center for Studies in Statistical Mechanics, University of Texas, Austin, Tex. 78712
Hoban, C. J., Australian Road Research Board, 500 Burwood Highway, Vermont South, Victoria, Australia, 3133
Hurdle, V. F., Department of Civil Engineering, University of Toronto, Toronto, Ontario, M5S 1A4 Canada
Jovanis, Paul P., Civil Engineering Department, Northwestern University, Evanston, Ill. 60201
Lee, Clyde E., Department of Civil Engineering, University of Texas, Austin, Tex. 78712
Lee, Fong-Ping, Department of Civil Engineering, University of Texas, Austin, Tex. 78712
Legere, Jay F., BDM Corporation, 7915 Jones Branch Drive, McLean, Va. 22102
Lin, Feng-Bor, Department of Civil and Environmental Engineering, Clarkson College of Technology, Potsdam, N.Y. 13676
Mahmassani, Hani, Department of Civil Engineering, University of Texas, Austin, Tex. 78712
May, Adolf D., Institute of Transportation Studies, University of California, Berkeley, Calif. 94720
McCoy, Patrick T., Civil Engineering Department, University of Nebraska-Lincoln, Lincoln, Neb. 68588-0531
Messer, Carroll J., Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843
Michalopoulos, Panos G., Department of Civil and Mineral Engineering, University of Minnesota, Minneapolis, Minn. 55455-0220
Mounce, John M., Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843
Ou, Fong-Lieh, Forest Service, U.S. Department of Agriculture, P.O. Box 2417, Washington, D.C. 20013
Payne, Harold J., VERAC, Inc., P.O. Box 26669, San Diego, Calif. 92126-0669
Percy, Martin C., New York State Department of Transportation, 317 Washington Street, Watertown, N.Y. 13601
Pienaar, Wessel J., Institute of Transportation Studies, University of California, Berkeley, Calif. 94720
Radwan, A. Essam, Civil Engineering Department, Arizona State University, Tempe, Ariz. 85287
Roess, Roger P., Polytechnic Institute of New York, 333 Jay Street, Brooklyn, N.Y. 11201
Rose, Cecil A., Institute of Transportation Studies, University of California, Berkeley, Calif. 94720
Sananez, Juan C., Institute of Transportation Studies, University of California, Berkeley, Calif. 94720
Shortreed, John H., Department of Civil Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1 Canada
Stokes, Robert W., Texas Transportation Institute, Texas A&M University, College Station, Tex. 77843
Tomizuka, Masayoshi, Department of Mechanical Engineering, University of California, Berkeley, Calif. 94720
Van Aerde, Michel, Department of Civil Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1 Canada
Williams, James C., Department of Civil Engineering, University of Texas, Austin, Tex. 78712
Yagar, Sam, Department of Civil Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1 Canada

Contents

LEVEL OF SERVICE CONCEPTS: DEVELOPMENT, PHILOSOPHIES, AND IMPLICATIONS
Roger P. Roess 1

PASSENGER CAR EQUIVALENTS FOR UNINTERRUPTED FLOW: REVISION OF CIRCULAR 212 VALUES
Roger P. Roess and Carroll J. Messer 7

LEFT-TURN EQUIVALENCIES FOR OPPOSED, SHARED, LEFT-TURN LANES
John H. Shortreed 14

ANALYSIS OF UNSIGNALIZED INTERSECTION CAPACITY
Adebayo B. Bakare and Paul P. Jovanis 21

USE OF THE NCHRP SIGNALIZED INTERSECTION CAPACITY METHOD–A SOUTH AFRICAN EXPERIENCE
Adolf D. May, Wessel J. Pienaar, and Cecil A. Rose 32

EVALUATING CAPACITIES OF ONE-LANE ROADS WITH TURNOUTS
Fong-Lieh Ou 40

COMPUTER SIMULATION TO COMPARE FREEWAY IMPROVEMENTS (Abridgment)
Robert W. Stokes and John M. Mounce 45

DEVELOPMENT AND APPLICATION OF A MACROSCOPIC MODEL FOR RURAL HIGHWAYS
Juan C. Sananez and Adolf D. May 49

CAPACITY, SPEED, AND PLATOONING VEHICLE EQUIVALENTS FOR TWO-LANE RURAL HIGHWAYS
Michel Van Aerde and Sam Yagar 58
Discussion
Myung-Soon Chang 65
Authors' Closure 65

DYNAMIC FREEWAY SIMULATION PROGRAM FOR PERSONAL COMPUTERS
Panos G. Michalopoulos 68

ROLE OF ADAPTIVE DISCRETIZATION IN A FREEWAY SIMULATION MODEL
Philip S. Babcock IV, David M. Auslander, Masayoshi Tomizuka, and Adolf D. May 80

SIMULATION STUDY OF GUIDELINES FOR RURAL ROAD IMPROVEMENTS
C. J. Hoban 93

SIGNALIZED INTERSECTION DELAY MODELS–A PRIMER FOR THE UNINITIATED
V. F. Hurdle 96

AUTOMATED COLLECTION OF VEHICULAR DELAY DATA AT INTERSECTIONS
Jay F. Legere and A. Essam Radwan 105

VEHICLE-DETECTOR INTERACTIONS AND ANALYSIS OF TRAFFIC-ACTUATED SIGNAL CONTROLS
Feng-Bor Lin and Martin C. Percy 112

INVESTIGATION OF NETWORK-LEVEL TRAFFIC FLOW RELATIONSHIPS: SOME SIMULATION RESULTS
Hani Mahmassani, James C. Williams, and Robert Herman 121

ANOTHER LOOK AT STORAGE REQUIREMENTS FOR BANK DRIVE-IN FACILITIES
John L. Ballard, John G. Goble, Richard J. Haden, and Patrick T. McCoy 130

SIMULATION OF TRAFFIC PERFORMANCE, VEHICLE EMISSIONS, AND FUEL CONSUMPTION AT INTERSECTIONS: THE TEXAS-II MODEL
Clyde E. Lee and Fong-Ping Lee 133

DISCONTINUITY IN EQUILIBRIUM FREEWAY TRAFFIC FLOW
Harold J. Payne 140

Level of Service Concepts: Development, Philosophies, and Implications

ROGER P. ROESS

ABSTRACT

The concept of level of service and its use in highway capacity analysis are explored and discussed. Development of the concept is traced from the 1950 Highway Capacity Manual through the 1965 manual and more recent work published in Circular 212 and in the final reports of research efforts. The relative complexity of recent level of service applications and the changing interpretations, needed to understand the implications of level of service analyses, are discussed. Measures of effectiveness used in the definition and determination of level of service are also presented and discussed.

The first edition of the Highway Capacity Manual (HCM) (1) appeared in 1950 and presented a series of empirical procedures for the estimation of the traffic-carrying capabilities of a variety of traffic facilities. Although this manual did not specifically refer to levels of service, it treated capacity under a number of conditions. Practical capacity, for example, was defined as the maximum number of vehicles passing a point or segment of highway under prevailing conditions, when reasonable operating conditions are maintained.

Practical capacity was the first attempt to address the quality of traffic service provided at specified volume levels, and although reasonable operating conditions are fairly loosely defined, the concept of relating maximum volume levels to operating characteristics was firmly established in the 1950 HCM.

The concept of level of service was formally introduced in the 1965 HCM and was defined as follows:

> Level of Service is a qualitative measure of the effect of a number of factors, which include speed and travel time, traffic interruptions, freedom to maneuver, safety, driving comfort and convenience, and operating cost (2,p.7).

The concept, as defined, relates solely to measures and characteristics that directly affect the quality of service provided to the driver. Measures included in the definition are those that are directly perceivable by the individual motorist, and are intended to describe, in relative terms, the quality of the driving experience.

It is from this straightforward definition that the level of service concept as currently used has arisen. The 1965 HCM defines six levels of service, with letter designations A-F. Level of service A designates the best quality of service and refers to virtually free flow in which the operation of an individual vehicle is not significantly affected by the presence of other vehicles. Level of service F designates the worst quality of flow and refers to conditions in which stop-and-go travel, long delays, and queued traffic exist.

Each level represents a range of operating conditions, and each is defined in terms of boundary values of appropriate parameters. Analysis procedures are intended to relate levels of service on specific types of facilities to the maximum volumes that can be accommodated at each level without causing the quality of service to fall below the defined limits for the given level of service. Such maximum volumes are referred to as service volumes.

LEVEL OF SERVICE IN THE 1965 HCM

The 1965 HCM deviated from the concept of level of service in one very important respect: Volumes, or surrogate volume measures, were often used to define level of service limits independent of the service quality parameters noted previously.

For uninterrupted flow facilities (freeways, multilane highways, two-lane highways), the 1965 HCM defined level of service in terms of two parameters, operating speed and volume-to-capacity ratio (V/C). The two were expressed as independent controls on level of service, and highway operations had to meet both the speed and V/C criteria to achieve a given level. Interestingly, the speed-V/C criteria defined were not consistent with the observed speed-volume relationships given in the manual. Particularly at levels C and D, minimum operating speed criteria given were lower than the speeds that regularly occurred for the given V/C criteria. This is discussed in detail elsewhere (3) and is important because virtually all level of service determinations made using the 1965 HCM criteria for uninterrupted flow facilities are controlled by the volume or V/C criteria.

Levels of service for other types of facilities are directly related to volume measures in the 1965 HCM. Signalized intersection level of service is based on the load factor, a parameter dependent on the demand volume. Criteria for ramp junctions are expressed only in terms of merge, diverge, and weaving volumes, and weaving area level of service is based on equivalent service volume estimates.

Only in procedures for arterials and downtown streets is level of service defined directly and solely in terms of performance measures--average overall travel speed. Despite the fact that volume is the single traffic parameter specifically excluded from the definition of level of service, it is the parameter most frequently used to define levels of service in the 1965 HCM. As a result, many public officials and practitioners have come to think of levels of service as being defined in terms of service volumes rather than in terms of the quality measures originally intended.

More recent capacity analysis procedures, however, have tended to adhere more closely to the original definition of level of service. A wider range of measures of effectiveness has been incorporated into procedures. Of greater importance, how-

ever, is the fact that the interpretation of level of service is becoming more complex and is not always the same as, or even similar to, the levels of service defined in the 1965 HCM. Subsequent sections of this paper explore some of the more recent developments in highway capacity analysis techniques and the approaches to level of service embodied therein.

RECENT HIGHWAY CAPACITY ANALYSIS DEVELOPMENTS

In January 1980 the Transportation Research Board published Circular 212, Interim Materials on Highway Capacity (4). It contained the results of two research efforts undertaken to prepare draft materials for a third edition of the Highway Capacity Manual. The first was sponsored by the National Cooperative Highway Research Program and conducted by JHK & Associates (5). It resulted in the preparation of capacity analysis procedures for signalized and unsignalized intersections, pedestrians, and transit facilities. The second was sponsored by the FHWA and conducted by the Transportation Training and Research Center of the Polytechnic Institute of New York (6). It produced capacity analysis procedures for freeways and freeway components. The publication of these materials in Circular 212 was intended to allow broad use and testing of the procedures by professionals, so that substantial feedback could be obtained before the publication of a new HCM.

Since 1980, two additional documents of major importance have been completed. A major revision to the signalized intersection procedure of Circular 212 was prepared by JHK & Associates under NCHRP sponsorship (7), and new procedures for the analysis of two-lane highways were developed by the Texas Transportation Institute of Texas A&M University System (8).

These recent documents contain several novel approaches to the level of service concept.

Uninterrupted Flow--Freeways

Levels of service for freeway segments in Circular 212 are defined in terms of density, the number of vehicles occupying a unit length of freeway lane, and average running speed. The concept presented, in which density is the primary measure of effectiveness, deviates from the 1965 HCM in three major ways:

1. Density (passenger cars per mile per lane) is used for the first time as a level of service parameter. It quantifies the proximity of other vehicles and is directly related to freedom to maneuver within the traffic stream.
2. Volume is not used as a defining criterion. Service volumes are related to density and speed criteria in accordance with calibrated speed-flow-density curves observed on modern freeways.
3. Criteria are applied to uniform 15-min rates of flow (expressed as equivalent hourly volumes) not to actual full peak-hour volumes.

These points are critical. Figure 1 shows the basic concept of freeway level of service. Levels are defined in terms of density and speed. Volumes that actually occur when these speeds and densities exist are tabulated based on observed correlations. The combinations of density, speed, and volume shown as level of service criteria are not independently selected but represent combinations actually expected to occur under ideal conditions on uninterrupted freeway segments.

Figure 2 shows the speed-flow relationships used

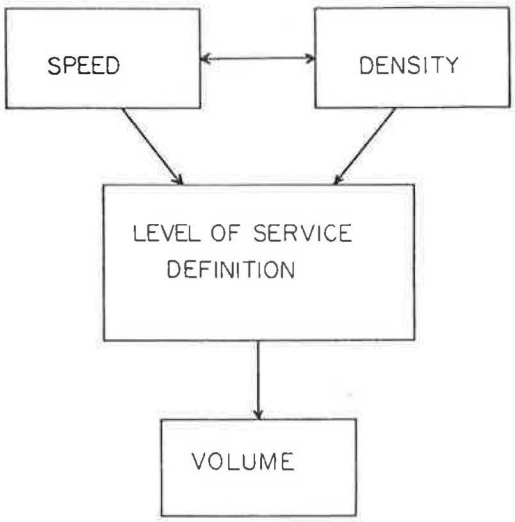


FIGURE 1 Level of service concept for uninterrupted flow.

in Circular 212, somewhat adjusted to reflect the higher average driving speeds observed under free flow conditions in recent years. Note that there is a substantial range of volumes [up to about 1,600 passenger cars per hour per lane (pcphpl)] over which speed is relatively stable followed by a range in which speed deteriorates rapidly with small increases in volume as capacity is approached. This characteristic led to the adoption of density as a primary parameter for defining level of service, because a speed definition of 50-55 mph for a given level of service, for example, would have covered the entire range of desirable volume levels.

The shape of these curves also influenced the selection of criteria for level of service boundaries for various average highway speeds (AHS) given in Table 1. Note that as the levels go from good (level A) to poor (level F) the range of densities encompassed by each level increases, as does the range of speeds. The range of volumes encompassed, however, gets progressively smaller as the levels get poorer. Thus, the range of service volumes in level E is only 75 pcphpl, and the range of density covered is 20 passenger cars per mile per lane (pc/mi/ln). This is in accordance with the shape of the observed speed-flow curves (i.e., as capacity is approached, a small change in volume will bring about a radical deterioration in service quality, as measured by speed and density). Critical comment on the narrow range of service volume in level E may eventually lead to some revision of these criteria, but it should be noted that level of service is defined by the performance parameters of density and speed. It is therefore important to maintain reasonable ranges for these values, despite the narrow service volume ranges that result.

The application of criteria to 15-min rates of flow is another important change from the 1965 HCM. The 1965 HCM more or less applied an average level of service over the full peak hour of operation, with the exception of levels C and D for freeways, which explicitly consider the peak hour factor (PHF) and the peak 5-min flow period. The Circular 212 method recommends the analysis of uniform periods of flow. Thus, if operations were at level of service C for 0.5 hr and at level of service E for the rest of the hour, they would be separately analyzed and labeled. The 1965 HCM would essentially label the full hour as level of service E, even though the condition exists for only 0.5 hr. The emphasis on

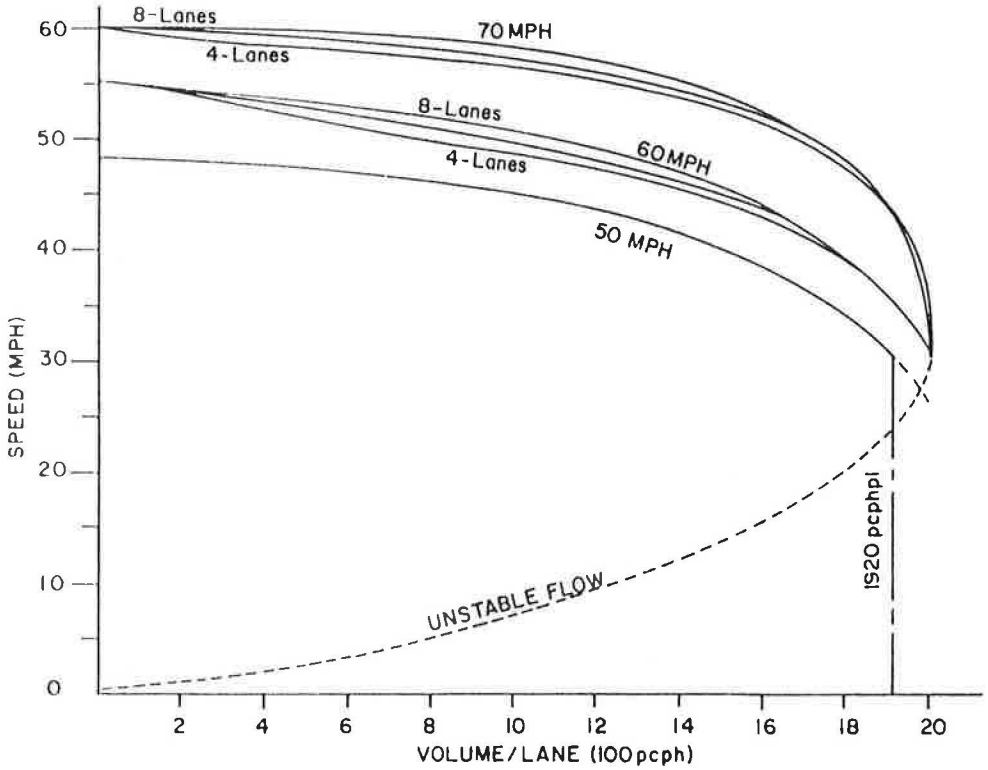


FIGURE 2 Speed-flow curves for freeway segments.

TABLE 1 Level of Service for Basic Freeway Segments (4)

LEVEL OF SERVICE	PERFORMANCE CRITERIA FOR LEVELS OF SERVICE		MAXIMUM SERVICE VOLUMES (ONE DIRECTION) FOR LEVELS OF SERVICE DURING UNIFORM PERIODS OF FLOW (PCPH)			
	SPEED MPH (Km/h)	DENSITY PC/mi/LN (PC/km/LN)	4-Lane (2 ea.dir)	6-Lane (3 ea.dir)	8-Lane (4 ea.dir)	EA. ADD LANE
AHS = 70 MPH (112 km/h.)						
A	≥50(80)	≤15(9.4)	1600	2400	3280	820
B	≥50(80)	≤25(15.6)	2500	3900	5400	1350
C	≥48(77)	≤35(21.9)	3400	5100	6800	1700
D	≥40(64)	≤47(29.4)	3850	5775	7700	1925
E	≥30(48)	≤67(41.9)	4000	6000	8000	2000
F	<30(48)	>67(41.9)	-	highly variable		-
AHS = 60 MPH (96 km/h.)						
A	*	*	*	*	*	*
B	≥45(72)	≤25(15.6)	2300	3525	4800	1200
C	≥43(69)	≤35(21.9)	3050	4575	6100	1525
D	≥38(61)	≤47(29.4)	3600	5400	7200	1800
E	≥30(48)	≤67(41.9)	4000	6000	8000	2000
F	<30(48)	>67(41.9)	-	highly variable		-
AHS = 50 MPH (80 km/h.)						
A	*	*	*	*	*	*
B	*	*	*	*	*	*
C	≥40(64)	≤35(21.9)	2800	4200	5600	1400
D	≥35(56)	≤47(29.4)	3300	4950	6600	1650
E	≥30(48)	≤67(41.9)	4000	6000	8000	2000
F	<30(48)	>67(41.9)	-	highly variable		-

* Level of Service not achievable due to reduced safety on highways with restricted AHS

peak 15-min rates of flow carries through to other procedures in the Circular and to the more recent procedures for two-lane highways and signalized intersections as well.

Freeway Components

Level of service criteria for weaving sections and ramp junctions are generally keyed to those for uninterrupted flow freeway segments in Circular 212.

The circular introduces a totally new approach to weaving area analysis in which levels of service for weaving and nonweaving traffic are separately rated. Because speed has been found to be sensitive to volume throughout the full range of volumes in weaving sections, speed is used as a direct measure of effectiveness for nonweaving vehicles. The level of service for weaving vehicles is based on the differential between the average speed of nonweaving vehicles and the average speed of weaving vehicles. It is assumed that weaving vehicles would tolerate somewhat slower speeds than nonweaving vehicles for any given level of service because of the lane-changing required.

Merge and diverge criteria at ramp junctions are keyed solely to volumes. Volume levels, however, have been set to allow freeway operations in the vicinity of the ramp to continue as defined by general freeway criteria. In general, the merge and diverge volume limits of the circular are lower than the corresponding criteria in the 1965 HCM. As do the general freeway criteria, weaving area and ramp junction criteria refer to peak 15-min flow rates and performance criteria.

Two-Lane Highways

The recently completed two-lane highway methodology prepared by Texas Transportation Institute for NCHRP (8) introduces a new measure to the level of service arena. As was the case for freeways, speed has been found to be relatively constant over a broad range of volumes. Although the methodology retains speed as a principal level of service parameter, it also introduces a secondary measure: percent of time delayed. This parameter is a measure of the proportion of time drivers must spend in a queue because of the inability to pass a leading vehicle or vehicles. Although this parameter is difficult to measure directly, the method assumes that headways of 5 sec or less indicate an involuntarily queued vehicle.

For two-lane highways, this is a most interesting parameter and is indicative of one of the most vexing driver frustrations encountered on such roads. Although the parameter is used more in a qualitative than a quantitative sense in the procedure, its introduction is an indication of developing philosophies of level of service.

Signalized Intersections

Circular 212 formalizes a U.S. analysis procedure based on critical movement analysis. The critical movement approach is hardly new. Greenshield's original work on signal timing is based on critical movement analysis. Work on applying critical movement techniques to capacity analysis has been extensive both in the United States (9,10) and abroad (11). Circular 212 relates level of service directly to the saturation ratio, which is the V/C ratio.

The updated procedure prepared by JHK & Associates for NCHRP (8), however, bases level of service on delay, specifically the average individual stopped delay. The V/C ratio does not enter into the criteria for level of service. This is consistent with the concept of level of service (LOS) but nevertheless introduces an entirely new interpretation of level of service labels. The following table gives the criteria presented in the JHK procedure.

Level of Service	Average Individual Stopped Delay (sec/veh)
A	< 10.0
B	10.1-20.0
C	20.1-30.0
D	30.1-40.0
E	40.1-60.0
F	> 60.0

Research conducted during development of the procedure has shown that delay and V/C ratio are not strongly correlated on a one-to-one basis. Signal progression, for example, has a major influence on delay. The JHK procedure essentially predicts average individual stopped delay assuming random vehicle arrivals. A correction factor is provided to account for progression and the type of signal controller. The factor, however, ranges from 0.5 to 1.5. Thus an intersection with a random arrival delay of, for example, 42 sec/veh could be adjusted to a low of 21 sec (LOS C) or to a high of 63 sec/veh (LOS F). The V/C ratio would remain unchanged, because it is not affected by progression.

Two interesting cases can arise when the level of service criteria in the JHK procedure are used:

1. A delay of > 60 sec (LOS F) may result when the V/C ratio is less than 1.00 and operations are entirely stable and
2. A delay of < 60 sec (LOS A-E) may result when the V/C ratio is marginally higher than 1.00 (i.e., marginal oversaturation exists).

Thus, level of service F does not necessarily connote "forced" or "breakdown" flow conditions, but it is an indication of excessive or unacceptable levels of delay. Further, V/C ratios marginally greater than 1.00, a condition that identifies breakdown flow, will not necessarily be rated as level of service F.

This concept of level of service is considerably different from that used throughout the 1965 HCM. The lower boundary (maximum service volume) for level of service E, for example, is no longer synonymous with capacity. In fact, its relationship to capacity is somewhat variable. In essence, the JHK procedure separates the consideration of level of service from capacity. Both the LOS and the V/C ratio are critical parameters in the overall analysis of an intersection, and both must be considered in evaluating present or future operations.

This will require a change in the way LOS terminology has been commonly used and understood in the past. Delay is a more relative measure than V/C ratio; what is unacceptable in a small urban area may be quite acceptable in a major metropolis. In midtown Manhattan, for example, delays of 60 sec/veh may be quite tolerable as long as oversaturation is avoided (i.e., V/C < 1.00). Traffic engineers and planners will have to defend cases in which a poor LOS is deemed acceptable and explain cases in which level of service F exists but in which excess capacity still remains. This may be a difficult task when dealing with groups accustomed to the more absolute LOS designations currently in use.

Unsignalized Intersections

The unsignalized intersection methodology presented in Circular 212 is a translation of a German procedure developed in 1974. In one of the few cases where a direct indexing of level of service criteria to volume still exists, the procedure uses the parameter "unused capacity" (i.e., the capacity of an approach or lane minus the volume on the approach or lane). Although not consistent with the general philosophy of LOS as defined in the 1965 HCM, unused capacity is presumed to generally correlate with approximate delay ranges that are verbally defined. Because this correlation is not numerically expressed or statistically verified, delay is clearly not a principal parameter in this methodology, nor is it used in defining the level of service boundaries.

Transit Facilities

Circular 212 presents a comprehensive overview of transit capacity and level of service. This is basically a new area for formal highway capacity analysis and introduces some interesting concepts. Transit level of service is a two-dimensional issue, dealing with the internal environment of the transit vehicle and the traffic environment of the vehicle itself. A bus, for example, can be traveling at free flow speed on an uncongested highway but be loaded with many standees, thereby providing a low level of service. Conversely, a lightly loaded bus can provide an excellent internal environment but be stuck in a traffic stream operating at level of service F. The transit material of Circular 212 emphasizes the internal environment of the transit vehicle as the principal LOS measure, using load factor as the determining parameter. Load factor is defined as the total number of passengers on the vehicle divided by the seating capacity of the vehicle.

The transit methodology of Circular 212 also introduces a new concept of capacity. Capacity is defined as the maximum loading of the transit vehicle including a "reasonable" number of standees. Crush load capacity is a larger number of people who can be accommodated under "crush" conditions--loading that is severely uncomfortable. Capacity is also defined in terms of persons per hour passing a point or segment of facility, rather than the more traditional vehicles per hour. The concept of person capacity as opposed to vehicle capacity is critical in the analysis and justification of various bus transit priority schemes.

Pedestrians

The treatment of basic flow characteristics of pedestrians in Circular 212 is quite similar to that of uninterrupted flow vehicular characteristics. Whereas for uninterrupted flow, density was used as the principal defining parameter for LOS, the pedestrian methodologies use the parameter "space" (square feet per pedestrian), which is the inverse of density. These criteria assume that pedestrians are most severely affected by their proximity to others, which has been shown to have a drastic impact on freedom to maneuver and on walking speed. Pedestrian capacity is approached in terms of person capacity.

Other Facilities

Materials on capacity analysis of other types of facilities are being developed. Indeed, revisions to the materials discussed previously are also under way. LOS criteria for suburban and rural multilane highways will be similar to those for freeways and will rely on density as a primary measure of effectiveness. Arterial LOS criteria are currently projected to be based on average overall travel speed. Different criteria for different classes or categories of arterials may be defined.

SOME OVERALL ISSUES

Relationship to Other Standards

Whether fortunate or unfortunate, a large number of formal federal, state, and local standards for such divergent areas as highway design, highway operations, noise, and air quality are written in terms of specific level of service criteria. Thus, a state may require that rural highways be designed to level of service B or that highways be redesigned and reconstructed when they reach level of service E. A planning board may require that air quality analysis be done for operations at a specific level of service.

Most of these standards have legal significance. As new procedures for capacity analysis are introduced and used, it is vital to remember that level of service definitions and criteria have changed, sometimes radically. Level of service C for a signalized intersection, for example, may bear little resemblance to level of service C as defined in the 1965 HCM. Because most LOS references in existing standards are to the 1965 HCM, it is critical that analysts and users make the proper adjustments in the application of new procedures. In time, as new procedures are included in a new Highway Capacity Manual, it is to be expected that these standards will be revised. Such revisions, however, may be slow in coming, and administrators, policymakers, legislators, and others involved will have to be carefully re-educated in the interpretation of LOS criteria and their meaning.

The level of service concept, as defined in the 1965 HCM, has become such a useful tool in describing traffic operations, and so generally accepted as an absolute quality scale, that changes in its framework will require a concentrated re-education effort.

Re-Education

When the 1965 HCM was published, it was followed by a multitude of professional short courses, offered by numerous consultants and universities, aimed at training practitioners in its use. Much the same can be expected with a third edition HCM. Such courses have already been offered on Circular 212. Given the relative complexity of newer techniques compared with the 1965 HCM, and given the rather substantial revisions in the use of and criteria for level of service, the re-education effort will be most important to a smooth transition. There are also more users to consider: Since the 1965 HCM, more and more professionals have found critical uses for the manual, including as input to environmental analyses.

Courses will have to focus not only on practitioners but on the policymakers and administrators who must act on the basis of capacity analyses and related information. It is at this level that the revisions in LOS criteria must be most clearly transmitted to avoid the misuse of new analysis output based on old criteria.

Level of Service F

The 1965 HCM uses level of service F to describe conditions in a queue of vehicles that forms behind a breakdown point or constriction. Such breakdowns or constrictions exist at points where the arrival volume exceeds the departure volume, leading to queue formation.

Because analysts are generally more interested in the point of the breakdown and its cause than in the operating conditions in the queue, many analysts have used the LOS F designation to describe the point of the breakdown. This is technically incorrect, because operations at the point of the breakdown are virtually always at or near capacity. Operations within the queue are what is properly described as LOS F. Nevertheless, LOS F designation of the breakdown point is useful to highlight such critical points.

Circular 212 and other new procedures do not clarify this point to a substantial degree, although sample problems indicate that the use of LOS F to designate points where demand exceeds capacity is accepted.

SUMMARY AND CONCLUSION

Level of service criteria and measures used in Circular 212 and other new capacity procedures indicate that the concept of LOS defined in the 1965 HCM is being more closely adhered to than is the 1965 HCM itself. Table 2, which gives criteria for LOS used in the 1965 HCM and in more recent procedures, illustrates this point.

Volume and volume-based measures were not intended to be among the defining criteria for LOS. Measures describing the basic service quality perceived by the user--travel time, freedom to maneuver, comfort and convenience, safety, and so forth--are the terms used in the definition. Volumes were to be specified (based on empirical observation) for various types of facilities and levels of service. This procedure has been generally followed in the development of new capacity analysis methodologies. However, the use and interpretation of LOS have been substantially altered, causing a problem that will have to be addressed by re-education of users and policymakers alike.

TABLE 2 Level of Service Criteria Compared: 1965 HCM and Recent Techniques

	Measures of Effectiveness Used to Define LOS	
Type of Facility	1965 HCM	Recent Methods
Freeways	Operating speed V/C ratio	Density Avg running speed
Multilane highways	Operating speed V/C ratio	Density Avg running speed
Two-lane highways	Operating speed V/C ratio	Avg running speed Percent time delayed
Ramps	Volumes	Volumes
Weaving areas	Volumes Operating speed	Avg running speed
Signalized intersections	Load factor	Delay
Unsignalized intersections		Unused capacity
Arterials	Operating speed	Avg overall travel speed
Transit		Load factor
Pedestrians		Space

REFERENCES

1. Highway Capacity Manual. Bureau of Public Roads, Washington, D.C., 1950.
2. Highway Capacity Manual 1965. HRB Special Report 87. HRB, National Research Council, Washington, D.C., 1965, 397 pp.
3. R.P. Roess, W.B. McShane, and L.J. Pignataro. Freeway Level of Service: A Revised Approach. In Transportation Research Record 699, TRB, National Research Council, Washington, D.C., 1979, pp. 7-16.
4. Interim Materials on Highway Capacity. Transportation Research Circular 212. TRB, National Research Council, Washington, D.C., 1980, 276 pp.
5. JHK & Associates. Development of an Improved Highway Capacity Manual. Final Report on Phase I, NCHRP Project 3-28. TRB, National Research Council, Washington, D.C., 1979.
6. R.P. Roess, W.R. McShane, E.M. Linzer, and L.J. Pignataro. Freeway Capacity Analysis Procedures. DOT-FH-11-9336. Polytechnic Institute of New York, Brooklyn, N.Y.; FHWA, U.S. Department of Transportation, 1978.
7. JHK & Associates. Urban Signalized Intersection Capacity. Procedural Report, NCHRP Project 3-28(2). TRB, National Research Council, Washington, D.C., Feb. 1983.
8. C.J. Messer. Two-Way, Two-Lane Rural Highway Capacity. Procedural Report, NCHRP Project 3-28(A). Texas Transportation Institute, Texas A&M University, College Station; TRB, National Research Council, Washington, D.C., Feb. 1983.
9. D.S. Berry and P.K. Gandi. Headway Approach to Intersection Capacity. In Highway Research Record 453, HRB, National Research Council, Washington, D.C., 1973, pp. 56-60.
10. C.J. Messer and D.B. Fambro. Critical Lane Analysis for Intersection Design. In Transportation Research Record 644, TRB, National Research Council, Washington, D.C., 1977, pp. 26-37.
11. X. Miller. Capacity of Signalized Intersections in Australia. Bulletin 271. Australian Road Research Board, Numawading, Victoria, March 1968.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

Passenger Car Equivalents for Uninterrupted Flow: Revision of Circular 212 Values

ROGER P. ROESS and CARROLL J. MESSER

ABSTRACT

As part of an overall federal effort to allocate road user taxes, a number of recent studies have been directed toward the calibration of passenger car equivalent (pce) values for trucks. These studies have provided the opportunity to review the pce values for uninterrupted flow contained in Transportation Research Board Circular 212, "Interim Materials on Highway Capacity." The results of these efforts and their implications for highway capacity analysis are reviewed. Specific recommendations for revisions of the pce values of Circular 212 are made.

A passenger car equivalent (pce), in highway capacity analysis terms, is the number of passenger cars that is roughly the equivalent of one truck, bus, or recreational vehicle under prevailing roadway and traffic conditions. The use of such equivalents is central to highway capacity analysis where mixed traffic streams are present, and the calibration of these values can have a significant impact on capacity analysis computations. As part of an overall federal effort to allocate road user taxes, there have been a number of recent studies specifically focused on the calibration of pce values for trucks. Although these efforts have provided a large quantity of useful data with which to compare pce values currently in use, they have not simplified this complex issue. Because many of these studies were directed to the development of economic equivalents for the purposes of road user tax allocations, there was no direct association with pce's for capacity analysis. Further, the exact definition of equivalent is not consistently interpreted, either in the capacity analysis literature or in these studies, making comparison of results difficult indeed.

An attempt is made herein to address the overall issue of equivalency and to use the available information to provide more realistic pce values for uninterrupted flow, specifically for use in capacity analysis.

APPROACHES TO PCE CALIBRATION FOR UNINTERRUPTED FLOW

There have been a wide variety of philosophies applied to the development of pce values. These are presented briefly here and are more fully discussed in the final report of a recent FHWA-sponsored study (1).

Equivalents for trucks in the 1965 Highway Capacity Manual (HCM) (2) are based on the Walker method, named for Powell Walker who developed the calibration methodology. This method relates pce's on two-lane highways to the relative number of passings of trucks by passenger cars versus the number of passings of passenger cars by other passenger cars. For multilane highways, the method was modified to account for the delay to other vehicles caused by trucks. Equivalent delays caused by trucks on grades were computed using grade-performance curves and were used as the basis for equivalents.

Since the 1965 HCM, numerous other techniques have been applied. The Institute for Research conducted a study of urban freeways and based pce calibrations on relative spatial headways between trucks and passenger cars versus those between pairs of passenger cars (3). Craus modified the equivalent delay principle of the Walker method for application to two-lane highways (4).

Uninterrupted flow pce values presented in Circular 212 (5) are based on the output of a multilane simulator developed at the Midwest Research Institute (6). Described in detail in a paper by Linzer et al. (7), the approach taken produced equivalents resulting in pce volumes consuming the same proportion of the roadway's capacity as the actual mixed traffic volume. In more analytic terms, the effective volume-to-capacity (V/C) ratio was held constant.

Cunagin and Messer (1), in one of the federally sponsored pce efforts for road user tax allocation, used a combination of the Walker method and the equivalent delay principle, with minor modifications in specific analytics to match the data collection effort. The difference between economic pce values and those used in capacity analysis is highlighted here. In the final report on NCHRP Project 3-28A, "Two-Lane, Two-Way, Rural Highway Capacity" (8), Messer uses different pce values than those calibrated for two-lane highways in the tax allocation study.

In the latter work, a new concept was used: Equivalents were calibrated to produce pce volumes that operate at the same average speed as the actual mixed traffic stream.

Because of the wide variance in pce philosophies adopted by researchers, it is difficult to directly compare numerical results. Unfortunately, there was no uniform understanding of what a pce meant before these studies were undertaken, and indeed the intended use of results also varied. Three of the concepts, however, appear to have direct relevance to highway capacity analysis:

1. The equal speed concept appears to be most relevant. Because level of service criteria for capacity analysis are based on performance parameters, it is logical that pce values should relate to those same performance parameters. For the two-lane highway work noted previously, speed is the principal criterion for designation of levels of service. Thus, conversions from mixed to pce volumes would not alter the performance parameters defining level of service. If this principle were to be extended to the uninterrupted flow procedure of Circular 212, it would suggest that pce's be based on equal densities, because density is the principal parameter defining level of service. It must be noted that none of the other concepts for pce's

reviewed here guarantees that the equivalent pce volume operates at the same performance levels as the actual mixed traffic stream.

2. The spatial headway approach is also of some interest. Average spacing and density are related on a one-to-one basis, and spatial headway could be argued to be a surrogate (more easily measured) parameter for density. Use of space is also of direct interest in capacity analysis.

3. The constant V/C ratio approach used in developing the Circular 212 values is relevant in its own right, because V/C values are related to speeds and densities. Further, the proportion of capacity used and the proportion still available are critical pieces of information. However, while V/C ratios are held constant, equivalent traffic streams may not operate at the same speed and density as the actual mixed traffic stream. The speed-flow-density relationships shown in Circular 212, which were used in the development of level of service criteria, are for ideal conditions (i.e., a pce traffic stream with ideal geometrics on the highway). Introduction of trucks or other non-passenger cars into the traffic stream alters this relationship. Thus, if V/C is held constant in calibrating pce's, speed and density are not. Conversely, if speed or density or both are held constant, V/C is not. In the Circular 212 approach, constant V/C values may be relevant, but the interpretation of level of service criteria, when applied to mixed traffic streams, is unclear.

Unfortunately, it will not be possible to reconcile these three approaches as new capacity techniques are developed in anticipation of a third edition of the Highway Capacity Manual. The data bases are incompatible and do not allow revision of the results of these studies in a single format. Thus, elements of all three principles will survive in new techniques.

AN APPROACH TO MULTILANE PCEs

The pce values of Circular 212 have been the subject of much discussion and lively debate since their appearance in 1980. These discussions have centered on whether the values are too high or too low, and whether the truck performance curves used in their calibration are representative for capacity analysis purposes.

Performance curves in the circular are for a truck with a weight-to-horsepower ratio of 300 lb/hp, and are taken from a study of such characteristics conducted at Pennsylvania State University (9). Although general agreement on the crawl speeds shown for 300 lb/hp trucks has been achieved, there is some dispute over the critical length of grade needed to reach that crawl speed. The Pennsylvania State curves agree closely with those adopted by St. John (10) in his studies at the Midwest Research Institute (when both are adjusted to reflect a 55 mph maximum speed). Studies by the California Department of Transportation (11), however, show crawl speeds significantly in excess of those used in Circular 212, as well as critical grade lengths approximately 500 feet longer than those in the circular. The California study is based on mixed traffic streams, with no single, characteristic, weight-to-horsepower value represented. The California curves, however, agree almost exactly with St. John's curves for a truck with a weight-to-horsepower ratio of 125 lb/hp.

The critical issue, therefore, appears not to be whether the performance curves for the 300 lb/hp truck adopted in Circular 212 are accurate but whether the selection of 300 lb/hp is appropriate for capacity analysis. The California study suggests that a 300 lb/hp truck is considerably less powerful that the average truck. Recent studies at the Texas Transportation Institute (8) and by St. John (12) produced similar results, reporting 170 lb/hp and 150 lb/hp as median values for multilane truck populations. The Pennsylvania State study focused on passing lane design and used the 300 lb/hp truck, explicitly noting that this represented a typical heavy truck.

It appears, therefore, that the average truck population on multilane uninterrupted flow facilities is in the 125-170 lb/hp range, not the 300 lb/hp used in Circular 212. A further issue remains: Is the average or median weight-to-horsepower ratio appropriate for use in capacity analysis? Cunagin and Messer (1) suggest that heavier trucks have a greater negative impact on operations than do lighter trucks. Although the pce values of Circular 212 can be analytically modified to reflect a different typical weight-to-horsepower ratio, they cannot be adjusted to reflect a truck population with a mix of weight-to-horsepower ratios. Thus, a mix of trucks with an average weight-to-horsepower ratio of 150 lb/hp would be expected to have a more negative impact on operations than a truck population in which all trucks had a ratio of 150 lb/hp.

Considering these points, the following recommendations are made:

1. The pce values for normal truck populations should be based on performance characteristics of a 200 lb/hp truck.
2. In keeping with the approach of Circular 212, pce values for nonstandard truck populations should also be provided--light truck populations would be represented by a 100 lb/hp truck, heavy truck populations by a 300 lb/hp truck.

Note that the 200 lb/hp truck is the assumed norm in the 1965 Highway Capacity Manual. It is emphasized, however, that the performance curves for such a truck in the 1965 manual indicate substantially poorer performance than current studies, and that the concept behind the development of pce values in the 1965 Highway Capacity Manual differs greatly from the three primary concepts presented herein.

It should also be noted that, as in the 1965 Manual, the same pce values are recommended for use for freeways and for normal multilane highways. All pce studies show no substantial difference for these types of facilities.

TRUCK PCEs ON INDIVIDUAL GRADES

Figures 1-3 show the performance characteristics of 100 lb/hp, 200 lb/hp, and 300 lb/hp trucks on extended upgrades. They were taken from St. John and Kobett (10) and modified to reflect a maximum truck speed of 55 mph, as in Circular 212.

Passenger car equivalent values, E_T, can be adjusted to reflect new typical weight-to-horsepower ratios in the following manner:

1. For a given percent grade and length of grade, the final speed of trucks is found from Figure 1, 2, or 3.
2. The performance curves of Circular 212 are entered to find (a) the length of grade of the same percent as in step 1, which results in the same final speed of trucks as in step 1, and (b) the percent grade of the same length as in step 1, which results in the same final speed of trucks as in step 1.

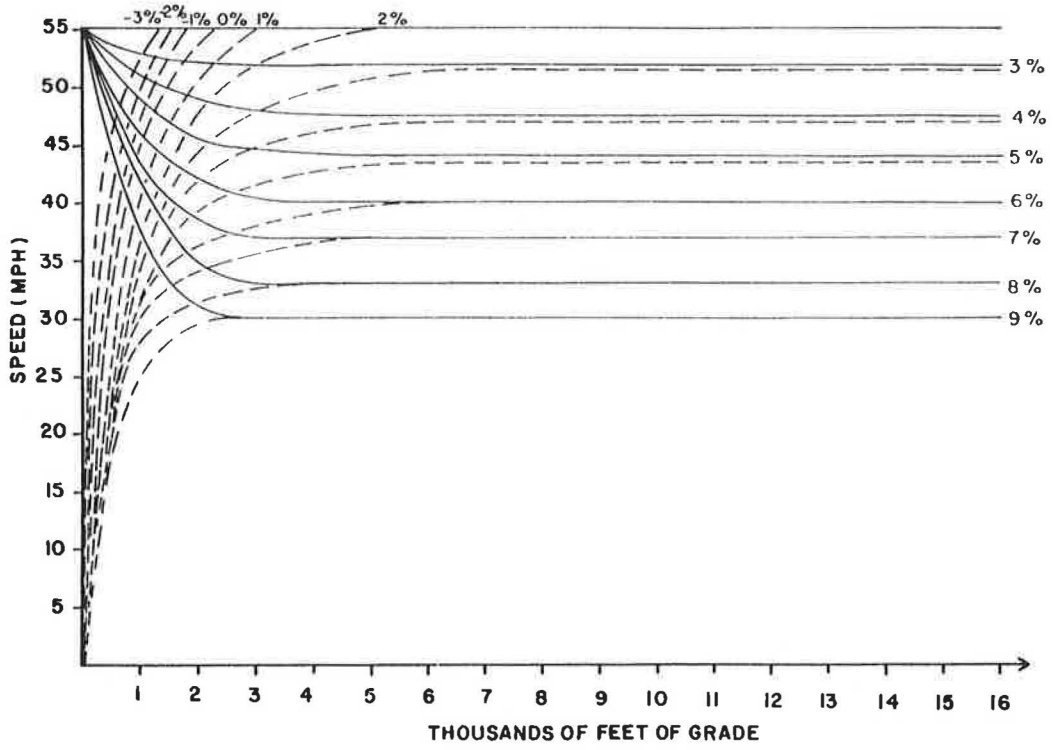


FIGURE 1 Performance curves for a 100 lb/hp truck (*10*).

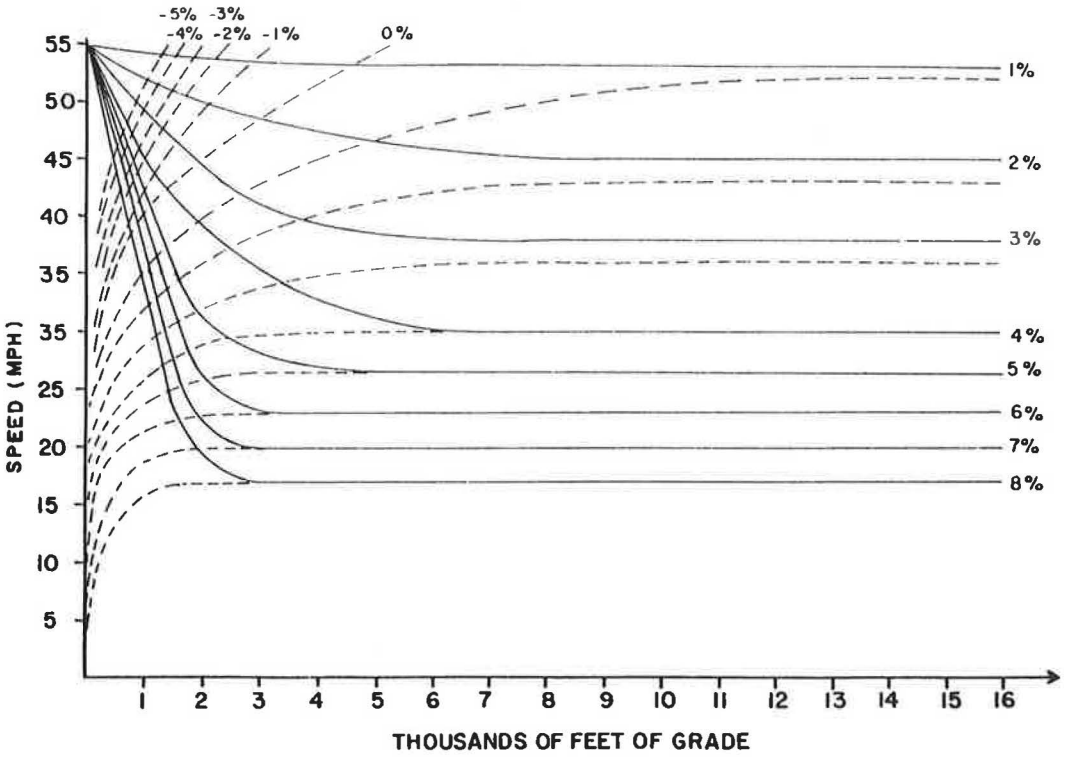


FIGURE 2 Performance curves for a 200 lb/hp truck (*10*).

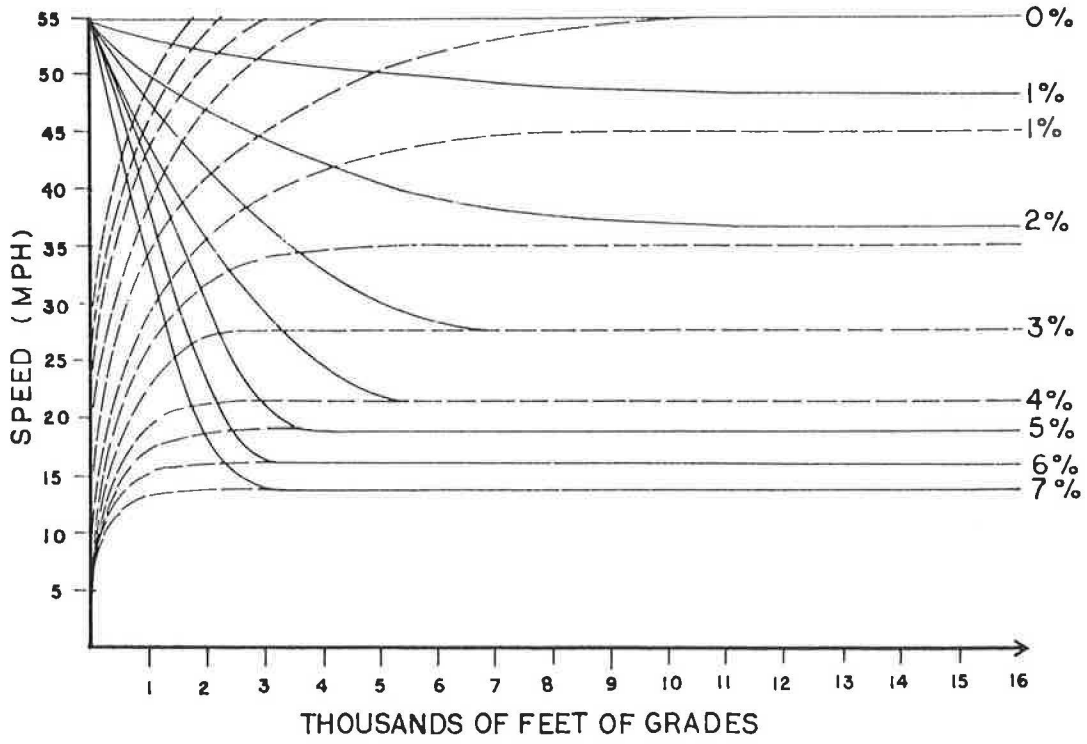


FIGURE 3 Performance curves for a 300 lb/hp truck (*10*).

3. The pce table of Circular 212 is entered with the grade-length conditions of 2(a) and 2(b). The two pce values thus found are averaged and rounded to the nearest integer, with the result being the revised equivalent for the condition described in step 1.

The pce values that result from this analysis for trucks are given in Tables 1-3 and are recommended by the authors for use in revised highway capacity analysis procedures.

RECREATIONAL VEHICLE PCEs ON INDIVIDUAL GRADES

The pce values for recreational vehicles (RVs) in Circular 212 were based on a typical weight-to-horsepower ratio of 60 lb/hp. Recent studies indicate that this also may have been too high a value. Canadian studies have reported average ratios for RVs in the 30-35 lb/hp range. Further, in their equivalence studies, Cunagin and Messer have found that pce's for RVs range from about 1/3 of the corresponding truck pce at high values of E_T to about

TABLE 1 Passenger Car Equivalent Values for 100 lb/hp Trucks

Grade (%)	Length (mi.)	E_T 4-Lane Freeways								E_T 6-8 Lane Freeways							
Percent Trucks		2	4	5	6	8	10	15	20	2	4	5	6	8	10	15	20
≤ 2	All	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	0-1/4	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	1/4-1/2	4	4	4	3	3	3	3	3	4	4	4	3	3	3	3	3
	1/2-3/4	4	4	4	4	3	3	3	3	4	4	4	3	3	3	3	3
	3/4-1	5	4	4	3	3	3	3	3	5	4	4	4	3	3	3	3
	>1	6	5	5	5	4	4	4	3	6	5	5	4	4	4	3	3
4	0-1/4	4	4	4	3	3	3	3	3	5	4	4	4	3	3	3	3
	1/4-1/2	5	5	5	4	4	4	4	4	5	4	4	4	4	4	4	4
	1/2-1	6	5	5	5	4	4	4	4	6	5	5	4	4	4	4	4
	>1	7	6	6	5	4	4	4	4	7	5	5	5	4	4	4	4
5	0-1/4	6	5	5	5	4	4	4	3	6	5	5	5	4	4	4	3
	1/4-1	8	7	7	6	5	5	5	5	8	7	7	6	5	5	5	5
	>1	9	7	7	6	5	5	5	5	8	7	7	6	5	5	5	5
6	0-1/4	7	5	5	5	4	4	4	4	7	5	5	5	4	4	3	3
	1/4-1	9	7	7	6	5	5	5	5	8	7	7	6	5	5	5	5
	>1	9	7	7	7	6	6	5	5	9	7	7	6	5	5	5	5

NOTE: If a length of grade falls on a boundary value, use the equivalent for the longer grade category. Any grade steeper than the percent shown must use the next higher grade category.

TABLE 2 Passenger Car Equivalent Values for Normal Truck Populations (200 lb/hp)

Grade (%)	Length (mi.)	E_T 4-Lane Freeways								E_T 6-8 Lane Freeways							
Percent Trucks		2	4	5	6	8	10	15	20	2	4	5	6	8	10	15	20
<1	All	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1	0-1/2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	1/2-1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	≥1	4	3	3	3	3	3	3	3	4	3	3	3	3	3	3	3
2	0-1/4	4	4	4	3	3	3	3	3	4	4	4	3	3	3	3	3
	1/4-1/2	5	4	4	3	3	3	3	3	5	4	4	3	3	3	3	3
	1/2-3/4	6	5	5	4	4	4	4	4	6	5	5	4	4	4	4	4
	3/4-1 1/2	7	6	6	5	4	4	4	4	7	5	5	5	4	4	4	4
	≥ 1 1/2	8	6	6	6	5	5	4	4	8	6	6	5	4	4	4	4
3	0-1/4	6	5	5	5	4	4	4	3	6	5	5	5	4	4	4	3
	1/4-1/2	8	6	6	6	5	5	5	4	7	6	6	6	5	5	5	4
	1/2-1	9	7	7	6	5	5	5	5	9	7	7	6	5	5	5	5
	1-1 1/2	9	7	7	7	6	6	5	5	9	7	7	6	5	5	5	5
	≥ 1 1/2	10	7	7	7	6	6	5	5	10	7	7	6	5	5	5	5
4	0-1/4	7	6	6	5	4	4	4	4	7	6	6	5	4	4	4	4
	1/4-1/2	10	7	7	6	5	5	5	5	9	7	7	6	5	5	5	5
	1/2-1	12	8	8	7	6	6	6	6	10	8	7	6	5	5	5	5
	≥ 1	13	9	9	9	8	8	7	7	11	9	9	8	7	6	6	6
5	0-1/4	8	6	6	6	5	5	5	5	8	6	6	6	5	5	5	5
	1/4-1/2	10	8	8	7	6	6	6	6	8	7	7	6	5	5	5	5
	1/2-1	12	11	11	10	8	8	8	8	12	10	9	8	7	7	7	7
	≥ 1	14	11	11	10	8	8	8	8	12	10	9	8	7	7	7	7
6	0-1/4	9	7	7	7	6	6	6	6	9	7	7	6	5	5	5	5
	1/4-1/2	13	9	9	8	7	7	7	7	11	8	8	7	6	6	6	6
	1/2-3/4	13	9	9	8	7	7	7	7	11	9	9	8	7	6	6	6
	≥ 3/4	17	12	12	11	9	9	9	9	13	10	10	9	8	8	8	8

NOTE: If the length of grade falls on a boundary value, use the equivalent for the longer grade class. Any grade steeper than the percent stated must use the next higher grade category.

1/2 at low values of E_T. Based on these results, reductions in the Circular 212 values for E_R are recommended, as given in Table 4.

PASSENGER CAR EQUIVALENTS FOR BUSES ON INDIVIDUAL GRADES

The subject of bus performance on grades has not received any research attention since the early 1960s, and there is no new data on which to base revisions to the pce values in the 1965 Highway Capacity Manual. These values were transferred to Circular 212 and will doubtless carry over to any additional revisions, barring new data. As in the past, where bus percentages are small compared with trucks or RVs, numerical procedures would recommend that buses be considered as trucks.

PASSENGER CAR EQUIVALENTS ON GENERAL FREEWAY SEGMENTS

The Institute for Research study (3) produced a set of pce values for a broad range of vehicle types on urban freeways essentially at level grade. The results of this study point to two important factors:

1. Passenger car equivalent values are shown to vary based on volume levels, increasing with increasing volume, and
2. Passenger car equivalent values on level grades appear to be generally lower than the 2.0 and 1.6 values in Circular 212 for trucks and buses (which were taken directly from the 1965 Highway Capacity Manual).

The first point is a vexing one. Values for pce's in the study ranged from 1.1 for several vehicle types at low volumes to 2.0 for tractor-trailers at high volumes. The adoption of pce values varying with volume would present serious problems in capacity analysis procedures, greatly complicating computations. Further, the study addressed level terrain but did not extend to rolling or mountainous terrain. None of the studies that address pce's on specific grades showed significant variation with volume. Because the design benefits of smaller pce values at low volumes would be minimal, it is recommended that constant pce values with volume be used for relevant vehicle types on level terrain.

The Institute for Research study does, however, suggest that the pce values used in the 1965 Highway Capacity Manual and in Circular 212 are higher than necessary. For example, the maximum pce value of 2.0 applies only to tractor-trailers under the highest volume conditions. Maximum pce values for single-unit trucks are 1.5 or 1.6, depending on the number of axles, and the maximum pce value for buses is 1.6. Pickup trucks and vans were found to be the same as passenger cars. On the basis of these results, slight reductions in the level terrain pce values of Circular 212 appear to be in order. The extension of these reductions to rolling and mountainous terrain is not automatically indicated. The current values are generally consistent with the revised pce's for grades of the percent and length usually present in such terrains, and it is therefore recommended that they be retained. Table 5 gives the recommended values for pce's on extended uninterrupted flow segments.

TABLE 3 Passenger Car Equivalent Values for 300 lb/hp Trucks

Grade (%)	Length (mi.)	E_T 4-Lane Freeways								E_T 6-8 Lane Freeways							
Percent Trucks		2	4	5	6	8	10	15	20	2	4	5	6	8	10	15	20
<1	All	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1	0-1/4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	1/4-1/2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	1/2-3/4	4	4	4	4	3	3	3	3	4	4	4	3	3	3	3	3
	3/4-1	5	4	4	4	3	3	3	3	5	4	4	4	3	3	3	3
	1-1 1/2	6	5	5	5	4	4	4	3	6	5	5	4	4	4	3	3
	>1 1/2	7	5	5	5	4	4	4	3	7	5	5	5	4	4	3	3
2	0-1/4	4	4	4	3	3	3	3	3	4	4	4	3	3	3	3	3
	1/4-1/2	7	6	6	5	4	4	4	4	7	5	5	5	4	4	4	4
	1/2-3/4	8	6	6	5	5	4	4	4	8	6	6	6	5	5	4	4
	3/4-1	8	6	6	6	5	5	5	5	8	6	6	6	5	5	5	5
	1-1 1/2	9	7	7	7	6	6	5	5	9	7	7	6	5	5	5	5
	>1 1/2	10	7	7	7	6	6	5	5	10	7	7	6	5	5	5	5
3	0-1/4	6	5	5	5	4	4	4	3	6	5	5	5	4	4	4	3
	1/4-1/2	9	7	7	6	5	5	5	5	8	7	7	6	5	5	5	5
	1/2-3/4	12	8	8	7	6	6	6	6	10	8	7	6	5	5	5	5
	3/4-1	13	9	9	8	7	7	7	7	11	8	8	7	6	6	6	6
	>1	14	10	10	9	8	8	7	7	12	9	9	8	7	7	7	7
4	0-1/4	7	5	5	5	4	4	4	4	7	5	5	5	4	4	3	3
	1/4-1/2	12	8	8	7	6	6	6	6	10	8	7	6	5	5	5	5
	1/2-3/4	13	9	9	8	7	7	7	7	11	9	9	8	7	6	6	6
	3/4-1	15	10	10	9	8	8	8	8	12	10	10	9	8	7	7	7
	>1	17	12	12	10	9	9	9	9	13	10	10	9	8	8	8	8
5	0-1/4	8	6	6	6	5	5	5	5	8	6	6	6	5	5	5	5
	1/4-1/2	13	9	9	8	7	7	7	7	11	8	8	7	6	6	6	6
	1/2-3/4	20	15	15	14	11	11	11	11	14	11	11	10	9	9	9	9
	>3/4	22	17	17	16	13	13	13	13	17	14	14	13	12	11	11	11
6	0-1/4	9	7	7	7	6	6	6	6	9	7	7	6	5	5	5	5
	1/4-1/2	17	12	12	11	9	9	9	9	13	10	10	9	8	8	8	8
	>1/2	28	22	22	21	18	18	18	18	20	17	17	16	15	14	14	14

NOTE: If the length of grade falls on a boundary value, the equivalent corresponding to the longer grade category is used. Any grade steeper than the percent shown must use the next higher grade category.

TABLE 4 Passenger Car Equivalent Values for Recreational Vehicles

Grade (%)	Length (mi.)	E_R 4-Lane Freeways								E_R 6-8 Lane Freeways							
Percent RV's		2	4	5	6	8	10	15	20	2	4	5	6	8	10	15	20
<2	All	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	0-1/2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	≥ 1/2	4	3	3	3	3	3	3	3	4	3	3	3	3	3	3	3
4	0-1/4	3	2	2	2	2	2	2	2	3	2	2	2	2	2	2	2
	1/4-3/4	4	3	3	3	3	3	3	3	4	3	3	3	3	3	3	3
	≥ 3/4	5	4	4	4	3	3	3	3	4	4	4	4	3	3	3	3
5	0-1/4	4	3	3	3	3	3	3	3	4	3	3	3	2	2	2	2
	1/4-3/4	5	4	4	4	4	4	4	4	5	4	4	4	4	4	4	4
	≥ 3/4	6	5	4	4	4	4	4	4	5	5	4	4	4	4	4	4
6	0-1/4	5	4	4	4	3	3	3	3	5	4	4	3	3	3	3	3
	1/4-3/4	6	5	5	4	4	4	4	4	6	4	4	4	4	4	4	4
	≥ 3/4	7	6	6	6	5	5	5	5	6	5	5	5	4	4	4	4

NOTE: If a length of grade falls on a boundary condition, the equivalent from the longer grade class is used. Any grade steeper than the percent shown must use the next higher grade category.

COMPUTATION OF ADJUSTMENT FACTORS FOR HEAVY VEHICLES

In Circular 212, values of E_T, E_R, and E_B are converted to a multiplicative adjustment factor that is applied to a maximum service volume under ideal conditions:

$$f_{HV} = 100/[100 + P_t(E_T-1) + P_r(E_R-1) + P_b(E_B-1)]$$

where f_{HV} is the multiplicative adjustment factor; P_t, P_r, and P_b are the percentages of trucks, RVs, and buses in the traffic stream; and E_T,

TABLE 5 Passenger Car Equivalent Values for Extended Freeway Segments

FACTOR	TYPE OF TERRAIN		
	Level	Rolling	Mountainous
E_T for trucks	1.7	4.0	8.0
E_B for buses	1.5	3.0	5.0
E_R for RV's	1.6	3.0	4.0

E_R, and E_B are the pce values for trucks, RVs, and buses, respectively.

Cunagin and Messer (1) suggest another approach that was applied to the two-lane highway work at Texas A&M. The point is made that the composite effect of various types and classes of non-passenger cars on the traffic stream may not be the same as the algebraic combination of individual impacts, as suggested in the preceding equation. Their suggested formulation is

$$f_{HV} = 100/[100 + P_{HV}(E_{HV} - 1)]$$

where P_{HV} is the total percentage of heavy vehicles in the traffic stream, and E_{HV} is the pce for all heavy vehicles in the traffic stream for the existing mix of traffic. There is some logic to this latter approach, but values of E_{HV} must be calibrated for various mixes of trucks, RVs, and buses. The rural two-lane highway procedure developed at Texas A&M for NCHRP tabulates the pce's for a standard mix of trucks and RVs (buses are treated as one or the other, depending on conditions). An equation is provided to adjust this value to a mix other than the norm.

The data bases used for multilane pce's in Circular 212 and herein do not permit the direct calibration of values for various mixes of heavy vehicles. For this reason the algebraic approach used in the circular will be needed until the latter approach is investigated further.

SUMMARY AND CONCLUSION

A review of the various approaches to calibration and interpretation of pce values in highway capacity analysis has been presented. The recommended revisions to pce values for multilane uninterrupted flow currently in use are based on an evaluation of the latest research and data and should result in improved accuracy of analysis procedures for these types of facilities.

REFERENCES

1. W.D. Cunagin and C.J. Messer. Passenger Car Equivalents for Rural Highways. Final Report, contract DTFH61-80-C00128. Herbert G. Wyte Associates; Texas Transportation Institute, Texas A&M University, College Station, 1982.
2. Highway Capacity Manual 1965. HRB Special Report 87. HRB, National Research Council, Washington, D.C., 1965, 397 pp.
3. E.L. Seguin, K.W. Crowley, and W.D. Zweig. Passenger Car Equivalents on Urban Freeways. Interim Report, contract DTFH61-80-C00106. Institute for Research, State College, Pa., Aug. 1982.
4. J. Craus, A. Polus, and I. Grinberg. A Revised Method for the Determination of Passenger Car Equivalencies. Transportation Research, Vol. 14A, No. 4, Aug. 1980.
5. R.P. Roess, E.M. Linzer, W.R. McShane, and L.J. Pignataro. Freeway Capacity Procedures. In Transportation Research Circular 212: Interim Materials on Highway Capacity, TRB, National Research Council, Washington, D.C., Jan. 1980, 276 pp.
6. A.D. St. John. Freeway Design and Control Strategies as Affected by Trucks and Traffic Regulations. Report FHWA-RD-74-42. Midwest Research Institute, Kansas City, Mo., 1975.
7. E.M. Linzer, R.P. Roess, and W.R. McShane. Effects of Trucks, Buses, and Recreational Vehicles on Freeway Capacity and Service Volume. In Transportation Research Record 699, TRB, National Research Council, Washington, D.C., 1979, pp. 17-26.
8. C.J. Messer. Two-Lane, Two-Way, Rural Highway Capacity. Final Report, NCHRP Project 3-28A. Texas Transportation Institute, Texas A&M University, College Station; TRB, National Research Council, Washington, D.C., Feb. 1983.
9. Pennsylvania State University. Review of Truck Weight to Horsepower Ratios as Related to Passing Lane Design. NCHRP Project 20-7. TRB, National Research Council, Washington, D.C., 1978.
10. A.D. St. John and R.D. Kobett. Grade Effects on Traffic Flow, Stability, and Capacity. NCHRP Report 185. TRB, National Research Council, Washington, D.C., 1978.
11. P.Y. Ching and F.D. Rooney. Truck Speeds on Grades in California. California Department of Transportation, Sacramento, 1979.
12. A.D. St. John. Truck Population on High-Type Rural Highways. Presented at 59th Annual Meeting of the Transportation Research Board, Washington, D.C., 1980.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

Left-Turn Equivalencies for Opposed, Shared, Left-Turn Lanes

JOHN H. SHORTREED

ABSTRACT

A study of intersection signal optimization in a medium-sized city indicated that there were a large number of approaches without an exclusive left-turn lane, but rather with a shared left and through lane. The study also found that the accurate determination of the capacity of a shared left-turn lane was critical to the analysis of intersection performance. Moreover a search of the literature suggested that there are no current methods of estimating the capacity of an approach containing a shared left-turn lane. A method is proposed for estimating the capacity of a shared left-turn lane. The method builds on the existing procedures for exclusive left-turn lanes first developed by the British in 1966 and since extended in Australia and the United States. The proposed method develops an estimate of the flow of through vehicles during the initial part of the green display when left-turn vehicles are blocked by the opposing flow. Subsequent gap acceptance behavior is modeled to reflect the length of the displayed-green time. Finally, the flow after green is modified to reflect the probability of the shared left-turn lane actually being blocked when the green ends. The resulting capacities compare favorably with results quoted in the literature for cases where conditions in a shared left-turn lane are similar to those in an exclusive left-turn lane situation. The results are given both in tabular form and as a set of iterative equations.

A procedure for calculating left-turn equivalencies for opposed, shared, left-turn lanes is presented. For approaches to an exclusive left-turn lane, the results are compared with observations reported in the literature.

It was observed that in the analysis and design of signalized intersections many solutions involve the use of a lane for both left and through traffic flow. The absence of explicit methods for estimating the capacity of this type of lane often leads to the exclusion of shared lanes from analysis in favor of exclusive left-turn lanes or exclusive signal displays for which analysis methods are available. This in turn often leads to nonoptimal design of signal timing.

PROPOSED METHOD

A method of estimating the capacity of a shared, opposed, left-turn lane is presented. The method is basically an extension of traditional methods for the analysis of an exclusive, opposed, left-turn lane (1-4). The main extensions of the traditional methods are

1. Through vehicles in a shared left lane are allowed to move during the period when left turns are blocked by the opposing flow;
2. The gap acceptance mechanism is modified to reflect the limitation on the possible size of gaps;
3. Flow after green is not allowed if the shared left lane experiences free flow with no blockage by left-turning vehicles; and
4. The allocation of through vehicles to the shared left lane and other lanes of the approach is modeled directly.

The results of the method are presented and where appropriate these results are compared with observed results reported in the literature. The proposed method has been incorporated in a basic computer program for the critical movement analysis of an intersection.

The opposed, shared, left-turn situation is shown in Figure 1. A critical movement analysis is assumed, and one phase is shown with an opposed, shared, left-turn lane. The flows are given in vehicles per hour (vph) but these are normally the design flow rates associated with the peak 15-min period. A movement (i.e., ( ) in Figure 1) is a lane or lanes that for analysis purposes can be separated from the total approach but cannot be further subdivided because of common directional traffic flows. The opposing flow is the movement [(1)(2)] that consists of through and right-turn flows in lanes

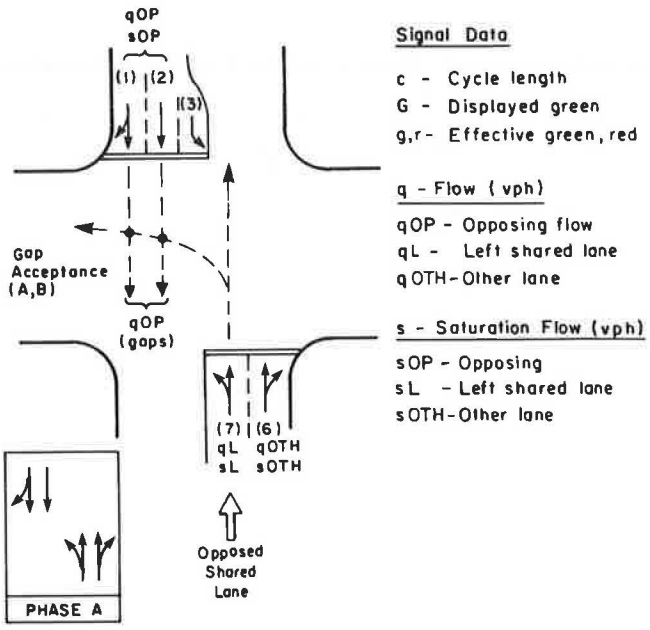


FIGURE 1 Typical opposed, shared, left-turn lane situation.

(1) and (2). This opposing flow (qOP) and its associated saturation flow (sOP) are used to determine the time when no left turns can be made from the shared lane (7). In Figure 1 qOP(gaps) is the opposing flow that interfers with the gap acceptance of the left-turn movement. Typically qOP(gaps) is just qOP less any opposing flows that do not interfere with left turns. If in Figure 1 the opposing right turns in lane (1) interfered with the left turns in lane (7) (e.g., if both turned into the same lane), the opposing right turns would be included in qOP(gaps); otherwise they would be excluded. This decision is left to the analyst.

The criterion for selecting the critical movement for the partial situation shown in Figure 1 is to select the movement with the greatest degree of saturation or the largest ratio of flow to capacity. Capacity is the saturation flow (s) times the ratio of effective green time to the cycle length (g/c). For purposes of explanation it is assumed that the movement lost times are all equal and that all movements in Figure 1 have the same g/c ratio. Critical movement for phase A is the largest degree of saturation (x) between

$$x[(1)(2)] = qOP/[sOP(g/c)]$$

and

$$x(6) = x(7) = qOTH/[sOTH(g/c)]$$

The shared left-turn lane (7) is a part of movement [(6)(7)], and it is assumed that at capacity both lanes will have an equal degree of saturation. That is, at capacity the through movement would be distributed between lane (6) and lane (7) so that both lanes would have the same degree of saturation. This condition results in a requirement for a trial-and-error solution for determining the saturation flow for the opposed, shared, left-turn lane because, as will be seen, its capacity is a function of the percentage of through vehicles in the lane.

In Figure 1 the critical movement in phase A will be one of movement [(1)(2)] or movement [(7)(6)]. Given the flows and signal timing, this largely depends on the estimated saturation flow. The calculation for the saturation flow is relatively straightforward except for the shared left-turn lane.

Saturation Flows

Saturation flows in Figure 1 are given in vph and are defined as

$$s = \sum_{\substack{\text{lanes}\\ \text{in}\\ \text{movement}}} [s(\text{lane type}) \cdot f(w) \cdot f(Lt) \cdot f(Rt) \cdot f(gr) \ldots] \quad (1)$$

where

- s(lane type) = the basic saturation flow measured for local drivers for a specified lane type,
- f(w) = lane-width adjustment factor,
- f(Rt) = right-turn adjustment factor,
- g(gr) = grade adjustment factor,
- f(Lt) = left-turn adjustment factor, and
- .... = other adjustment factors (e.g., trucks).

The left-turn factor is defined as

$$f(Lt) = 100/[(\%Lt \cdot LTE) + (1-\%Lt)] \quad (2)$$

where %Lt is the percentage of left turns in the movement and LTE is the left-turn equivalent (the number of through cars that would use up just as much capacity as one left-turning vehicle).

In Equation 1 there is a right-turn adjustment factor, f(Rt), based on a right-turn equivalent (RTE) analogous to f(Lt) and LTE. The LTE in Equation 2 is found by estimating the capacity of lane (7) with left turns and then comparing this with the capacity with no left-turning vehicles in the lane.

Capacity of Opposed, Shared, Left Lanes

Figure 2 shows the basis for calculating the per cycle capacity of an opposed, shared, left lane. In Figures 2a and 2b the opposing flow builds up a queue during the effective red phase. This queue discharges at the saturation flow and then the opposing flow occurs at the average level of qOP. It is assumed that the opposing flow arrives at random, that the intersection is not oversaturated, and that there are usually no vehicles left over from the previous cycle. The time taken for the opposing queue to discharge (i.e., the saturation time for opposing flow) is

$$stOP = qOP\ [(c - g)/(sOP - qOP)] \quad (3)$$

In period 1, which is stOP seconds long, only through vehicles usually can proceed. The average number of vehicles that can proceed per cycle is

$$T1 = \sum_{i=1}^{n} [1 - (\%Lt/\%L)]^i \quad (4)$$

where

- T1 = through vehicles per cycle in time period 1;
- %Lt = percentage left turns in the total movement (e.g., movement [(6)(7)] in Figure 1);
- %L = percentage of the movement traffic in the left lane; that is, 100 [qL/(qL + qOTH)]; and
- n = maximum number of through vehicles in stOP (i.e., stOP/sL where sL is the unopposed saturation flow rate for the left lane).

In Equation 4 (%Lt/%L) is the proportion of left-turning vehicles in the shared lane, and one less this is the proportion of through vehicles. The probability of there being one through vehicle per cycle is just the probability that the first vehicle in the left lane is a through vehicle and this is equal to the proportion of through vehicle in the left lane. The probability of a second through vehicle is the joint probability that the first vehicle is a through vehicle and the second vehicle is also a through vehicle (i.e., the proportion of through vehicles squared). This continues for the number of through vehicles per cycle that could proceed in stOP seconds. As shown in Figure 2 the flow rate for through vehicles falls off as there is a higher and higher probability of having the lane blocked by a left-turning vehicle.

In time period 2 of Figure 2 there will be a number of queued left-turning vehicles that will flow at the left-turn saturation flow. The flow in time period 2 then is

$$L2 = T1\ \{[\%Lt/\%L]/[1 - (\%Lt/\%L)]\} \quad (5)$$

where

- L2 = left-turning vehicles per cycle in time period 2 and the time taken for L2 is t2 = L2(3600/sLgap) (Equation 6) where t2 is time for period 2 (sec) and

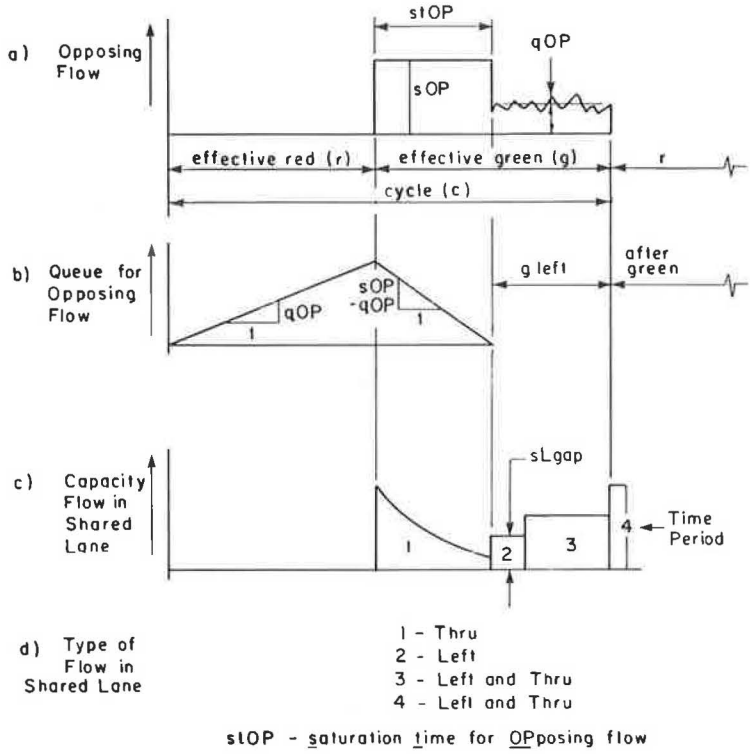


FIGURE 2 Basis for calculating capacity of an opposed, shared, left lane.

sLgap = saturation flow for left-turning vehicles through gaps in the opposing flow.

In time period 3 a mixture of through and left-turning vehicles will proceed with a capacity of

$$LORT3 = [(g - stOP - t2)/3600]/s3 \quad (7)$$

where

LORT3 = left or through vehicles per cycle for time period 3,
s3 = saturation flow rate for period 3 found from the average headway for left-turning and through vehicles, and

$$s3 = 3600/[(\%Lt/\%L) \cdot (3600/sLgap)] + [(1-\%Lt/\%L) \cdot (3600/sL)]$$

The other variables are as before.

Finally, in time period 4, the after-green period, left-turning or through vehicles proceed (LORT4). The maximum number depends both on the geometry of the intersection and on the characteristics of the drivers. Maximum values are 1.0 to 2.0 vehicles per cycle and a typical value is 1.5. The minimum value of LORT4 is zero, which occurs either when there are very few left-turning vehicles or when there is very little opposing traffic. Under these conditions the traffic in the left lane is not blocked by left-turning vehicles and the traffic proceeds smoothly for the whole of the effective green period. As a result there are no blocked vehicles to move after the green. Thus, LORT4 is

$$LORT4 = 1.5 \quad \text{if } [(3600 \times \%Lt)/(sLgap \times \%L)] > 1.8 \quad (8)$$

$$LORT4 = 1.5\ [(3600 \times \%Lt)/(sLgap \times \%L)] \quad \text{otherwise}$$

$$(\text{s.t. } LORT4 \geq 0)$$

where

LORT4 = left-turning or through vehicles per cycle in period 4,
%Lt = percentage left turns in the movement, and
%L = percentage of movement in the left lane.

In Equations 6, 7, and 8 the value of sLgap, the opposed left-turning flow, must be estimated. This is usually based on the following formula (2,5):

$$sLgap = qOP(gap) \times \{[\exp(-A \times qOP'(gap))]/[1 - \exp(-B \times gOP'(gap))]\} \quad (9)$$

where

sLgap = saturation flow for opposed left turns in vehicles per hour,
qOP'(gap) = opposing flow for gaps in vehicles per second,
qOP(gap) = opposing flow in vehicles per hour,
A = average critical gap acceptance for the first left-turning vehicle in any gap, and
B = average critical gap acceptance for any following left-turning vehicles in any gap.

Equation 9 is for the situation where there is no limit to the size of gaps in the opposing flow. However, in the most extreme case, a gap in the opposing traffic cannot be greater than (g-stOP) or (gleft) seconds. Moreover, when the opposing flow is in only one lane, the opposing flow cannot be less than a minimum value. When this is recognized, sLgap is less than the value given by Equation 9, and, assuming either an exponential or shifted exponential distribution, sLgap can be estimated as follows:

$$sLgap = qOP(gap)\{\exp[-(A/qOP')]$$

$$+ \exp[-(A + B)/qOP'] + \exp[-(A + 2B)/(qOP')] + \ldots\} \quad (qOP \text{ is multilane})$$

$$sLgap = qOP(gap)\ \{\exp[-(A - Hmin)/(qOP' - Hmin)] + \exp[-(A + B - Hmin)/(qOP' - Hmin)] + \ldots\} \quad (qOP \text{ is 1 lane}) \qquad (10)$$

where

sLgap = saturation flow for opposed left turns,
qOP(gap) = opposing flow for gaps (vph),
qOP' = opposing flow for gaps (sec/vehicle),
A,B = gap acceptance parameters [A = 4.5 sec for one lane (+ 0.5 sec/additional lane) and B = 3600/sL for one lane (+ 0.5 sec/additional lane)],
Hmin = minimum headway for single lane flow 2 (seconds), and
+ ... = terms continue while A + nB $\leq$ gleft.

Given T1, L2, LORT3, and LORT4, the capacity of the left lane per cycle or per hour can be found. At capacity for the combined movement (e.g., movement [(6)(7)] of Figure 1), the capacity of the left lane must correspond to %L, the percentage of the approaching traffic choosing the left lane. If this is not the case, then %L is changed and Equations 4-8 and 10 must be re-evaluated with the new value of %L. This iterative procedure continues until %L does not change. The capacity of the left-turn lane has then been estimated and a check is made to see that there is capacity for left turns:

$$\%L > \%Lt \qquad (11)$$

Finally, the LTE can be calculated by

$$LTE = \{[sL \times (g/3600) - (T1 + L2 + LORT3 + LORT4)]/[(\%Lt/\%L) \times (T1 + L2 + LORT3 + LORT4)]\} + 1 \qquad (12)$$

where LTE = left-turning vehicles' equivalent in terms of through cars.

Figure 3 shows an example calculation of an LTE. In Figure 3 the initial estimate of the percentage of traffic in the left lane (%L) was very accurate and no further iterations were necessary. Had this not been the case, further calculations would have been needed. This process is not onerous if a computer program is available, but for hand calculations it is clearly not very useful and some other procedures are needed. In the remainder of this paper, tabulated results of LTE values as calculated by the proposed equations for opposed, shared, left-turn lanes are presented, and these results are compared with those in the literature.

Tabulated Values of LTE

At the outset it should be recognized that LTE values will depend on the gap acceptance parameters and the number of vehicles after green. These will vary among situations. Table 1 gives the LTE for some typical situations calculated using Equations 4, 5, 10, 6, 7, and 8. In addition, Table 1 provides the LTE estimates for opposed left turns in exclusive left-turn lanes given by both the old Australian method (2) and the new Australian method (1).

The new Australian method suggests that "when the opposed turning vehicles and through vehicles share the same lane, the saturation flow be calculated

Data

qOP	= 600 vph	A	= 5 sec.
qOP(gaps)	= 600 vph	B	= 3600/1440
sL	= 1440 vph	g	= 24 sec.
sOTH	= 1500 vph	c	= 60 sec.
SOP	= 1550/lane	RTE	= 1.0
%Lt	= 10%	2	Opposing lanes
%Rt	= 10%		

Calculations

Equ (3) $stOP = 600\ \frac{(60 - 24)}{(3100 - 600)} = 8.6$ sec.

Assume %L = 41%

Equ (4) $T1 = \sum_{i=1}^{3} (1-(.1/.41))^i = 1.76$ thru vehicles/cycle

Equ (5) $L2 = 1.76\ (10/41)/(1-(10/41)) = .57$ left veh./cycle

Equ (10) $sLgap = 600(e^{-5/6} + e^{-8/6} + e^{-11/6} + e^{-14/6}) = 559$ vph

Equ (6) $t2 = .57 \times (3600/559) = 3.7$ sec.

Equ (7) $s3 = 3600/(\frac{10}{41} \times \frac{3600}{559}) + ((1 - \frac{10}{41}) \times \frac{3600}{1440}))$

$= 1040$ vph

$LORT3 = \frac{(24-8.6-3.7)}{3600} \times 1040 = 3.38$ veh/cycle

Equ (8) $\left\{3600 \times \frac{\%Lt}{SL*\%L}\right\} = \frac{3600x10}{559x41} = 1.57$, thus

$LORT4 = 1.5(1.57/1.8) = 1.31$ veh/cycle

Left lane capacity = T1 + L2 + LORT3 + LORT4

= 1.76 + .57 + 3.28 + 1.31 = 7.02 veh/cycle

or = 427 vehicles/hour

Capacity of Other lane = (24/60) 1500 = 600

Check % in left lane = (422/(422/600))x100 = 41.3% (close enough to 41% assumed)

Check capacity for % left turns : 41.3% > 10% - o.k.

Equ (12) $LTE = \frac{1440(24/3600) - (1.76+.57+3.28+1.31)}{(10/41)(1.76+.57+3.28+1.31)} + 1$

$= 2.51$

FIGURE 3 Example calculation of the capacity of shared, opposed, left-turn lane and LTE.

. . . [in] the same method as . . . exclusive lanes" (1,p.15). The method is to adjust the left-turning vehicles only by the LTE found from

$$LTE = [(sL/3600)g]/[(sLgap \cdot gleft) + LORT4]$$

where

sLgap = [qOP'(gap) exp(-A qOP'(gap))]/[1 - exp(-B qOP'(gap))] (vehicles/second) (see Equation 9),
g = effective green time for the movement, and
gleft = g - stOP - t2.

The values of sL, sOP, A, and B are the same for all values of LTE in Table 1. This removes any variation in results caused by different characteristics of drivers. For example, Michalopoulos (3) in studying exclusive left-turn lanes did not control for the parameters used to reflect driver characteristics in different countries when this was in fact the only difference between several of the models he was evaluating. The LTE results given in Table 1 for Miller's procedure (2) are based on the same parameters for driver characteristics as the proposed method. Miller's LTE values were developed for exclusive left-turn lanes but are recommended for use for left-turning vehicles only in shared left-turn lanes. The recommended equation (2) is

TABLE 1 Comparison of Left-Turn Equivalencies for Opposed, Shared, Left-Turn Lanes (sL = 1440; sOP = 1550; A = 4.5, 5.0; B = 2.5, 3.0)

	ONE opposing lane (A=4.5; B=2.5)		TWO opposing lanes (A=5; B=3)		
c = 60	(g/c)=.4	(g/c)=.6	(g/c)=.2	.4	.6
qOP = 400	2.9 2.7 2.6 2.0 2.4	1.8 1.6 1.7 1.6 1.9	3.2 2.7 2.7 1.8 1.9	1.9 1.8 1.9 1.7 1.7	1.7 1.6 1.7 1.6 1.7
600	6.2 6.0 6.0 5.4 5.6	3.0 2.8 2.6 2.3 2.7	2.4 3.1 3.2 1.8 1.9	2.5 2.5 2.4 2.2 2.3	2.1 2.0 2.0 2.0 2.1
800	*	6.2 5.2 5.2 4.1 4.7	*	3.3 3.3 3.2 3.0 3.0	2.6 2.6 2.6 2.5 2.5
c = 90					
qOP = 400	3.4 2.8 2.6 2.3 2.8	2.0 1.8 1.8 1.7 2.1	3.6 3.4 3.2 2.3 2.3	2.0 1.9 1.9 1.8 1.9	1.7 1.7 1.7 1.7 1.8
600	8.7 9.4 9.4 7.4 7.8	3.4 3.0 2.9 2.5 3.0	4.3 4.7 4.7 4.4 4.4	2.8 2.7 2.6 2.5 2.5	2.3 2.2 2.2 2.2 2.2
800	*	7.2 6.2 6.1 4.8 5.6	*	4.1 3.6 3.6 3.5 3.5	2.8 2.8 2.8 2.8 2.8
c = 120					
qOP = 400	3.7 2.9 2.7 2.4 3.0	2.2 1.9 1.8 1.7 2.1	3.9 3.6 3.5 2.6 2.7	2.1 2.0 2.0 1.9 2.0	1.9 1.8 1.8 1.8 1.9
600	12.4 12.4 12.4 9.2 9.8	3.8 3.2 3.0 2.6 3.2	6.3 6.2 6.2 5.7 5.7	3.1 2.8 2.7 2.7 2.7	2.4 2.3 2.3 2.3 2.3
800	*	7.6 6.9 6.9 5.2 6.2	*	4.6 3.9 3.9 3.9 3.9	3.1 3.0 2.9 2.9 2.9

LEGEND

* opposing flow exceeds capacity

(2.7) - not enough capacity for left turns

Table Entries

Proposed Equation	
10% lefts	
20% lefts	new Australian method *(1)*
30% lefts	old Australian method *(2)*

LTE = [(sL/3600) x g]/[((f/B) x gleft)

+ LORT4] (13)

f was determined by simulation (i.e., where qOP is 0, 200, 400, 500, or 800, f is 1.0, .81, .65, .54, or .45, respectively)

where

B = average unopposed headway for the lane (e.g., 3600/sL),
g = effective green time for the movement, and
gleft = g - stOP.

The results given in Table 1 for the proposed method and the Australian results differ because of three considerations:

1. The proposed method considered the movement of through vehicles during stOP (saturation flow period for the opposing traffic);
2. The proposed method estimates sLgap only for the available gap sizes; and
3. The proposed method discounts LORT4 (flows after green) when the flow in the left lane approaches a free flow condition because of a low percentage of left turns or a low opposing flow.

Because the two Australian methods (Table 1) were derived for exclusive left-turn lanes, they do not vary with the percentage of left-turning vehicles for the approach. The proposed method with 30 percent left turns is the closest result to an exclusive left-turn lane. All three methods give similar values, with Miller's method generally giving values closer to those of the proposed method. For higher cycle times and large g/c ratios all methods are essentially equivalent for heavy left-turn movements in the left lane.

With lower, more typical values of the percentage of left turns (Table 1) the use of LTE from an exclusive left-turn lane analysis underestimates the effective LTE of the left-turning vehicles. The use of the proposed LTE values given in Table 1 is preferred.

Gap Acceptance Parameter Values

In the proposed method, the LTE given in Table 1 will vary with sL, the saturation flow for the shared left lane and the gap acceptance parameters A (initial vehicle) and B (following vehicles). The suggested values were taken from average North American values given in Table 2 (1). The proposed method varies gap acceptance parameters A and B as the number of opposing lanes is varied.

TABLE 2 Gap Acceptance Parameters Used in the Literature (*1*)

	Initial Gap A (sec)	Saturation Flow sL (vph)	Following Headway B (sec)	LORT4 After Green
Gordon and Miller (*8*)	5	1,200	3.0	1.5
Webster and Cobbe (*4*)	5 and 6 (1 & 2 lanes)	1,400	2.5	
Fambro et al. (*5*)	4.5	1,440	2.5	1.6
Peterson et al. (*9*)	4.8-5.8		0.54A	

Michalopoulos (3) gives two regression equations for finding sLgap based on observed data for five intersections in upstate New York. The equations for signalized intersections are

$$sLgap = -1.245\ qOP + .000\ 014\ qOP^2\ A + 1165 \quad \text{(1 lane)} \tag{14}$$

$$sLgap = -0.875\ qOP + .000\ 012\ qOP^2\ A + 1145 \quad \text{(2 lanes)} \tag{15}$$

Table 3 gives some selected comparisons of sLgap as estimated by the proposed method with those predicted by Equations 14 and 15. The sLgap estimates of Michalopoulos have a constant value and do not vary with the percentage of left turns. There is good agreement between the two results for two-lane opposing traffic and longer cycle lengths. The calculated results vary with the situation, and the Michalopoulos values do not. No information is given about signal conditions; however, it appears that in general the proposed typical gap acceptance parameters produce results that are comparable with those observed by Michalopoulos.

OTHER ANALYSIS METHODS

Not much data has been found in the literature related to left-turn equivalencies for shared left-turn lanes. The 1965 Highway Capacity Manual (HCM) (6) provides an adjustment factor for an approach with a shared left-turn lane. If a 12-ft lane width is assumed, then Table 6.5 of the HCM implies a constant value for LTE of 4.0 for a one-lane approach containing the shared left lane and a value of 2.0 for a two-lane approach. The present study has investigated only a two-lane approach for the shared left lane.

For an exclusive left-turn lane the HCM (6) suggested a value of sLgap:

$$sLgap = 1200 - qOP(gap) \tag{16}$$

where the opposing flow for gaps assumes 5 percent trucks. Moreover, if the left-turn capacity due to sLgap was less than LORT4, then the LORT4 capacity was used. However, LORT4 and sLgap would not be used in combination.

The proposed TRB critical lane analysis (7) deals with shared left turns in two different ways. At step 7 in the procedure for checking for critical lanes an LTE for either a shared or an exclusive left-turn lane is proposed:

Flow qOP	LTE
1-299	1
300-599	2
600-999	4
>1000	6

Comparison with the data in Table 1 suggests that these values are of appropriate magnitudes, but these values do not respond to changes in the percentage of left turns, the number of opposing lanes, and so forth.

In the proposed TRB critical lane analysis (7) at step 4 there is a left-turn check to establish the adequacy of the left-turn capacity. It is similar to the HCM 1965 in that it uses the maximum of LORT4 or the capacity due to sLgap but not both. The proposed formula for sLgap (7) is

$$sLgap = [(g/c)(1200)] - qOP(gap) \tag{17}$$

The original British procedure for saturation flow analysis suggested two ways to deal with shared, opposed, left-turning vehicles (4). The first suggestion was to use an LTE of 1.75 for all left-turning vehicles. The second procedure was to use sLgap for a period of gleft and then set up an extended green phase to handle any excess traffic. This method did not deal with through movements

TABLE 3 Comparison of Vehicles per Hour as Calculated and as Observed (*3*)

	One Lane		Two Lane	
qOP ↓ / g/c =	.4	.6	.4	.6
c = 60				
200	764/898	1006/898	652/992	823/992
600	271/421	684/421	589/662	657/662
1000	*	*	316/350	440/350
c = 90				
200	949/898	1118/898	803/992	924/992
600	52/421	691/421	643/662	662/662
1000	*		394/350	441/350
c = 120				
200	1053/898	1159/898	886/992	962/992
600	67/421	691/421	657/662	663/662
1000	*	*	423/350	441/350

as calculated/Equ (14) or (15)

* opposing flow greater than capacity

during stOP nor did it explicity suggest the use of LORT4, the flow after green.

The British method was extended by Miller (2) who developed LTE for exclusive left-turn lanes and then recommended its use for shared lanes. The left-turn capacities were estimated by sLgap for gleft and to this was added LORT4 for the after-green flow. The value of sLgap was found from simulation, and Table 1 gives the resulting LTE. Recently the Australian procedure has been revised in that sLgap is estimated by Equation 9 rather than Equation 13, and gap acceptance parameter and saturation flow values were changed. The basic procedure was unchanged.

Fambro et al. (5) applied the Australian procedure to traffic in Texas. They estimated gap acceptance parameters and LORT4 after-green flows and found there was good agreement between the observed LTE for exclusive left-turn lanes and the Australian procedure. They also found that below ultimate capacity levels traffic had a tendency to concentrate in one lane of a multilane approach. This tended to increase the estimate of the time of saturation flow for the opposing movement (stOP). This refinement has not been included in the proposed procedure. Fambro et al. (5) estimated the capacity for a shared left-turn lane by assuming a 50-50 split of through and left-turning vehicles in the shared lane. This assumption is very restrictive and is not representative of many situations.

Fambro et al. (5) found, for limited data, that the after-green LORT4 flow was 1.41 vehicles per cycle for separate left-turn lane and 1.03 vehicles per cycle when there was no separate left-turn lane. The proposed method suggests 1.5 vehicles for a shared left lane and these can be left-turning or through vehicles (LORT4).

CONCLUSIONS

A proposed procedure is presented for estimating left-turn equivalents (LTE) for left-turning vehicles in shared left-turn lanes. The method is an extension of the Australian procedure for estimating LTE for exclusive left-turn lanes.

It is proposed that the opposed saturation flow for left-turning vehicles (sLgap) be reduced to reflect the length of available green time (gleft) left after the time of saturation flow for the opposing flow (stOP). These results appear to be consistent with observed results reported in the literature.

It is proposed that the vehicles after green (LORT4) are reduced under conditions where blockage of the shared left lane by left-turning vehicles is not likely to occur.

The results indicate that, for higher cycle times and heavy percentage of left turns, Equation 13 can also be used to estimate LTE, but for other situations it is better to use the proposed method.

It is recommended that field studies be carried out to gather more data on shared lanes because there are few results in the literature. The results indicate that many shared left lanes can have quite high capacities and should be considered as a viable alternative to other approaches such as the use of an exclusive left-turn lane or an exclusive phase.

ACKNOWLEDGMENT

This research was made possible by a grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. R. Akcelik. Workshop on Area Traffic Control, Part 1: Signalized Intersection Capacity Workshop. Internal Report AIR-1094-1. Australian Road Research Board, Numawading, Victoria, 1979.
2. A.J. Miller. Signalized Intersections--Capacity Guide. Report 79. Australian Road Research Board, Numawading, Victoria, 1968.
3. P.G. Michalopoulos, J. O'Connor, and S.M. Novoa. Estimation of Left-Turn Saturation Flows. In Transportation Research Record 667, TRB, National Research Council, Washington, D.C., 1978, pp. 35-41.
4. F.V. Webster and B.M. Cobbe. Traffic Signals. Road Research Laboratory Technical Paper 56. Her Majesty's Stationery Office, London, England, 1966.
5. D.B. Fambro, C.J. Messer, and D.A. Anderson. Estimation of Unprotected Left-Turn Capacity at Signalized Intersections. In Transportation Research Record 644, TRB, National Research Council, Washington, D.C., 1977, pp. 113-119.
6. Highway Capacity Manual 1965. HRB Special Report 87. HRB, National Research Council, Washington, D.C., 1965, 397 pp.
7. Interim Materials on Highway Capacity. Transportation Research Circular 212. TRB, National Research Council, Washington, D.C., Jan. 1980, 276 pp.
8. I.D. Gordon and A.J. Miller. Right Turn Movements at Signalized Intersections. Proc., 3rd Australian Road Research Board Conference, Vol. 3, No. 1, 1966, pp. 446-459.
9. B.E. Peterson, A. Hansson, and K.L. Bang. Swedish Capacity Manual. In Transportation Research Record 677, TRB, National Research Council, Washington, D.C., 1978, pp. 1-28.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

Analysis of Unsignalized Intersection Capacity

ADEBAYO B. BAKARE and PAUL P. JOVANIS

ABSTRACT

The conceptual bases of unsignalized intersection capacity procedures from the Swedish Capacity Manual and from the Interim U.S. Capacity Manual are compared, and the procedures are evaluated with field data from four intersections in the Chicago, Illinois, region. The Swedish method, based on principles of queueing theory and using parameters reflecting Swedish conditions, grossly overestimated capacity and underestimated delay. The empirically based U.S. method substantially underestimated capacity, resulting in level of service values that were approximately one level of service too low. Causes of the poor estimation included critical gaps that were too small in the Swedish method and too large in the U.S. method and inappropriate definitions of dominant volumes. Based on the conceptual and empirical comparisons, critical gaps were revised and dominant volumes were redefined in both methods. Evaluations of these revised procedures revealed much closer correspondence to field-measured values for capacity and delay. With revised critical gaps, the interim U.S. method appears to correspond fairly well with field data. The Swedish method, however, contained errors in capacity and delay estimates that were traced to probable differences in driver performance and traffic conditions in Sweden and in the United States. Preliminary tests of the Swedish delay model indicate that it can provide very accurate delay estimates if it is revised to reflect U.S. drivers and driving conditions.

Intersections of streets at grade in urban regions are critical portions of highways because they are primary sites of traffic accidents and points of considerable congestion and delay. The efficiency and capacity of the entire street system is generally dependent on the characteristics of the intersections in the system. Although there has been considerable research conducted on the operation of signalized intersections, comparatively few studies have examined the operation of unsignalized intersections (1).

The Highway Capacity Manual (2) published in 1965 treated unsignalized intersections summarily. It recommended that where the volumes on the two minor streets at a two-way stop-signed intersection are low and most vehicles arriving at the STOP sign can enter or cross without substantial delay, an approximate method can be used to determine service volumes. "[A] signalized condition in which the signal split is prorated directly on the basis of the relative volumes on the intersecting streets, and inversely on the basis of their relative widths..." is assumed (2,p.156).

Since 1965 research studies in Germany (3), Sweden (4), and Australia (5) have developed calculation procedures for capacities at unsignalized intersections. The German procedure was translated into English as Capacity of At-Grade Intersections (6) and after some modifications was adopted and published in Transportation Research Circular 212 (7,pp.37-72).

The TRB Subcommittee on Unsignalized Intersections submitted a memorandum report on the Circular 212 procedure at the TRB Annual Meeting in Washington, D.C., in January 1982. The report compared the Circular 212 procedure with the Swedish and Australian capacity manuals and applied it to some field data. Comments from other users of the procedure were received. The subcommittee concluded that the Circular 212 procedure used critical gap times that were too long, resulting in unusually low calculated capacities and level of service, especially for left turns and through movements from the minor road. Based on sight distance considerations, the subcommittee recommended the use of a maximum critical gap of 8.5 sec. Critical gap values given in the procedure were adjusted downward by 0.5 to 1.0 sec to allow for differences between German and U.S. driving conditions and drivers.

The subcommittee acknowledged that the procedure in Circular 212 represented a major improvement in the evaluation of performance of nonsignalized intersections. Furthermore, it called for carefully structured research to study driver behavior and the various effects of geometric features. The group recommended that carefully constructed problems be worked out using this procedure as well as the Swedish and Australian methods.

RESEARCH OBJECTIVES AND METHODOLOGY

The overall objective of this research was an empirical comparison of procedures to estimate unsignalized intersection capacity and level of service. Before the empirical analysis was conducted, each major capacity analysis method was reviewed in terms of theoretical structure, applicability to a broad range of traffic conditions, and computational procedure. Procedures from Australia (5), Great Britain (8), Sweden (4), and the United States (7) were reviewed and compared as part of a recently completed master's thesis at Northwestern University (1). Based on this comparison and the recommendations of the TRB subcommittee report, the Swedish method and the method described in Circular 212 were selected for further testing.

An overview of the study methodology for the empirical testing is shown in Figure 1. First, each major component of the procedures (critical gaps, capacity, and level of service) is analyzed individually. Based on the comparison of theoretical and computational structure as well as the empirical results, each model is revised and again examined with respect to field data. The methodology sought to isolate the prediction errors that were due to model structure from those that were due to differences in drivers and driving conditions, which were captured in the Swedish and German data.

The analysis of the two unsignalized intersection capacity procedures includes the following specific steps.

1. Comparison of the conceptual and computational structure of the methods;

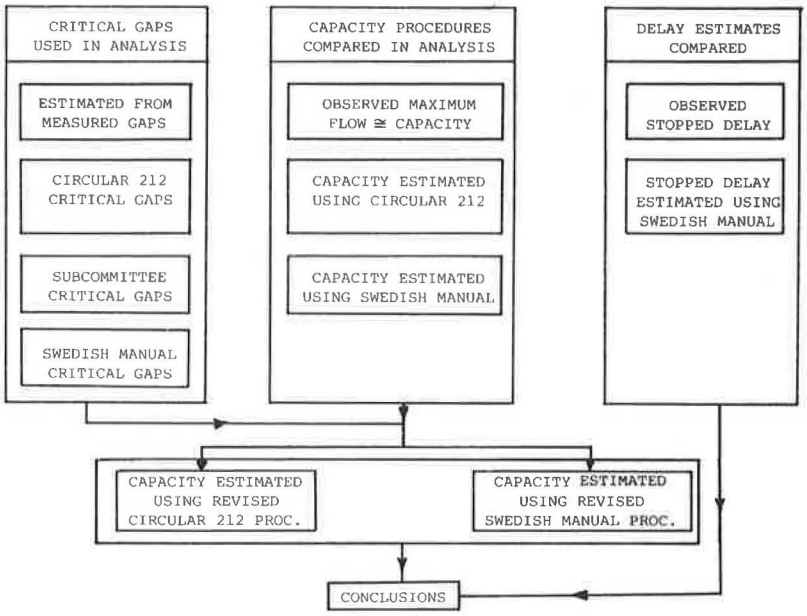


FIGURE 1 Flow chart of study methodology.

2. Collection of data on gaps at four two-way stop-sign controlled intersections and analysis of these data to estimate the mean critical gaps for the minor road movements;
3. Comparison of the critical gaps obtained from field data with those in Circular 212, the subcommittee report, and the Swedish manual;
4. Determination of the capacities of minor road approaches and the level of service to minor road vehicles at the studied intersections using two capacity methods followed by comparison of these results with observed capacities and delays; and
5. Recommendation of revisions to the Circular 212 and Swedish procedures that would make the procedures more closely reflect U.S. drivers and driving conditions.

CAPACITY CALCULATION PROCEDURES

Transportation Research Circular 212

The procedure in Circular 212 (1) is a method for computing the capacity and level of service of priority-type nonsignalized intersections (i.e., those intersections controlled by two-way STOP or YIELD signs). The computational steps involve identifying the nonpriority movements (all movements from the minor approaches and left turns from the major road), determining the traffic streams conflicting with each of the nonpriority movements, and then applying the appropriate critical gaps. The maximum (or potential) capacity for each movement is read from the graph of maximum capacity versus conflicting traffic stream. These capacities are then adjusted to account for intersection congestion caused by the servicing of conflicting movements.

When a lane is shared by traffic making more than one movement, the capacity of such lanes is determined using a proportioning equation of the form:

$$M_{134} = [C_R + C_T + C_L]/[(C_R/M_1) + (C_T/M_3) + (C_L/M_4)] \quad (1)$$

where

M_{134} = capacity of all streams using the shared lane;
C_R, C_T, C_L = demand of the right, through, and left movements, respectively; and
M_1, M_3, M_4 = capacity of the right, through, and left individual streams, respectively, in passenger car equivalents per hour (pch).

The existing (or projected) traffic demand for the movement is converted to passenger car equivalents per hour (pch) to account for approach grade and traffic mix. This demand is deducted from the calculated capacity to give what is called the reserve capacity. Reserve capacity determines what level of service (A through F) is assigned to the movement and describes the traffic delay that will be expected. If minor street delays are a function of critical gap and major street volume, reserve capacity should correlate fairly well with delay. This capacity model assumes that

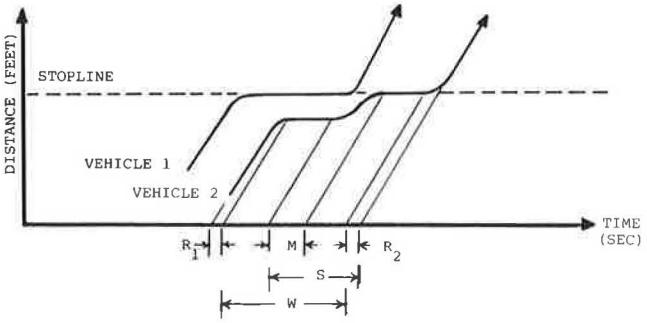
1. Vehicles on the major road arrive at random;
2. Minor road vehicles do not force their way into the intersection; and
3. Priority is given to traffic movements in the following sequence: main street through and right-turn vehicles, minor street right turns, major street left turns, minor street vehicles crossing the major street, and minor street left turns.

It is important to recognize that the model structure implies specific rules about right-of-way allocation. For example, assumptions 2 and 3 imply that no minor road crossing or left-turning vehicle will move halfway across the major street, stop (blocking major street left turns), and then proceed. The assumptions also imply that minor road crossing traffic is not impeded by minor road left turns.

Swedish Manual

The Swedish manual (SM) (4) estimation procedure is based on a theoretical queueing model that con-

siders each lane in the approach controlled by a STOP or YIELD sign as a service position. Service time is the time it takes a vehicle from the minor approach to move from the second position in queue, arrive at the stop line, and start to enter the intersection. This time and the rest of the total delay experienced by minor road vehicles are shown in Figure 2.



M = MOVE UP TIME

S = SERVICE TIME

R = R_1 + R_2 = RUNNING DELAY

W = WAITING TIME

D = R+W = TOTAL DELAY

FIGURE 2 Illustration of delay-service time relationship.

Service time for the movement was determined as a function of primary flow rates and critical gap by using the relation

$$\text{Service time} = 1/\left\{q_1 \sum_{i=1}^{\infty} [a + (i-1)a_m \int^{(a+i)a_m} f_i(h_i)dh_i]\right\} \quad (2)$$

where

q_1 = dominant volume in vehicles per second,
a = critical gap in seconds,
a_m = 0.6a or move-up time in seconds, and
$f_i(h_i)$ = headway distribution of the major road flow (three headway distributions were assumed).

The Swedish researchers used Equation 2 to develop a series of curves relating critical gap, conflicting volume, and service time that are very similar to those in Circular 212. By assuming values for major road headway distribution that are other than exponential, the Swedish were able to develop two specific tables for nonrandom major street arrivals. Capacity is estimated by taking the inverse of service time. After the estimation of service time, a series of calculations is performed for level of service factors such as queue lengths, delay, and proportion of stopped vehicles.

The implications of this procedure can best be understood by examining Figure 2, which shows the trajectory of a minor road vehicle as it approaches a stop line, waits in a queue, and then proceeds through the intersection. When major road flows are very light, there is likely to be no queue when vehicles approach the intersection. Thus service time (with a constant move-up time) will underestimate capacity in these conditions.

The authors of the Swedish manual recognized some of these problems in their discussion of the manual during the 1981 Transportation Research Board Annual Meeting. Peterson and Hannson (9) comment that U.S. traffic densities are much higher than the Swedish data base, raising concerns that U.S. conditions are beyond the range of data used to calibrate the model. Differences in driver characteristics also may affect values of move-up time (M) and gap acceptance. The empirical analyses discussed later specifically compare the Swedish and U.S. values for these variables.

Discussion

The basic steps in the two procedures are similar. The input into either calculation model is critical gaps required by the nonpriority movement and the conflicting (or, more appropriately, dominant) traffic volume to the movement. The output from either model is maximum capacity, which is adjusted for congestion impedance and shared lane to give practical capacity.

In addition to the different models used in the two procedures, Circular 212 defines conflicting streams to a nonpriority movement differently than does the SM. Furthermore, critical gaps listed in the SM are lower than those in Circular 212. A summary of other similarities and differences in the procedures is given in Table 1.

The Swedish manual procedure is very detailed, with many graphs and tables for use in corrections for variables like critical gaps, number of lanes, and major road speeds. The computational steps are difficult to follow and understand, due in part to the unedited translation into English.

FIELD DATA COLLECTION AND PREPARATION

Site Locations

The films used in this study were originally collected by Carl N. Swerdloff (April 1962), Claude Yvon Gagnon (June and July 1962), and researchers with the firm of De Leuw Cather and Company (1963). Data from a study by A. Graham Bullen (1964) are also used. Swerdloff (10) collected his information at two two-way STOP controlled intersections in Skokie, Illinois: Niles Center Road and Howard Street, and Howard Street and Kostner Avenue. Gagnon (11) also observed the intersection of Niles Center Road and Howard Street. De Leuw Cather (12) studied the intersection of Kostner Avenue and Kirk Street. Bullen (13) obtained data at the McCormick Boulevard-Bridge Street-Grey Street junction in Evanston, Illinois.

All approaches had good sight distance and speed limits of 30 mph. In addition, the sites were all located approximately 0.5 mile from the nearest traffic signal along any of the approach legs. Howard-Kostner and Kostner-Kirk intersections had two lanes on the major road, and Niles Center-Howard and McCormick-Bridge-Grey had four lanes. All the intersections were similar with respect to their crossing angle and had equal sight distance conditions from all approaches as given in Table 2.

The capacity model used in Circular 212 assumed a random or Poisson arrival of major road vehicles. This arrival pattern would result in an exponential distribution of headways (or gaps) in the major road traffic. A chi-square goodness-of-fit test was conducted comparing the actual gap distribution at the four- and two-lane, two-directional major roads with an exponential distribution. The test could not reject the null hypothesis that the gaps were exponentially distributed (α = .05).

TABLE 1 Comparison of Circular 212 and Swedish Procedures

- BASIC STEPS ARE THE SAME IN BOTH PROCEDURES
- SIMILARITIES
  - BASED ON GAP ACCEPTANCE (CRITICAL GAP).
  - RELATES CAPACITY TO VOLUME OF CONFLICTING STREAM TO THE MOVEMENT.
  - ADJUSTS MAXIMUM CAPACITY FOR CONGESTION IN THE INTERSECTION.
  - CONSIDERS LANES USED BY MORE THAN ONE MOVEMENT.
- DIFFERENCES
  - 212 ESTIMATES CAPACITY DIRECTLY, SWEDISH MANUAL (SM) ESTIMATES SERVICE TIME (A PART OF STOPPED DELAY) THEN CAPACITY.
  - 212 ESTIMATES MAXIMUM CAPACITY THEN ADJUSTS FOR CONGESTION TO FIND PRACTICAL CAPACITY: SM ESTIMATES VOL/CAP RATIO, ADJUSTS, THEN ESTIMATES CAPACITY.
  - SM ADJUSTS FOR SHORT LANE, 212 DOES NOT.
  - SM PRACTICAL CAPACITY IN VEH/HR, 212 IN PAX CARS/HR (TO ALLOW FOR VEHICLE MIX AND APPROACH GRADE).
  - SM PROCEDURE LIMITED TO VOL/CAP RATIO LESS THAN 0.8, 212 NOT SO.
  - 212 RELATES CALCULATED CAPACITY (RESERVE CAPACITY) TO LEVEL OF SERVICE: SM USES THE CALCULATED CAPACITY TO ESTIMATE DELAYS AND QUEUE LENGTHS.
  - 212 APPLICABLE ONLY TO 2 WAY STOP AND YIELD CONTROLS, SM APPLICABLE TO ALL UNSIGNALIZED INTERSECTIONS.

TABLE 2 Summary of Study Intersection Characteristics

Intersection[1]	Type	Width Traveledway		Appr. Peak Hr Vol (vph)		Parking Conditions	
		Major	Minor	Major	Minor	Major	Minor
Kostner-Kirk	2 lane 2 way / 2 lane 2 way	20'	16'	400	60	Both sides	Both sides
Howard-Kostner	2 lane 2 way / 2 lane 2 way	22'	22'	900	130	Both sides	Both sides
McCormick-Bridge	4 lane 2 way / 2 lane 2 way	42'	24.5'	600	300	No parking	No parking
Niles Ctr.-Howard	4 lane 2 way / 2 lane 2 way	52'/44'	22'	600	335	Parking on 44' appr.	Both sides

[1]Name of major street listed first.

Data Collection

All of the sites were investigated using a time-lapse camera to collect data. The camera was elevated 12-20 feet above the ground and ran, with one exception, at a film speed of 100 frames per minute (Bullen used a film speed of 60 frames per minute). In all, six films, each containing 40 min of continuous filming in the peak period, were decoded. A total of 1,152 vehicles were observed being offered 2,988 gaps (including lags), and stopped delay experienced by 587 vehicles was measured. Flow and delay data from one other film (of 1-hr duration) made by Bullen were also used.

To simplify the comparisons at all study locations, each approach was described relative to the position of the filming equipment when set up rather than to the cardinal directions. The minor road on

which the camera was set up to view the intersection was called the near side and the other minor road the far side. The major road approaches were referred to as being from the left or right (of someone looking through the camera). Further details of data collection procedures are given elsewhere (1).

The events and times decoded from the films are summarized as follows.

1. Main street vehicles' arrival times;
2. Side street vehicles' arrival times in queue at stop line;
3. Gap (and lag) acceptance or rejection;
4. Traffic volumes by approach, turning direction, and vehicle type; and
5. Service time for side street vehicles.

Critical Gap Estimation

The minimum acceptable time gap for a driver varies from one member to another of the driving population. Thus it was necessary to consider the distribution of critical gaps across the population studied. Bissell (14) showed that the common logarithm of the critical gaps has a normal distribution and suggested the application of the probit transformation. This transformation from percentages to probits forces the normal sigmoid curve of the untransformed data into a linear relationship.

The probit of the proportion (P) is defined as the abscissa that corresponds to a probability of P in a normal distribution having a mean of 5.0 and a variance of 1.0. A normalizing transformation for the gap is required so that the transformed measure (X) of the time (t) is normally distributed. The probit of the expected proportion accepting a gap is related to the gap by the following linear equation.

$$Y = 5.0 + [(1/\sigma)(X - \mu)] \quad (3)$$

where

Y = probit of the proportion accepting the gap,
X = logarithm of the gap,
μ = population mean of logarithm of critical gap distribution, and
σ = standard deviation of logarithm of critical gap distribution.

Gap sizes are aggregated into 1-sec ranges (0.0-0.9, 1.0-1.9, and so on). For each gap size, the total number of gaps offered and those accepted were tabulated according to the minor road movements--left, through, and right--as well as left turns from the major road. No gap shorter than 2 sec was accepted and all gaps longer than 10 sec were accepted.

For the estimation of mean critical gap, each movement was treated separately. The mean critical gap and the standard deviation (σ) of the critical gap distribution were then estimated by linear regression using the method proposed by Finney (15). Next, Ashworth's (16) correction for bias was applied to correct for having the gaps of cautious drivers overrepresented in the data. The correction involved subtracting $q\sigma^2$ from the estimated mean to obtain an unbiased mean of the critical gap distribution (q is the major road flow).

Comparison of Capacity Results with Observed Capacity Flows

Because this was a field validation of the capacity calculation procedures, it was desirable to compare the estimated capacities from the Circular 212 and SM procedures with observed capacity flows at the intersections. To do this, minor street approaches operating at or near capacity must be observed. Ideally, intersections operating at capacity should be observed over a 1-hr period. Unfortunately, flows close to capacity were difficult to find and observe because any such flows over a period longer than about 15 minutes would often call for the consideration of traffic signal installation at the intersection.

Nonetheless, it was possible to observe flows near capacity for a rather short period of 5 min (4:40 to 4:45 p.m.) at the Niles Center-Howard intersection. All vehicles from the far-side minor approach arriving at the intersection waited in queue to get to the stop line, and more than 70 percent of vehicles arriving from the near-side minor approach waited in queue. All major road gaps greater than 6.0 sec were used (or accepted) during this short period. Without doubt, this flow was very close to capacity, and only a few more vehicles, mainly right turns from the near-side minor approach, could have entered the major road during this 5-min period. The number of minor road vehicles that entered the intersection in the 5-min period was 64, which, if sustained over a 1-hr period, would give 768 vehicles.

During the peak 5-min period, the major road flow was 864 vph and was not appreciably higher than the mean major road flow of 820 vph. Hence, the 64 vehicles that entered or crossed the major road during the 5-min period may be, on average, sustained over all 5-min periods in the entire 1 hr.

Discussion

The new capacity procedures are expected to reflect driving conditions during the 1980s and 1990s, yet this study examines the accuracy of the methods with data from the 1960s. It is important to consider how changes in the vehicle fleet and driving conditions in general may affect this study's findings.

It is clear that there have been important changes in the vehicle fleet: trucks are now longer and heavier and automobiles are smaller and lighter. It is not believed that the differences in vehicle size would substantially change gap acceptance phenomena. In virtually all cases, the size of the accepted gap will be much larger than the size of the vehicles that define it; therefore a change in vehicle size is likely to have a minor effect on gap acceptance. Performance characteristics of automobiles and trucks, particularly acceleration capability, could have a substantial effect on gap acceptance.

Data collected by PRC Voorhees in the 1980s in their study of passenger car equivalents included data on typical vehicle accelerations from a stop line at a traffic signal to a target speed of approximately 30 mph (17). Mean accelerations for automobiles varied from 4.4 mph/sec for large automobiles (longer than 15 ft) to 4.7 mph/sec for automobiles shorter than 15 ft. Although the truck sample size was limited, mean accelerations for trucks with three or more axles ranged from 3.2 to 2.8 mph/sec.

A comparable set of figures for the 1960s vehicle population is not directly available. The 1976 Transportation and Traffic Engineering Handbook (18, pp.22-24) lists normal and maximum acceleration rates for various vehicles. The value for passenger cars was 3.3 mph/sec, less than the automobile values in the PRC study and nearly equal to their truck accelerations. Some data on truck and automobile

accelerations for safe crossing from a stop sign are contained in the U.S. geometric design guide for rural highways (19). Acceleration rates from this reference are even lower, on the order of 2 to 3 mph/sec.

It would appear that drivers in the 1980s, because of better vehicle acceleration capability, would be able to accept smaller gaps than drivers in the early 1960s. Whether drivers actually do accept smaller gaps is not at all clear. One of the difficulties in comparing acceleration rates is that little is known about the driving conditions or driver population that form the basis for the data from the 1960s (18,19). Because gap acceptance is a function, in part, of a driver's perception of a potential collision, it is not clear that the improved performance of automobiles will necessarily result in smaller critical gaps. The best comparison that can be offered at this time is a comparison of the critical gaps estimated in this study with those recommended by the TRB subcommittee; such a comparison is made in the next section.

COMPARISON OF CRITICAL GAPS, CAPACITY, AND LEVEL OF SERVICE

Critical Gap

Despite the more than 1,100 gaps available for analysis, critical gaps could be estimated reliably only for three movements: right turns from the minor road with a four-lane major road, and through movements crossing both a two- and a four-lane major road (1). The R^2 values for the transformed data ranged from 0.64 to 0.98.

Table 3 gives a summary of the critical gap values obtained from field data and the gaps suggested in Circular 212, the Swedish manual, and by the TRB subcommittee. Statistical tests of significant difference at the 5 percent level were performed between the estimated mean critical gap from this study and those from other sources. At the intersection with a four-lane major road, the mean critical gap estimate of 5.4 sec for right turns from the minor road is not significantly different from the TRB subcommittee and SM values of 5.5 and 5.3 sec, respectively, but is significantly less than the Circular 212 value of 6.0 sec. For the through movement across two- and four-lane major roads, the estimated mean critical gaps of 6.0 and 6.4 sec are significantly more than SM values of 5.6 and 5.9 sec and significantly less than Circular 212 values of 7.0 and 7.5 sec.

The estimated critical gaps are surprisingly close to critical gaps recommended by the TRB subcommittee. Consistent with review comments received about Circular 212, the critical gaps contained in the draft chapter appear to be too high. Swedish values, particularly for major road crossing traffic, appear to be too low. In considering these results, one must remember that there are several potential sources of variation in critical gaps including characteristics of the drivers in the sample (e.g., peak period versus off-peak drivers, urban versus rural), site, and vehicle characteristics. Although the findings reported here do not include all these sources of variation, the consistency of the present results with the TRB subcommittee recommendations is encouraging.

Capacity Results

Table 4 gives a summary of the results of the two capacity calculation procedures (see numbers in parentheses). Note that the Swedish manual consistently predicted capacities that were 80 percent higher than Circular 212 values. In fact, the capacity for the far side of Niles Center and Howard Street differed by a factor of nearly 4. These results are quite consistent with the recommendation of the TRB subcommittee to decrease the critical gaps in Circular 212. Reduced critical gaps would have increased the capacity estimates substantially.

In going through the two procedures, the following causes for differences were found.

1. The Circular 212 procedure resulted in higher partial volume-to-capacity (V/C) ratios for each minor road movement and hence higher V/C ratio for the entire approach than the SM.
2. Circular 212 considered more traffic movements as conflicting streams to a particular minor road movement than the SM.
3. All flows (in vph) exiting into the same approach as the minor street movement under consideration are divided into the number of lanes in that exit in the SM, whereas Circular 212 assumed that all such traffic streams conflict with the minor road movement.
4. Both procedures allow for impedance due to left-turning traffic blocking the minor road movements, but the SM adjustments reduce the calculated capacities more than do those of Circular 212.

The first three differences in procedures made the Swedish manual calculated capacities much higher than those of Circular 212, and the fourth difference only slightly reduced the calculated maximum capacity in SM relative to Circular 212.

Capacity (both minor approaches) calculated using both methods as well as observed capacity for Niles Center and Howard are given in Table 5. In addition to the capacities from each procedure individually, a separate set of calculations was made using critical gaps estimated from field data at the site. The use of the estimated gaps substantially increased the Circular 212 capacity and only slightly reduced the Swedish estimate.

The conclusion of these capacity comparisons is that neither procedure gave very accurate estimates of capacity for Niles Center and Howard. After the critical gaps had been adjusted, however, Circular 212 gave a capacity that was much more reasonable (within 10 percent of the field-measured flow).

Level of Service

An important feature of all calculation methods in the U.S. Highway Capacity Manual is the estimation of level of service (LOS). Table 6 gives a summary of the reserve capacity and level of service obtained from Circular 212. The Swedish manual does not compute level of service but contains estimates for stopped delay, which can be summed with running delay to give an estimate of total delay. For reference, the table includes a column for measured stopped delay that was obtained from the films.

The LOS values in column 3 of Table 6 can be compared with measured stopped delay to see if the LOS estimates are reasonable. The level of service A values (corresponding to delays of 5.0 and 6.5 sec) appear to be correctly predicted. There is considerable ambiguity for levels of service C, D, and E, however, because stopped delay ranges from 10.0 to 16.9 sec but is not consistent within service levels. The capacity for the far side at Niles Center is totally wrong. It yields a negative reserve capacity. These results suggest that the reserve capacity values used for intermediate levels of

TABLE 3 Critical Gap Comparisons

Vehicle Maneuver	Source	Major Road (prevailing speed 30 mph)	
		2 lanes	4 lanes
Right Turn from Minor Road	Circular 212	6.0	6.0
	TRB Subcommittee	5.5	5.5
	Swedish	5.3	5.3
	This Study	*	5.4
Left Turn from Major Road	Circular 212	5.0	5.5
	TRB Subcommittee	5.0	5.5
	Swedish	4.8	4.8
	This Study	*	*
Crossing Major Road	Circular 212	7.0	7.5
	TRB Subcommittee	6.0	6.5
	Swedish	5.6	5.9
	This Study	6.0	6.4
Left Turn from Minor Road	Circular 212	7.5	8.0
	TRB Subcommittee	6.5	7.0
	Swedish	5.8	6.1
	This Study	*	*

*Insufficient data

TABLE 4 Comparison of Calculated Capacities Using Original and Revised Procedures

Intersection	Minor Approach	Capacity (veh/hr)[1]	
		Circular 212	Swedish Manual
Kostner-Kirk	Near	720 (572)	675 (900)
	Far	633 (477)	660 (825)
Howard-Kostner	Near	325 (235)	368 (536)
	Far	335 (248)	417 (577)
McCormick-Bridge-Grey	Near	384 (288)	410 (565)
	Far	401 (301)	450 (576)
Niles Center-Howard	Near	460 (335)	497 (621)
	Far	379 (109)	416 (422)

[1]Capacity estimates using original procedures are in parentheses.

service may have to be revised and that all levels of service are too low.

The Swedish estimates of stopped delay are consistently less than measured values; in some cases the errors exceed 200 percent. Clearly, these delay estimates are inadequate considering any reasonable level of accuracy. These results, however, are consistent with the overestimates of capacity discussed in the previous section. Because the Swedish method uses the V/C ratio to compute delay, a high estimate of capacity will result in low V/C values and thus low delays.

Table 6 illustrates a fundamental difficulty of using reserve capacity as a level of service indicator: Reserve capacity cannot be measured independently in the field for comparison with calculated values to assess model accuracy. Level of service variables for freeways (density), arterials (average speed), and signalized intersections (stopped delay) can be measured in the field and compared with model predictions. Reserve capacity is a function of model estimates and therefore cannot be independently measured.

TABLE 5 Comparison of Estimates of Minor Approach Capacity with Observed Capacity

Intersection: Niles Center-Howard	
Practical Capacity Methods	Capacity of Both Minor Approaches (vph)
1. Circular 212*	
using 212 tabular critical gaps	443
using estimated critical gaps	835
2. Swedish Manual Procedure	
using Swedish manual critical gaps	1043
using estimated critical gaps	1006
Field-Measured Flow	
Based on 5 minute near-capacity flow (v/c > 0.9)	768
Major road flow in peak/hr = 820 vph	
Equivalent major road flow during 5 minute near capacity (minor road) flow = 864 vph	

*Actually pass. cars per hr., slightly less if vph.

TABLE 6 Comparison of Calculated and Field-Measured Level of Service

Intersection		Circular 212		Average Stopped Delay (sec/veh)	
		Reserve Capacity (pch)	LOS	Swedish Manual	Measured
Kostner-Kirk	N	545	A	4.0	5.0
	F	444	A	3.3	6.5
Howard-Kostner	N	176	D	4.9	10.9
	F	173	D	4.8	16.1
McCormick-Bridge-Grey	N	79	E	7.2	11.0
	F	220	C	6.2	10.0
Niles Center-Howard	N	186	D	4.8	11.1
	F	-204	-	26.5	38.2

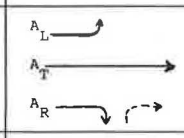
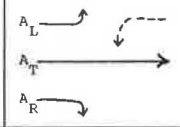
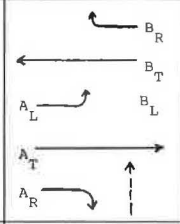
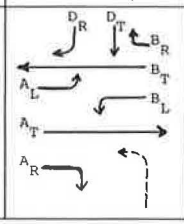
EVALUATION AND TESTING OF REVISED PROCEDURES

Model Revisions

The comparison of model structure as well as the empirical results indicate that substantial revisions are needed in both the Circular 212 and Swedish procedures. Changes were made in critical gaps as well as dominant traffic streams in an attempt to make the procedures more reflective of U.S. drivers and driving conditions.

Because Circular 212 critical gaps appeared to be generally too high and Swedish gaps too low, the revised procedure uses critical gaps recommended by the TRB subcommittee. To allow consistent comparison with both the Circular 212 and Swedish methods, the same set of critical gaps was adopted for both procedures.

Several of the definitions of dominant flows used in the existing procedures were modified (see Figure 3). Contrary to the Swedish definition (9,p.6), right turns from the major road do affect the minor road movements (by not signaling their intention to turn, or doing so when it is already too late for

MOVEMENT		CIRCULAR 212	SWEDISH MANUAL	REVISED
RIGHT TURN INTO MAJOR STREET		$\frac{1}{2}A_R + A_T$	$\frac{A_T}{N_b}$	$\frac{1}{2}A_R + A_T + \frac{1}{2}A_L$
LEFT TURN FROM MAJOR STREET		$A_R + A_T$	$\frac{A_T + A_R}{N_c}$	$A_R + A_T + \frac{1}{2}A_L$
CROSSING MAJOR STREET		$\frac{1}{2}A_R + A_T + A_L + B_L + B_T + B_R$	$\frac{A_T + A_L}{N_d} + B_L + B_T + \frac{B_R}{N_d}$	$\frac{1}{2}A_R + A_T + A_L + B_L + B_T + B_R$
LEFT TURNS INTO MAJOR STREET		$\frac{1}{2}A_R + A_T + A_L + B_L + B_T + B_R + D_T + D_R$	$A_T + A_L + \frac{B_L + B_T}{N_a} + D_T + \frac{D_R}{N_a}$	$\frac{1}{2}A_R + A_T + A_L + B_L + B_T + \frac{1}{2}B_R + D_T + D_R$

N = NUMBER OF LANES IN THE EXIT DIRECTLY OPPOSITE APPROACH IN QUESTION.

FIGURE 3 Original and revised definitions of dominant traffic flows.

the gap to be used by the nonpriority stream). Definitions of dominant streams in the two procedures were also changed where any of the exits from an intersection had more than one lane.

The suggested definitions of dominant flows included one-half of the right-turn traffic from the major road (A_R) as dominant to the right turns into the major road (like Circular 212). For the same reason that right-turning drivers from the major road may not signal their intention to turn and hence influence nonpriority (or subordinate) movements, left-turning major road vehicles (A_L) likewise sometimes do not show their intention to turn left; hence one-half of such traffic flows is included in the definitions suggested. In addition, and for the same reason, only one-half of the right-turning traffic approaching from the left of left-turning minor road vehicles (B_R) is dominant to this subordinate movement.

Evaluation of Revised Method

The new procedures were used to calculate capacities at the four intersections (Table 4). The modifications increased the capacities estimated using Circular 212 by 100 to 150 vehicles per hour. Although the calculated capacities for the Swedish method were generally reduced by 125 to 165 vehicles per hour, the values still generally exceeded the Circular 212 estimates.

The reasons for the much higher capacity estimate using the Swedish procedure were examined in detail for the Niles Center-Howard intersection. Table 7 links major components of the Swedish method calculations by summarizing service times and stopped delay for different sets of critical gaps and calculation procedures.

Looking at the results for the near-side approach first, the service time entries illustrate the importance of critical gap. When service time was estimated using the revised procedure the time for through and left-turn movements increased by 2.5 and 3.5 sec, respectively. Even with the revised procedure, however, the through and left-turn service times were 1.7 and 3.6 sec less than the respective times measured directly from the films. These results appear to reflect fundamentally different behavior of U.S. and Swedish drivers in the data sets. In Figure 2, service time is made up of both gap acceptance time (S - M) and move-up time (M). Both values would appear to be different for U.S. and Swedish drivers, particularly for left turners.

The increase in service time estimated with the revised procedure is reflected in the substantially higher delay estimated for the near-side approach; the stopped delay increases from 7.2 to 12.1 sec when field-measured values are used. Interestingly, the 12.1 sec average delay is an error of only 1 sec compared with field-measured delay.

Results for the far side were, unfortunately, not completely comparable because vehicles crept into the intersection from the minor road as a result of a high percentage of major road right turns (1). As a result, critical gaps and service times for the approach are atypical. Notice, however, that field-measured service times are again somewhat higher than the estimated times, particularly for through movements. The implications of the service time differences were again reflected in the delay values: The adjusted Swedish delay is 10.5 sec less than field values. Field-measured service times, however, give an average vehicle delay that is only 3.5 sec less.

These detailed comparisons, although only for two intersection approaches, seem to indicate that the Swedish delay model may be valid if it is provided with accurate service times. Much more research on U.S. traffic is needed to revise the Swedish service time charts to better reflect U.S. drivers and conditions. The procedure does seem to have significant promise for delay prediction if updated accurately.

Summary

Table 8 gives a summary of four sets of level of service calculations: Columns 2 and 3 give values from the revised Circular 212; column 4 is the delay from the revised Swedish procedure; and the last column is field-measured total delay that is developed by adding 10 sec of running delay (1) to field-measured stop delay.

A comparison of Circular 212 LOS and the measured delays shows good correspondence at the extremes, A and E. The results for LOS B-D are less clear and differentiable: A measured delay of 20 sec is assigned LOS C and B, and 21 sec gives LOS D. The Swedish delay values correspond very well with the measured delay except at LOS E and appear to track fairly consistently with the LOS estimates from Circular 212.

The poor results for LOS E are not totally surprising given the comments by Swedish researchers that traffic densities in their data set were much lower than U.S. conditions. It is characteristic of queueing theory models that delays increase most dramatically and variably at high V/C ratios. It is likely that the far-side Niles Center-Howard approach, which is very heavily congested, had traffic conditions that were beyond the range of most of the Swedish data.

CONCLUSIONS AND RECOMMENDATIONS

A study has been completed of the conceptual structure and field validity of two procedures to estimate unsignalized intersection capacity and level of service. Among the major findings of the analysis were

1. The procedure in Transportation Research Circular 212 (7) contains critical gaps that are too large for U.S. conditions. The critical gaps proposed by the Transportation Research Board Subcom-

TABLE 7 Comparison of Swedish Manual Service Time Model with Direct Field Measurements

		Service Time (sec)			Average Stopped Delay (sec/veh)			
		Critical Gaps from:						
Minor Approach	Movement	SM	Revised SM	Field Measured	SM	Revised SM	Using Field Measured Service Times	Field Measured
	R	4	5	4.1				
Far	T	7.8	7.8	10.1	*	27.7	34.6	38.2
	L	9.7	9.3	9.1				
	R	4	4.5	4.7				
Near	T	8.5	11.0	12.7	7.2	7.7	12.1	11.1
	L	7	10.5	14.1				

*v/c ratio greater than 0.8; procedure not applicable

TABLE 8 Comparison of Estimated and Field-Measured Level of Service Measures for Revised Procedures

Intersection		Circular 212		Average Total Delay (sec/veh)	
		Reserve Capacity	LOS	Swedish	Measured
Kostner-Kirk	N	695	A	18.0	15.0
	F	622	A	17.3	16.5
Howard-Kostner	N	256	C	24.2	20.9
	F	297	C	23.1	26.1
McCormick-Bridge-Grey	N	181	D	25.0	21.7
	F	320	B	19.7	20.0
Niles Center-Howard	N	311	B	19.3	21.1
	F	62	E	36.9	48.2

mittee on Unsignalized Intersections appeared to be more rational and better reflect the data in this study.

2. When critical gaps were revised and some modification was made to the definitions of dominant volumes, the Circular 212 procedure gave satisfactory estimates for LOS A and E but could not clearly differentiate levels B-D.

3. The procedure in the Swedish Capacity Manual (4) consistently overestimated capacities and underestimated delays. The critical gaps proposed in the method were generally too small for U.S. drivers.

4. When critical gaps and dominant flows were redefined, the Swedish procedure still gave higher capacities than Circular 212. Delays from the revised procedure corresponded fairly well with field-measured values.

5. Detailed comparison of field data with values in the Swedish manual indicated that fundamentally different driver behavior in the United States and in Sweden contributes to some of the method's inaccuracy. When the Swedish delay model was used with capacity (service time) data obtained directly from the field, very accurate estimates of delay were obtained. The results imply that the delay model is conceptually sound but needs to be used with data that more accurately reflect U.S. drivers and conditions.

It is recommended that further field testing be conducted using the Circular 212 procedure. Although the number of intersections in this study was limited, it is clear that critical gaps should be changed and that dominant volumes may need revision. Further field tests are needed to generalize these research results.

It is strongly recommended that a research study be conducted to determine the applicability of the Swedish method to U.S. conditions. A major potential advantage of the Swedish method is its use at four-way stop signs and at two-way stops with platooned major street volumes. The Circular 212 procedure is valid only for two-way stop or yield control with random major street arrivals, a condition that does not exist in many urban areas where controls are located near signalized crossings.

The other major advantage of the Swedish method is the output of estimated delay per vehicle. If the method can be adapted to U.S. conditions, a procedure would be available to estimate minor street delays for traffic signal warrants. Obviously a major field data collection and analysis effort will be required to revise the Swedish manual. In the long term, this is probably a worthwhile investment. The Circular 212 procedure, with revised critical gaps and possibly revised dominant flows, would seem to perform adequately in the meantime.

REFERENCES

1. A.B. Bakare. Field Tests of Procedures for Calculating Capacity at Unsignalized Intersections. M.S. thesis. Northwestern University, Evanston, Ill., 1983.
2. Highway Capacity Manual 1965. Special Report 87. HRB, National Research Council, Washington, D.C., 1965, 397 pp.
3. Merblatt for Lichtsignalanlogen an Landstrassen, Ausgabe 1972. Forschungshesellschaft fur das Strassenwessen, Koln, Federal Republic of Germany, 1972.
4. Berakning av kapacitet, kolangd, fordronjnnig i vagtarfikanlaggningar. TV 131. Statens Vagverk, Stockholm, Sweden, (unpublished English translation, 1982).
5. Road Design Manual for Rural and Urban Roads other than Freeways. County Roads Board, Victoria, Australia, Jan. 1974.
6. Capacity of At-Grade Junctions. Organization for Economic Co-operation and Development, Paris, France, 1974.
7. Unsignalized Intersections. In Interim Materials on Highway Capacity, Transportation Research Circular 212. TRB, National Research Council, Washington, D.C., Jan. 1980, pp. 37-72.
8. R.M. Kimber and R.D. Coombe. The Traffic Capacity of Major/Minor Priority Junctions. Supplemental Report 582. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1980.
9. B.E. Peterson and A. Hannson. Swedish Capacity Manual. In Transportation Research Record 667, TRB, National Research Council, Washington, D.C., 1978, pp. 1-10, 21-28.
10. C. Swerdloff. A Study of Gap Acceptance at a Stop Sign Location. M.S. thesis. Northwestern University, Evanston, Ill., 1962.
11. C.Y. Gagnon. Effect of Size of Gaps in Line on Acceptance of the Lag at a Stop Sign. M.S. thesis. Northwestern University, Evanston, Ill., 1962.
12. De Leuw, Cather and Company. Effect of Control Devices on Traffic Operations. NCHRP Report 41. HRB, National Research Council, Washington, D.C., 1967, 84 pp.
13. G.R. Bullen. Field Test of an Intersection Simulation Model. M.S. thesis. Northwestern University, Evanston, Ill., 1964.
14. H.H. Bissell. Traffic Gap Acceptance from a Stop Sign. Graduate research report. University of California Institute of Transportation and Traffic Engineering, Berkeley, 1960.
15. D.J. Finney. Probit Analysis, A Statistical Treatment of the Sigmoid Response Curve. Cambridge University Press, Cambridge, England, 1974.
16. R. Ashworth. A Note on the Selection of Gap Acceptance Criteria for Traffic Simulation Studies. In Transportation Research 2, Pergamon Press, London, England, 1968, pp. 171-175.
17. R. Sumner, D. Hill, and S. Shapiro. Passenger Car Equivalence on Urban Arterial Roads. PRC Voorhees, McLean, Va., Nov. 1981.
18. Institute of Traffic Engineers. Transportation and Traffic Engineering Handbook. Prentice Hall, Englewood Cliffs, N.J., 1976.
19. A Policy on Geometric Design of Rural Highways. AASHTO, Washington, D.C., 1965.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

Use of the NCHRP Signalized Intersection Capacity Method--A South African Experience

ADOLF D. MAY, WESSEL J. PIENAAR, and CECIL A. ROSE

ABSTRACT

The primary objective of this study was to assess the applicability of the proposed NCHRP operations method for signalized intersection capacity analysis to South African traffic conditions. Secondary objectives were first, to review the reactions of 90 engineers who received a 5-day intensive course on the use of the NCHRP method and second, to draw a comparison between the predictions of this new method, those that could be obtained using the Highway Capacity Manual (HCM) method, and the relevant field-measured data. The analysis was based on applying the proposed NCHRP Operations Method to ten South African intersection data sets for which not only input data but also field measurements of lane saturation flows, stopped delays, queue lengths, and percent vehicles stopped were available. The ten intersections varied from simple geometrically designed intersections with pretimed two-phase signals to more complicated intersections with actuated multiphase signals. The NCHRP method is validated by evaluating its predictive accuracy under conditions experienced at the ten South African intersections. The validation is supplemented by the following analyses: a comparison of lane saturation flows using the proposed NCHRP and HCM methods, factors influencing saturation flows, the possibility of a default value for saturation flow, an assessment of stopped delay predictions using measured saturation flow as input, and an assessment of permissive left-turn stopped delays. The conclusions include the identification of the strengths and weaknesses of the NCHRP method under South African traffic conditions and the identification and quantification of alternative procedures and default values that would enhance the method's application in that country. In general, the proposed NCHRP method providse accurate results, and its ability to estimate saturation flow is better than that of the HCM method. One of the findings was that the use of a base saturation flow of 1,900 vehicles per hour of green per lane, combined with heavy vehicle adjustment factors having twice the magnitude of the NCHRP factors, would provide saturation flow estimates with roughly 60 percent smaller mean errors under South African traffic conditions.

As part of NCHRP Project 3-28 to update the 1965 Highway Capacity Manual (HCM) (1), work has been carried out by JHK & Associates, in cooperation with the Traffic Institute, Northwestern University, leading to a proposed new method for signalized intersection capacity analysis. This work is still in progress, and a number of draft documents describing the proposed method have been made available. The three drafts of interest in this study are those dated December 1981 (2), May 1982 (3), and February 1983 (4).

In July 1982 the principal author visited South Africa at the invitation of the Southern Africa Road Federation to conduct two courses on the application of the NCHRP method. At each course location a coinstructor assisted with the presentation, and a coauthor participated in each course. At the first course, offered in Pretoria, John Sampson acted as coinstructor and coauthor Wessel Pienaar participated as a delegate. At the second course, held in Stellenbosch near Cape Town, coinstruction was offered by John Jones, and coauthor Cecil Rose participated as a delegate. As part of the courses, data were collected at five intersections at each location to allow course members to apply the method under South African conditions. These data were also used to relate certain measures of effectiveness (MOEs) predicted by the NCHRP operations method to those measured at the intersections. It is with this comparison of estimated and field-measured MOEs that this study is primarily concerned.

SCOPE OF STUDY

The primary objective of this research was to assess the applicability of the draft NCHRP signalized intersection capacity operations method to South African traffic conditions. Secondary objectives were first, to review the reactions of approximately 90 course delegates to the use of the method and second, to draw a comparison between the predictions of this new method, those that could be obtained using Chapter 6 of the 1965 HCM, and the relevant field-measured data.

The primary objective was addressed by a detailed analysis of ten intersections, supplemented by a comparison of the predicted values of four MOEs with the field-measured values where these were available. The four MOEs examined were saturation flow, average stopped delay, percent vehicles stopped, and maximum queue length. The drafts of the method used in this analysis were those dated May 1982 (3) and February 1983 (4), although the December 1981 (2) draft was used as the basis of instruction and workshop sessions during the two courses. At the completion of each course, delegates were asked to comment on the deficiencies and difficulties of the method. It was hoped that this would point to areas that require simplification, those that need further clarification in the descriptions, and those on which course instructors should place greater emphasis.

A description of courses and a summary of delegates' comments are supplied in the next section. The third section is a presentation of characteristics of the ten analyzed intersections, and a brief overview of South African urban travel conditions is supplied. In the fourth section, a validation of the method under South African conditions, supplemented

by additional saturation flow and stopped delay analyses, is presented. The fifth section contains concluding remarks and a summary of the most significant results.

THE COURSES

The topic covered by the courses was the analysis of signalized intersection capacity. The subject matter was based on the proposed new NCHRP method as contained in the December 1981 draft of the documents prepared by JHK & Associates (2).

The first half of each course was intended to instruct students in the background, supporting theory, and use of the method. Both the planning and the operations applications were covered. The examples supplied in the text of the document were used in the instruction sessions. To give delegates further hands-on experience in the use of the method, five sample intersections located in the San Francisco Bay region were analyzed by groups of students using data obtained in advance by the principal author.

The latter half of each course was a workshop designed to bring delegates into contact with a particular local intersection and to have them undertake field measurements of the four major MOEs and analyze the performance of the intersection using the method. The delegates worked in groups analyzing the five local intersections at each location. Details of the ten intersections are given in the third section.

Each intersection was observed by the designated groups for a period, which included the evening peak period, of 1.5 to 2 hours. Geometric data, traffic flow data, traffic conditions, and control details--in effect steps 1 to 4 of the operations application--were recorded before the visit to the sites. Groups were required to confirm the correctness of the previously collected data and then take in situ measurements of saturation flows, stopped delays, percent vehicles stopped, and maximum queue length in each lane group. Lane groups were selected by the delegates before the site visits. The site observations, therefore, served a dual purpose: first, to give the delegate groups practical experience in the operating conditions at an intersection and second, to collect the MOE data required for comparative purposes later in the analyses.

Delegate Comments on Procedural Deficiencies

On the basis of the intensive instruction sessions and the experience gained from the use of the method on at least two example intersections and one locally analyzed intersection, delegates were asked to note the difficulties and deficiencies they perceived in the use of the method under South African operating conditions. Only those statements made by two or more persons have been included. The comments made by the course delegates on the procedures included in the operations application, in order of decreasing frequency of occurrence, were

1. The procedure for determining left-turn (S.A. right) delay was inadequate;
2. The application was long, tedious, and cumbersome considering the perceived accuracy;
3. Saturation flows were underestimated because of adjustment factors relating to area type;
4. Greater clarity is needed in lane group selection guidelines;
5. Stopped delay calculated by the Appendix C procedure was unreasonably high;
6. Results are too heavily dependent on user's judgment;
7. The treatment of heavy left turns, especially in double turn lanes, is inadequate;
8. Improvement is needed in the handling of left turns from single lane approaches;
9. Confident use of the method requires considerable learning and practice time;
10. It is difficult to determine critical lane groups in the event of overlapping phases; and
11. The two applications use different level of service concepts.

In the ongoing development and refinement of the proposed method, several issues raised in this section have in fact been receiving further attention. Subsequent drafts of the document (3,4) have incorporated changes in response to similar comments from other sources.

Comments 1-6 are dealt with to an extent in the fourth section of this paper where comparisons are drawn between measured saturation flows and stopped delays and those predicted by using different procedures included in the operation application.

DESCRIPTION OF STUDY AREA

The location and characteristics of the ten intersections to which the new NCHRP operations method, taught at this course, was applied are described. In addition, a brief overview is supplied of South African urban travel conditions, which could have influenced the accuracy of the method and possibly underlay certain comments made by delegates.

Location and Characteristics of Intersections

The ten intersections analyzed were selected before the courses began. A variety of design and control features as well as relative proximity to the course locations were considered. The five intersections analyzed at the Pretoria course are located within the municipal area of that city. Two of the intersections have four approaches, one is formed by the intersection of two one-way streets, and the remaining two intersections are in the form of skewed T-junctions. The five intersections analyzed at the Stellenbosch course are situated within the Cape Town metropolitan area. All the Cape Town intersections have four approaches. The locations of the ten intersections are given in Table 1. A summary containing ranges of the most pertinent intersection input data is given in Table 2.

South African Travel and Vehicle Characteristics

Driving conditions in South Africa differ from those experienced in the United States, where the method

TABLE 1 Location of the Ten Analyzed Intersections

Center	No.	Street Location
Pretoria	1	Duncan Street and Lynnwood Road
Pretoria	2	Burnett Street and University Road
Pretoria	3	Pretorius Street and Schoeman Street
Pretoria	4	Curson Street and Walton Jameson Avenue
Pretoria	5	Curson Street and Kirkness Street
Cape Town	1	Darling Street and Buitenkant Street
Cape Town	2	Hertzog Boulevard and Pirow Street
Cape Town	3	Main Road and Campground Road
Cape Town	4	Weltevreden Road and Duinefontein Road
Cape Town	5	Modderdam Road and 35th Street

TABLE 2 Range of Values of Input Variables

Input Variable	Range		Input Variable	Range	
	Pretoria	Cape Town		Pretoria	Cape Town
Approach Widths (meters)	5.0-11.0	6.1-14.8	Approach Volume (vph)	97-1776	336-2196
Lane Widths (meters)	3.0-5.0	2.6-4.0	Left Turning Volume	0-216	25-412
Lanes per Approach	1-3	2-4	Right Turning Volume	0-492	0-683
Grade of Approach	0	0 to 4	Percentage Heavy Vehicles	0-7	1-17
Peak Hour Factor	0.85	0.71-0.98	No. of Buses	0-4	0-7
Bus Stops	1Y - 14N	1Y - 19N	Pedestrian Volumes (P/hr)	<10 - 65	<10 - 519
Parking	None	1Y - 19N	Arrival Type	3 and 5	1 - 5
No. of Parking Maneuvers	None	None	Phases/Cycle	2 and 3	2,3 and 4
Area Types	5 other	2BD 3 other	Cycle Lengths (secs.)	55-60	72-143

is being developed and validated. First, the typical South African car has manual transmission and the exterior dimensions of a compact car. These features might lead to greater maneuverability at intersections with tight geometric design. In addition, observations indicate that South Africans drive at faster cruising speeds, accelerate faster after stops, allow for shorter headways, and are less inclined to yield the right-of-way to pedestrians than are their U.S. counterparts. In combination, these factors have the potential to increase vehicle throughput (volume) at South African intersections.

Second, signal timing and law enforcement practices differ in the two countries. South African practice favors short amber periods and common use of all red periods. Further, South African road ordinances require that an intersection be clear at the commencement of red. Turning on red is not permitted. Collectively, these control practices not only reduce effective green time but also inhibit the capacity of turning lanes.

Third, there are differences in design principles that can affect intersection capacity:

1. South African practice favors midblock bus stops instead of the commonly found near-side or far-side bus stops at intersections in the United States.
2. With the exception of certain central business districts, on-street parking in South Africa is clearly divorced from intersections.

Both of these design principles should lead to greater intersection capacity, and hence increased saturation flows, in South Africa.

Although these differences in travel conditions are not exhaustive, they should set the scene for the validation attempt described in the next section.

VALIDATION OF THE NCHRP OPERATIONS APPLICATION UNDER SOUTH AFRICAN CONDITIONS

The groups of course delegates were assigned to measure the values of the four MOEs at the previously mentioned intersections. After taking the measurements, each group supplemented the values it obtained with estimates made using the method. These group estimates were subsequently revised by the two teaching assistants. Their revisions were again reviewed and verified by the authors. The validation discussed here makes use of these values.

Two of the MOEs, average queue length and percent vehicles stopped, are not analyzed because they are derivatives of average stopped delay and do not require any other independent variable as input.

It should be noted that, although South African rules of the road require driving on the left-hand side, all references to left and right are reversed to coincide with U.S. terminology. All units of length are metric. In the following analysis, it is assumed that the field-measured values are correct and that any differences between predicted and measured values are ascribable to errors in the methods.

Validation of Saturation Flow Estimates

Field-measured saturation flow and directly comparable saturation flow predicted by the NCHRP method and the HCM method are available for 35 of the 74 originally analyzed lane groups. Because of low flows, saturation flow measurements could not be made at the remaining lane groups. A summary of these saturation flows is given in Table 3. The intersection number, direction of movement, permitted traffic movement, and number of lanes per lane group are identified. The columns represent the previously mentioned saturation flows. The error of estimate per lane group is also supplied for both the NCHRP and the HCM methods. Statistical analyses were performed and their results are given in Table 4. The table gives data on four classes of lane groups: single lane exclusive left turn (S.A. right), single lane through and shared through and turn, two-lane through and shared through and turn, and three-lane through and shared through and turn.

The table also gives the total sample of lane groups collectively. The following statistics are calculated: overall mean saturation flows, standard deviation, mean error, and linear regression analyses between measured and estimated values. To assess the degree of similarity between calculated and measured flows on a qualitative basis, it was decided to regard all estimates closer than 10 percent to the measured values as "good." Those predictions further than 10 percent but up to 20 percent are regarded as "fair," and those further than 20 percent astray from the measured values are regarded as "poor." A summary of this qualitative assessment of the NCHRP method's ability to predict saturation flow per lane group is given in Table 5. This table shows that approximately 50 percent of the method's lane group saturation flow estimates are good predictions, and 75 percent are fair estimates or better. The method very slightly underestimates saturation flow.

Comparative Analyses of Lane Saturation Flow Estimates Using the NCHRP and HCM Methods

Saturation flow estimates obtained using the NCHRP method and those obtained using the 1965 HCM method are compared, and these saturation flow estimates are related to the field-measured saturation flows. The reason for attempting this analysis is twofold. First, because the NCHRP method is meant to replace the HCM method, it would be useful to know whether it is superior to the method it is to replace. Second, although the course delegate comment that the method underestimates saturation flow cannot be

TABLE 3 Lane Group Saturation Flow Results (vphgl)

Intersec. Number	Direction Movement	Movement	No. Lanes Per Group	Measured Saturation Flow	Predicted Saturation Flow			
					NCHRP Method		1965 HCM Method	
					Calculated	Error	Calculated	Error
P1*	NB		2	2449	3335	+886	2600	+151
P1	SB		2	2975	2863	-112	2270	-705
P1	EB		2	3302	3801	-001	2560	-742
P3	NB		1	1860	1782	-078	1583	-277
P3	NB		1	1880	1782	-098	1583	-297
P3	NB		1	1640	1782	+142	1583	-057
P3	WB		1	1860	1755	-105	1535	-325
P3	WB		1	1800	1755	-045	1535	-265
P3	WB		1	1800	1755	-045	1535	-265
P4	SB		1	1072	1120	+148	835	-237
P4	WB		1	1455	1676	+221	1300	-155
P4	WB		1	906	1592	+686	1300	+394
P5	SB		1	1800	1728	-072	2154	+354
P5	WB		2	3600	3492	-108	3109	-491
C2**	NB		3	6056	5324	-732	4585	-1471
C2	NB		1	103	88	-015	185	+082
C2	SB		1	1154	1024	-130	808	-346
C2	SB		3	5562	4879	-683	4023	-1539
C2	EB		2	3858	3452	-406	3156	-702
C2	WB		1	270	88	-182	248	-022
C2	WB		3	4018	4331	+313	4681	+663
C3	NB		1	1420	1463	+043	1430	+010
C3	SB		1	1700	1516	-184	1859	+159
C3	EB		1	1976	1475	-501	1385	-591
C3	WB		1	1993	1494	-499	1681	-312
C4	SB		2	3066	3326	+260	3050	-016
C4	EB		2	2696	3143	+447	2644	-052
C4	EB		1	338	558	+220	592	+254
C4	WB		2	2620	3259	+639	2584	-036
C5	NB		2	2660	2908	+248	2441	-219
C5	SB		1	1856	1226	-630	1197	-659
C5	EB		2	3103	3420	+317	3111	+008
C5	EB		1	1765	1454	-311	1140	-625
C5	WB		1	1869	1287	-582	1287	-582
C5	WB		2	3600	3449	-151	3357	-243
Total				80082	78982	-1100	70926	-9156

*P denotes Pretoria **C denotes Cape Town

TABLE 4 Statistical Analysis of Lane Group Saturation Flow Results

Sample and Calculation	Measured	Estimate NCHRP	Estimate 1965 HCM
Single lane exclusive left turn (S.A. right)			
Number of groups	7	7	7
Mean saturation flow	1051	818	780
Standard deviation	802	573	454
Mean error	N/A	296	367
Single lane through and shared			
Number of groups	14	14	14
Mean saturation flow	1654	1627	1521
Standard deviation	330	171	297
Mean error	N/A	205	264
Lane groups with two lanes			
Number of groups	11	11	11
Mean saturation flow	3084	3268	2807
Standard deviation	463	214	356
Mean error	N/A	325	306
Lane groups with three lanes			
Number of groups	3	3	3
Mean saturation flow	5212	4845	4430
Standard deviation	1063	497	355
Mean error	N/A	576	1224
All lane groups			
Number of groups	35	35	35
Mean saturation flow	2288	2257	2027
Standard deviation	1313	1268	1117
Mean error	N/A	293	380
Linear regression			
Y-intercept	N/A	+53	+43
Slope	N/A	+0.991	+1.108
Corr. coef. (r)	N/A	0.96	0.94

Note: N/A = not applicable.

TABLE 5 Qualitative Assessment of the Method's Ability to Predict Saturation Flow on a Lane Group Basis

	Good		Fair		Poor		Total	
	No.	%	No.	%	No.	%	No.	%
Overestimate	6	17.2	3	8.6	4	11.4	13	37.1
Underestimate	11	31.4	6	17.1	5	14.3	22	62.9
All	17	48.6	9	25.7	9	25.7	35	100.0
Pretoria	10	28.6	2	5.7	2	5.7	14	40.0
Cape Town	7	20.0	7	20.0	7	20.0	21	60.0

supported statistically in this study, it would be interesting to establish the performance of the NCHRP method relative to the HCM method. Lane group saturation flow results of both methods are summarized in Table 3, and a statistical analysis thereof is given in Table 4. A comparative analysis of saturation flow results for through and shared through and turn groups, expressed on a per lane basis, is given in Table 6. Despite the bias in flow volumes between the cities, Tables 4 and 6 show that the NCHRP method not only estimates saturation flows more accurately per city and per type of lane group, but in each case it also predicts values higher than does the HCM method.

TABLE 6 Comparative Analysis of Through and Shared Lane Group Saturation Flow Results

Calculation	Saturation Flow (vphg)								
	Measured			NCHRP			HCM		
	Pret.	Cape Town	Total	Pret.	Cape Town	Total	Pret.	Cape Town	Total
No. of Groups*	14	14	28	14	14	28	14	14	28
No. of Lanes	18	27	45	18	27	45	18	27	45
Avg. Sat. Flow	1578	1650	1621	1657	1592	1617	1416	1497	1465
Std. Dev.	297	275	284	152	117	131	276	176	216
Mean Error	N/A**	N/A	N/A	128	201	172	262	223	239
Over-ests.	N/A	N/A	N/A	7	14	21	4	7	11
Under-ests.	N/A	N/A	N/A	11	13	24	14	20	34
Range of Errors per Sgl. Ln.	N/A	N/A	N/A	1 - 686	43 - 501	1 - 686	57 - 371	8 - 659	8 - 659

* Single lane exclusive turning movement groups are summarized in the top row of Table 4.

** N/A denotes "not applicable."

Factors Influencing Saturation Flows

In the determination of saturation flow the NCHRP method considers the effect of nine variables. In this analysis it was necessary to eliminate five of these variables because of an insufficient range of pertinent data. Also, because essentially only through and shared through and turn lane groups were being analyzed, the left-turn factor (S.A. right) was also omitted. The only remaining factors were lane width, heavy vehicle, and right-turn factors (incorporating pedestrian volumes and percentage of right turns).

Table 7 gives the ratio of measured to calculated saturated flows (M/C) for both areas and the values of the four remaining variables analyzed. The table is arranged in order of decreasing M/C ratio of each area. The data for the Pretoria group show less variation in the ratio, when the extremely low value of 0.57 is disregarded, than is shown by the Cape Town data. There is no discernible correlation between the value of the ratio and any of the values of the four input variables. Note also that in the case of percent heavy vehicles these data range only between 0 and 7 percent, with eleven of the values between 2 and 5 percent. This compares with a range of 1 to 17 percent heavy vehicles in Cape Town. The Pretoria data also lack sufficient range for the right-turn percentage and pedestrians, leaving little opportunity to draw conclusions.

Examination of the Cape Town data set of 14 lane groups reveals no apparent correlation between the M/C saturation flow ratios and lane widths, percentage right turns, or pedestrians. This is despite the fact that the values of the variables show greater variation than do those for the Pretoria intersections. It can, however, be observed that there is a correlation between the variation of the M/C ratio and the percentage of heavy vehicles in the traffic stream. It was found that as the M/C ratio decreases, the percentage of heavy vehicles increases.

As a further check, the adjustment factors for the other eight variables influencing saturation flow were aggregated to see whether in combination they displayed any trend that could have an effect on the M/C ratio. It was found, however, that the values of combined adjustment factors, minus the heavy vehicle factors, are uniformly distributed, yielding an average of 0.926 with a standard deviation of only 0.073. A regression analysis of the value of the M/C ratio versus the percentage of heavy vehicles (%HV) yields the following equation:

$$M/C = 1.20 - 0.022(\%HV) \quad : \quad r = -0.67 \qquad (1)$$

TABLE 7 Factors Affecting Saturation Flow

I/S #	Appr.	# Lanes	Sat. Flow / Lane			Lane Width (M)	% HV	% RT	Peds.
			Meas.	Calc.	M/C				
P3*	WB	1	1860	1755	1.06	3.7	5	0	0
P3	NB	1	1880	1782	1.05	3.7	2	0	0
P1	SB	2	1487	1431	1.04	3.0	3	55	50
P3	NB	1	1860	1782	1.04	3.7	2	0	0
P5	SB	1	1800	1728	1.04	3.0	7	0	0
P5	WB	2	1800	1746	1.03	3.7	7	0	0
P3	WB	1	1800	1755	1.03	3.7	5	0	0
P3	WB	1	1800	1755	1.03	3.7	5	0	0
P1	EB	2	1651	1650	1.00	3.0	3	7	65
P3	NB	1	1640	1782	0.92	3.7	2	0	0
P4	SB	1	1072	1220	0.88	3.0	7	45	25
P4	WB	1	1455	1676	0.87	3.0	4	0	10
P1	NB	2	1225	1667	0.73	3.0	0	23	25
P4	WB	1	906	1592	0.57	3.0	4	19	10
C3**	EB	1	1976	1475	1.34	3.3	2.4	21	20
C3	WB	1	1993	1494	1.33	3.0	5.6	0	18
C2	NB	3	2019	1775	1.14	3.7	7.3	0	490
C2	SB	3	1854	1626	1.14	3.7	2.4	35	10
C2	EB	2	1929	1726	1.12	3.7	1.6	17	183
C3	SB	1	1700	1516	1.12	3.0	3.9	0	156
C5	WB	2	1800	1725	1.04	3.6	7.5	2	10
C3	NB	1	1420	1463	0.97	3.3	8.5	10	96
C2	WB	3	1339	1444	0.93	3.7	1.1	28	516
C4	SB	2	1533	1663	0.92	3.7	7.7	16	10
C5	NB	2	1330	1454	0.91	3.6	13.4	48	10
C5	EB	2	1551	1710	0.91	3.6	9.8	0	15
C4	EB	2	1348	1571	0.86	3.7	17.1	21	36
C4	WB	2	1310	1630	0.80	3.7	14.9	16	10

* P denotes Pretoria. ** C denotes Cape Town.

From Equation 1 it can be deduced that when the M/C ratio equals 1.2, the percentage heavy vehicles is zero. Likewise, when M/C = 1.0, the percentage heavy vehicles is 9.1. Among the 14 lane groups investigated, three sizes of groups can be identified: four groups consist of one lane each; seven groups contain two lanes; and three groups have three lanes each. For each lane group size a regression analysis of the field-measured versus calculated saturation flows was performed. Using these regression analyses, the expected flows were calculated for each lane group size that would yield M/C ratios of 1.0 and 1.2, respectively. It was found that for groups with one, two, and three lanes, an M/C of 1.0 can be expected when the actual flows equal 1,368, 1,750, and 1,502 vehicles per hour of green per lane (vphgl). When these flows are multiplied by 1.2 to represent the saturation flows where the expected M/C ratio would equal 1.2, a weighted average base saturation flow of 1,906 vphgl (≈1,900) is found. At 0 percent heavy vehicles, the average combined adjustment factor of 0.926 will reduce the re-estimated base saturation flow of 1,900 vphgl to an adjusted saturation flow of 1,760 vphgl.

With a weighted average actual saturation flow of 1,588 vphgl containing 9.1 percent heavy vehicles, the adjusted flow of 1,760 vehicles will have to be scaled down by a factor of 0.902 to represent the actual effect these heavy vehicles have on saturation flow. This compares with an adjustment factor of 0.955 used in the NCHRP method for 9.1 percent heavy vehicles. This leads to the conclusion that heavy vehicles in the study area tend to affect the flow of cars twice as adversely as do heavy vehicles in the United States. Re-estimated heavy vehicle factors for South African conditions are given in Table 8.

TABLE 8 Re-estimated Heavy Vehicle Saturation Flow Adjustment Factors for South Africa

%HV	0	2	4	6	8	10	20
$f_{HV(US)}$	1.00	0.99	0.98	0.97	0.96	0.95	0.91
$f_{HV(SA)}$	1.00	0.98	0.96	0.94	0.92	0.90	0.81

TABLE 9 Analysis of Through Lane Saturation Flow

Calculated Statistic	Saturation Flow/Lane			
	Field Measured	Calculated (NCHRP Method)	Default Value of 1800	Calculated (Using base flow of 1900 vehicles and reestimated heavy vehicle factor.)
Median	1800	1755	1800	1792
Mean	1797	1721	1800	1758
Standard Deviation	172	86	0	94
Mean Error	N/A	156	127	63
Maximum	2019	1782	1800	1860
Minimum	1455	1494	1800	1526

Default Value for Adjusted Saturation Flow

In Appendix E of the May 1982 draft of the NCHRP operations method it is stated that the median saturation flow for through lanes under fair to good geometric and traffic conditions was found to be approximately 1,600 vphgl. For comparative purposes all 13 lane groups (representing 17 lanes) in the sample, which involved exclusive through movements and for which field-measured saturation flows are available, were analyzed to determine whether a reliable default value could be established for South African conditions. The results of this analysis are given in Table 9.

From the data in Table 9 it can be concluded that the use of a default value of 1,800 vphgl to represent adjusted saturated flow for through lanes with good geometric conditions, in the absence of more detailed knowledge of traffic conditions, is likely to provide accurate results. Note that the use of a base saturation flow of 1,900 vphgl, and the use of the re-estimated heavy vehicle factors, will provide saturation flow estimates with 60 percent smaller mean errors than do the values used by the NCHRP method.

Validation of Stopped Delay Results

To validate stopped delay, 45 sets of the originally analyzed 74 sets of lane group data are available for analysis. Intersections 3, 4, and 5 of Cape Town are omitted from the analysis because stopped delays were not measured at those intersections. Table 10 gives the field-measured and estimated stopped delay results for the lane groups analyzed. These estimated stopped delays were calculated using predicted saturation flows as input. An assessment of stopped delay predictions using measured saturation flows as input is presented hereafter.

To assess the degree of correlation between calculated and field-measured stopped delays on a qualitative basis, it was decided to regard all estimates closer than 5 sec to the measured values as "good." Those predictions that deviate between 6 and 10 sec are judged "fair," and all predictions further than 10 sec astray from the observed delays are judged "poor." A summary of this qualitative assessment of the NCHRP method's ability to predict stopped delay per lane group is given in Table 11. The data in this table indicate that 65 percent of all stopped delay estimates are good predictions, and approximately 82 percent are fair estimates or better. Although the sample contains 10 permissive-only left-turn (S.A. right) lane groups constituting 22 percent of the sample, they contribute 71 percent of all poor estimates. This issue was also raised as a matter of serious concern at the courses and is therefore dealt with separately later. The sample contains only two protected or permissive left-turn lane groups. One estimate was good and one was fair, and they do not warrant further investigation. The sample contains no protected-only left-turning movements.

A statistical analysis of through and shared through and turn lane group stopped delay results is given in Table 12. The following statistics are calculated: overall mean, standard deviation, and mean error of stopped delays.

Assessment of Stopped Delay Predictions Using Measured Saturation Flows as Input

Of the 45 lane groups used for stopped delay analysis there are 18 lane groups for which saturation flows were field measured. Comparative analysis of stopped delays for these 18 lane groups, using measured and estimated saturation flows as inputs, is

TABLE 10 Measured and Estimated Stopped Delays (sec)

Int. No.	Approach	Direction	Stopped Delay		
			Field Measured	Calculated Using Estimated Sat. Flows	Calculated Using Measured Sat. Flows
P1	NB		99	41	74
P1	NB		10	14	25
P1	SB		11	121	*
P1	SB		41	40	40
P1	EB		14	47	*
P1	EB		11	16	42
P1	WB		15	14	*
P1	WB		4	7	*
P2	NB		18	15	*
P2	NB		21	15	*
P2	EB		7	6	*
P2	EB		9	12	*
P2	WB		4	8	*
P2	WB		5	8	*
P3	NB		13	13	11
P3	NB		18	13	11
P3	NB		8	13	21
P3	WB		10	8	8
P3	WB		6	8	8
P3	WB		9	8	8
P4	NB		6	12	*
P4	SB		13	13	19
P4	EB		2	4	*
P4	EB		4	4	*
P4	EB		9	11	*
P4	WB		9	11	*
P4	WB		4	4	4
P4	WB		4	4	4
P5	SB		16	15	15
P5	WB		1.3	1.5	1.5
C1	NB		10	8	*
C1	NB		5	16	*
C1	SB		23	16	*
C1	SB		16	8	*
C1	EB		5	8	*
C1	EB		16	23	*
C1	WB		8	25	*
C1	WB		9	11	*
C2	NB		9	9	9
C2	NB		35	26	26
C2	SB		6	12	13
C2	EB		20	49	22
C2	EB		34	148	66
C2	WB		10	47	48
C2	WB		7	16	16

P denotes Pretoria. C denotes Cape Town.
* denotes that saturation flow was not measured.

given in Table 10. A statistical analysis detailing the effect of measured saturation flows on stopped delay estimates is given in Table 13. The table also shows the effect that the application of estimated saturation flows to a base saturation flow of 1,900 vphgl and the re-estimated heavy vehicle factors have on stopped delay estimates. The data in this table indicate that, for the 18 lane groups where it was possible to calculate stopped delays based on measured saturation flows, there was no significant change in the estimates. It therefore appears that there exists a need for further research to develop stopped delay correction factors for both through and shared through and turn lanes under South African conditions.

TABLE 11 Qualitative Assessment of the Method's Ability to Predict Saturation Flow on a Lane Group Basis

	Good		Fair		Poor		Total	
	Number	%	Number	%	Number	%	Number	%
Overestimates	13	29	4	8	7	16	24	53
Estimates deviating less than 0.5 second	7	16	--	--	--	--	7	16
Underestimates	9	20	4	9	1	2	14	31
Pretoria	25	56	2	4	3	7	30	67
Cape Town	4	9	6	13	5	11	15	33
ALL	29	65	8	17	8	18	45	100

TABLE 12 Statistical Analysis of Lane Group Stopped Delay Results: Through and Shared Through and Turning Lane Groups

Calculation	Pretoria		Cape Town		Total	
	Meas.	Calc.	Meas.	Calc.	Meas.	Calc.
Number of Lane Groups	22	N/A*	10	N/A	32	N/A
Mean Delay (seconds)	9.9	10.7	11.0	15.3	10.2	12.1
Standard Deviation	8.7	7.8	6.4	12.3	8.0	9.5
Mean Error	N/A	2.3	N/A	7.7	N/A	4.0

* N/A denotes "not applicable."

Assessing Exclusive Left-Turn (S.A. right) Stopped Delays

The foremost procedural deficiency identified at the courses was the technique for determining permissive exclusive left-turning (S.A. right) stopped delay. It was contended that the results obtained by scaling up through delay with a factor of 3 to represent permissive left-turn delay are suspect. The argument underlying this concern is probably that left-turn delay is a function of opposing flow rather than of adjacent through delay.

The May 1982 (3) and February 1983 (4) drafts of the NCHRP method do, however, in their Appendix G, suggest an alternative technique for determining such delays. This alternative technique was used and its results were compared with those obtained using the original December 1981 (2) procedure and with measured delays. Table 14 details the results obtained for exclusive or permissive left-turning movements.

The data given in Table 14 clearly indicate that the alternative Appendix G technique for estimating left-turning stopped delay supplies substantially better predictions for lanes with a volume-to-capacity (V/C) ratio of less than 1.0 than does the December 1981 procedure. However, the last estimate given in the table indicates that refinement of the alternative technique is still necessary in cases where the V/C ratio exceeds a value of 1.0.

TABLE 13 Statistical Comparison of Stopped Delay Estimates Using Predicted Versus Measured Saturation Flows as Input

Calculation	Stopped Delay			
	Field Measured	Estimated Using Predicted Saturation Flow	Estimated Using Measured Saturation Flow	Estimated Using Base Sat. Flow of 1900 Veh. and Reestimated % HV Factors
Number of Lane Groups	18	18	18	18
Mean Delay (seconds)	11.5	14.3	15.4	12.8
Standard Deviation	8.9	11.9	11.4	9.3
Mean Error	N/A*	3.9	5.5	2.8

* N/A denotes "not applicable."

TABLE 14 Permissive Left-Turning (S.A. Right) Stopped Delay Results

Intersection	Approach	V/C Ratio	Stopped Delay in Seconds		
			Measured	December 1981	Appendix G
Pretoria					
1	NB	1.00	99	41	102
1	SB	0.23	11	121	19
1	EB	0.76	14	47	21
4	EB	0.06	9	11	10
4	WB	0.20	9	11	11
Cape Town					
1	EB	0.82	16	23	30
1	WB	0.29	8	25	20
2	NB	0.77	35	26	35
2	EB	0.85	34	148	33
2	WB	1.43	40	47	352
Overall mean			27.5	50.0	63.3
Standard deviation			27.9	46.8	104.8
Mean when V/C < 1.0			26.1	50.3	31.2
Standard deviation			29.3	49.6	28.0
Mean error			N/A	39.1	5.3

Note: N/A = not applicable.

CONCLUSIONS

The NCHRP operations method estimates approximately 50 percent of saturation flows under South African conditions within a 10 percent range from measured values, and 75 percent of all estimates lie closer than 20 percent from the real saturation flows.

The NCHRP method very slightly underestimates saturation flows but still predicts them higher than the HCM method. The mean error of all saturation flow estimates using the NCHRP method is 172 vphgl compared with a mean error of 239 vphgl using the HCM method.

No evidence was found that the NCHRP method overestimates the effect of pedestrians on saturation flows. However, the method does not accurately take account of the effect of heavy vehicles on saturation flows under South African conditions. It was found that the use of a base saturation flow of 1,900 vphgl and the use of heavy vehicle factors with approximately twice the magnitude of the NCHRP factors (suggested for the United States), would provide saturation flow estimates with roughly 60 percent smaller mean errors.

The use of a default saturation flow of 1,800 vphgl for through lanes with good geometric conditions, in the absence of more detailed knowledge of conditions, is likely to provide accurate results.

Just over 64 percent of all stopped delay estimates range within 5 sec of the measured values, and approximately 82 percent range closer than 10 sec. In general, the NCHRP method very slightly overestimates average stopped delays.

The use of measured saturation flows, rather than estimated saturation flows, as input to calculate average stopped delays does not improve the accuracy of the delay estimates. It appears that there exists a need for further research to develop a stopped delay correction factor for both through and shared through and turn lanes in South African conditions.

Using Appendix G of the May 1982 and February 1983 versions of the NCHRP method, rather than the December 1981 scaling technique, markedly improves permissive left-turning stopped delays for lanes with V/C ratios of less than 1.0. The mean error of estimate for such lanes dropped from 39.1 to 5.3 sec.

A comment made at the course was that the application of the NCHRP method was too long, tedious, and cumbersome considering its perceived accuracy. The authors of this paper believe, however, that when the method has been mastered it gives better results for saturation flow than does the HCM method, which is currently in widespread use in South Africa; provides a broader range of MOEs; and allows greater diversity in the selection of the analysis unit.

These positive features allow for the possibility of making analyses that extend further than the conventional traffic engineering and planning analyses previously done on intersections. One such example is the ability of the NCHRP method to estimate stopped delay. The disadvantages of delays are wasted fuel, lost time, and increased air pollution. These three aspects are receiving increasing attention in South Africa and their inclusion would certainly enhance the application of the method in that country.

ACKNOWLEDGMENTS

The authors wish to thank the following organizations and persons who contributed to the course proceedings and data inputs used in this study: NCHRP for releasing the December 1981 draft of the method for use at the courses; the Southern Africa Road Federation for convening the courses; the two coinstructors, John Sampson and John Jones, for their invaluable assistance during the course proceedings and careful analysis of the studied intersections; and the 90 course delegates for their valuable analyses and constructive criticism of the applications.

REFERENCES

1. Highway Capacity Manual 1965. HRB Special Report 87. HRB, National Research Council, Washington, D.C., 1965, 397 pp.
2. JHK & Associates and Traffic Institute, Northwestern University. Urban Signalized Intersection Capacity. Draft Report, NCHRP Project 3-28(2). TRB, National Research Council, Washington, D.C., Dec. 1981.
3. JHK & Associates and Traffic Institute, Northwestern University. Urban Signalized Intersection Capacity. Draft Report, NCHRP Project 3-28(2). TRB, National Research Council, Washington, D.C., May 1982.
4. JHK & Associates and Traffic Institute, Northwestern University. Urban Signalized Intersection Capacity. Draft Report, NCHRP Project 3-28(2). TRB, National Research Council, Washington, D.C., Feb. 1983.

Publication of this paper sponsored by Committee on Highway Capacity and Quality of Service.

Evaluating Capacities of One-Lane Roads with Turnouts

FONG-LIEH OU

ABSTRACT

Speed-flow relationship models for one-lane roads with two-way traffic are developed. Each model considers a composite variable of speed divided by the traffic distribution ratio as a dependent variable and both traffic distribution ratio and volume as independent variables. The traffic distribution ratio represents the degree of traffic conflict and is measured as the percentage of one-way traffic on the heavy-traffic direction to total traffic. A 1982 traffic survey of four study sites in the Mount St. Helens Monument region forms the data base. The following are specific findings of the study: (a) Model specification and coefficients, including elasticities, are stable. (b) The capacity of a single-lane road with turnouts may exceed 400 vehicles per day without reaching the congested-flowing situation when the majority of traffic is controlled by citizen band radios. (c) Speed is more sensitive to traffic distribution than to volume. (d) The predictive ability of the developed models has been validated at nine study sites with satisfactory results. The results of this study provide road engineers and managers some guidelines for selecting the most cost-effective design standard and management strategy for one-lane roads with turnouts.

The Highway Capacity Manual (1,p.5) defines capacity as "the maximum number of vehicles which has a reasonable expectation of passing over a given section of a lane or a road in one direction (or in both directions for a two-lane or a three-lane highway) during a given time period under prevailing roadway and traffic conditions." This definition is not applicable to a one-lane road with turnouts that is managed for two-way traffic. The reason for building one-lane roads with two-way operation is that low traffic demand cannot economically justify building multilane roads. Multilane roads provide a high level of service, but they require a great amount of traffic demand to offset high construction and maintenance costs. The configuration of a typical one-lane road with turnouts is shown in Figure 1. The width of one lane ranges from 12 to 14 ft.

The U.S. Department of Agriculture Forest Service is probably the largest organization in the world to promote and manage two-direction traffic on one-lane roads with turnouts. Over the years, the Forest Service has built a 270,000-mile forest road system. More than 72 percent of the system consists of one-lane roads. The design standards of one-lane roads were determined by either speed or travel-time delay based on the 1960 Logging Road Handbook: The Effect of Road Design on Hauling (2). It was not until 1981 that volume was considered one of the criteria for evaluating traffic service (3). However, because the mathematical relationship between volume and traffic performance has not been defined, there is difficulty implementing this new concept. Volume is used primarily for determining long-term traffic demand (such as daily, seasonal, and annual traffic) rather than short-term system supply (such as hourly volume) in terms of capacity.

Defining capacity of forest roads is difficult because the traffic on them rarely reaches capacity. In 1982 it was expected that certain road segments in the Mount St. Helens Monument region might exceed their design capacity because, as a result of the May 18, 1980, volcanic eruption, approximately 900 million board feet of salvage timber were scheduled to be hauled to market in two seasons. The Forest Service took this opportunity to select 22 sites for a traffic study. Although the preliminary results concerning speed related to design standards have been reported elsewhere (4), the data collected in this study also permit an analysis of the relationship between volume and traffic performance to assess the capacity of low-volume roads.

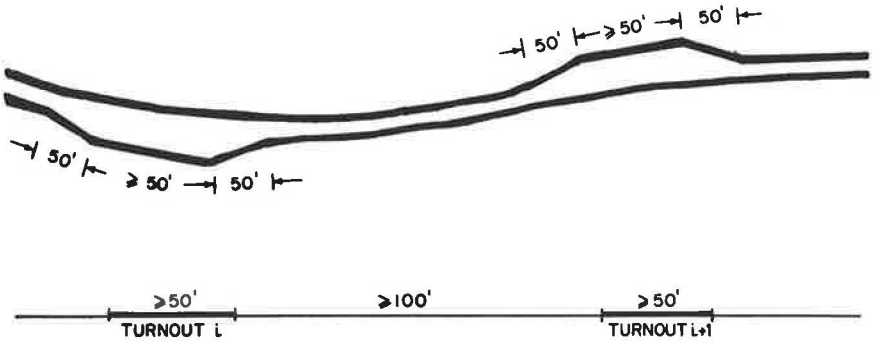


FIGURE 1 Configuration of one-lane roads with turnouts.

The purpose of this study was to compare lane capacity of single-lane roads with various design standards. The results of this study should not be used to determine the flow, average speed, or density at capacity for low-volume roads unless the traffic is controlled by citizen band (CB) radios.

BASIC CONCEPTS

The principal characteristics of traffic are flow, speed, and density. The fundamental characteristics are dependent on the geometric design of the roadway, the composition of the traffic stream, the consistency of road maintenance, and combinations of the three. The importance of these characteristics is manifested by the need for specific indications of impending traffic congestion, which can be used either in control processes intended to maintain optimum efficiency of an existing road system or in designing processes intended to obtain optimum cost-effectiveness of a new road system.

The relationship between flow (q), average speed (u), and density (k) can be shown by

$$q = k \times u \quad (1)$$

The functional relationships between any pair of variables are shown in Figure 2 and may be mathematically expressed by

$$u = u(q) \quad (2)$$

$$u = u(k) \quad (3)$$

$$q = q(k) \quad (4)$$

Of these functional relationships, the speed-flow relationship shown in Figure 2(a) is usually the basis for highway capacity analysis. The portion of the curve with the solid line denotes free-flowing conditions, and the portion of the curve with the dotted line represents congested-flowing conditions. During free-flowing conditions, speed decreases corresponding to the increase of flow. On the other hand, in congested-flowing conditions, flow decreases from capacity flow corresponding to the decrease of speed.

A number of studies have been devoted to evaluating the capacity of multilane high-volume roads (1,5). Capacity in terms of the speed-flow relationship plays an important role (along with other parameters such as speed, travel time, and freedom to maneuver) in defining the level of service of multilane high-volume roads. Level of service represents the quality of service being provided to the drivers who use the facility. The capacity of one-lane urban highways has been investigated in construction and reconstruction work zones (6,7). The results of these studies indicated that the lane capacity of one-lane roads differs significantly from that of multilane roads. The studies also found that on a two-lane highway, one-way and two-way traffic would result in different degrees of freedom to maneuver and thus affect lane capacity (7).

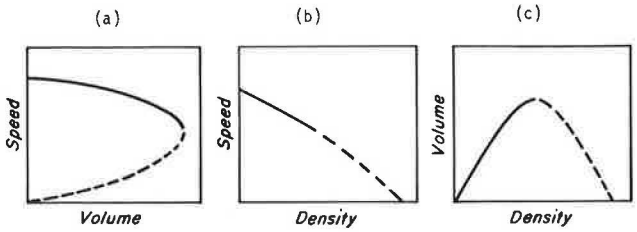


FIGURE 2 General relationships between traffic volume, speed, and density.

DATA SOURCE

The procedure of data collection and the road characteristics of study sites have been reported elsewhere (4). All road segments were used for speed-volume analysis, but only four of them were selected for capacity evaluation. Road characteristics of these four selected sites are given in Table 1. Except for grade, the design standards of these road segments are different. Sites 17 and 20 are paved, and sites 26 and 33 are gravel. Both sites 17 and 26 have good alignment and sight distance. Site 33 has good alignment and fair sight distance, and both design criteria for site 20 are fair.

The data are characterized by traffic volume, traffic composition, and weather condition (Table 2). The average volume ranges from 13 vehicles per hour (vph) at site 26 to 30 vph at site 33. This figure is relatively low compared with that of an urban highway system. The single-lane capacity of an urban highway (maximum 5-min flow) could be as high as 1,600 vph (7). The majority of the traffic was

TABLE 1 Road Characteristics of Study Sites

Characteristics	Site 17	20	26	33
Type of road	Single lane	Single lane	Single lane	Single lane
Type of surface	Paved	Paved	Gravel	Gravel
Alignment	Good	Fair	Good	Good
Grade (percent)	4.1	5.9	3.9	5.8
Sight distance	Good	Fair	Good	Fair
Template for passing	Yes	Yes	Yes	Yes
Length of road segment (miles)	0.93	1.20	0.58	2.10

TABLE 2 Data Characteristics

Site	Date of Observation	Duration of Traffic Count	Average Hourly Volume	Traffic Composition (%)						Weather Condition
				Light Vehicles	Forest Service Light Vehicles	Recreational Vehicle with Trailer	Empty Log Truck	Loaded Log Truck	Other Truck	
17	June 17, 1982	10:57 a.m.-2:00 p.m.	31	17	2	0	28	20	33	Clear, warm
	Aug. 25, 1982	10:15 a.m.-2:10 p.m.	34	29	4	0	18	49	0	Clear, warm
	Average		32	22	3	0	24	32	19	
20	June 29, 1982	12:44 p.m.-3:20 p.m.	30	36	8	0	12	32	12	Cloudy, cool
	July 28, 1982	6:20 a.m.-10:20 a.m.	16	37	16	0	13	30	4	Patchy
	Aug. 25, 1982	5:50 a.m.-8:20 a.m.	17	24	8	0	23	40	5	Clear, warm
	Average		20	32	11	0	17	34	6	
26	June 30, 1982	7:53 a.m.-10:53 a.m.	10	23	10	0	30	30	7	Cloudy, rain
	July 28, 1982	12:01 p.m.-2:10 p.m.	13	41	31	0	7	21	0	Clear
	Aug. 26, 1982	5:20 a.m.-8:50 a.m.	17	44	1	0	29	21	5	Clear, warm
	Average		13	38	11	0	24	23	4	
33	July 1, 1982	8:00 a.m.-11:20 a.m.	37	17	7	0	45	28	3	Cloudy
	July 26, 1982	2:00 p.m.-3:10 p.m.	37	40	3	0	7	16	0	Partly cloudy
	July 28, 1982	11:00 a.m.-1:35 p.m.	43	18	6	0	31	41	4	Overcast with fog
	Aug. 26, 1982	5:50 a.m.-8:40 a.m.	38	19	1	0	27	47	6	Clear
	Average		39	21	0	0	32	39	3	

empty log trucks and loaded log trucks, which accounted for from nearly 50 to more than 70 percent of the total traffic. The portion of other trucks differs greatly, ranging from 3 percent at site 33 to 19 percent at site 17. This discrepancy was caused by the reconstruction of a road segment near site 17 on June 17, when the traffic count took place. The portion of Forest Service light vehicles varied from 3 percent at site 17 to 11 percent at sites 20 and 26. No recreational vehicles were observed because the road segments in the study area were not open to the public during the study period.

Using the data collected from each site, the duration of the traffic count was broken down into several intervals, with each interval containing an independent, continuous flow. The gaps between intervals are longer than 4 min, and the intervals range from 5 to 35 min. The average hourly volume was expanded from the traffic of each interval.

Three variables for the analysis are speed, volume, and the degree of traffic conflict. The degree of traffic conflict is represented by the percentage of traffic in the heavy-traffic direction relative to total traffic. The ratio is 50 percent when the traffic in two directions is equal. On the other hand, the ratio is 100 percent if the flow becomes one-way traffic. Most vehicles were equipped with CB radios. When a traffic conflict occurred, the loaded log truck had the right-of-way. Other vehicles were required to yield the way to loaded log trucks by using turnouts as safe bay areas.

Design standards such as alignment and site distances were not selected for defining the speed-flow relationship for two reasons. First, the sample of study sites is too small. Second, the design standards for this study were not quantitatively defined. Using qualitative variables for modeling would require a sizable sample to produce meaningful results.

MODEL CALIBRATION

The first step in defining the capacity of a road is to examine the relationship between speed and volume. The speed-volume relationship for 13 study sites is shown in Figure 3. Each dot represents the result of a traffic count during a period ranging from 2 to 3 hours. Dots that apply to the same site are connected by solid lines. The ultimate practical capacity is expressed by the dashed line. Note that the ultimate practical capacity is defined as the maximum practical speed corresponding to a particular volume.

The figure reveals several remarkable features of speed-volume relationships. First, speed-volume relationships vary among study sites. This variation is affected by design standards and traffic characteristics as reported elsewhere (4). Next, the speed-volume relationships observed at a given site are inconsistent. Traffic performance at most sites follows the rule of thumb that an increase in volume tends to decrease speed. However, at some sites speed was not sensitive to volume, whereas at others speed increased as a result of high volume. This inconsistency indicates that capacity cannot be defined without consideration of other traffic characteristics such as traffic conflict. The third feature shown in Figure 3 is that the operation of two-way traffic on one-lane roads with turnouts is governed by the ultimate practical speed-volume curve shown by the dashed line. This indicates that one cannot expect a log truck to travel on one-lane roads at 30 mph when the volume reaches 40 vph. Finally, the figure shows that the majority of traffic can operate at a speed ranging from 20 to 30 mph with the volume between 15 and 40 vph, or 150 and 400 vehicles per day (vpd). This finding is vitally important because the Forest Service Transportation Engineering Handbook requires the construction of double-lane roads to meet the demand when traffic

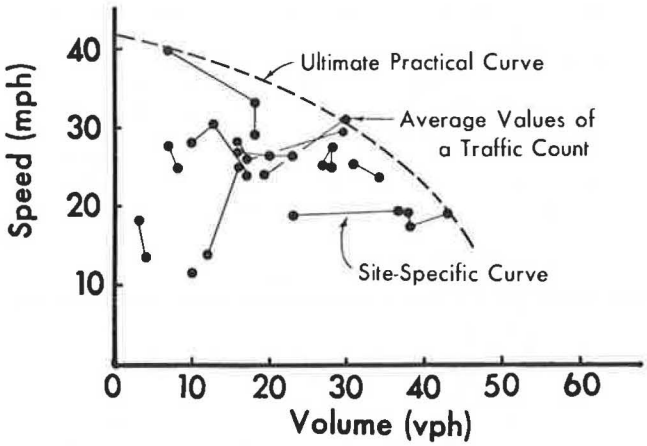


FIGURE 3 Speed-volume relationship of one-lane roads with turnouts.

volumes are greater than 250 vpd (3). Use of 400 vpd as a benchmark for determining the need for double-lane roads could save a considerable amount of transportation cost annually.

The foregoing discussion indicates that defining capacities of single-lane roads is difficult but manageable. Using the data derived from each of four selected sites for capacity analysis may exclude the impact of design standards on speed and, hence, reduce the difficulty to a minimum. Graphic analysis of these derived data revealed that speed-flow relationships between one-way and two-way traffic flows are significantly different. As expected, the traffic distribution between the two ways of travel plays an important role in affecting traffic performance. The greater the existing traffic conflict, the lower the speed. Although the impact of traffic conflicts on speed was reduced to a minimum because most vehicles were equipped with CB radios, psychologically drivers did not treat two-way traffic as one-way traffic (6). Thus the degree of freedom to maneuver for two-way traffic is lower than that for one-way traffic. For example, two-way traffic requires twice the sight distance acceptable for one-way traffic.

Because the speed-flow relationship is nonlinear, several nonlinear curves (e.g., product, logistic, exponential, logarithmic, and mix-logarithmic forms) have been used to simulate the relationship of speed with traffic volume and distribution ratio. By trial and error, a mix-logarithmic model has been found to fit most of the data well:

$$Z = a_0 - a_1C' - a_2V \quad (5)$$

where

Z = a composite variable equal to U/C,
U = average speed (mph),
C = traffic distribution ratio (percent of one-way traffic on heavy-traffic direction to total traffic that ranges from 13 to 33 vph),
C' = logarithmic value of C,
V = volume (vph), and
a = constant to be estimated.

Four models based on Equation 5 have been developed. The composite variable of speed to traffic distribution is highly related to both explanatory variables, with the coefficient of determinant, R^2, ranging from 0.7073 to 0.8527 (see Table 3). However, it should be noted that such a relationship is valid only within the range of base conditions from which models were developed. Such a relationship is subject to further investigation to see whether it holds when the base conditions of a new environment are out of this range.

STABILITY ANALYSIS

A comparison of the variable coefficients of four models (Table 3) shows that the speed-flow relationship is stable. When the composite variable of speed divided by traffic distribution ratio is considered as the dependent variable, the constants of these models range from 1.41573 for site 33 to 1.79683 for site 26. The coefficients of traffic distribution ratio range from 0.2637 for site 33 to 0.3197 for site 20, and the coefficients of volume fall in the range of 0.000282 for site 33 to 0.000371 for site 26. The differences are small compared with those of the design standards and traffic characteristics of the study sites.

Model stability can be further examined by traffic characteristics elasticity expressed as the change in the quantity of traffic performance by drivers in response to a 1 percent change in traffic characteristics. When the elasticity is 1.0, it is called unit elasticity. The elasticity is unelastic if it is less than 1.0 and is elastic if it is greater than 1.0. Two types of elasticities can be derived from Equation 1: volume elasticity

$$\eta_U = a_2VC/U \quad (6)$$

and traffic distribution elasticity

$$\eta_{TD} = C/U(U-a_1) \quad (7)$$

In accordance with Equations 6 and 7, elasticities for the four sites were calculated (Table 4). Table 4 reveals that all derived elasticities are inelastic (i.e., the value is less than unity). The result can be explained by using site 17 as an example. If a traffic management policy is to impose a limited control on one-way traffic to improve site 17 road system efficiency, reducing the traffic distribution ratio (percent of one-way traffic with heavy traffic to the total traffic) 1 percent will result in a 0.1228 percent speed increase. On the other hand, if the policy is to reduce volume, a 1 percent decrease in traffic volume will increase speed by 0.0665 percent. Elasticity can be a useful tool for decision makers who compare costs and benefits to determine a traffic management strategy.

TABLE 3 Speed-Flow Models and Base Conditions

Site	Speed-Flow Models	Base Conditions: Speed (mph) Avg	Max	Min	Hourly Volume (vph) Avg	Max	Min	Traffic Distribution (% of one-way to total) Avg	Max	Min
17	$U = (1.53354 - 0.27094C' - 0.000354V)C$ $R^2 = 0.8527$; F = 92.64 D.F. = 32; D.W. = 2.4410	26.2	31.5	18.8	59	200	5	83.4	100	50
20	$U = (1.79683 - 0.31966C' - 0.000292V)C$ $R^2 = 0.7115$; F = 51.79 D.F. = 42; D.W. = 1.5517	29.8	44.2	20.9	80	250	3	80.4	100	50
26	$U = (1.72195 - 0.30365C' - 0.0003713V)C$ $R^2 = 0.7073$; F = 41.08 D.F. = 34; D.W. = 1.7391	29.6	37.3	22.1	55	160	6	79.1	100	50
33	$U = (1.41573 - 0.26369C' - 0.000282V)C$ $R^2 = 0.8296$; F = 114.42 D.F. = 47; D.W. = 1.2695	19.3	25.5	12.2	71	260	15	79.2	100	50

Note: U = average speed (mph), C = percentage of one-way traffic on the heavy-traffic direction to total traffic (range from 50 to 100), C' = logarithmic value of C, and V = average hourly volume (vph).

TABLE 4 Traffic Characteristics Elasticities

Site	Elasticity	
	Flow	Traffic Distribution
17	-0.0665	0.1228
20	-0.0764	0.0915
26	-0.0520	0.0927
33	-0.1031	0.2156

Examination of Table 4 indicates that the derived elasticities are stable. Both elasticities for sites 17, 20, and 26 are identical. The flow elasticities range from -0.0520 for site 26 to -0.0764 for site 20, and the traffic distribution elasticities are within the range of 0.0915 to 0.1228. The elasticities for site 33 are higher than those for the other sites. They are -0.1031 and 0.2156 for flow and traffic distribution elasticities, respectively. This differentiation indicates that the traffic management policy can be a more effective option when applied to site 33 than when applied to other study sites. Comparison of elasticities for both traffic characteristics indicates that the effect of traffic distribution on speed is more sensitive than is that of volume. Traffic distribution elasticity is approximately twice volume elasticity.

MODEL VALIDATION

To validate the developed models, nine sites from the study area were selected for application. These sites are given in Table 5. As indicated in the table, the first four sites are the study sites where the models were developed. These four sites were selected to check whether the speed-flow models developed from a site can be used to predict the average speed of that site. The data in the table indicate that at sites 17, 20, 26, and 33 the discrepancies between the observed speeds and the speeds estimated by their own models are less than 1 percent. The result indicates that a model developed from a particular site is capable of predicting speed at that site. Note that the data used for model development were derived from two to four traffic counts, whereas the data for validation are based on one traffic count.

The next concern of model validation is to examine the spatial transferability of the developed models. The data given in Table 5 reveal that the result of applying a model developed from one site to other sites is mixed. The models developed from sites 17, 20, and 26 can predict the speed at most sites with less than 15 percent error. However, the difference between the observed speed and the speed estimated by these three models can be as great as 60 percent. On the other hand, the site 33 model underestimated the speed at all sites except site 34. The error in estimation could amount to 40 percent.

The result of transferability analysis indicates that there is no warrant for applying a model developed from one site to predict speed at another site. This result was expected because the design standard of one-lane roads varies from one road segment to another and the design standard has not been considered in the model development. Use of the design standard as one of the variables to develop speed-flow models is beyond the scope of this study. However, by classifying the nine study sites into three groups in accordance with the design standard it appears that sites 18 and 21 belong to the high-standard group, sites 17, 19, 20, 25, and 26 constitute the medium-standard group, and sites 33 and 34 fall into the low-standard group. High-standard roads are defined as paved roads with good alignment, good sight distance, and flat grade; low-standard roads are gravel or dirt with poor surfacing, poor alignment, and poor sight distance, as well as steep grade. Medium-standard roads are represented by paved or gravel roads with a fair rating of alignment, grade, and sight distance.

Based on this classification, the models of sites 17, 20, and 26 can predict the speed for the sites in the medium-standard group with errors of estimate ranging from 2.2 to 13.8 percent, and the site 33 model can predict the speed at site 34 with less than 10 percent error. As expected, these four models developed from medium- and low-standard site groups underestimated the speed at sites included in the high-standard group. On the other hand, the models developed from sites with medium design standards overestimated the speed of low-standard roads.

CONCLUSIONS

Although the study of capacities of single-lane roads with two-way traffic is limited, the results reported here provide convincing evidence that the capacity of low-volume roads can be defined. Specific findings of this study are the following: (a) Four speed-flow relationship models developed from four study sites have been found to be stable in terms of model structure and coefficients, including elasticities. (b) The capacity of a single-lane road with turnouts may exceed 400 vpd without reaching the congested-flowing situation when the majority of traffic is controlled by CB radios. (c) The impact

TABLE 5 Comparison of Observed and Estimated Speed for Nine Selected Sites

	Speed									Base Conditions	
		Forecast by Models									
		Site 17		Site 20		Site 26		Site 33			
Site	Observed mph	mph	Difference (%)	mph	Difference (%)	mph	Difference (%)	mph	Difference (%)	Average Hourly Volume (vph)	Traffic Distribution (% of one-way traffic to the total)
17	25.2	25.0	-0.8	24.6	-2.2	28.3	+12.2	19.8	-21.5	27	61
20	28.2	24.3	-13.8	28.1	-0.3	27.5	-2.6	19.5	-30.9	16	55
26	28.3	25.3	-10.6	29.1	+2.8	28.5	+0.7	20.0	-29.3	10	60
33	19.9	24.9	+25.1	28.9	+45.2	28.2	+41.7	19.7	-0.1	37	62
18	31.6	23.5	-25.6	27.3	-13.6	26.6	-15.8	19.0	-39.9	30	52
19	26.0	24.9	-4.2	28.8	+10.8	28.2	+8.5	19.7	-24.2	20	60
21	29.5	24.3	-17.6	28.1	-4.7	27.4	-7.1	19.5	-33.9	18	55
25	24.9	23.5	-5.6	27.2	+9.2	26.6	+6.8	19.1	-22.0	8	50
34	18.4	25.3	+37.5	29.2	+58.7	28.6	+55.4	20.1	+9.2	37	62

Note: + = overestimated, - = underestimated.

of traffic distribution between two ways on speed is twice that of volume. (d) The developed models can yield site-specific speed estimates with margins of error of less than 1 percent. (e) The model developed from one site is capable of predicting the speed at other sites with similar design standards. The results of this study provide some general guidelines for road engineers and managers to use in selecting cost-effective road design standards and developing cost-effective road management programs for one-lane roads with turnouts.

ACKNOWLEDGMENTS

The research reported in this paper was sponsored by the Forest Service, U.S. Department of Agriculture. A steering committee established to direct this research included Al J. Hessel, Lee W. Collett, and Jerry Knaebel. Although the author is the major contributor to the study results reported here, the other members of the study team, which included David R. Nordengren, Robert Keeney, Clarence Petty, and Lonnie Gray, collaborated closely during the study period.

REFERENCES

1. Highway Capacity Manual 1965. HRB Special Report 87, HRB, National Research Council, Washington, D.C., 1966, 397 pp.
2. J.J. Byrne, R.J. Nelson, and P.H. Googins. Logging Road Handbook: The Effect of Road Design on Hauling Costs. Agriculture Handbook 183. Forest Service, U.S. Department of Agriculture, 1960.
3. Forest Service Handbook 7709. U.S. Department of Agriculture, Chapter 11, Dec. 1981.
4. F.-L. Ou, A.J. Hessel, L.W. Collett, and D.R. Nordengren. Effect of Road Design on Timber Hauling Speed in the United States. In Transportation Research Record 898, TRB, National Research Council, Washington, D.C. 1983, pp. 61-65.
5. Interim Materials on Highway Capacity. Transportation Research Circular 212, TRB, National Research Council, Washington, D.C., Jan. 1980, 276 pp.
6. C.L. Dudek and S.H. Richards. Traffic Capacity Through Urban Freeway Work Zones in Texas. In Transportation Research Record 869, TRB, National Research Council, Washington, D.C., 1982, pp. 14-18.
7. R.E. Dudash and A.G.R. Bullen. Single Lane Capacity of an Urban Freeway During Reconstruction. In Transportation Research Record 905, TRB, National Research Council, Washington, D.C., 1983, pp. 115-117.

Publication of this paper sponsored by Committee on Low Volume Roads.

Abridgment

Computer Simulation To Compare Freeway Improvements

ROBERT W. STOKES and JOHN M. MOUNCE

ABSTRACT

Use of a simulation program, FREQ6PE, to compare proposed improvements for the Southwest Freeway (US-59) in Houston, Texas, is described. The simulation model was calibrated using actual field data and was then used to identify the best of a number of proposed geometric improvements. The proposed improvements were evaluated by comparing key simulated measures of effectiveness for the proposed systems with comparable measures for the base (do-nothing) system. Based on the experience gained in using the program, it is concluded that the program can be an effective and economical tool for studying the dynamic response of a freeway to a variety of input specifications.

The Southwest Freeway (US-59) bisects one of the fastest growing corridors in the Houston region. Traffic demands on the freeway outside of I-610 (see Figure 1) have increased 45 percent over the past 5 years to an average daily volume of about 194,000 vehicles. Depressed levels of service often extend from 6:00 to 9:00 a.m. and from 4:00 to 7:00 p.m., with trip times frequently tripling from off-peak to peak periods (1).

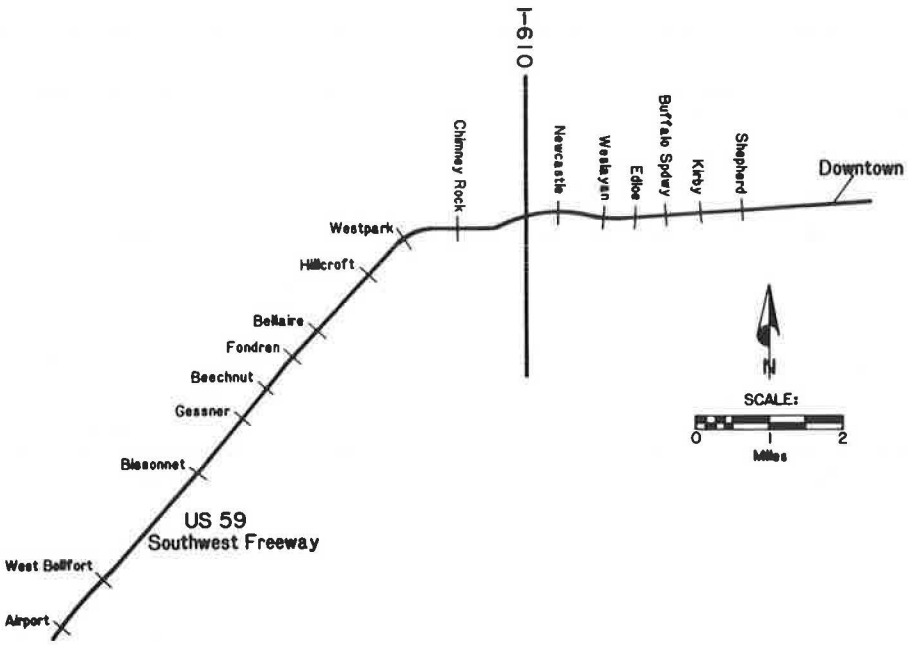


FIGURE 1 Southwest Freeway study corridor.

As a consequence of the continuing growth within the corridor and the corresponding declines in freeway levels of service, a number of geometric improvements have been proposed for the corridor. The FREQ6PE simulation program (2) developed at the University of California at Berkeley was used to evaluate the proposed geometric designs. A summary of the improvements evaluated is given in Table 1.

The objective of the computer simulations was to quantitatively assess the proposed design configurations and identify the configuration or configurations that best satisfied the objective of maximizing person throughput on the facility. The basis of comparison for the improvements was the 1995 do-nothing alternative.

TABLE 1 Improvements Evaluated

Improvement	Description
Do Nothing	No Improvement.
Add One Lane	Add one freeway lane over the entire length of the study corridor.
Add Two Lanes	Add two freeway lanes over the entire length of the study corridor.
Add Shoulder Lane	Add a shoulder lane from Westpark Entrance to I-610 Exit.
Stack Ramps	Stack (elevate) ramps at: a) Buffalo Speedway Entrance/Kirby Exit; and b) Kirby Entrance/Greenbriar Exit.

APPLICATION

Model Calibration

Freeway main-lane demands were input to the FREQ6 model in the form of mode-specific origin-destination (O-D) tables by 15-min time segments for the morning peak period (6:00-9:30 a.m.). Additional input included pertinent geometric data and speed-flow curves. The calibration process involved graphically comparing measured travel times and observed queue length with model output. The capacities (service volumes) of critical (bottleneck) sections were adjusted until the model output approximated observed main-lane queueing patterns. The calibration procedures used were clearly subjective in nature. However, because the primary concern was the differential effects of the design configurations rather than the performance of a particular system, the calibration procedures employed seemed consistent with the level of precision desired.

Freeway capacity was found to be a particularly sensitive parameter in the calibration process. Even modest changes in capacity produced substantial differences between simulated and observed conditions.

The merging analysis subroutine of the model posed some initial difficulties during the calibration process. For example, the merging analysis option should be engaged only when theoretical ramp capacities are used. The merging analysis subroutine then simulates ramp operations by metering ramp volumes. If the merging analysis option is engaged when measured volumes are used, the metering effect of the merging analysis subroutine will tend to inflate the ramp delay values.

Summary of Results

The design year (1995) simulations were performed for two traffic growth scenarios. A low-growth scenario assumed a 3 percent annual increase in freeway traffic demand, and a high-growth (or worst-case) scenario assumed the existence of a base year (1981) freeway latent travel demand of 40 percent (3). Tables 2 and 3 give summaries of the key operational measures of effectiveness developed from the simulations. It should be noted that the measures of effectiveness given in Tables 2 and 3 are based on

TABLE 2 Summary of Simulated Measures of Effectiveness Assuming Existence of Freeway Latent Travel Demand *(3,4)*

Analysis Period: 6 a.m. to 9:30 a.m.

Alternative	Traffic Year	Freeway Travel Time		Total Travel Time (Incl. ramp delay)		Total Travel Distance		Average Vehicle Speed (mph)	Gasoline Consumption (gallons)
		Veh-Hr.	Pass-Hr.	Veh-Hr.	Pass-Hr.	Veh-Mi.	Pass-Mi.		
Do Nothing	1981	4,400	6,000	4,400	6,000	190,600	260,300	43	11,700
	1995	9,900	13,500	61,100	83,400	186,700	225,100	20	43,400
Stack Ramps	1995	8,000	10,900	55,500	75,600	183,300	250,300	23	40,900
Add One Lane	1995	10,900	14,900	46,900	62,400	270,600	369,700	26	39,800
Add Shoulder Lane	1995	9,300	12,800	52,900	72,100	187,700	256,400	20	39,400
Add Two Lanes	1995	9,900	13,500	37,600	48,200	328,000	448,100	35	37,500

TABLE 3 Summary of Simulated Measures of Effectiveness Assuming No Freeway Latent Travel Demand *(3,4)*

Analysis Period: 6 a.m. to 9:30 a.m.

Alternative	Traffic Year	Freeway Travel Time		Total Travel Time (Incl. ramp delay)		Total Travel Distance		Average Vehicle Speed (mph)	Gasoline Consumption (gallons)
		Veh-Hr.	Pass-Hr.	Veh-Hr.	Pass-Hr.	Veh-Mi.	Pass-Mi.		
Do Nothing	1981	4,400	6,000	4,400	6,000	190,600	260,300	43	11,700
	1995	9,500	12,900	33,500	45,700	185,500	253,500	21	27,600
Stack Ramps	1995	7,700	10,500	31,300	42,600	181,200	247,500	25	26,800
Add One Lane	1995	9,400	12,800	19,800	25,900	273,600	373,900	33	24,100
Add Shoulder Lane	1995	9,100	12,400	29,500	40,100	182,200	248,900	21	25,600
Add Two Lanes	1995	5,000	6,800	15,100	18,900	279,000	381,100	56	23,500

information from the simulation summary tables and, as such, refer to the entire length of the freeway. Conditions on individual subsections of the freeway may vary considerably from these system averages.

To get a feel for the potential localized effects of the system configurations evaluated, the queueing contours output by FREQ6 were examined. Figure 2 shows a summary of these contours for the "with latent demand" scenario. The queueing contours depict those freeway subsections operating at or below 35 mph. As shown in Figure 2, the "stacked ramps" alternative and the "add two lanes" alternative could have significant clearing effects on those sections of the freeway downstream of the I-610 interchange. Note, however, that only the "add two lanes" alternative appears to have the potential to produce any substantial improvement in system operating speeds (see Table 3). Consequently, in terms of the study objective of maximizing person throughput, the "add two lanes" alternative appears to be the best of the improvements evaluated.

CONCLUSIONS

Given the growth rates and travel patterns that characterize the freeway corridor, none of the improvements evaluated (with the possible exception of the "add two lanes" alternative) appears capable of doing much to alleviate the peak-period congestion problems that plague the corridor. Because of their propensity to attract traffic from other routes and to generate additional (or at least previously unserved) vehicle trips, improvements directed at the supply side of the problem do not appear, in this case, to be particularly effective. The simulation results suggest that operational improvements or demand management strategies that address the problems of main-lane congestion by diverting traffic to alternate routes or other travel modes (e.g., high-occupancy vehicles) may be more productive.

Based on the experience gained using the FREQ6 model it is concluded that the model can be an effective tool for comparing freeway improvements. The measures of effectiveness output by the program are extensive and provide the analyst a good deal of flexibility in selecting an evaluation scheme. Though the calibration process can be tedious, the macroscopic nature of the program requires a minimum amount of input data and computer time. For the simulations reported in this paper, the simulation of 3.5 hr of clock time (14 time segments) required only 8 sec of central processing unit time at a cost of about $2.50 per run. Thus, when properly calibrated, the FREQ6 model can be a highly effective and economical tool for studying the dynamic response of a freeway to a variety of input specifications.

ACKNOWLEDGMENT

The contributions of Dick McCasland, Gene Ritch, and Danny Morris of the Texas Transportation Institute were of great value to this effort and are gratefully acknowledged.

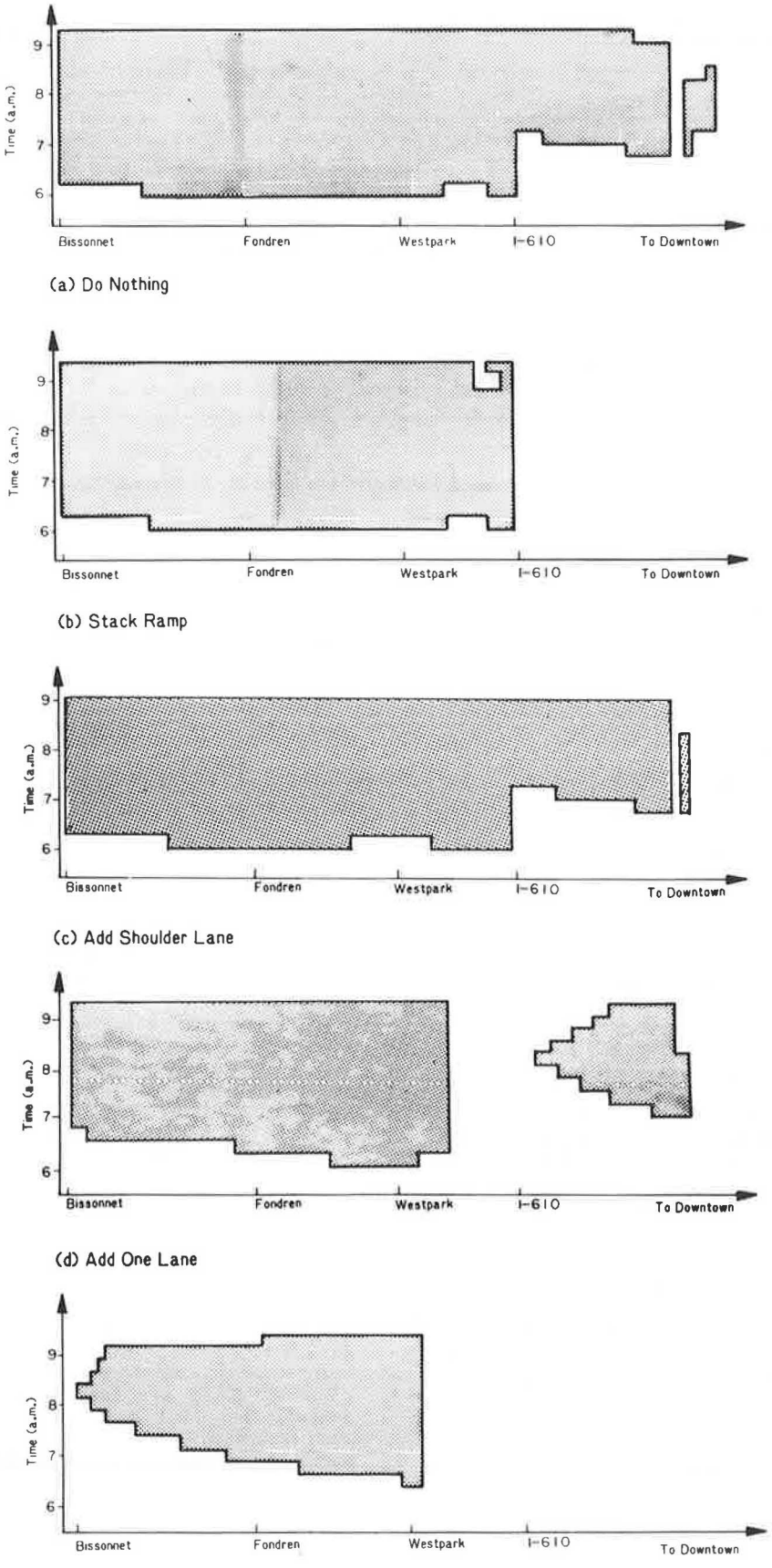


FIGURE 2 Simulated 1995 a.m. peak-period queueing contours (with latent demand).

REFERENCES

1. R.W. Holder, D.L. Christiansen, and C.A. Fuhs. Houston Corridor Study. Final Report. Texas Transportation Institute, Texas A&M University, College Station, 1979.
2. P.P. Jovanis, W.-K. Yip, and A.D. May. FREQ6PE--A Freeway Priority Entry Control Simulation Model. Research Report UCB-ITS-RR-78-9. Institute of Transportation Studies, University of California, Berkeley, 1978.
3. R.W. Stokes, J.M. Mounce, D.E. Morris, and R.L. Peterson. Simulation Analyses of Proposed Improvements for the Southwest Freeway (US-59), Houston, Part 1: Long-Term, Capital-Intensive Improvements. Final Report. Texas Transportation Institute, Texas A&M University, College Station, 1982.
4. R.W. Stokes, D.E. Morris, and W.R. McCasland. Simulation Analyses of Proposed Improvements for the Southwest Freeway (US-59), Houston, Part 2: Demand Management/Operational Improvements. Texas Transportation Institute, Texas A&M University, College Station, Jan. 1983.

This paper is based on research conducted as part of an interagency project entitled "Southwest Freeway Improvement Evaluations" sponsored by the Texas State Department of Highways and Public Transportation. The views, interpretations, and conclusions expressed are those of the authors. They are not necessarily those of the Texas State Department of Highways and Public Transportation.

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

Development and Application of a Macroscopic Model for Rural Highways

JUAN C. SANANEZ and ADOLF D. MAY

ABSTRACT

The development of a macroscopic computer simulation model is presented. The simulation model, RURAL1, calculates traffic performance given road supply (geometrics) and demand (traffic) information. The model can analyze four types of subsections: freeway, multilane, two-lane, and passing-lane. To perform the simulation, the roadway must first be divided into subsections; users can specify up to 100 subsections. Subsection boundaries are established on the basis of changes in road geometrics, or traffic demand characteristics, or both. RURAL1 calculates traffic performance measures, such as average speed, travel time, and vehicle delay, on a directional basis for each subsection and summarizes performance results for the entire roadway section. The simulation model was applied to an actual field site where the existing condition was evaluated against two additional cases.

In recent years road maintenance budgets have increased substantially, affecting the availability of funding for new construction. Thus, state transportation agencies have been looking for new ways of managing the existing transportation system more effectively, using an approach called transportation system management (TSM). This approach is now being applied to the rural road system.

New techniques are needed to evaluate the cost-effectiveness of rural road improvements and to provide planners and decision makers with more accurate information on which to base their decisions. Sophisticated computer models, all microscopic, have been developed to study traffic operations on rural roads. Although these models offer great capabilities for analyzing traffic behavior, their applications are limited. Particularly important is the restriction these models impose on the size of the road section that can be simulated. Because of their simpler structure and logic, macroscopic models can easily be used to study longer sections of roadway where several improvements are to be implemented. However, as a result of their simplified logic, macroscopic models offer less detail and precision in performance and measures of effectiveness (MOEs) than microscopic models do and should be treated as a supplement to microscopic models rather than as a replacement.

Historically, macroscopic models have been derived after the development of, and with the use of, microscopic models. In freeway corridors, for example, the development of the macroscopic FREQ model followed early microscopic models used in the analysis and study of freeway operations. The need to consider additional impacts and to study more control strategies played an important role in the creation of FREQ. Another example is TRANSYT, a macroscopic model for the analysis and optimization

of a coordinated set of intersections. TRANSYT followed earlier microscopic models for the study of isolated intersections. The complexity of the system (a group of intersections) and additional impacts such as fuel consumption and emissions were the primary reasons for the creation of TRANSYT. As a system grows in complexity (long sections with many geometric elements) and as the study of more impacts is required, macroscopic models will be employed because of the relative simplicity with which they handle more complicated situations.

A study, sponsored by the California Department of Transportation (Caltrans) and FHWA, is being conducted at the University of California, Berkeley, to develop a macroscopic model for the analysis of traffic operations on rural roads. The development and initial application of RURAL1, the first version of this model, are described herein.

BACKGROUND

Since 1976 a series of traffic simulation models has been developed at the Institute of Transportation Studies (ITS) at the University of California, Berkeley, to assist in the evaluation of traffic performance on rural highways. SIMTOL, the first of these models, was developed by W.A. Stock (1) for two-lane, two-way rural roads. Allowing for detailed modeling of vertical alignment data, SIMTOL assumed a high standard horizontal alignment that did not affect driver behavior, except with respect to no-passing zones. The major drawback of the model was that it simulated one direction of traffic, making assumptions about the gaps in the opposite flow. Despite this limitation, the model provided interesting information on the spatial characteristics of two-lane traffic.

In 1980 Botha and May (2) developed a computer program (TWOMIC2-CL) for the microscopic simulation of traffic operations on rural roads with climbing lanes. This model was a modification of a sophisticated simulation model for two-lane, two-way rural roads developed at the Midwest Research Institute--the TWOWAF model, which has been thoroughly validated and used by many research organizations across the United States. Botha and May incorporated the capability of simulating climbing-lane operations in the TWOWAF model's logic. TWOMIC2-CL was then used to derive guidelines for the optimal length and location of a climbing lane on a specific grade. Improvements of TWOMIC2-CL, which were identified by Botha, were undertaken in a follow-up research project.

A recently completed research project (3) produced an improved version of TWOMIC2-CL. The new model, TWOMIC3-CL, has revamped merging logic and additional input and output refinements; it also introduces a new measure of effectiveness (MOE)--accident potential. TWOMIC3-CL needs additional validation of the maneuver multipliers used in the derivation of the accident potential MOE. One aspect of this research was the development of an approach to the construction of a macroscopic model.

Recent simulation models have been developed elsewhere. First, the North Carolina State University (NCSU) model (4) was derived from an earlier simulation model developed by the Franklin Research Institute (the FIRL model) (5). The NCSU model incorporated a detailed truck-passing performance model and a routine to generate speed and headway data points, individual travel times, and so forth. The NCSU model was later modified to simulate roadway intersections (6).

Second, the Swedish National Road and Traffic Research Institute developed an event-based simulation model for rural highways, the Swedish-VTI model (7). This model, written in SIMULA-67, is probably one of the most thoroughly validated two-lane simulation models currently available. Several classes of vehicles can be specified with stochastically selected desired speeds. The model has a comprehensive passing logic derived from empirical passing observation studies conducted in Sweden. The model has been used by the National Swedish Road Board to investigate improvement options on primary roads and is now being adapted for use in several countries, including the United Kingdom and India.

Third, St. John and Kobett developed a microscopic simulation model, known as the TWOWAF model (8), which is able to simulate two-way traffic operations on two-lane highways for a wide variety of configurations. The model can simulate several vehicle classes; passing maneuvers and vehicle performance are also simulated in great detail. Some refinements made to the TWOWAF model are improved vehicle generation routines, reduced number of vehicle classes, and program redimensioning. A group of researchers at the Texas Transportation Institute, headed by Carroll Messer in cooperation with KLD Associates (9), have used a slightly modified version of the TWOWAF model in deriving the relationships for the two-lane chapter of the new Highway Capacity Manual (HCM) (10).

Finally, two simulation models have been developed in Australia. The first model was developed by the Australian Road Research Board (ARRB) and it is called the TRARR model (11). It has the flexibility to specify up to 18 vehicle classes. Its passing logic is based on a set of deterministic decision rules and passing safety values. The TRARR model requires an extensive input data file of road characteristics and traffic parameters. ARRB staff are calibrating the TRARR model in conjunction with case study applications. The second Australian model was developed by Hoban (12) and consists of a set of programs based on the results of previous studies of traffic behavior and simulation. The model was used to investigate the effect of passing lanes on the traffic performance of rural roads and was deliberately restricted to fairly level terrain and uncongested flow rates.

Most of the research conducted at ITS and elsewhere has been concerned with the development of microscopic simulation models; apparently, there have been no macroscopic models available for evaluating traffic performance on rural highways. Because of the limitations of existing microscopic models, it was believed that a simulation model capable of studying traffic operations over long stretches of roadway was needed, and the idea of developing a macroscopic model, RURAL1, was generated.

RURAL1

RURAL1 is a macroscopic deterministic model for the analysis of traffic on rural roads. The structure of the rural highway model is rather simple. Given supply information (road geometrics and subsection breakdown) and demand information (traffic characteristics), RURAL1 calculates performance and relevant MOEs. Subsection boundaries should be established any time there are changes in supply or demand characteristics; for example, changes in gradient or design speed are considered in the criteria for subsection specification, and an inter-

section is treated as a boundary condition because traffic volumes change at this location. Each of the four submodels comprising RURAL1 analyzes a corresponding subsection type; these submodels are shown in Figure 1. A fifth submodel dealing with subsection dependency is also shown in Figure 1. The submodels are processed independently; RURAL1 uses them when it encounters the particular type of subsection analyzed by each submodel. The submodels can be used several times during a single run, depending on the number of subsections of a specific type.

This simplified model structure allows (a) ease of constructing and programming the model; (b) flexibility for change and incorporation of program submodels; and (c) a short amount of computer time expended in a single simulation, which allows the user to evaluate many alternatives at a low cost. The five submodels are described hereafter.

Freeway Analysis Submodel

This submodel evaluates capacity and performance for freeway subsections on a directional basis. The calculations are based on the principles of the draft report on capacity and level of service evaluation for basic freeway segments (13). The submodel first calculates subsection capacity using Equation 1, which includes a driver factor (which accounts for a driver's knowledge of the road) and truck mix, grade and grade length, lane width, and other factors.

$$C = c_i \cdot N \cdot f_w \cdot f_{hv} \cdot f_t \quad (1)$$

where

C = capacity in vehicles per hour (vph) one way;

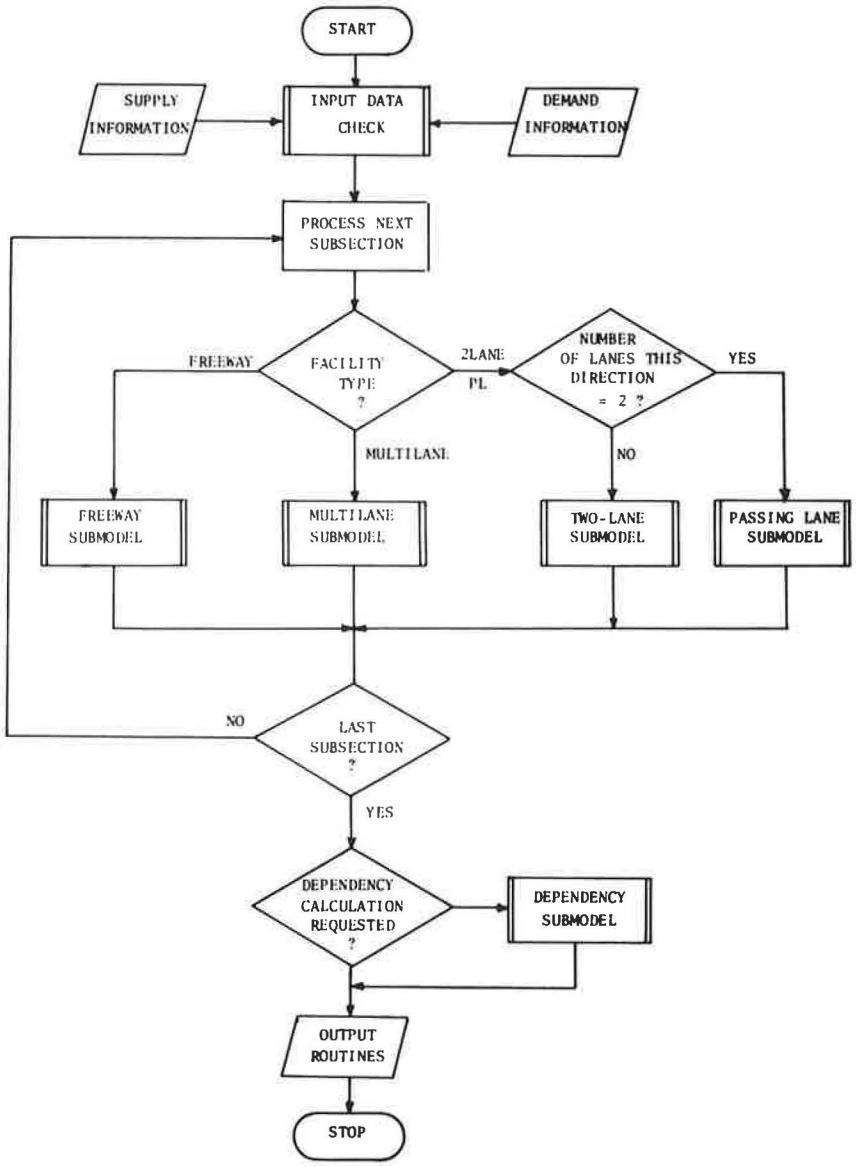


FIGURE 1 Schematic representation of RURAL1 with program submodels.

c_i = capacity under ideal conditions in passenger cars per hour per lane (pcphpl);
N = number of lanes in one direction;
f_w = lane and shoulder width factor;
f_{hv} = heavy vehicle factor; and
f_t = driver population factor.

After capacity is calculated, the submodel calculates subsection volume-to-capacity (V/C) ratio using volume adjusted by peak-hour factor as shown in Equation 2.

$$V/C = (V/PHF) \cdot (1/C) \quad (2)$$

where V/C is the volume-to-capacity ratio, V is the actual one-way volume in vph, PHF is the peak-hour factor, and C is as defined previously.

When the V/C has been computed, MOEs are calculated using the V/C, speed, and density relationships presented in the draft freeway chapter of the new Highway Capacity Manual. Downgrade analysis is performed in a similar fashion but with modified truck equivalent values related to the steepness of the downgrade.

Multilane Analysis Submodel

Like the freeway submodel, the multilane submodel calculates capacity and performance for multilane subsections on a directional basis. It makes use of the principles of the draft report for capacity and level of service evaluation for multilane subsections (14). The freeway submodel and multilane submodel are similar. Major differences in multilane analysis are the introduction of a divider factor to account for the type of median divider available and different factors for lane width and side obstructions based on the median divider type. Equations 3 and 4 show the multilane analysis.

$$C = c_i \cdot N \cdot f_{hv} \cdot f_w \cdot f_{te} \cdot f_t \quad (3)$$

$$V/C = (V/PHF) \cdot (1/C) \quad (4)$$

where f_{te} is a factor for type of multilane highway and c, c_i, N, f_w, f_{hv}, f_t, V, V/C, and PHF are as in Equations 1 and 2.

Two-Lane Analysis Submodel

Based on the capacity and level of service procedures developed in the draft report for two-lane highways (9) the two-lane submodel uses the specific grade approach proposed in this draft report. The calculations required to determine traffic performance are different from those of the other two submodels. Instead of calculating capacity and obtaining a V/C value, the two-lane submodel iterates, beginning with level of service A, until the corresponding level of service is found by comparing actual volume to calculated maximum service volumes. Equation 5 shows the calculation of service volumes for a given level of service.

$$SV_L = 2{,}800 \cdot D_D \cdot H_L \cdot W_L \cdot G_L \cdot A_L \cdot T_L \quad (5)$$

where

SV = service volume at a given level of service (LOS), L, in passenger cars per hour (pcph) two ways;
2,800 = ideal capacity in pcph (two way);
D_D = directional distribution factor;
H_L = peak-hour factor, LOS L;
W_L = lane-width factor, LOS L;
G_L = terrain factor, LOS L;
A_L = passenger car factor (upgrade), LOS L; and
T_L = truck factor, LOS L.

When level of service has been determined, MOEs are calculated using an approach similar to that of the other two submodels: linear interpolation using service volumes rather than V/C values to calculate average speed and other MOEs. Downgrade analysis is performed using the same methodology used for the upgrade but with modified truck equivalent values according to the steepness of the grade.

Passing-Lane Analysis Submodel

This submodel calculates traffic performance for three-lane sections. Figure 2 shows a flow chart of this submodel. Traffic performance for the direction that has the passing lane is calculated using the equations and factors of the multilane submodel with a reduced value for ideal capacity per lane. The basic assumption of the passing-lane submodel is that lane distribution is a function of traffic volume and that heavy vehicles [trucks and recreational vehicles (RVs)] have a tendency to occupy the right lane. Adjusted truck and RV percentages are calculated for each lane and corresponding MOEs

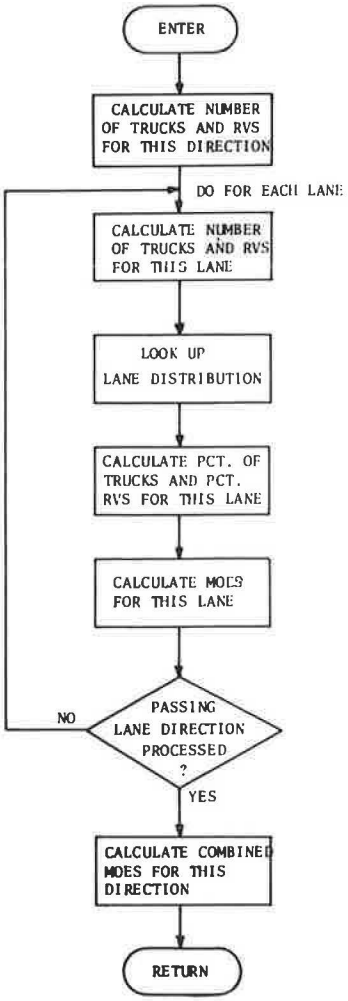


FIGURE 2 Passing-lane submodel.

are also calculated separately and are later combined for the passing-lane direction. Passing lanes longer than 2 miles are calculated as multilane sections and traffic performance for the downgrade direction is calculated as a single direction of a two-lane section with 100 percent restricted passing.

Subsection Dependency Submodel

The main idea of this submodel is to adjust the performance calculations on the basis of the interrelations between adjacent subsections. Traffic performance for any subsection is a function not only of subsection geometric characteristics and traffic volumes; it also depends on the platooning of vehicles in the preceding roadway sections. For example, if a flat subsection is preceded by a steep upgrade, the traffic performance on the flat section will be different than that on the same flat subsection if it were preceded by a downgrade subsection. The subsection dependency submodel tries to quantify the effect of upstream subsections on the subsection being calculated. Figure 3 shows the submodel's logic. The dependency submodel is perhaps the most critical component of RURALl and efforts are being made to calibrate it successfully.

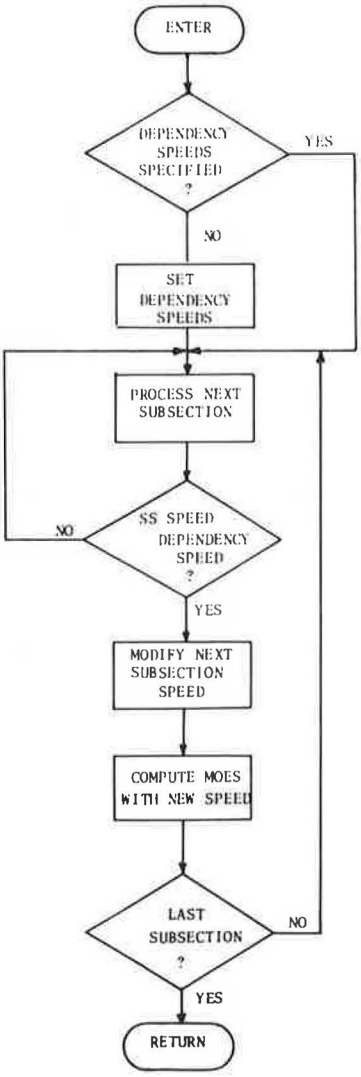


FIGURE 3 Subsection dependency submodel.

Model Assumptions

RURALl is a macroscopic deterministic model; thus the behavior of individual vehicles is ignored in favor of the average behavior of platoons or groups of vehicles. No special allowance is made for the actual randomness of demand input parameters. This may appear to be a considerable simplification of real life, but the comparative simplicity (relative to a microscopic model) affords a rapid cost-effective way of obtaining reasonable estimates of the performance of many design configurations.

The assumptions of RURALl can be categorized as general or specific. General assumptions hold for the entire section of roadway and are related to the program logic. Specific assumptions are those that relate to the specific type of subsection analyzed or the type of calculation performed.

General Assumptions

General assumptions of RURALl are

1. Calculations are made for a 1-hr time slice; demand remains constant within this time period.
2. The program deals only with uncongested flow; demand cannot exceed capacity.
3. The road is broken into subsections and subsection performance is calculated independently. The subsection dependency submodel will correct this deficiency.
4. The program evaluates performance only for a given demand and supply configuration.
5. The program uses the principles of the draft reports on capacity and level of service evaluation for freeway, multilane, and two-lane sections.
6. Four types of subsections can be analyzed: freeway, multilane, two-lane, and passing lane (F, M, 2L, PL).

Specific Assumptions

Specific assumptions of RURALl are

1. A typical truck table [200 lb/nominal horsepower (NHP)] is used in calculating passenger car equivalents (F, M, 2L, PL).
2. The gradient for the second direction is assumed to have the same magnitude as the first direction but reverse sign (F, M, 2L, PL).
3. Horizontal curvature is taken into consideration in the specification of no-passing zones (2L).
4. Linear interpolation is used to determine the level of service and corresponding MOEs (F, M, 2L, PL).
5. Performance calculations are conducted separately for each direction (F, M, 2L, PL).
6. Downgrade treatment is done using the modified truck equivalent values according to the steepness of the grade (F, M, 2L, PL).

Input Data Description

Input data are divided into two categories: common input and subsection input. Common input data are the relevant information that remains unchanged for the entire road section or for a particular type of subsection. Subsection input data, the information pertinent to a particular road subsection, are divided into geometric and traffic data and can be specified for both directions or only one direction.

Common Input Data

Common input data are specified in the first card of the input deck. The fields are

1. Problem description: a 60-character alphanumeric field describing the simulation run.
2. Number of subsections: the total number of subsections into which the road section has been divided. A maximum of 100 subsections can be specified.
3. Calculation type: the type of calculation RURAL1 will conduct. Three types of calculations can be specified: performance, demand, and supply. RURAL1 can handle only type 1 calculation (i.e., it can only calculate performance at the present time).
4. Main direction specification: cardinal orientation used only for identification. (RURAL1 assumes that direction 1 is the main direction.)
5. Desired speed: the speed at which a motorist would drive the entire road if not restricted by geometrics or traffic. This value is used in the calculation of vehicle delay.
6. Driver factor: used in the performance calculations for freeway and multilane sections. It will remain constant for all freeway and multilane sections.
7. Output level: can take a value of 0 for a summary output (default) or 1 for a detailed output (see next section for output description).

Subsection Input Data

Subsection input data require a card for each of the road subsections and include information for both directions. Data are divided into geometric data and traffic data. Geometric data refer to the physical attributes of the road subsection, and traffic data refer to the characteristics of traffic for each subsection. Geometric data and traffic data can be specified for both directions or for one direction if constant.

Geometric Data

Geometric data include

1. Subsection type: type 1, 2, or 3, corresponding respectively to freeway, multilane, or two-lane section. If the section is type 2, multilane, it should be specified whether the subsection is divided. The numbers 2-0 correspond to undivided multilane and 2-1 to divided multilane sections. Passing-lane sections should be coded as type 3.
2. Number of lanes: the number of subsection lanes per direction.
3. Percent grade: gradient for direction 1; program automatically reverses the sign for direction 2.
4. Subsection length: the length in hundredths of a mile of a particular subsection.
5. Design speed: the subsection design speed expressed in mph; specified for all four types of subsections. The freeway and multilane submodels use the design speed in the interpolation of V/C, and the two-lane submodel uses it through the specification of no-passing zones.
6. Lane width: subsection lane width in feet; assumed constant for both directions.
7. Obstruction distance: median and side obstruction for freeway and multilane sections and side obstruction for two-lane and passing-lane sections, expressed in feet.
8. Percent no-passing zones: specified for both directions; only relevant for two-lane sections. The single lane of a passing-lane subsection is assumed to have 100 percent restricted passing. Field should be left blank for freeway and multilane sections.

Traffic Data

Traffic data include

1. Level of service: should be left blank because RURAL1 calculates performance only.
2. Traffic volume: expressed as two-way hourly volume.
3. Directional split: percent of two-way traffic traveling in direction 1; used in calculating directional volumes.
4. Percent trucks: percent of single-unit and tractor-trailer combinations with six or more tires on the pavement excluding recreational vehicles; specified for each direction.
5. Percent recreational vehicles: percent of vehicles having six or more tires not included in earlier classification; specified for each direction.
6. Peak-hour factor: defined as the rate of the peak-hour volume to the maximum flow rate for a specified period, usually 15 min, within the peak hour; specified for each direction; field can be left blank for two-lane sections and program will select default value.

Output Description

RURAL1 presents two output options that are selected in one of the fields of the first input card. A value of 0 in this field specifies that the user would like a summary output. A value of 1 indicates a request for a detailed output.

The summary output option provides the user with a replication of the input deck that can be used to check the values of the different fields. If one of the field values is out of range, RURAL1 will not process the information and will underline the corresponding wrong field. The second part of the summary output provides a direction summary of the calculations performed by RURAL1. Some geometric and traffic features are given in this summary along with the MOE calculations. In addition, a direction total with average and maximum values for some MOEs is printed. Finally, a table is provided showing the simulation results for each direction and the totals for both directions.

The detailed output consists of the summary output plus a detailed subsection output with the factors calculated by the program for each subsection. One page is used to show both directions for any road subsection.

Use of the Model

RURAL1 can be used in many ways to evaluate traffic performance on rural roads. A first application of RURAL1 is the study of traffic behavior along a road section with variable geometric characteristics for a given hourly traffic volume. A single run must be made with road subsection specifications that include road and traffic characteristics.

A before-and-after study could be conducted using RURAL1 to evaluate changes in traffic performance resulting from different design strategies. Several design alternatives can be evaluated by comparing them with each other or with the no-action alterna-

tive. The number of simulation runs depends on the number of design alternatives to be studied.

Another possible application of RURAL1 is the evaluation of traffic performance on the same road section over a predetermined time period when traffic demand varies. Two simulation runs must be conducted: one with the road geometrics and the base traffic characteristics as input data, and the other with unchanged road features and the new data on traffic characteristics. Users could find RURAL1 helpful in the prediction of traffic performance on a road section where no actions are taken over fixed time intervals (e.g., 5-year, 10-year intervals). Using the detailed output option, RURAL1 can evaluate several road sections that are not connected. If the model is used this way, performance summary tables should be ignored.

RURAL1 has been derived using the preliminary draft information on capacity and level of service evaluation for freeway, multilane, and two-lane sections; to date, it is the only computerized method for using these draft procedures. One possible application of RURAL1 is the evaluation of the relationships proposed in these capacity procedures. It is important to mention that RURAL1 has been constructed in a modular fashion that makes future changes to these procedures easy to handle.

MODEL APPLICATION

A practical application of RURAL1 to a case study is presented here to evaluate the effects of changes in geometric and traffic characteristics on two-lane roads. Three of the model uses described before are demonstrated in this application: road section evaluation, demand change evaluation, and design change evaluation.

Site Selection

A rural highway section was selected from several candidate locations meeting the selection criteria. Table 1 gives a list of rural highway sections in California that were considered. After further investigation of these sites, including field visits in some cases, a 54-mile-long section on CA-20 in Lake and Colusa counties was selected. Adequate information on both geometric and traffic characteristics is available for this site. Figure 4 shows a map of the study section with major highways that connect to it or are nearby.

Data Collection

RURAL1 requires geometric and traffic data to calculate road section performance. Geometric data were reduced from available road plans between post miles

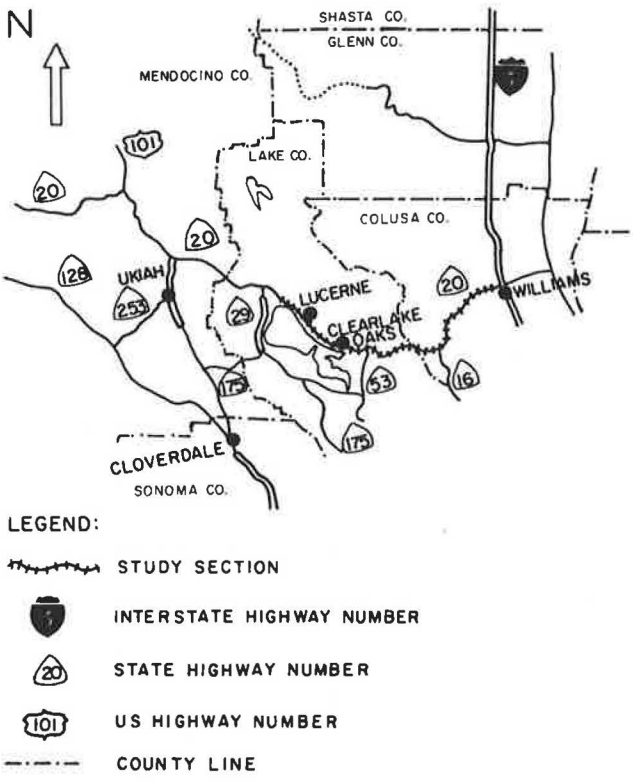


FIGURE 4 Map showing study section.

TABLE 1 Candidate Locations for Initial Model Application

ROUTE	COUNTY	DISTRICT	LENGTH (MILES)	CL	ADT[a]	PERCENT[a] TRUCKS	TERRAIN[b]
20	LAK/COL	01/03	54	YES	8,900	6.8-14.2	R-M
41	MAD	06	46	YES	11,600	6.2-11.7	F-R-M
44	SHA	02	55	YES	2,800	6.6-24.1	F-R-M
70	PLU	02	100	YES	8,300	2.3-19.4	F-R-M
108	TUO	10	59.5	YES	9,300	3.5-6.0	R-M
128	MEN	01	30	NO	4,050	9.0-15.5	R-M
178	KER	09	80.4	YES	5,200	1.7-21.7	M
299	TRI	01/02	72.3	YES	5,800	3.8-22.9	M

a - 1981 FIGURES

b - TERRAIN: F = FLAT
R = ROLLING
M = MOUNTAINOUS

23 and 46 of Lake County. Road plan information was used to determine horizontal curvature, profile data, and lane-width data. No-passing zone data were determined from the state highway photolog. Because no road plans were available for post miles 10 through 23 in Lake County and post miles 0 through 18 in Colusa County, geometric information was approximated using the data given in the Caltrans Route Segment Report (15) for this route. One simplification, the elimination of passing lanes, had to be made to apply RURAL1 to this section; initially RURAL1 could not handle this type of subsection. To illustrate the model application, the basic section was considered as a two-lane section throughout its entire length.

Traffic data were taken from Caltrans 1981 traffic volume and truck volume reports (16,17). Traffic volumes were adjusted to 1983 and 1988 figures using a 4 percent growth factor, and truck data were assumed to be constant for the same period of time. Because of a lack of recreational vehicles and directional distribution percentages, these values were approximated.

Design of Simulation Experiment

The application of the model was designed to demonstrate three possible uses of RURAL1 for a practical case study: the evaluation of a road section under existing traffic conditions, the evaluation of the same section under changes in traffic demand, and the evaluation of the incorporation of design improvements. Table 2 gives the road configuration for these experiments.

Example 1 in Table 2 gives the road configuration for the first application. The 54-mile road section was divided into 31 homogeneous subsections (SS), with criteria for section division being changes in either demand or supply characteristics. Subsection lengths and percent grade are also shown. Traffic volume corresponds to adjusted 1983 values using a 4 percent annual growth rate. The second simulation run was made with the same roadway configuration used in example 1 because no geometric changes were specified. Traffic volumes were adjusted to 1988 values using the 4 percent growth rate.

Finally, example 2 in Table 2 gives the road configuration including some design improvements in the study section. Subsections 12 and 13 were replaced with two undivided multilane subsections, subsections 14 through 16 were replaced with three divided multilane subsections, and subsections 28 through 31 were replaced with four freeway subsections. Traffic volumes used in this simulation run correspond to adjusted 1988 values.

TABLE 2 Road Section Configuration for Simulation Experiment

SS #	LENGTH (MILES)	PERCENT GRADE	EXAMPLE 1 BASIC CASE SS. TYPE	EXAMPLE 2 DESIGN CHANGE SS. TYPE
1	3.00	+1.0	2LANE	2LANE
2	4.00	-1.0	2LANE	2LANE
3	3.00	-1.5	2LANE	2LANE
4	3.00	+2.0	2LANE	2LANE
5	2.00	+0.8	2LANE	2LANE
6	1.00	+1.2	2LANE	2LANE
7	1.00	+2.0	2LANE	2LANE
8	1.10	+0.3	2LANE	2LANE
9	1.75	+1.1	2LANE	2LANE
10	0.48	+6.0	2LANE	2LANE
11	1.20	+2.0	2LANE	2LANE
12	1.20	+7.0	2LANE	4LANE DIVIDED
13	1.00	-5.4	2LANE	4LANE DIVIDED
14	0.40	-2.5	2LANE	4LANE UNDIV.
15	0.90	-6.0	2LANE	4LANE UNDIV.
16	1.70	-0.4	2LANE	4LANE UNDIV.
17	2.40	+1.8	2LANE	2LANE
18	0.60	+4.9	2LANE	2LANE
19	1.00	+2.9	2LANE	2LANE
20	1.00	+5.7	2LANE	2LANE
21	1.00	-2.3	2LANE	2LANE
22	1.00	+5.3	2LANE	2LANE
23	1.00	+4.9	2LANE	2LANE
24	1.10	+2.2	2LANE	2LANE
25	0.40	+5.8	2LANE	2LANE
26	1.30	-5.0	2LANE	2LANE
27	2.00	-4.0	2LANE	2LANE
28	3.00	-3.0	2LANE	4LANE FREEWAY
29	4.00	-2.0	2LANE	4LANE FREEWAY
30	3.60	+2.0	2LANE	4LANE FREEWAY
31	4.00	-1.0	2LANE	4LANE FREEWAY

Simulation Results

Table 3 gives a comparison of the results of the three simulation runs. The first line corresponds to simulation 1, road section evaluation under existing traffic conditions; the second line corresponds to simulation 2, evaluation of a road section with changes in traffic demand for a 5-year period; and the third line corresponds to simulation 3, evaluation of design improvements.

Results shown for each of the simulation runs include average speed, maximum V/C ratio, vehicle travel time, vehicle delay, vehicle-traveled hours and vehicle-traveled miles. Results are presented separately for each direction and totals are given for both directions. Vehicle delay was calculated any time average vehicle speed fell below the user-specified desired speed. Delay was computed as the difference in travel times between these speeds times the amount of traffic traveling the section.

Because the volumes are different, the first two simulation runs can be compared only on the basis of average speed and travel time. There are slight changes in average speed and travel time. Run 1 has a higher average speed and lower travel time than run 2.

Simulation runs 2 and 3 can be compared on the basis of total delay and total traveled hours because traffic volumes are the same. Run 3, which includes the design modifications, shows an improvement in overall average speed of 53 mph compared to 51 mph in run 2. Vehicle delay is reduced by 13.8 veh-hr, a 16 percent change from the delay value calculated in run 2. Total traveled hours are reduced by 28 veh-hr, a 3 percent change from the calculated value in simulation 2. It can be concluded from this analysis that design modifications resulted in a modest improvement in traffic operations for the study section.

PROGRAM LIMITATIONS

Limitations of RURAL1 include

1. RURAL1 can handle up to 100 subsections. The program was designed to simulate at least a 50-mile road section.

TABLE 3 Simulation Results

SIMULATION #	DESCRIPTION	DIRECTION	AVERAGE SPEED (MPH)	MAXIMUM V/C	TRAVEL TIME (HR:MIN)	VEHICLE DELAY (VEH-HRS)	TOTAL TRAVELED (VEH-HRS)	TOTAL TRAVELED (VEH-MILE)
1	ROAD SECTION EVALUATION	DIR. 1	52	0.73	1:3	31.2	432	22000
		DIR. 2	51	0.74	1:3	22.0	321	16431
		BOTH	52	0.74	2:6	53.2	753	38341
2	DEMAND CHANGE EVALUATION	DIR. 1	51	0.84	1:4	49.0	537	26822
		DIR. 2	50	0.84	1:4	34.9	395	19919
		BOTH	51	0.84	2:8	83.9	932	46741
3	DESIGN CHANGE EVALUATION	DIR. 1	53	0.84	1:2	43.3	524	26822
		DIR. 2	53	0.84	1:2	26.8	380	19919
		BOTH	53	0.84	2:4	70.1	904	46741

2. RURAL1 treats each subsection independently without allowing for interactions between adjacent subsections. A new logic is being developed to eliminate this problem.

3. RURAL1 cannot handle oversaturated situations (where demand exceeds capacity). The program has no way of knowing in advance if this situation will arise. If a subsection is indeed oversaturated, the program does not calculate its MOEs and cannot calculate direction summary totals. RURAL1 gives an error message in this case.

4. The program can handle four types of subsections: freeway, multilane, two-lane, and passing-lane. Other subsection types (e.g., intersections, speed-limit zones) cannot be evaluated using RURAL1. An error message is given if other subsections are input to the program.

5. The program cannot determine if input data are correct if data are entered within RURAL1 accepted ranges; however, the program does give error messages for out-of-range data.

6. RURAL1 handles performance calculations only.

7. RURAL1 uses a 1-hr time slice and performance calculations are summarized by hour of operation.

CONCLUSIONS AND RECOMMENDATIONS

RURAL1 is a macroscopic deterministic simulation model that can be used for the evaluation of traffic performance on rural highways. Its very simple logic is based on the capacity and level of service procedures presented in the draft reports for freeway, multilane, and two-lane sections of the new Highway Capacity Manual.

RURAL1 lets users study longer sections of roadway at a lower cost than was possible with earlier microscopic models. However, some detail and precision in the calculations are sacrificed compared with calculations by microscopic models.

It is recommended that further research be conducted to eliminate some of the assumptions and limitations of RURAL1. It is also recommended that field studies be conducted to validate model components.

ACKNOWLEDGMENT

The authors would like to thank Fred Rooney of Caltrans Division of Transportation Operations and Ron Nelson of Caltrans District 3 office in Marysville for their valuable comments and cooperation in data collection and analysis. We also thank Laura Wingerd for her excellent programming work with RURAL1.

REFERENCES

1. W.A. Stock. Simulation of Traffic Flow Phenomena on Two-Lane, Two-Way Highways. Ph.D. dissertation. University of California, Berkeley, 1977.
2. J.L. Botha and A.D. May. A Decision-Making Framework for the Evaluation of Climbing Lanes on Two-Lane, Two-Way Rural Roads. Final Report, FHWA-CA-TO-UC-82. California Department of Transportation, Sacramento, Oct. 1982.
3. J.C. Sananez, J.L. Botha, R.S. Bryant, M. Abbas, and A.D. May. Further Investigation of Traffic Operations on Two-Lane, Two-Way Rural Highways: Research Summary. Research Report UCB-ITS-RR-82-15. University of California, Berkeley, Dec. 1982.
4. C.L. Heimbach, J.W. Horn, S. Khasnabis, and G.C. Chad. A Study of Non-Passing Zone Configurations on Rural Two-Lane Highways in North Carolina. Project ERSD-110-69-3. North Carolina State University Highway Program, Raleigh, March 1973.
5. A. Cassel and M.S. Janoff. A Simulation Model of a Two-Lane Rural Road. In Highway Research Record 257, HRB, National Research Council, Washington, D.C., 1968, pp. 1-16.
6. S. Wu and C.L. Heimbach. Simulation of Highway Traffic on Two-Lane, Two-Way Rural Highways. Presented at the 59th Annual Meeting of the Transportation Research Board, Washington, D.C., 1981.
7. G. Gynnerstedt, A. Carlsson, and B. Westerlund. A Model for the Monte-Carlo Simulation of Traffic Flow along Two-Lane Single-Carriageway Rural Roads. Report 44. Swedish National Road and Traffic Research Institute, Stockholm, 1977.
8. A.D. St. John and D.R. Kobett. Grade Effects on Traffic Flow Stability and Capacity. NCHRP Report 185. TRB, National Research Council, Washington, D.C., 1978, 110 pp.
9. C.J. Messer. Two-Lane, Two-Way Rural Highway Capacity. Draft Procedural Report, NCHRP Project 3-28A. Texas Transportation Institute, Texas A&M University, College Station; TRB, National Research Council, Washington, D.C., Feb. 1983.
10. Highway Capacity Manual 1965. HRB Special Report 87. HRB, National Research Council, Washington, D.C., Jan. 1965, 397 pp.
11. G.K. Robinson. A Model for Simulating Traffic on Rural Roads. Internal Report AIR-290-1. Australian Road Research Board, Numawading, Victoria, 1980.
12. C.J. Hoban. Overtaking Lanes on Two-Lane Rural

Highways. Ph.D. dissertation. Monash University, Clayton, Victoria, Australia, 1980.
13. Polytechnic Institute of New York. New Highway Capacity Manual, Chapter 3: Basic Freeway Segments. Draft Report, NCHRP Project 3-28B. TRB, National Research Council, Washington, D.C., March 1983.
14. Polytechnic Institute of New York. New Highway Capacity Manual, Chapter 7: Multilane Highways. Draft Report, NCHRP Project 3-28B. TRB, National Research Council, April 1983.
15. Route Segment Report, Volume 2: Route Segment Listing. California Department of Transportation, Sacramento, Oct. 1979.
16. 1981 Traffic Volumes on California State Highways. California Department of Transportation, Sacramento, 1981.
17. 1981 Annual Average Daily Truck Traffic on the California State Highway System. California Department of Transportation, Sacramento, 1981.

This paper is based on work conducted as part of a Highway Planning and Research (HPR) project sponsored by the California Department of Transportation (Caltrans) and the FHWA. However, the contents of this paper reflect the views of the authors who are responsible for the facts and accuracy of the data presented. The contents do not necessarily reflect the official views or policies of the sponsors. This paper does not constitute a standard, specification, or regulation.

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

Capacity, Speed, and Platooning Vehicle Equivalents for Two-Lane Rural Highways

MICHEL VAN AERDE and SAM YAGAR

ABSTRACT

Passenger car equivalents (pce's), derived for purposes of capacity, speed, and platooning analyses, are examined using literature sources and traffic data analyzed for 37 different two-lane rural highway sites in Ontario. Speed pce's for trucks and recreational vehicles were found to be considerably higher than those presently used for most types of standard capacity analyses. Truck pce values were found to be 11.4, 6.1, and 3.8 for the 10th, 50th, and 90th percentile speeds, respectively. Corresponding pce values for recreational vehicles were determined as 3.9, 3.7, and 2.6, and the opposing direction pce was found to be 0.5 for all percentiles. The relative effects of trucks and recreational vehicles, in terms of the creation of platoon followers, were found to be much smaller than the corresponding equivalents for speed. Platoon follower pce's for trucks, recreational vehicles, and opposing direction vehicles were 1.23, 1.23, and 0.06 for low traffic volumes and 1.20, 1.07, and 0.07 for high traffic volumes. Platoon leader pce's were 1.55 for recreational vehicles, and 2.0 and 1.35 for trucks on recreational and commuter routes, respectively.

Volumes of different vehicle types and different directions of travel affect the operational characteristics of two-lane two-way rural highways in different ways and to different extents. The analysis of a nonhomogeneous stream of vehicles is therefore often simplified if the relative effect of each vehicle type can be expressed in terms of passenger car equivalent (pce) units. Passenger car equivalents have been quoted for different vehicle types, terrains, levels of service, and rural and urban settings because they can vary with these factors.

Past and current use of vehicle equivalents for trucks, recreational, and opposing direction vehicles on rural two-lane highways on relatively level terrain is reviewed. Specifically, the current practice and literature on pce's are surveyed and these findings are compared with pce values derived from a comprehensive data collection project in Ontario. Because pce's differ for capacity, speed, and platooning analyses, pce values are examined separately for each of these measures. Some of the reasons for these discrepancies are examined.

CURRENT PRACTICE AND LITERATURE ON PCE VALUES

A variety of pce derivations based on capacity analysis are found in the literature, and others pertain to service volumes, speed reduction, or platooning. Separate literature reviews were carried out for each, and the most relevant and significant of these findings are summarized in this section. A comparison of literature estimates and those found in Ontario is provided at the end of the paper.

Capacity-Based Vehicle Equivalents

Vehicle equivalents have most commonly been used for analyses of capacity and level of service. Capacity-

based pce estimates have been made by the Highway Capacity Manual (1), the Organization for Economic Co-operation and Development (OECD) (2), Werner et al. (3), Werner and Morrall (4), Walton and Lee (5), Cunagin and Messer (6), Yagar (7), and Krumins (8).

The Highway Capacity Manual (HCM) (1) assumes that both the level of service and capacity of a two-lane rural highway are directly related to the combined two-way volume, regardless of directional split. This implies a vehicle equivalent of 1 for opposing direction traffic. The HCM also provides estimates of truck and bus equivalents for different grade and level of service considerations. However, an OECD report (2) indicates that several countries, including the United States, Denmark, and the Federal Republic of Germany, have evidence that many of the equivalents given in the HCM may be too high.

Werner et al. (3) determined that, for rolling or mountainous terrain, recreational vehicles had a greater impact on capacity than passenger cars but a smaller impact than trucks. Werner and Morrall (4) also found average passenger car equivalents for trucks, buses, and recreational vehicles for two-lane Alberta highways on level terrain. Walton and Lee (5) explained the reduction in estimates of truck equivalency factors in terms of changes in truck engine performance and typical truck weight-to-power ratios. Cunagin and Messer (6) determined pce's for flat terrain, two-lane rural highways based on a combination of the Walker spacial headway and equivalent delay methods. They found truck equivalents to range from 1.5 to 1.7, for 5 percent trucks, and from 1.5 to 2.0, for 25 percent trucks. The ranges indicate the variability due to differences in volume levels from levels of service A to E.

Because of a universal shortage of data, little quantitative research has been done on vehicle equivalents at ultimate capacity. Two-lane highways are seldom allowed to reach their ultimate two-way capacity, and data are therefore very difficult to obtain.

Curves 1-3 in Figure 1 show the capacity relationships that have been proposed by the HCM (1), Krumins (8), and Yagar (7), respectively. The HCM curve considers a constant opposing direction vehicle equivalent of 1.0, regardless of directional split, and Krumins (8) and Yagar (7) propose much smaller magnitudes of opposing volume impacts. The relationship proposed by Krumins (8) implies an increasing marginal impact as opposing volumes increase, and the Yagar (7) curve suggests that opposing direction pce values decrease in a continuous fashion toward zero as opposing volume increases, indicating that high volumes can be achieved simultaneously in both directions. This is supported by traffic volume data from Australia and Canada, which are plotted in Figure 1. Note that the characteristics of the Canadian highway on which the lower volumes in Figure 1 were recorded were far from ideal, with one-way capacity considerably fewer than 2,000 vehicles per hour. Thus both data points suggest very low capacity pce's for opposing volumes as the directional split approaches 50 percent.

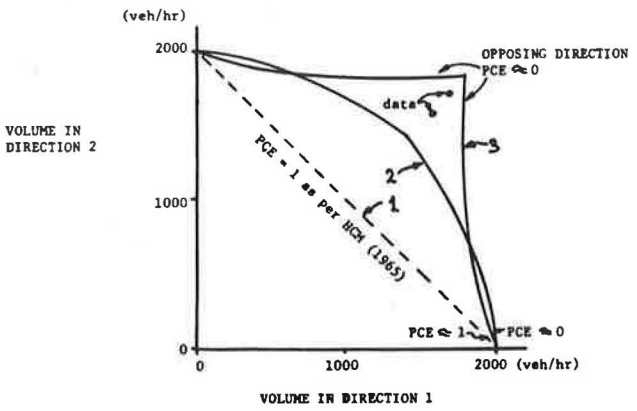


FIGURE 1 Comparison of opposing direction pce's for an ideal two-lane highway.

Speed Reduction-Based Equivalents

Passenger car vehicle equivalents have also been estimated from the relative sizes of the speed reductions caused by equal volumes of each vehicle type. Passenger car equivalent estimates, based on ratios of speed reduction coefficients for different vehicle types and different directions of travel, have been made by Normann (9), Duncan (10), Craus et al. (11), and Krumins (8).

In a study of highway capacity in 1934-1935, Normann (9) found a ratio of opposing direction to main direction speed reduction coefficients of about 0.7. Duncan (10) carried out a similar investigation of speed-volume relationships, in terms of light and heavy vehicles in the main and opposing directions, for 17 two-lane roads. The ratio of the average speed-reduction coefficients indicated a heavy vehicle pce of 15, and an average of coefficient ratios produced a pce ratio of 7.5. The ratio of the opposing-to-main-line effect was about 2:3. In Canada, Krumins (8) used a linear regression model to predict main-line speeds and estimated these speeds to be about seven times more sensitive to main-line volumes than to opposing direction traffic volumes, which differs markedly from the findings of Normann and Duncan.

Craus et al. (11) reviewed current approaches to pce determination and suggested a revised method based on the ratio of delay caused by one truck to the delay caused by one passenger car. They found the main tendencies and fluctuations of pce's, as a function of level of service and truck speed, to be similar to those in the HCM (1).

Because of the difficulty in setting up a controlled experiment on the highway, computer simulation techniques have been used to estimate the effects of different vehicle types and to establish speed pce values. The most significant contributions in this area have been made by Taylor et al. (12), Stock and May (13), St. John (14), and St. John and Kobett (15).

Taylor, Miller, and Ogden (12) suggest, based on simulation techniques, that the proportion of trucks in the flow does not significantly affect speeds for gradients below 3 percent. They also suggest an upper limit for the proportion of trucks in the traffic flow above which the effect of an increasing proportion of trucks is not as great. This critical proportion was found to lie between 0.05 and 0.08, the higher value corresponding to higher gradients.

Using the simulation model SIMTOL, Stock and May (13) also found that the HCM may overestimate the detrimental effect of trucks on steeper grades. Using a microscopic simulation model, A.D. St. John (14) proposed that the truck factor, currently of linear form, should be nonlinear. He reasoned that each incremental addition of slow vehicles to the traffic flow affects the speed less than the former one, because speeds have already been somewhat depressed. St. John and Kobett (15) found that the current form of the truck factor neglects nonlinear effects and inaccurately estimates the effects of heterogeneous truck populations.

Platooning-Related Equivalents

Although platooning and platoon follower equivalents are not quoted specifically in the literature, some vehicle-type effects on platooning can be derived from studies of overtaking or passing behavior.

Troutbeck (16) conducted an extensive study of overtaking behavior on Australian rural highways and found that large vehicles are more difficult to pass than small ones and are therefore more likely to become platoon leaders. He found mean overtaking times to increase with the length of the overtaken vehicle but concluded that further differences in overtaking behavior could not be attributed to either the height or the configuration of heavy vehicles of a given length. Morrall and Werner (17) studied the opposing traffic flow effect on platooning and suggested that not only the magnitude of the opposing volume but also its spacing was critical. They indicated that an opposing hourly volume of only 200 vehicles could prevent all safe overtaking if all vehicles traveled at uniform headways of 18 sec.

Summary of Literature Review

There have been a number of efforts to determine pce values for various conditions, but most of these efforts have focused on recalibrating or slightly modifying the methods outlined in the HCM (1), which in turn go back to Normann (9) and Wardrop (18). This strict adherence to the HCM procedure has resulted in most findings being inadequate in two major areas:

1. Passenger car equivalents have generally been assumed to be similar for capacity, speed, platooning, and other types of analysis. This notion appears to be incorrect and is perhaps one of the main sources of discrepancies among the various pce studies.
2. Capacity, level of service, and pce analyses have been based on combined two-way volume counts. However, recent findings (7,8,19,20) have shown the operational characteristics of two-lane rural highways to be mainly a function of directional volumes, suggesting the need for a re-evaluation of all pce's on a directional basis.

ESTIMATING PCEs FOR TWO-LANE RURAL HIGHWAYS IN ONTARIO

The literature indicated there is no universal pce equivalent that can be used for all purposes. Rather, each analysis should be based on the pce's determined for that particular application. This is illustrated by estimating the respective types of pce's for Ontario's two-lane rural highways. Because there were not enough data available for estimating capacity pce's, the Ontario analyses, which are described, consider only speed and platooning pce's. These are compared with the various values obtained from the literature at the end of the paper. The data used for this purpose, and the details of the analyses, are described.

Data Collection

A total of 267,536 passenger cars, 14,021 trucks, 10,804 recreational vehicles, and 3,035 other vehicles were monitored on two-lane highways in Ontario between July 1 and October 8, 1980, using the radar-platoon technique. At 37 different sites a total of 441 hours of data were obtained for a total of 5,292 5-min time slices of speed-volume information. The data and the data collection sites are described in detail by Van Aerde and Yagar (21).

All data were collected using the radar-platoon technique (22), which records the speed of each platoon traveling in the main direction along with the vehicle types of the platoon leader and all of the followers in the platoon. Given that all vehicle speeds in a platoon are equal, this procedure yields a 100 percent sample of traffic volumes, and corresponding speed and platooning data. Data were collected in this manner in a series of 5-min time slice records, which also contained corresponding 5-min counts of vehicles traveling in the opposing direction.

ESTIMATING VEHICLE EQUIVALENTS IN TERMS OF SPEED REDUCTION

Passenger car equivalents are determined in this section based on the relative rates of speed reduction for each type of vehicle traveling in the main direction and for all vehicles combined traveling in the opposing direction.

Speed-volume relationships were compared for various study sections to establish patterns, special trends, and recurring general shapes. This analysis identified a general speed-volume curve shape consisting of two distinct parts: a linear section, which represents the normal operating conditions, and a nonlinear section, which represents a transition to a breakdown in flow as capacity is approached. This general shape is shown in Figure 2 using data for a typical site, Location 400S1.

Because the nearly linear section of the speed-volume curve represents the entire range of practical operating volumes, further study was focused on it. A linear approximation was found to fit the data at each of the locations that were studied, and was quantified by Van Aerde and Yagar (19) for each of the 10th, 50th, and 90th speed percentiles.

For purposes of analysis, a multiple linear regression model was structured as follows:

$$\begin{aligned}\text{Percentile speed} = \text{Free speed} &+ (C1 \cdot \text{Number of cars}) \\ &+ (C2 \cdot \text{Number of trucks}) \\ &+ (C3 \cdot \text{Number of recreational vehicles}) \\ &+ (C4 \cdot \text{Number of other vehicles}) \\ &+ (C5 \cdot \text{Number of opposing vehicles})\end{aligned}$$

to estimate the free speed and the speed-reduction coefficients C1-C5. Coefficients C1-C5 indicate the relative sizes of speed reductions for each vehicle type (or direction of travel) and permit pce values to be determined as follows:

$$\text{PCE for vehicle type n} = Cn/C1$$

However, the lack of large volume counts for some vehicle types at certain locations renders the estimates of some of these coefficients insignificant or unstable or both. Speed-reduction coefficients were therefore aggregated across all sites to obtain more significant and stable average results.

A simple average over all sites for a speed-reduction coefficient would include several insignificant values, and averaging only the statistically significant values would produce an estimate that is biased in favor of the more extreme positive values. Therefore, the variability in levels of significance

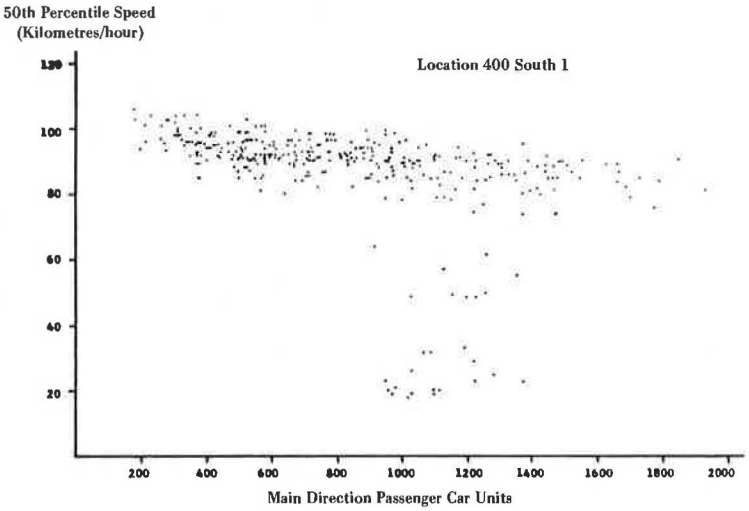


FIGURE 2 Generalized speed-volume relationship.

of the speed-reduction coefficients was incorporated in the analysis through the use of a weighted average, where each speed-reduction coefficient was weighted inversely proportional to its variance. The results of this aggregation are given in Table 1, with a free speed intercept and set of speed-reduction coefficients quoted for each percentile speed. In each case, the average estimate is provided along with an estimate of its standard error. Also, the calculated pce's (Cn/Cl) are quoted for trucks, recreational vehicles, and other vehicles in the direction being monitored and the total number of vehicles in the opposing direction.

Table 2 gives recommended pce values that were derived from Table 1 after accounting for statistical errors. Equivalents for trucks and recreational vehicles were left unchanged, and the equivalent for other vehicles was arbitrarily set to a default value of 1 (i.e., equal to cars) because there were insufficient data to estimate any significant value different from the 1.0 used for cars. Because there was only a slight variation of opposing direction pce's, they were set to 0.5 for all percentiles.

TABLE 1 Estimated Speed-Reduction Coefficients Aggregated Across All Sites

		Percentile Speeds		
		10th	50th	90th
Free speed intercept	Estimate	81.3	90.1	101.7
	Std. error	0.3	0.2	0.3
Speed reduction coefficients				
Main line				
Passenger Car	Estimate	-3.2	-5.2	-8.4
	Std. error	0.4	0.3	0.3
Trucks	Estimate	-36.4	-31.5	-30.7
	Std. error	0.4	2.9	3.2
	Equivalent	11.4	6.1	3.6
Recreational vehicles	Estimate	-12.4	-19.1	-22.1
	Std. error	3.6	2.7	2.9
	Equivalent	3.9	3.7	2.6
Other vehicles	Estimate	-3.6	5.6	7.7
	Std. error	7.5	5.4	5.7
	Equivalent	1.1	-1.1	-0.9
Opposing				
Total count	Estimate	-1.6	-2.4	-3.7
	Std. error	0.8	0.5	0.6
	Equivalent	0.5	0.5	0.4

TABLE 2 Estimated pce's for Speed Reduction Based on Ontario Data

		Percentile Speeds		
Direction	Vehicle Type	10th	50th	90th
Main line	Trucks	11.4	6.1	3.8
	Recreational	3.9	3.7	2.6
	Other	1.0	1.0	1.0
Opposing	All types	0.5	0.5	0.5

ESTIMATING VEHICLE EQUIVALENTS IN TERMS OF PLATOON LEADERS AND FOLLOWERS

In general, platooning is caused by fast vehicles catching up with slower vehicles and not being able to pass. Trucks, recreational vehicles, and buses often have lower desired speeds and poorer acceleration capabilities than standard passenger cars and are therefore more likely to be caught by faster vehicles than are passenger cars. In addition, once caught, the larger vehicles are more difficult to pass because their height, width, or length impairs the follower's sight and necessitates longer passing distances.

Traffic in the opposing direction also influences platooning on two-lane highways because any main-line passing maneuvers require acceptable gaps in the opposing traffic stream. Such acceptable gaps decrease, but at a decreasing rate, as vehicles are added to the opposing traffic stream. Platooning is also dependent on the distribution of the opposing traffic, which in turn depends on with-flow volumes. Platooning is therefore a function of an interaction term of the volumes in the two directions.

The following analysis will examine the impact on the extent of platooning in the main direction of different vehicle-type flows in the main direction and the total opposing direction vehicle flow. Although there can be several measures of the extent of platooning, the impact of each vehicle type is examined here in terms of platoon leaders and platoon followers.

The investigation of platoon leaders estimates the propensity of a vehicle type to become a platoon leader, and the number of followers provides a measure of how many vehicles are held up in platoons by the presence of such a vehicle in the traffic stream.

Platoon Creation

Large vehicles, such as trucks, buses, and recreational vehicles, have a higher individual propensity to become platoon leaders than do passenger cars. These leader propensities are analyzed using the ratio of percent leads, by vehicle type, to percent of total main-line traffic count, by vehicle type. Table 3 gives a summary of these ratios and then lists them normalized with respect to the original ratio for passenger cars to obtain pce's in terms of platoon leadership.

The summary in Table 3 indicates that trucks and recreational vehicles are, respectively, about 1.8 and 1.5 times as likely to be leaders as passenger cars, whereas other vehicles are not significantly different from passenger cars in this regard. The normalized ratios, which represent a form of pce's, were calculated for each location and are plotted separately for trucks and recreational vehicles in Figures 3(a) and 3(b), respectively.

Figure 3(a) shows the truck ratios clustered in two groups. For the upper group, which represents significant locations on commuter routes (7 and 85), trucks are 1.35 times as likely to lead as are passenger cars. For the second group, which represents the recreational routes (400 and 35), the truck ratio is nearly 2.0. This difference may be attributed to differences in drivers on these routes, to a decreasing marginal effect of trucks as the percentage of trucks in the traffic stream increases, or to a larger average size of trucks on the recreational highways, which tend to carry longer trips.

Estimates of generalized recreational vehicle ratios are made only for recreational routes because the percentage of recreational vehicles on commuter roads is too small to provide statistically significant results. Figure 3(b) indicates that, on recreational routes, recreational vehicles are more likely to lead than passenger cars by a factor of approximately 1.55.

TABLE 3 Platoon Leader Ratios by Vehicle Type

Vehicle Type	Ratios	
	Original[a]	Normalized[b]
Total vehicles	1.000	1.056
Passenger cars	0.946	1.000
Trucks	1.716	1.813
Recreational	1.386	1.464
Other	1.023	1.082

[a] Original ratio = percentage of leads by vehicle type divided by percentage of total count by vehicle type.

[b] Normalized ratio = original ratio for vehicle type divided by original ratio for passenger cars.

Follower Creation

Although platoons are identified by their leaders, leaders experience little frustration or reduction in safety as a result of being in a platoon. They are generally not delayed and, except for some pressure from the following vehicles, experience virtually perfect service. It is principally the followers who suffer the largest reduction in service; they may be delayed, frustrated, and even caused to attempt unsafe passing maneuvers.

The increase in the number of followers as a function of main-line traffic volume is shown in Figure 4 for two different locations. Because the relationship in nearly linear, except for a slight curvature at a volume of approximately 650 veh/hr, the number of followers can be modeled using piecewise linear functions with separate linear models fitted for the low- and high-volume regions. This is discussed in detail elsewhere (20,23). Based on a "knee" (change in slope) in the relationship at a volume of approximately 650 veh/hr (in the main-line direction) the number of followers was modeled using separate multiple linear models for traffic volumes between 100 and 650 veh/hr (low-volume range) and 650 to 2,000 veh/hr (high-volume range) as follows:

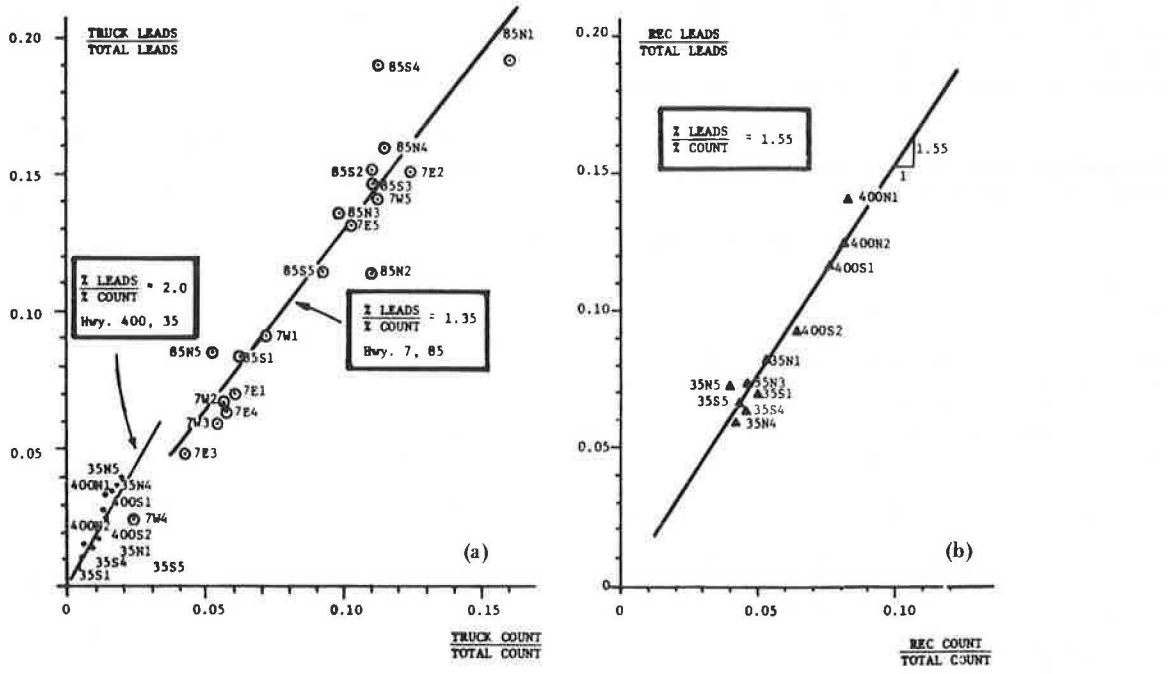


FIGURE 3 Ratios of platoon leaders to traffic volume: (a) trucks on commuter and recreational routes, (b) recreational vehicles on recreational routes.

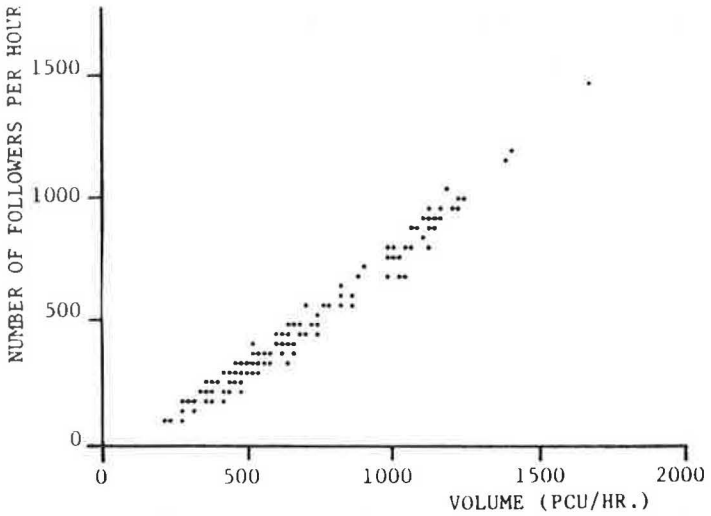


FIGURE 4 Typical relationship between number of followers and main-line volume.

$$\text{Number of followers} = A1 + B1 \cdot \text{Cars} + B2 \cdot \text{Trucks} + B3 \cdot \text{Recreational vehicles} + B4 \cdot \text{Other vehicles} + B5 \cdot \text{Opposing vehicles} \cdot \text{main-line vehicles}$$

In the models, coefficients B1-B5 estimate the rate at which the number of followers increases for each traffic volume component. Main-line vehicle-type coefficient (B1-B4) represent the number of additional followers produced per vehicle, and the opposing volume coefficient (B5) is quoted as the number of follower produced per opposing vehicle at a main-line volume of 1,000 veh/hr. The interaction term for opposing direction traffic volumes makes any opposing volume effect proportional to the total main-line volume. Therefore, the follower production due to traffic in the opposing direction is larger or smaller than the quoted coefficient when the main-line traffic volumes are larger or smaller, respectively, than 1,000 veh/hr.

Table 4 gives average vehicle-type follower coefficients and normalized follower production rates (i.e., quoted with respect to the car rate of 1.0) for the high- and low-volume region. Because the relative size of these rates represents the relative effect of different vehicles on platooning, the normalized ratios can be considered estimates of platooning pce's. For the low-volume region, pce's were calibrated using data from all two-lane highway sites, and those for the high-volume region were calibrated based on data from the mainly recreational routes (most of Ontario's commuter roads are upgraded before very high traffic volumes are reached).

TABLE 4 Passenger Car Equivalents in Terms of Number of Followers Produced

		Low-Volume Range		High-Volume Range	
Direction	Parameter	Avg Value	pce	Avg Value	pce
Main line	A Intercept	-81.1		-246.8	
	B1 Car	0.74	1.00	1.01	1.00
	B2 Truck	0.91	1.23	1.21	1.20
	B3 Recreational	0.91	1.23	1.08	1.07
	B4 Other	0.72	0.98	1.14	1.13
Opposing	B5 Total count	0.042	0.06	0.066	0.07

Note: Low-volume range, 100-650 veh/hr in main-line direction; high-volume range, 650-2,000 veh/hr in main-line direction.

In the low-volume range, trucks or recreational vehicles produce about 0.91 followers compared with cars, which produce only 0.74. This results in a pce of 1.23 (i.e., 0.91/0.74) and indicates that trucks and recreational vehicles produce, on the average, approximately 23 percent more followers than do passenger cars under the same conditions.

Follower production rates were higher in the high-volume region for all vehicle types and both directions of travel. The estimated truck pce decreased slightly, and standard error increased significantly, due mainly to the much smaller representation of trucks on the high-volume recreational routes. The recreational vehicle pce decreased from 1.23 to 1.07, but its standard error remained virtually the same. The smaller pce might be due to the increased representation of recreational vehicles on high-volume recreational routes, whereby they become an integral part of the traffic stream and have a significantly reduced marginal impact. Other vehicles also appear to have a larger pce in the high-volume region, although their estimate is based on fewer data than those for cars, trucks, recreational vehicles, or opposing flows.

The opposing direction coefficients indicate that opposing volume affects main-line platooning for all ranges of main-line traffic by merging a number of smaller platoons into fewer, but larger, platoons. Specifically, the presence of 1,000 opposing direction vehicles is estimated to increase the number of followers by approximately 35, 70, and 135 veh/hr for main-line traffic volumes of 500, 1,000, and 1,500, respectively.

COMPARISON OF ONTARIO'S RESULTS WITH PCEs FROM THE LITERATURE

The findings of this study for speeds were found to agree with some sources but were drastically different from others. Findings from the literature for capacity and speed analyses are given in Table 5, and Table 6 gives percentile speed, platoon follower, and platoon leader pce's estimated for Ontario's two-lane rural highways. Note that the speed percentiles in Table 6 do not correspond with levels

TABLE 5 Average Generalized pce's Found in the Literature

pce and Vehicle Type	Levels of Service A	B	C	D	E	Avg	Source
Capacity							
Truck	3.0	2.5	2.5	2.0	2.0		HCM *(1)*
	2.0	2.2	2.2	2.0	2.0	2.1	Werner and Morrall *(4)*
	1.5	1.6	1.6	1.6	1.7	1.6	Cunagin and Messer *(6)*
Bus	2.0	2.0	2.0	2.0	2.0	2.0	HCM *(1)*
	1.8	2.0	2.0	1.6	1.6	1.8	Werner and Morrall *(4)*
Recreational vehicle	2.2	2.5	2.5	1.6	1.6	2.1	Werner and Morrall *(4)*
Opposing vehicles						1.0	HCM *(1)*
	1±				0+		Yagar *(7)*
Speed							
Truck						15.0	Duncan *(10)*
Opposing vehicles						1.00	HCM *(1)*
						0.7	Normann *(9)*
						0.66	Duncan *(10)*
						0.14	Krumins *(8)*

TABLE 6 Average Generalized pce's for Ontario's Two-Lane Highways

	Speed Equivalents (speed percentile)			Platooning Equivalents			
				Followers		Leaders	
Vehicle Type	10th	50th	90th	Low Vol.	High Vol.	Recreation	Commuter
Truck	11.4	6.1	3.8	1.23	1.20	2.00	1.35
Recreational vehicles	3.9	3.7	2.6	1.23	1.07	1.55	
Other	1.0	1.0	1.0	0.98	1.13	1.464	
Opposing vehicles	0.5	0.5	0.5	0.06	0.07		

Note: Low volume = main direction traffic volumes of 100-650 veh/hr; high volume = main direction traffic of 650-2,000 veh/hr.

of service, as drivers choose their own percentiles within the available level of service.

Speed Prediction Equivalents

Speed prediction vehicle equivalents for trucks and recreational vehicles traveling in the main-line direction were found to decrease for higher speed percentiles. The absolute magnitude of the opposing volume effect was larger for higher speed percentiles, but corresponding increases in the main-line speed reduction coefficients resulted in a constant opposing direction vehicle equivalent.

In Ontario, trucks were found to have a speed-reduction coefficient of approximately 30 km/hr per 1,000 trucks/hr. This compares with Duncan (10), who determined an average speed reduction due to heavy vehicles of 62 km/hr, assuming a composition of 15 percent heavy vehicles. The Ontario truck equivalents of 11.4, 6.1, and 3.8 for the 10th, 50th, and 90th percentile speeds, respectively, are much higher than truck pce's (Table 6) but lower than Duncan's (10). He found that the ratio of average speed-reduction coefficients indicated a heavy vehicle pce of 15, whereas an average of coefficient ratios produced a pce ratio of 7.5.

Ontario speed pce's for recreational vehicles ranged from 3.9 to 2.6, compared with truck pce's that ranged from 11.4 to 3.8. Werner, Morrall, and Halls (3) determined that the effect of recreational vehicles on service volumes was more than that of standard passenger cars but less than that of trucks.

The speed pce for vehicles traveling in the opposite direction was consistently found to be 0.5 for all speed percentiles. This compares with Normann (9) and Duncan (10), who estimated this effect as 0.7 and 2/3 of main line, respectively. In contrast, the HCM (1) uses a default of 1.0, and Krumins (8) found an opposing volume equivalent of 1/7.

Platooning Equivalents

Platooning pce's were determined in terms of both platoon leadership and follower creation. However, comparisons with other sources are difficult because there are no other quantitative estimates of vehicle-type impacts on platooning in the literature. Therefore, platooning pce's are compared with estimates for speed and capacity.

Main-line platoon follower pce estimates are generally lower than comparable estimates for speed and capacity. As the data in Tables 5 and 6 indicate, capacity pce's for trucks generally range from about 1.5 to 3.0, and speed pce's can vary from 3.8 to 15.0. These values are generally larger than the Ontario platoon leadership values of 2.00 for recreational roads and 1.35 for commuter roads or the follower creation values between 1.20 and 1.23. Similarly, recreational vehicle pce's range from 1.6 to 2.5 for capacity and from 2.6 to 3.9 for speed, but they were only 1.55 for platoon leadership and between 1.07 and 1.23 for follower creation.

Follower creation pce estimates for traffic in the opposing direction were 0.06 and 0.07 for the low- and high-volume ranges, respectively. Although these numbers are small, they represent additional followers without adding main direction vehicles (i.e., opposing traffic causes platoons to merge).

The effect of opposing direction platooning on main-line platooning was not examined, although the extent of platooning in the opposing direction affects main-line platooning. The marginal effect that is not explained by the opposing volume was not considered to be commensurate with the effort required to examine it.

CONCLUSIONS

The relative effect of trucks on the lower percentile drivers is particularly pronounced. The truck

pce ranged from 3.8 for 90th percentile speed to 11.4 for 10th percentile speed in Ontario. The recreational vehicle pce ranged from 2.6 to 3.9. The speed pce for vehicles traveling in the opposite direction was consistently 0.5 for all speed percentiles.

Passenger car equivalent values for platooning are lower than corresponding equivalents for speed. Follower creation pce's for trucks and recreational vehicles averaged 1.22 and 1.15, respectively, in Ontario, and corresponding platoon leader pce's averaged 1.8 and 1.4, respectively. The follower creation pce of traffic in the opposite direction averaged 0.065 for Ontario.

Although the various types of vehicle equivalents are related, pce's are not interchangeable, and only the appropriate pce's should be used in any application.

Discussion

Myung-Soon Chang*

Van Aerde and Yagar have made a good hypothesis that passenger car equivalents may be different for various criteria such as capacity, speed, platooning, and directional volume on two-lane rural highways. However, their data collection and analysis overlooked important considerations that should be recognized.

DATA COLLECTION

Van Aerde and Yagar collected platoon data using the radar-platoon technique (22). However, this method is very sensitive to the definition of what headway separation constitutes a different platoon. The first criterion in studying platooning would be the discrimination of platoon formation, and this varies among researchers. Miller, for example, separated platoons if the headway was longer than 8 sec (24). Edie et al. defined a vehicle as platooning if its headway was less than 4 to 5 sec, depending on its speed (25). Keller considered a vehicle as platooning if its headway was less than 2 sec (26).

If platooning involves a headway criterion, the radar-platoon technique Van Aerde and Yagar used is not sufficient and consequently needs more sophisticated data collection procedures and equipment. It should be recognized that their data collection method is subjective, depending on the perception of the observer, although it has an advantage of not requiring sophisticated equipment.

ANALYSIS

The authors used a multiple linear regression model to define pce values for different vehicle types. They defined pce for vehicle type n as the coefficient of vehicle type n (C_n) over coefficient of passenger car (C). However, this approach is only valid when no intercorrelation between the pair of independent variables (i.e., pair of vehicle types) exists. The coefficients given in Table 1 make it evident that there is intercorrelation or multicollinearity (i.e., high correlation between independent variables) between independent variables. Some of the algebraic signs are reversed, resulting in the negative value of pce's.

When there is multicollinearity between independent variables, not only will the sign often be reversed but the coefficients are also changed by the introduction and deletion of one or the other variable (27,28). When independent variables are correlated, the regression coefficient of any independent variable depends on which other independent variables are included in the model. Thus, a regression coefficient does not reflect any inherent effect of the particular independent variable on the dependent variable but has only a marginal or partial effect, given whatever other correlated independent variables are included in the model.

In short, the estimates of individual regression coefficients are very unreliable (28) when independent variables are intercorrelated among themselves.

CONCLUSIONS

Although the authors explored a potential use of regression models for deriving pce values, the suggested values should be modified using the approaches (28) that can eliminate intercorrelation or multicollinearity between independent variables.

*Texas Transportation Institute, Texas A&M University, College Station, Texas 77843

Authors' Closure

First we would like to thank Myung-Soon Chang for providing us with the opportunity to discuss some of the theoretical problems involved in applying multiple linear regression, which we had adopted as the only real choice for analyses of this type. This selection was based on an exhaustive consideration of the various statistical techniques for analyses of this nature. Chang suggested that we had overlooked these theoretical issues in the paper, but we had thought that a discussion of such theoretical issues was really peripheral to the central theme of our work. Theoretical issues are treated in the appropriate statistical literature. However, because the issues have been raised they are addressed. Because Chang's discussion has addressed data collection and analysis separately, we shall respond in these corresponding categories, so that the reader can relate our responses to Chang's respective questions.

First, we shall provide the reasons for selecting the manual radar-platoon data collection technique used for identifying and quantifying platoons. We will then address Chang's questions and concerns regarding the effect of possible multicollinearity on our multiple linear regression analysis.

DATA COLLECTION

Because the entire radar-platoon data collection procedure is presented in detail in the earlier literature, we have limited our response to specifically addressing Chang's concerns regarding our selection of a platoon discrimination procedure.

An exact quantitative definition of a platoon would involve a complex and even stochastic combina-

tion of relative positions, speeds, and accelerations of a potentially large number of vehicles, rather than simply the relative positions and speeds of only two vehicles. Therefore we thought it impractical to even attempt to fit a function for our purpose. Those who have attempted to automate this process have generally had to simplify their definition of a platoon to consider only successive pairs of vehicles. Also, platoon characteristics will vary among different highway types and situations. It is therefore not surprising that Miller (24), Edie et al. (25), Keller (26), and others have found different critical headways for defining limits for platoons under their very different sets of conditions, which Chang may have unfortunately confused as representive of the same process. For example, Keller's extreme headway of less than 2 sec was not intended as a critical headway for two-lane rural highways, which our paper addresses exclusively. We had measured an average headway for platooned vehicles on a two-lane rural highway of 2.0 sec, which is even greater than the low critical value that Chang attributes to Keller. This average value is not to be confused with a follower discrimination criterion as Chang has apparently done in Keller's case.

With these considerations in mind, we thought that the traffic could best be divided into platoons by a trained observer, who would have the benefit of observing long traffic streams as they approached. Being concerned about the subjectivity of the process, we had a number of people observe the same traffic: a physicist, a politician, a student, and a traffic engineer. There was consistently virtually unanimous agreement about where the platoons started and ended. This indicated that the error due to subjective selection of platoons referred to by Chang was less than that due to differences among researchers. Our results are not dependent on any mathematical descriptions of platoons that different researchers might speculate on or even find. Because we were interested only in sorting the traffic into platoons and not in finding a model for describing this process, the human observer method served our needs better than any mathematical model we could find or create.

ANALYSIS

Criticism of our statistical analysis centered around multicollinearity and reversed algebraic signs. We are well aware of the problems of multicollinearity, but cannot apologize for our use of multilinear regression (MLR). After careful consideration of the nature of our data and the available analysis techniques we believed that we could live with the implications of MLR. We are satisfied with our "ball park" estimates, and feel that they have supported our main hypothesis that pce's are different for capacity, speed, and platooning, respectively. We have also quoted levels of statistical significance, and only draw conclusions concerning the types of vehicles that were represented in sufficient sample size to allow meaningful statistical inference. Chang suggested that he can alter our findings by eliminating multicollinearity. However, a closer examination of our paper would show that this has already been done, as we shall discuss. It is also noted that the models and data used for estimating respective pce's for speed, capacity, and platooning were consistent. This, along with the statistical significance obtained in estimating these different types of pce's, indicates that they are reasonably accurate, subject only to their statistical significance, represented by the ratio of their relative magnitudes and errors of estimation given in Table 1. In each case, a ratio of about 2 or more represents reasonable statistical significance.

Chang referred to Mullet (27) who discusses the causes of coefficients having the wrong sign, quoting the following four reasons:

1. Range of independent variable not fully covered by data,
2. Exclusion or omission of important predictor variables,
3. Multicollinearity between independent variables, and
4. Computational error.

We respond to each of these as follows:

1. It should be noted that the entire range of traffic volumes, up to and including capacity, was covered by our data. The mix of locations included recreational, commuter, and combined highway types, providing mixtures that included both high and low volumes of trucks in combination with a similar range of recreational vehicle volumes. This therefore includes virtually the entire possible range of traffic condition variables that were of interest to our study.
2. The MLR analysis drew on all the component traffic volumes present in the traffic stream and was therefore comprehensive without being overspecified. Because the counts for all vehicles were included, no important predictor variables were excluded and no variable types were redundant.
3. Issues of multicollinearity are treated specifically by Mason et al. (28), and are therefore addressed in detail hereafter.
4. Finally, computational error in the estimate of our regression coefficients was minimized through the use of the SAS statistical analysis program. SAS is one of the most widely applied and therefore tested mainframe statistical analysis packages. Its results were verified using an independent regression package.

Chang referred to Mason et al. (28) who discuss issues of multicollinearity in great detail and suggest methods to be used to eliminate or to allow for its presence. Specifically, they suggest three solutions:

1. To reduce or eliminate collinearity, they recommend that the data be augmented with additional data. Having monitored nearly half a million vehicles, we believe that it would be impractical to increase the size of the sample, which is, if anything, larger than necessary.
2. As an alternative they suggest that the regression should be attempted subject to restrictions on some of the independent variables. Such restrictions were explored and are documented by Van Aerde and Yagar (21) as Model II. Briefly, these restricted regressions fixed the relative sizes of the vehicle type speed reduction coefficients such that the regression was forced to focus on the relative sizes of the main-line and opposing direction speed effects. The resulting reduction in degrees of freedom provides more stable and often more reliable results.
3. The final recommendation involves selection of variables. They suggest that no important predictor variable should be omitted and that no two variables should be included if they represent virtually the same thing. This issue is similar to item 2 from Mullet and the same arguments apply here.

These discussions should satisfy any skeptic that we have not only carefully avoided multicollinearity

problems to the extent possible, but also have followed the appropriate remedial measures. It is therefore not necessary to modify our values as suggested by Chang.

It should be noted that in Table 1 our average regression coefficients were quoted with corresponding measures of statistical significance. All significant coefficients, which usually correspond to vehicle types with a large count, are of the correct sign and order of magnitude. Only the coefficients that had a larger variance associated with them ended up with a wrong sign or an incorrect magnitude, and in this case we explicitly stated that we do not recommend their use.

CONCLUSIONS

We believe that the data bank that we used for our analyses was more than adequate. Although we could easily have lived with a smaller sample, a sample of this size was made possible by the use of the efficient radar-platoon data collection technique. This technique is especially useful in providing data for analysis of platoons, because it provides data in terms of platoons.

Any automated platoon discrimination technique will have to be calibrated with the use of human observers, who were used in the radar-platoon technique. It is therefore not possible to develop an automated technique that can provide better platoon discrimination than competent, trained human observers.

Multiple linear regression analysis was found to be the best analysis approach because it permitted a similar model structure for the various platoon and speed measures that were explored. We have taken the necessary precautions to avoid collinearities or minimize their effects, and are confident about the reliability of our results.

REFERENCES

1. Highway Capacity Manual 1965. HRB Special Report 87. HRB, National Research Council, Washington, D.C., 1965, 397 pp.
2. Two-Lane Rural Roads: Design and Traffic. Organization for Economic Co-operation and Development, Paris, France, 1972.
3. A. Werner, J.F. Morrall, and G. Halls. Effect of Recreational Vehicles on Highway Capacity. Traffic Engineering, May 1975, pp. 20-25.
4. A. Werner and J.F. Morrall. Passenger Car Equivalencies of Trucks, Buses, and Recreational Vehicles for Two-Lane Rural Highways. In Transportation Research Record 615, TRB, National Research Council, Washington, D.C., 1976, pp. 10-17.
5. C.M. Walton and C.E. Lee. Characteristics of Trucks on Grades. In Transportation Research Record 631, TRB, National Research Council, Washington, D.C., 1977, pp. 23-30.
6. W.D. Cunagin and C.J. Messer. Passenger Car Equivalents for Rural Highways. Texas A&M University, College Station, 1983.
7. S. Yagar. Capacities for Two-Lane Highways. Report 13(1). Australian Road Research Board, Numawading, Victoria, March 1983, pp. 3-9.
8. I.V. Krumins. Highway Capacity and Level of Service. M.S. thesis. University of Calgary, Calgary, Alberta, Canada, 1981.
9. O.K. Normann. Highway Capacity. In Proc., HRB, Vol. 21, HRB, National Research Council, Washington, D.C., 1941, pp. 379-392.
10. N.C. Duncan. Rural Speed/Flow Relations. TRRL Laboratory Report 651. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1974.
11. J. Craus, A. Polus, and I. Grinberg. A Revised Method for the Determination of Passenger Car Equivalencies. Transportation Research: Part A, General, Vol. 14A, Pergamon, New York, 1980, pp. 241-246.
12. M.A.P. Taylor, A.J. Miller, and K.W. Ogden. Aspects of Traffic Flow on Grades. ARRB Proc., Vol. 6, Part 3, Australian Road Research Board, Numawading, Victoria, 1972, pp. 232-248.
13. W.A. Stock and A.D. May. Capacity Evaluation of Two-Lane Two-Way Highways by Simulation Modeling. In Transportation Research Record 615, TRB, National Research Council, Washington, D.C., 1976, pp. 20-27.
14. A.D. St. John. Nonlinear Truck Factor for Two-Lane Highways. In Transportation Research Record 615, TRB, National Research Council, Washington, D.C., 1976, pp. 49-53.
15. A.D. St. John and D.R. Kobett. Grade Effects on Traffic Flow Stability and Capacity. NCHRP Report 185. TRB, National Research Council, Washington, D.C., 1978.
16. R.J. Troutbeck. Analysis of Free Speeds. ARRB Proc., Vol. 8, Australian Road Research Board, Numawading, Victoria, 1976.
17. J.F. Morrall and A. Werner. A Measurement of Level of Service for Two-Lane Rural Highways. Canadian Journal of Civil Engineering, Vol. 9, No. 3, 1982, pp. 385-398.
18. J.G. Wardrop. Some Theoretical Aspects of Road Traffic Research. Proceedings Part II: Research and Theory, Institution of Civil Engineers, London, England, 1952.
19. M. Van Aerde and S. Yagar. Volume Effects on Speeds of 2-Lane Highways in Ontario. Transportation Research, Vol. 17A, 1983.
20. S. Yagar and M. Van Aerde. Platooning Relationships on 2-Lane Rural Highways. Submitted for publication, Jan. 1984.
21. M. Van Aerde and S. Yagar. Efficient Provision of a Large Data Bank for Speeds on 2-Lane Highways. Presented at Annual Conference of Road and Transportation Association of Canada, Halifax, Nova Scotia, Canada, Sept. 1982.
22. S. Yagar and M. Van Aerde. Radar-Platoon Technique for Efficient and Complete Speed Measurements. In Transportation Research Record 841, TRB, National Research Council, Washington, D.C., 1982, pp. 36-41.
23. M. Van Aerde. Operational Characteristics of 2-Lane 2-Way Rural Highways. M.S. thesis. University of Waterloo, Waterloo, Ontario, Canada, 1983.
24. A.J. Miller. A Queueing Model for Road Traffic. Journal of the Royal Statistics Society, Vol. B23, 1961, pp. 64-75.
25. L.C. Edie, R.S. Foote, R. Herman, and R. Rothery. Analysis of Single Lane Traffic Flow. Traffic Engineering, Jan. 1963, pp. 21-27.
26. H. Keller. Effects of a General Speed Limit on Platoons of Vehicles. Traffic Engineering and Control, Vol. 17, No. 7, July 1976, pp. 300-303.
27. G.M. Mullet. Why Regression Coefficients Have the Wrong Sign. Journal of Quality Technology, July 1976, pp. 121-126.
28. R.L. Mason, R.F. Gunst, and J.T. Weber. Regression Analysis and Problems of Multicollinearity. Communications in Statistics, Vol. 4, No. 3, 1975, pp. 277-292.

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

Dynamic Freeway Simulation Program for Personal Computers

PANOS G. MICHALOPOULOS

ABSTRACT

An interactive menu-driven macroscopic freeway simulation program with graphic capabilities is presented. The program is written in UCSD-Pascal language and runs on IBM personal computers. Recently developed flow models that describe complex phenomena such as lane changing, merging, and weaving are employed. The computational effort is minimized by using fast and efficient numerical methods for implementing these models. Input to the program is conventional traffic parameters, freeway and ramp characteristics (e.g., capacity, free-flow speed, jam density), demands, and origin-destination information. Output includes dynamic description of speed, flow, and density (both numerical and graphic); estimation of the most common measures of effectiveness; and graphic representation of flow conditions and congestion levels.

Improving the operating conditions of freeway flow during periods of congestion is one of the major problems in traffic engineering practice. Before implementation of an improvement, estimation of the effects of the various alternatives and comparison with the existing conditions are desirable. This usually entails determination of the measures of effectiveness (such as total travel, total travel time, delays, stops, energy consumption, and so forth) associated with a given situation, which may include a traffic management scheme. Often, a more detailed analysis is needed, which may require dynamic description of flow (i.e., in time and space) as well as the formation and dissipation of congestion. Such analysis could assist in comparing alternative geometric configurations; estimating the effects of incidents; and, in general, assessing the impacts of improvements, control strategies, and system changes. Finally, it may be necessary to determine only whether a given facility and management scheme combination can accommodate the demand in a satisfactory fashion.

Despite recent theoretical developments, simple tools for analyzing situations such as these are still lacking. Because analytical methods for describing the freeway flow process in sufficient detail and accuracy are impractical, field engineers often must turn to simulation to obtain answers to the previously mentioned problems. A number of freeway simulation programs are available for design and analysis purposes; among the most widely known programs are FREQ6PE (1), FREQ7PE (2), and INTRAS (3). The first two are very similar and are macroscopic in nature; the latter is microscopic. At present, existing program packages can run only on large computers that are not always easily accessible; furthermore, using a large computer usually implies familiarity with its operating system for compiling the programs for tape, disk, input-output (I/O) operations, and so forth. Finally, existing programs are cumbersome and data hungry, and they require reasonable familiarity with the program.

Difficulties of this nature make existing packages unattractive to potential users. Previous experience (4) suggests that the use of large-scale freeway simulation packages in at least one agency was essentially abandoned shortly after an initial trial period. Some of the difficulties related to the use of large computers can be resolved by recent advances in microcomputer technology. The low cost and anticipated widespread use of personal computers combined with the opportunity of employing less sophisticated menu-driven, user-friendly interactive programming make development of microcomputer software increasingly attractive and desirable.

In this paper an interactive, user-friendly macroscopic simulation program named KRONOS-1 is presented; the program is written in UCSD-Pascal language and runs on IBM personal computers. Because of the limitations of microcomputers, the models employed had to require minimal storage and computational effort and be fairly realistic and reasonably accurate. The last two requirements were satisfied by taking compressibility into account, as well as acceleration and deceleration characteristics of a traffic mass. Storage and computational effort were reduced by developing simple finite-difference schemes for solving the governing equations of the system. To further improve realism, merging or diverging areas are not treated as dimensionless points (as in most of the existing programs), but they have a finite length. The generation (or dissipation) of flow in these areas is a function of the relative speed between the freeway and the acceleration (or deceleration) lane, as well as the ramp demands (or exiting volumes) and freeway flow conditions in the merging (or diverging) area.

Perhaps the major advantage of the program presented here is that it allows treatment of phenomena not previously described by existing macroscopic programs. Such phenomena include lane changing, merging or diverging, and weaving. This was made possible through earlier extensive model development and experimentation. Testing and validation were performed by comparing the results of the program with a data base generated by microscopic simulation using the INTRAS program (3). This was necessitated by the need to test a wide range of ramp and freeway demand combinations as well as a number of alternative geometric configurations. Stated otherwise, cost and time considerations dictated experimentation under a controlled environment that was ensured by a previously tested and validated microscopic simulation program. It should be noted, however, that more extensive testing against field data, as well as program extensions, is under way.

MODELING AND ANALYSIS METHODOLOGY

Because of the limitations of personal computers, bookkeeping and storage requirements, as well as computational effort, should be minimized. This led

to the selection of macroscopic flow models for the KRONOS-1 simulation program. Previous experience (4,5) suggests that macroscopic simulation can lead to satisfactory accuracy at least for some of the problems described earlier. Existing macroscopic flow models fall into three general categories: (a) input-output (I/O), (b) simple continuum, and (c) high-order continuum. The models of the first category are rather simplistic in that they do not include space explicitly nor do they take compressibility into account. High-order continuum models, on the other hand, are the most sophisticated but they have not as yet gained wide popularity or proved truly superior to the simple-continuum alternative. In a recent study (6) these models performed about the same as the simple-continuum alternative. For this reason, the main model employed by KRONOS-1 is the simple-continuum one, presented here, which is based on the conservation equation. It should be noted, however, that the program also allows use of the most widely known high-order model (7).

According to the simple-continuum model (8), freeway flow can be described by the conservation equation that has the general form:

$$(\partial q/\partial x) + (\partial k/\partial t) = g\ (x,t) \tag{1}$$

where

$q = q(k) = ku$ is the flow rate of the traffic stream;
$k = K(x,t)$ and $u = u(k)$ are the density and speed, respectively;
t and x = time and space; and
g = the generation rate.

The latter represents generation or dissipation of cars at entrance or exit ramps, respectively. Therefore, g as well as the basic flow variables (k, q, and u) vary with time and space. Naturally, in freeway sections where there are no entrances and exits $g(x,t) = 0$. The continuum model assumes that flow is a function of density; this implies that Equation 1 is a nonlinear (i.e., hyperbolic) partial differential equation having density as the only unknown. It follows that this equation can be solved for a particular freeway to obtain the value of density (and therefore flow and speed) at each time and space point of the t-x domain. Analytical solution of this problem (i.e., estimation of k, q, and u in time and space) is only possible for continuous single-regime q-k relationships, simple arrival and departure patterns at the boundaries of the freeway in question (boundary conditions), zero generation terms, and simple flow patterns along the road at the beginning of simulation (initial conditions).

To circumvent the mathematical complexities of analytical solutions, to improve realism (by relaxing simplifying assumptions), and to allow further practical extensions, numerical treatment of Equation 1 (governing equation) was sought. The numerical solution of Equation 1 begins by discretization in both time and space. Figure 1 shows space discretization of a freeway section that consists of J segments of length Δx. It can be easily verified that, to keep the solution within reasonable bounds, the time and space increments Δt, Δx must obey the rule:

$$(\Delta x/\Delta t) > u_f$$

where u_f represents the free flow speed. Numerical solution of the conservation equation requires knowledge of the arrival and departure patterns (boundary conditions) at each end of the freeway section under consideration (nodes 1 and J+1 in Figure 1). When numerical methods are employed, no specific assumptions need to be made concerning these patterns (i.e., either deterministic or statistical arrival and departure distributions can be assumed).

Dynamic estimation of density, flow, and speed (i.e., estimation of these values on every node j at each Δt increment) is obtained from (6):

$$k_j^{n+1} = [(1/2)(k_{j+1}^n + k_{j-1}^n)] - [(\Delta t/2\Delta x)(G_{j+1}^n - G_{j-1}^n)] + (\Delta t/2)(g_{j+1}^n + g_{j-1}^n) \qquad \forall_j \tag{2}$$

$$u_j^{n+1} = u_e\ (k_j^{n+1}) \tag{3}$$

$$q_j^{n+1} = k_j^{n+1}\ u_j^{n+1} \tag{4}$$

where

k_j^n, u_j^n, q_j^n = density, speed, or flow, respectivity, of node j at $t = t_o + n\Delta t$,
t_o = initial time,
$u_e(k_j^{n+1})$ = equilibrium speed corresponding to the value of density k_j^{n+1},
$G_j^n = K_j^n u_j^n$, and
$g_j^n = g(x_j, t_n)$ = generation rate of node j at $t = t_o + n\Delta t$.

When the simplest equilibrium speed density model is assumed (9),

$$u_e\ (k_j^{n+1}) = u_f\ [1 - (k_j^{n+1}/k_o)] \tag{5}$$

and

$$G_j^n = k_j u_f\ [1 - (k_j^n/k_o)] \tag{6}$$

where k_o and u_f represent jam density and free flow speed, respectively.

This numerical solution allows use of any speed-density model, including discontinuous ones; in such case, Equations 5 and 6 can be altered accordingly. It should be noted that the generation rate g is either given or it can be derived dynamically from ramp demands and freeway flow and density at the previous time step. In the simplest case, where the entire ramp equals the segment length Δx, an average value of g can be assumed; however, as the ramp length increases or as Δx decreases, this assumption is unrealistic. In such a case, solution proceeds by considering the conservation equation of the acceleration lane separately and solving it simultaneously with the conservation equation of the freeway (10). This new equation must take into account lane changing effects that are described next.

The modeling presented to this point does not consider lane changing effects (i.e., all lanes are aggregated). Lane changing can be described in two ways (10); the first, and simplest, is discrete in the sense that it considers each lane separately. The second is continuous (i.e., it explicitly includes street width). Because of space limitations only the first option is described here.

A simple-continuum model for describing flow along two or more homodirectional lanes can be obtained by considering the conservation equation of each lane. This is accomplished by observing that the exchange of flow between lanes represents gener-

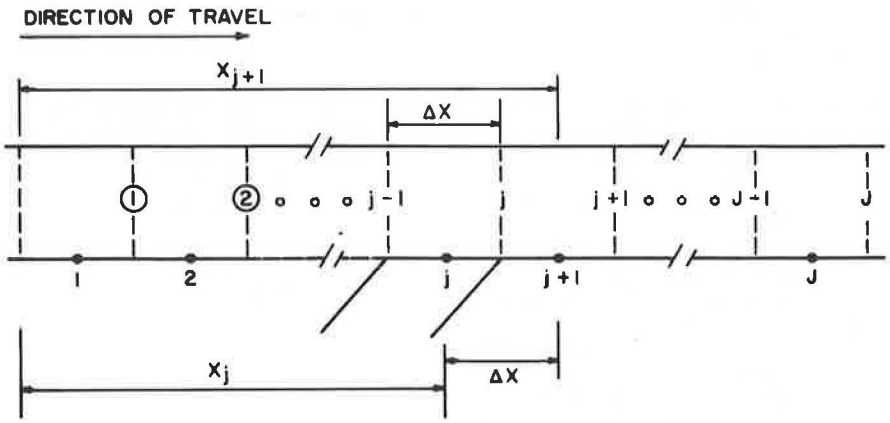


FIGURE 1 Space discretization of a freeway section.

ation (or loss) of cars in the lane under consideration. Following Gazis et al. (11), Munjal and Pipes (12), and Michalopoulos et al. (10), the simple-continuum system of equations describing flow on a two-lane freeway is

$$(\delta q_1/\delta x) + (\delta k_1/\delta t) = g + Q_1 \quad (7)$$

$$(\delta q_2/\delta x) + (\delta k_2/\delta t) = Q_2 \quad (8)$$

where

t,x = time and space coordinates, respectively
$q_i(x,t)$ = flow rate of the ith lane (i=1,2)
$k_i(x,t)$ = density of the ith lane (i=1,2)
$Q_i(x,t)$ = lane changing rate (i=1,2)
$g(x,t)$: generation rate of lane 1; at exit ramps g is negative

$$Q_1 = \alpha\{[k_2(x,t-\tau) - k_1(x,t-\tau)] - (k_{20} - k_{10})\} \quad (9)$$

$$Q_2 = \alpha\{[k_1(x,t-\tau) - k_2(x,t-\tau)] - (k_{10} - k_{20})\} \quad (10)$$

α = a sensitivity coefficient describing the intensity of interaction between lanes. In the simplest case α can be assumed constant; alternatively (10):

$$\alpha = \begin{cases} 0 & |k_2(x,t-\tau) - k_1(x,t-\tau)| \leq k_A \\ [\alpha_{max}/(k_o - k_A)]\,[|k_2(x,t-\tau) - k_1(x,t-\tau)| - k_A] & |k_2(x,t-\tau) - k_1(x,t-\tau)| > k_A \end{cases} \quad (11)$$

k_A = a constant value below which no exchange of flow occurs
τ = an interaction time lag (a value of zero could be assumed for simplicity)
k_{io} = an equilibrium density value that, if exceeded, will result in lane changing; i=1,2
k_o = jam density
g = generation term in lane 1 due to merging or diverging; this term also appears in Equation 1.

As before, the system of governing equations (Equations 7 and 8) can be solved numerically by discretizing in time and space. Figure 2 presents space discretization of a two-lane freeway section that includes an entrance ramp. It should be noted that for estimating average q, k, and u of any segment j, the nodes in Figure 2 are placed in the middle of each segment. This slight modification can also be made when all lanes are aggregated. The numerical solution allowing estimation of k, u, and q at each node and time increment is (10)

$$k_{1,j}^{n+1} = [(1/2)(k_{1,j+1}^n + k_{1,j-1}^n)] - [(\Delta t/2\Delta x)(G^n{}_{1,j+1} - G_{1,j-1}^n)] + [(\Delta t/2)(g_{j+1}^n + g_{j-1}^n)] + (\Delta t/2)(Q_{1,j+1}^n + Q_{1,j-1}^n) \quad (12)$$

$$k_{2,j}^{n+1} = [(1/2)(k_{2,j+1}^n + k_{2,j-1}^n)] - [(\Delta t/2\Delta x)(G_{2,j+1}^n - G_{2,j-1}^n)] + [(\Delta t/2)(Q_{2,j+1}^n + Q_{2,j-1}^n)] \quad (13)$$

where

$k_{i,j}^n$: density of the ith lane in the jth segment at $t = n\cdot\Delta t$ $(i = 1,2)$

$Q_{1,j}^n = \alpha_{1,j}^{n-s}\,[(k_{2,j}^{n-s} - k_{1,j}^{n-s}) - (k_{20} - k_{10})]$

$Q_{2,j}^n = \alpha_{2,j}^{n-s}\,[(k_{1,j}^{n-s} - k_{2,j}^{n-s}) - (k_{10} - k_{20})]$

$s\cdot\Delta t$ = interaction time lag $(=\tau)$

$\alpha_{i,j}^{n-s} = f(k_{1,j}^{n-s} - k_{2,j}^{n-s})$ as suggested by Equation 11; $\alpha_{i,j}^{n-s}$ could also be assumed constant

$G_{i,j}^n = k_{i,j}^n \cdot u_{i,j}^n = k_{i,j}^n \cdot u_e(k_{i,j}^n)$ $(i = 1,2)$

$u_e(k_{i,j}^n)$ = equilibrium speed corresponding to $k_{i,j}^n$; assuming the simple equilibrium model of Greenshields (9), it can be easily verified that $G_{i,j}^n = k_{i,j}^n \cdot u_f[1 - (k_{i,j}^n/k_o)]$ where u_f and k_o represent the free flow speed and jam density, respectively. However, more realistic u-k models are recommended.

Following computation of density at each time step the flow rate $q_{i,j}^{n+1}$ and speed $u_{i,j}^n$ are obtained from

$$u_{i,j}^{n+1} = u_e(k_{i,j}^{n+1})$$

and

$$q_{i,j}^{n+1} = k_{i,j}^{n+1}\, u_{i,j}^{n+1}$$

Extension of the simple-continuum model to more than two lanes is straightforward. Generalization to any number of lanes and further details concerning

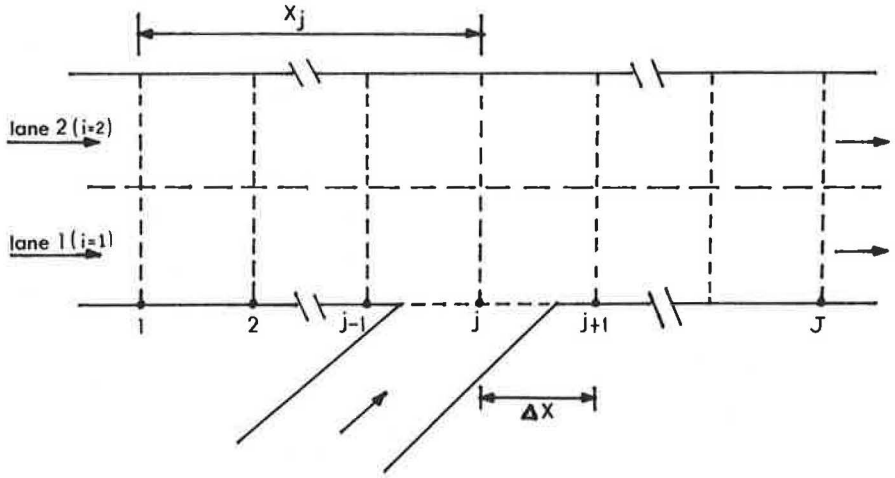


FIGURE 2 Space discretization of a two-lane freeway section.

the treatment of long merging and diverging areas, as well as weaving, are incorporated in KRONOS-1. Before concluding this section, it should be mentioned that a similar formulation and solution of the multilane problem can be extended to high-order models such as those proposed by Payne (7) and Phillips (13). Finally, it should be evident that when speed, flow, and density are known at every x and t, calculation of total travel (TT), total travel time (TTT), delays, stops, and energy consumption can easily follow.

GENERAL PROGRAM CAPABILITIES

The KRONOS-1 simulation program employs the previously summarized modeling for calculating speed flow and density dynamically on every node lying in the middle of each segment Δx. Delay is calculated from the difference between the travel times obtained from the actual speed and a user-specified minimum speed u_{min}; that is, cars are assumed to be delayed if they travel at a speed lower than u_{min}. Average values of q, k, u, total travel (TT), total travel time (TTT), and delay are found by zone (defined later in this section) or for the entire facility. At the end of simulation the program provides a summary of the number of cars arriving and departing each boundary of the freeway, as well as the ramps, the number of cars remaining on the freeway, and the queue length and size on the ramps at that time. Average and maximum values of the last two state variables are also calculated. Finally, the program produces a number of plots showing the change of q, k, and u in time and space, and the formation and dissipation of congestion, either by lane or in an aggregate fashion. Extensions are currently being made for calculating energy consumption and pollution levels.

In its present form, KRONOS-1 allows employment of user-specified speed-flow models (including discontinuous ones); alternatively, a default model can be employed. Further, arrival patterns can be constant or time dependent, deterministic or stochastic. Parameters of the selected 4-K model can also be specified by the user, and the flow model employed can be the simple or high-order continuum or a combination of the two (hybrid model). The user has the option of using a short input version; further, the results can be aggregated over all lanes or they can be presented in a lane-specific fashion. Finally, the merging, diverging, and weaving patterns can either be specified by the user or determined dynamically as a function of freeway volumes, capacity, geometrics, and ramp demands.

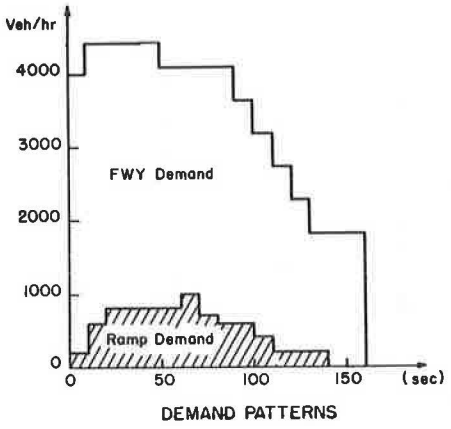
INPUT REQUIREMENTS

Because of the macroscopic nature of the program and the simplicity of the model, input requirements were kept to a minimum. Input is generally entered interactively in seven stages; alternatively, input could be retrieved from a disk file. Depending on the amount of detail desired, a reduced or an extended input version can be selected. The former employs a predetermined set of values for the model parameters; in the extended input version these parameters can be user specified as described in this section. From experience gained to this point it was found that input for the extended version can be entered in about 5-10 minutes per freeway mile, depending on the complexity of the situation, and the short version can save 25 to 50 percent of this time.

The first set of information required by the program includes the geometrics of the facility under consideration. The geometrics are found by dividing the facility into zones as shown in the lower part of Figure 3. The length and number of lanes in each zone must be specified as well as the length of each ramp.

Following the geometric input, the program requests the user to enter the freeway characteristics. These include a speed-density model that can be entered in numerical form (i.e., pairs of values of k and u) for each zone or it can be the same for the entire freeway. This model can be discontinuous. Alternatively, Greenshield's (9) model can be used as a default option. In addition to the u-k relationship, freeway characteristics include the free flow speed, the minimum speed for estimating delays, the jam density, and the capacity of each zone and ramp. It should be noted that ramp capacity should be given in terms of saturation flow, which depends on the geometrics of the ramp. The actual number of cars entering or leaving the ramps is determined (dynamically) by the program according to the user-specified arrival and merging patterns and the freeway flow conditions.

The third set of input data contains information related to simulation parameters; that is, it includes the simulation time, the initialization period, the time increment Δt, and the segment length Δx. The latter can be uniform for the entire freeway or it can vary in each zone.



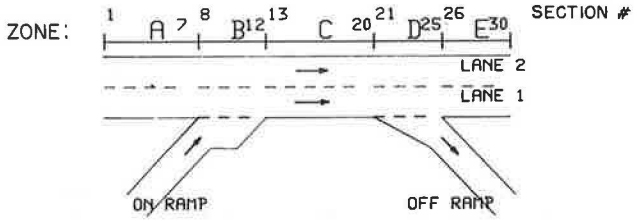


FIGURE 3 Demand patterns and division of freeway in zones.

Next, the program requests the initial conditions (i.e., the q, or k, or u distribution along the facility at t=0), which can be variable or uniform. An empty freeway can be selected as a default option.

The fifth input set is related to the arrival and departure patterns. These include the upstream end of the freeway and the ramps. Arrivals on the freeway and ON ramps can be constant, time varying as specified by the user or stochastic. In the latter case, the user must specify the mean and the variance of the arrivals. With respect to the departure patterns at the downstream end, it is assumed that there is no congestion so that $k^n_{j+1} = k^n_j$; alternatively, the departure pattern can be user specified. The percentage of cars exiting at the OFF ramps must be specified. In Figure 3 this refers to the percentage of cars in zones exiting at the OFF ramp. This percentage could be constant or time varying and is related to the origin-destination patterns. The latter are also needed in weaving sections.

Following definition of the boundary conditions the user must specify the particular model to be used as well as the merging pattern that best fits the ramp under consideration. This pattern is a function of ramp demands and freeway flow conditions and can be uniformly distributed along the acceleration lane; alternatively, it can be linearly or exponentially decreasing with distance. Finally, as a default option, the program can determine the merging pattern dynamically based on the freeway and ramp flow conditions as well as the ramp demands.

Next, the program asks the user to select the particular traffic model to be employed in the simulation. As mentioned earlier, either the simple or the high-order models can be used; in addition, KRONOS-1 allows use of a third hybrid model that is a combination of the two (6). The final set of input is related to the type and format of the output desired. This includes the type and frequency of the 2-D and 3-D plots desired, averaging of the results over a user-specified time increment, the unit where output is to be displayed (i.e., CRT, printer, plotter, graphics screen), whether hard copies of both input and output are desired, and so forth.

In the short input version of the program it is assumed that (a) Δx = 100 ft, (b) Δt = 1 sec, (c) the road is empty at t = 0, (d) the simple continuum model is employed, and (e) the merging pattern is determined by the program. Therefore, the user must enter only the simulation time, the geometrics and freeway characteristics, the arrival flow patterns, the percentage of volumes exiting, and the type of output desired. Following completion of the input, the program checks for errors and gives the user the opportunity to make changes or store the input data in a disk file for further modifications in a later run. Input information such as the geometrics and demand patterns is also plotted for quick visual inspection. The upper part of Figure 3 shows the demand pattern used for the typical entrance and exit ramp configuration shown in the lower part of the figure. The numbers adjacent to the zones represent the segments assigned to the boundaries of each zone.

PROGRAM OUTPUT

The program is executed in a time-scan fashion; that is, time is advanced by Δt and from the boundary conditions, and the solutions at t-Δt, k, u, and q are computed in the middle of each segment Δx; similar computations are made for the ramps, including the acceleration lanes. Subsequently TT, TTT, delay total arrivals and departures are updated, as well as queue size and queue length on each ramp along with the remaining statistics of the system (total and average stops, average q, k, and u by zone, and so forth). As time elapses, u, k, or q is plotted as a function of distance every NΔt seconds where N is specified by the user. Inspection of these plots assists in visualizing the formation and dissipation of congestion, as well as the evolution of queues in x and t. Figure 4 shows the density distribution per lane in space at t = 100 sec for the situation shown in Figure 3. The value of density is also printed; the symbol + indicates that density in both lanes 1 and 2 is about the same for a particular segment; symbols r, m, and v represent density at the ramp proper and the acceleration and deceleration lanes, respectively.

In addition to the detailed plots, 3-D plots of similar nature are produced. Such a plot, representing density on each lane as a function of x and t, is shown in Figure 5. This figure can be used for general inspection of the results, and further details can be found in Figure 4. As mentioned earlier, plots similar to those of Figures 4 and 5 can also be generated for flow and speed.

During the simulation, density of every segment is compared with a threshold value representing capacity; if this value is exceeded, the segment is assumed to be congested, and at the end of simulation a plot is produced showing congestion as a function of time and space. Such a plot for the situation of Figure 3 is shown in Figure 6, which depicts the duration and extent of congestion in lane 1. A better visualization of the formation and dissipation of congestion in time and space, as well as of the dynamic change of density along the road, can be seen on the graphics screen that depicts in color the value of density along the road. Because the screen is repainted every Δt, a very vivid representation of the flow dynamics and the evolution of congestion is realized.

At the end of simulation a summary table is produced showing the average value of q, k, and u for

```
Time= 100.0    (Step=100)    DENSITY
SEC.                0          50         100        150        200(Veh/Mil)
 # lane#1 lane#2 ramp+----+----+----+----+----+----+----+----+----+-
 1   54.0   52.4      ¦          21
 2   59.4   54.9      ¦           21
 3   68.4   59.2 12.9¦  r          2 1
 4   78.2   63.2 12.9¦  r           2  1
 5   90.5   69.1 12.9¦  r            2    1
 6  100.3   73.4 12.8¦  r             2     1
 7  109.3   78.4 12.3¦ r               2       1
 8  115.1   81.6  9.9¦ m               2        1
 9  117.0   84.6  7.6¦ m                2       1
10  115.4   86.3  5.5¦m                 2       1
11  112.6   87.4  3.3¦m                 2       1
12  107.7   87.7  2.4m                   2    1
13  101.5   87.5      ¦                   2 1
14   96.6   86.9      ¦                  2 1
15   92.3   85.9      ¦                  21
16   88.3   84.7      ¦                  21
17   84.3   83.3      ¦                  +
18   79.8   81.7      ¦                 +
19   74.4   80.0      ¦                12
20   67.3   78.1      ¦              1  2
21   56.1   76.2  6.4¦v           1    2
22   49.0   74.4 11.7¦ v         1     2
23   45.5   72.6 14.7¦  v      1       2
24   44.0   71.1 16.5¦  v      1      2
25   43.5   69.7 17.5¦   v     1      2
26   44.2   68.5 17.5¦   r     1      2
27   44.9   67.5 17.5¦   r     1      2
28   45.5   66.7 17.5¦   r     1    2
29   46.0   66.1 17.5¦   r     1    2
30   46.4   65.7      ¦        1    2
```

FIGURE 4 Density distribution by lane at t = 100 sec for the situation shown in Figure 3.

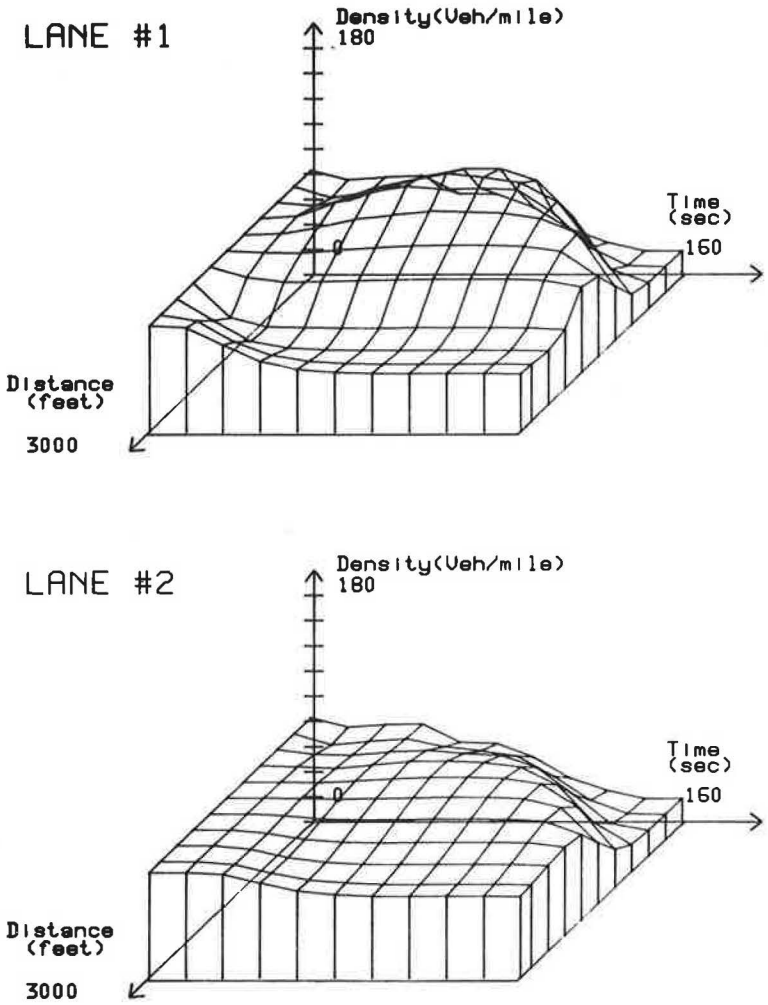


FIGURE 5 Graphic presentation of density in each lane as a function of time and space

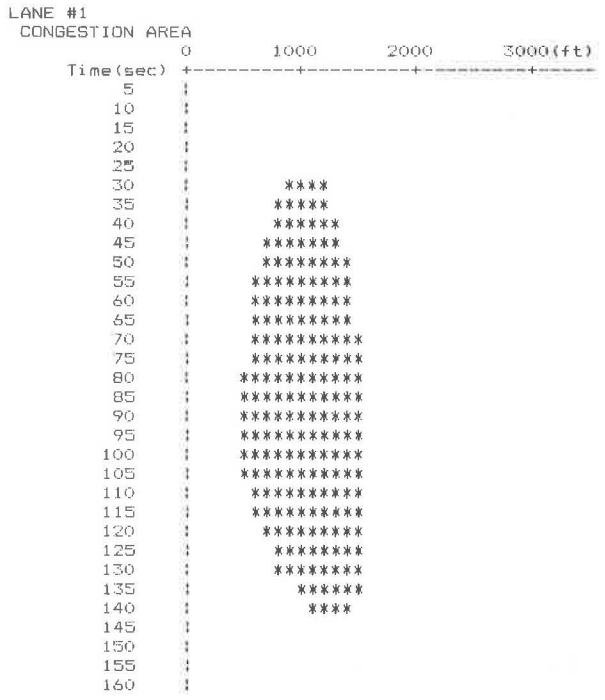


FIGURE 6 Graphic representation of congestion by lane in x and t.

the entire system and by zone as well as TTT, TT, delay, arrivals, departures, queue size and length, and the remaining measures of effectiveness and system statistics as described under General Program Capabilities. Extensions of the program allowing estimation of fuel consumption and pollution levels are currently under way.

HARDWARE AND PROGRAM LIMITATIONS

The memory of the computer currently used is 128K bytes; however, the usable memory with the existing IBM software (P-system) is actually only 34K bytes. Although new software is currently being prepared by IBM to correct this problem, the current 34K space restricts the size of the freeway sections that can be simulated. More specifically, this memory allows employment of up to 60 nodes; assuming a segment length Δx of 100 to 500 ft, this space allows simulation of freeway sections from 1.14 to 5.68 miles long.

Because the memory of the IBM-PC can easily be extended to 256K and possibly to 512K it is clear that the program's capability could be improved substantially even with existing IBM software (i.e., it could be quadrupled). Therefore, the maximum freeway length could be extended to 22.73 miles although in most practical applications the maximum length that needs to be analyzed rarely exceeds 10-12 miles. Longer segments could also increase the maximum length but this would imply lower accuracy. Further, as the number of lanes increases and separate statistics per lane are desired, maximum length decreases accordingly. For instance, for producing detailed statistics for two lanes, maximum length decreases by 50 percent, and conversely, if the results are aggregated, maximum length doubles. Incidentally, in its present form the program can handle up to three lanes; after this limit, all lanes must be combined or the program must be extended. This does not really present a problem if sufficient memory is available.

Execution time is a second consideration that might concern the potential user. Because of the limited capabilities of personal computers and the computations and bookkeeping required in simulation, execution time is generally long. Despite this, the effectiveness of the models and the solution algorithms result in satisfactory execution time especially when the magnitude of the tasks accomplished by the program is considered. For instance, for simulating a two-lane freeway section with one entrance ramp and 30 nodes per lane for 240 sec, initial execution time was 50 min. Although this time includes lane changing and separate statistics per lane, as well as the drawing of graphs on the graphics screen while presenting all computational details on the CRT, this timing performance was not considered satisfactory. Execution time was reduced by 50 percent when the compiled program was executed.

Further improvements with respect to execution time are currently being sought by employing the Intel 8087 processor that is designed to add fast and accurate floating point capability to the Intel 8088 processor, currently used by the IBM personal computers. The two chips with appropriate software (available from independent vendors) can reduce computation time from 5 to 20 times (14). Clearly such improvement is significant and should make using the program more attractive.

Although the maximum freeway length that can be simulated by KRONOS-1 may be somewhat restricted by the available memory size, no such restrictions exist with respect to the simulation period; that is, it can be as long as desired.

PROGRAM TESTING AND VALIDATION

Because of budget and manpower limitations, comparisons with field data were not made. However, the program was implemented to a number of situations that covered a wide range of the speed, flow, and density domain and included both entrance and exit ramps as well as multiple lanes. Subsequently, the results were compared against those obtained from microscopic simulation using INTRAS (3), a recently developed, well-documented, tested, and calibrated program. Microscopic simulation for generating a data base was further justified by the need to allow demands to fluctuate sufficiently over a controlled and frequently wide range in relatively short time intervals. Such intervals were sought for reducing the comparison effort. Another consideration justifying data base generation by INTRAS was the need to impose tractable initial and boundary conditions to allow intuitive inspection of the results.

Four of the initial test situations representing both uninterrupted and interrupted flow are shown in Figures 7, 8, and 9. All cases in these figures represent single-lane flow because it was thought that initial testing under the simplest flow conditions should prevent distortion of true model performance; distortion could be introduced from lane changing, especially in merging areas. Unsatisfactory program performance, even in this simple case, would imply that there is little reason to expect better results as the number of lanes increases. In case 1 (Figure 7a) demands are nearly constant approaching maximum flow. In case 2 (Figure 7b) flow starts at about one-third of its maximum value; subsequently, it gradually increases to the maximum flow rate where it remains for some time and then it gradually decreases to its initial level. In case 3, merging flows are introduced for a short time to

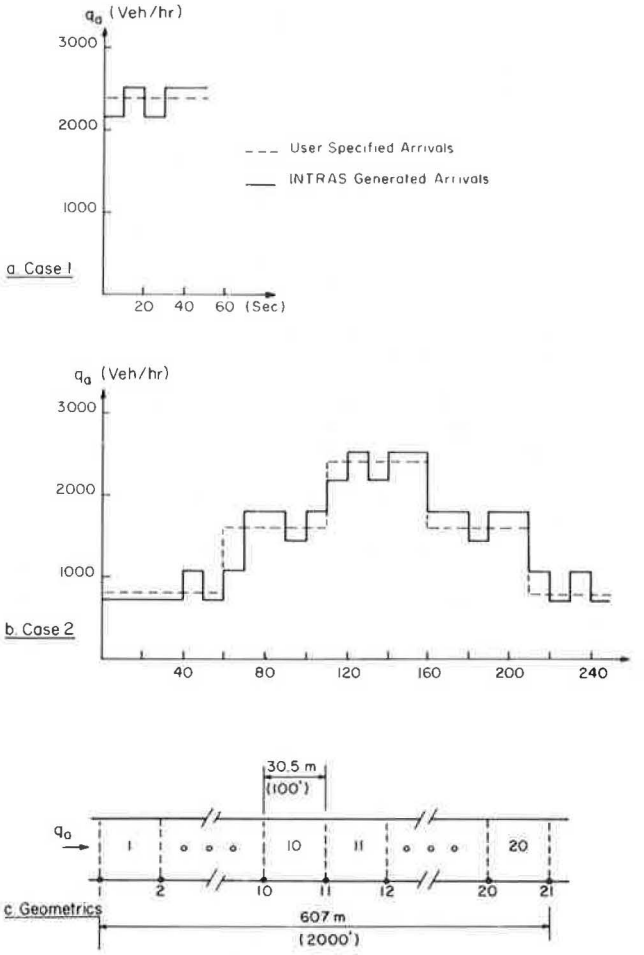


FIGURE 7 Arrival patterns and geometrics for uninterrupted flow testing (cases 1 and 2).

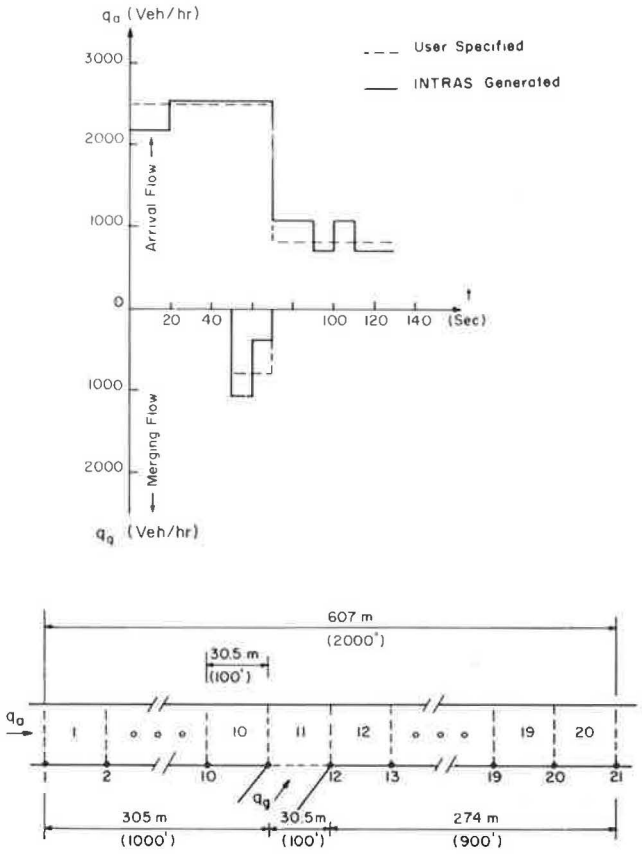


FIGURE 8 Input patterns and geometrics for interrupted flow (light congestion, case 3).

create light congestion. The demands drop substantially after congestion sets in for quick dissipation. Finally, case 4 is similar to case 3 but represents a longer freeway section and higher ramp demands for heavier congestion.

It should be noted that the merging flows shown in Figures 3 and 4 are the actual ones entering the freeway (i.e., not simply the ramp demands) and were obtained from INTRAS. Maximum freeway uninterrupted flow suggested by INTRAS was approximately 2,400 passenger car units per hour. This value may appear high compared with the maximum of 2,000 cars per hour usually employed in practice. It should be pointed out, however, that under nearly ideal conditions maximum flows of up to 2,300 vehicles per hour have been measured in practice (4). Because this figure includes some trucks and other heavy vehicles, it appears that the value obtained by INTRAS for passenger cars only is not unreasonable. Finally, in the KRONOS-1 simulations Δx and Δt were assumed to be 100 ft and 1 sec, respectively, and the free-flow speed was assumed to be 55 mph. Incidentally, the INTRAS estimates of q, k, and u were averaged over 10-sec intervals in every Δx increment to make meaningful comparisons. Final estimates (i.e., at the end of simulation) of total travel time (TTT) for the entire section under consideration were also compared. For the latter measure of effectiveness its percentage difference (PD) from the data was estimated and used as a criterion of close agreement; a negative PD indicates that the model overestimates TTT and positive PD indicates underestimation. In the comparisons of speed and density the mean absolute error (MAE) and mean square error (MSE) were employed. The former was defined as

$$MAE = (1/nxm) \sum_{i=1}^{m} \sum_{j-1}^{n} | z_j^i - \hat{z}_j^i |$$

where

- n = the number of space increments Δx (=100 ft),
- m = the number of 10-sec intervals corresponding to the entire simulation period,
- z_j^i = the average INTRAS estimate of speed flow and density over segment j and 10-sec interval i, and
- $\hat{z}_j^i$ = the average model estimate of speed flow and density over segment j and 10-sec interval i.

Using this notation the MSE is defined as

$$MSE = (1/nxm) \sum_{i=1}^{m} \sum_{j=1}^{n} (z_j^i - \hat{z}_j^i)^2$$

Finally TTT and TT were estimated by numerical integration of the expressions

$$TTT = \int_o^T N(t)dt$$

$$TT = \int_o^T \left[\sum_{\ell=1}^{N(t)} u_\ell(t) \right] dt$$

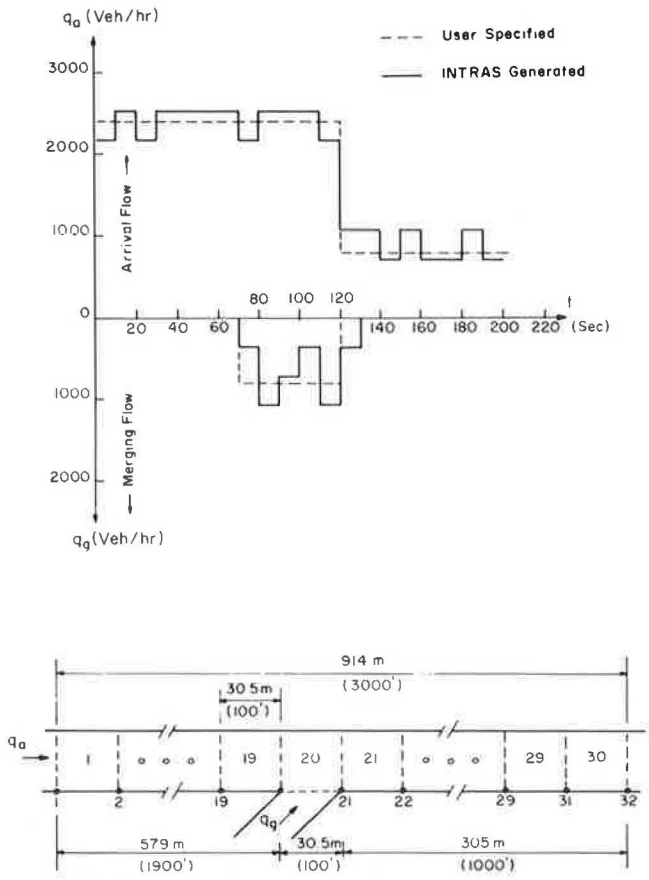


FIGURE 9 Input patterns and generation rates for case 4 (interrupted flow, heavier congestion).

where

T = the simulation time and
N(t) = the number of cars in the entire freeway section at an instant = $\int_0^L k(x)dx$.

In addition to MSE, MAE, and PD, visual comparisons of the KRONOS-1 results with the data base were made from the plots of q, k, and u versus time and space. Overall, the comparisons of the KRONOS-1 and INTRAS data were very satisfactory. This can be seen from the data in Table 1, which indicate that the error indices MSE, MAE, and PD are very small. For instance, the error in estimating TTT by KRONOS-1 ranges from 7.1 to 9.5 percent using the simple-continuum model (column 1). This error range is reduced to 3.6 to 6.7 percent by the hybrid model (column 2) and to 0.3 to 5.5 percent by employing a discontinuous v-k relationship (column 3). It is worthwhile noting that in further testing the error indices were reduced from 40 to 60 percent when the step size Δx and Δt was also reduced by one-half. A partial visualization of these comparisons is shown in Figure 10, which shows the distribution of density in case 4 (interrupted flow) at five time intervals during the simulation.

Further comparisons with multiple lanes and more than one entrance ramp were also made and the results were found equally satisfactory. Table 2 gives the comparisons of two additional situations representing uninterrupted and interrupted flow conditions on two-lane facilities, described in detail elsewhere (10). The first column corresponds to the simple-continuum model and the second to the high-order models; the third column is slightly different from the second in that the parameter T (reaction time) of the momentum equation of the high-order model was assumed to be a function of density. In the fourth column the results of the hybrid model are presented. According to the hybrid model, the momentum equation in a particular segment is dropped when congestion sets in. Column 5 corresponds to a two-dimensional model that explicitly includes street width (10). Columns 6-10 correspond to the case of interrupted flow and are similar to the first five.

As the data in Table 2 indicate, the error indices are very small and the differences among the various alternatives are not substantial. Figure 11 shows a comparison of the results obtained from the simple-continuum model in the uninterrupted flow case, and is similar to Figure 10. The INTRAS data correspond to the average of both lanes (INTRAS does not output lane-specific information). The per lane density estimates of KRONOS-1 are also shown, as well as the average value, which is very close to the INTRAS results. As before, the error levels could be substantially reduced by reducing the step size. Additional testing and evaluation results for multilane facilities are available (10) and they suggest close agreement with the INTRAS microscopic simulations.

CONCLUSIONS AND RECOMMENDATIONS

Perhaps the most interesting feature of the KRONOS-1 simulation program is that it can run on a microcomputer while its results closely agree to those of microscopic simulation, especially when the step size is reduced. This was made possible by the simplicity of the particular finite-difference methods used that allow quick calculation of the flow parameters k, u, and q in both time and space. The improved modeling employed by the program allows treatment of merging, diverging, and simple weaving sections. Therefore, despite the macroscopic nature of the program, acceleration, deceleration, and auxiliary lanes can be taken into account. This, combined with the simplicity of running the program, should encourage its use in even less complex problems such as estimating ramp capacity or level of service.

Although present execution time and the size of the freeway should be satisfactory for most practical applications, further improvements are possible. For instance, by incorporating the Intel 8087 processor into the personal computer it is anticipated that execution time will be reduced substantially (5-20 times); that is, it will be only a fraction of the simulation time. The size of the freeway should not present a problem, especially when expanded memory is employed. Enhanced memory will necessitate program extensions to allow a larger number of ramps and more than three main lanes when the lane changing option is selected. However, such extensions can easily be made. Further, it should be observed that software improvements currently under way at IBM should allow larger usable space without additional memory.

KRONOS-1 in its present form is only a prototype and as such it requires more rigorous testing and calibration. Such testing using field data and simulation results is currently under way. Practical extensions are also being implemented in a new version of the program called KRONOS-2. These extensions include simulation of ramp metering, estimation of fuel consumption and pollution levels, inclusion of trucks and other heavy vehicles, improved

TABLE 1 Error Indices of KRONOS-1 for the Simple Continuum and Hybrid Models

MOE \ Case / Model		Simple Continuum	Hybrid Model (Discont. u-k)	Simple Continuum (Discont. u-k)
k: Density (Veh./Mi)	1	7.69 (91.04)*	8.48 (10.9)	7.41 (83.39)
	2	6.89 (80.72)	6.24 (64.37)	6.21 (63.97)
	3	11.0 (234)	11.76 (238)	5.67 (49.20)
	4	11.8 (270)	11.53 (254)	9.68 (177)
u: Speed (Mi./Hr.)	1	3.27 (17.83)	4.71 (37.13)	2.94 (12.60)
	2	4.13 (27.89)	3.39 (16.72)	3.39 (16.78)
	3	7.21 (86.62)	10.4 (180)	2.69 (11.02)
	4	6.47 (77.62)	10.1 (169)	4.81 (43.14)
TTT (Veh.-Min.)	1	11.24 (-7.1)	10.11 (3.6)	10.46 (0.3)
	2	48.62 (-9.5)	46.02 (-3.7)	47.08 (-6.1)
	3	30.09 (7.9)	35.83 (-6.7)	33.82 (-3.5)
	4	89.65 (-8.2)	88.28 (-6.5)	87.39 (-5.5)

*Numbers indicate MAE (MSE) for k and u; in last MOE the numbers show the actual value of TTT and % difference from data.

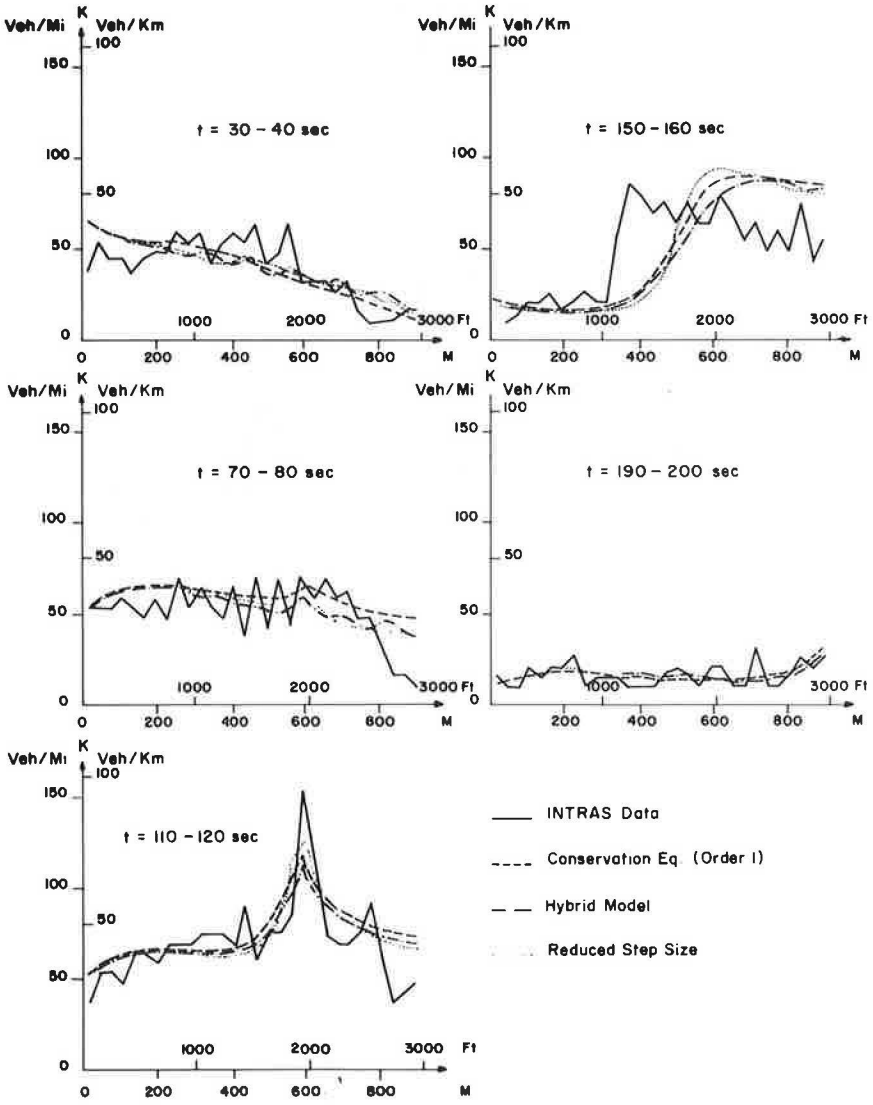


FIGURE 10 Comparison of KRONOS-1 and INTRAS results (case 4).

TABLE 2 Error Indices of Modeling Alternatives at Two-Lane Facilities When per Lane Estimates Are Averaged

Situation / Model / MOE	1. Uninterrupted Flow					2. Interrupted Flow				
	1 Simple Continuum	2 High Order T-cons	3 Modified High Order T=T(k)	4 Hybrid Model	5 2-D Model	6 Simple Continuum	7 High Order (Const. T)	8 High Order (Variable T)	9 Hybrid Model	10 2-D Model
Density (veh/mi)	2.06[1] (14.50)	2.07 (14.74)	2.21 (16.45)	2.22 (16.48)	2.38 (21.48)	6.76 (181)	7.30 (207)	6.91 (179)	6.81 (179)	6.84 (162)
Speed (mi/hr)	1.69 (8.88)	1.74 (9.46)	2.08 (12.63)	2.07 (12.59)	2.17 (13.65)	4.02 (55.3)	4.19 (60.0)	3.70 (48.8)	3.76 (48.5)	3.75 (47.5)
TTT (veh-min)	40.88[2] (-0.15)	40.53 (0.72)	38.86 (4.81)	38.88 (4.81)	36.85 (9.48)	230.1 (-2.56)	229.2 (-2.17)	204.1 (9.03)	217.3 (3.10)	228.5 (-1.85)

[1] Numbers outside of parenthesis in k, u and q rows indicate MAE; # in parenthesis indicates MSE.

[2] Number indicates estimated TTT while the number in parenthesis the % difference from INTRAS.

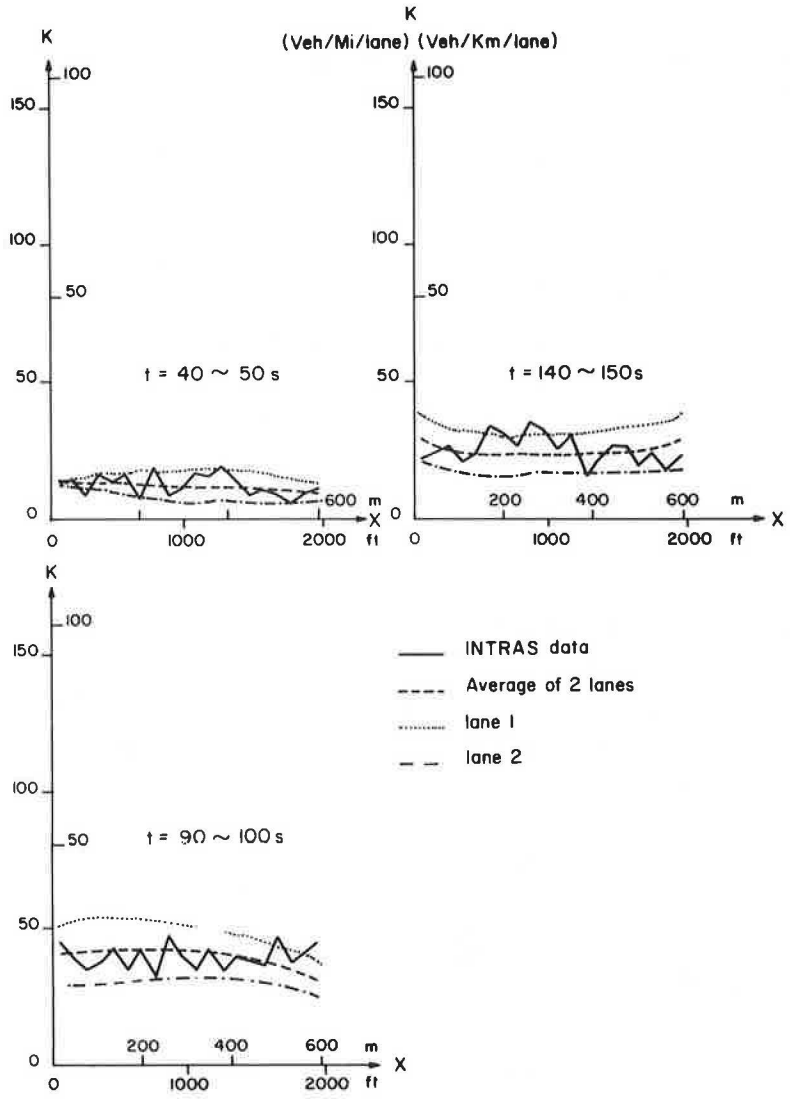


FIGURE 11 Comparison of KRONOS-1 and INTRAS results in a two-lane situation (uninterrupted flow, simple continuum model).

graphics and I/O operations, and so forth. KRONOS-2 is being developed for the Minnesota Department of Transportation and it should be completed in 1984.

ACKNOWLEDGMENT

The theoretical part of this research was supported by NSF (grant CEE 8210189). Computer programming was performed by J. Lin and Y. Yamauchi.

REFERENCES

1. P.P. Jovanis, W.K. Yip, and A.D. May. FREQ6PE: A Freeway Priority Entry Control Simulation Model. Report UCB-ITS-RR-78-9. Institute of Transportation Studies, University of California, Berkeley, 1978.
2. D.B. Roden, W. Okitsu, and A.D. May. FREQ7PE: A Freeway Corridor Simulation Model. Report UCB-ITS-80-4. Institute of Transportation Studies, University of California, Berkeley, 1980.
3. D.A. Wicks and E.B. Lieberman. Development and Testing of INTRAS, A Microscopic Freeway Simulation Model, Vol. 1-4. FHWA Report DOT-FH-11-8502. FHWA, U.S. Department of Transportation, 1980.
4. P.G. Michalopoulos and R. Plum. I-394 Alternative Design Analysis: Phase II, Final Report. Minnesota Department of Transportation Project 59232. Minnesota Department of Transportation, St. Paul, 1981.
5. P.G. Michalopoulos and R. Plum. Selection and Evaluation of Optimal Freeway Design by Computer Simulation. In Transportation Research Record 773, TRB, National Research Council, Washington, D.C., 1981, pp. 40-47.
6. D.E. Baskos, P.G. Michalopoulos, and Y. Yamauchi. Dynamic Modelling and Numerical Treatment of Freeway Flow. Submitted to Transportation Science, 1983.
7. H.J. Payne. Models of Freeway Traffic and Control. Mathematics of Public Systems, Vol. 1, 1971, pp. 51-61.
8. M.H. Lighthill and G.B. Witham. On Kinematic Waves II: A Theory of Traffic Flow on Long Crowded Roads. Proc., Royal Society, Series A229, No. 1178, London, England, 1971, pp. 317-345.
9. B.D. Greenshields. A Study of Traffic Capacity. Proc., 14th Annual Meeting of the Highway Research Board, HRB, National Research Council, Washington, D.C., 1934, pp. 448-477.
10. P.G. Michalopoulos, D. Baskos, and Y. Yamauchi. Multilane Traffic Flow Dynamics: Some Macroscopic Considerations. Transportation Research, 1984, in press.
11. D.C. Gazis, R. Herman, and G.H. Weiss. Density Oscillations Between Lanes of a Multilane Highway. Operations Research, Vol. 10, 1962, pp. 658-667.
12. P.K. Munjal and L.A. Pipes. Propagation of On-Ramp Density Perturbation on Unidirectional Two- and Three-Lane Freeways. Transportation Research, Vol. 5, 1971, pp. 241-255.
13. H.J. Phillips. A New Continuum Traffic Model Obtained from Kinetic Theory. Transactions of Automatic Control, Vol. AC-23, No. 2, 1978, pp. 1032-1036.
14. Softalk for the IBM Personal Computers. International Business Machines Corp., Armonk, N.Y., March 1983, pp. 70-72.

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

Role of Adaptive Discretization in a Freeway Simulation Model

PHILIP S. BABCOCK IV, DAVID M. AUSLANDER, MASAYOSHI TOMIZUKA, and ADOLF D. MAY

ABSTRACT

Preliminary studies with the FREFLO freeway simulation model showed it to have difficulty depicting congested flow situations. This problem is traced to the model's transformation from the continuous to the discrete domain. Methods of determining the proper discretization intervals are shown along with simulation examples of a simple freeway. The properly discrete model requires an excessive amount of computer time for a real freeway simulation. Hence, two adaptive schemes that reduce the computing time to a manageable level are presented. The resulting model, FRECON, is then calibrated and validated using five peak-period data sets from the Santa Monica Freeway in Los Angeles.

Simulation models can serve many purposes in the understanding and design of traffic systems. They allow experiments to be conducted conveniently and safely. Insights into the functional relationships of various system elements can be gained through modeling, and future conditions may be predicted based on current system behavior.

This work in simulation models was motivated by the need to compare freeway performance under different control schemes and to aid in the design of new control strategies (1). For these uses a freeway simulation model must give an accurate representation of freeway speeds, densities, and flow rates as a function of demands and geometry. The wide range of time scales used in the control strategies studied (ranging from less than 1 minute to 15 minutes) required a dynamic model. Because freeway on-ramp control usually uses data derived from bulk properties of the resulting freeway flow, macroscopic freeway models would provide sufficient resolution for this work.

The FREFLO model was selected as the basis for the simulation study. The model, developed by Payne (2,3), is macroscopic and dynamic. Preliminary work with the model, along with a discussion presented elsewhere (4), revealed a serious deficiency in the model's ability to simulate congested flows in a realistic fashion. This problem was traced to the transformation of the model's formulation from the continuous to the discrete domain required by a computer. For the discrete version to be an accurate representation of the continuous model or the real world, extreme care must be used in selecting the spatial and temporal discretization intervals.

To demonstrate the effects of various discretization intervals, a simple program will be presented that simulates a congested bottleneck. The effects of proper and improper discretization will be shown. Two methods of reducing the excessive computing load that often accompanies dynamic simulation models will also be discussed. The first of these methods, a heuristic adaptation scheme, is used in the FRECON simulation model. Finally, the calibration and validation of the FRECON model will be presented using data from 5 days on the Santa Monica Freeway in Los Angeles.

ORIGINS: CONTINUOUS FORMULATION

A dynamic, macroscopic model of traffic flow on a freeway may be derived from various standpoints. Payne derived the FREFLO model (formerly the MACK model) from car-following theory (2). The resulting model contains three basic equations: conservation, momentum (or dynamic), and equilibrium state. The state variables are density and speed. Phillips arrived at a similar formulation through a statistical derivation (5,6). Both models resemble the traditional hydrodynamic formulations of laminar, compressible flow.

The similarities between the Payne and Phillips dynamic equations are reassuring. The fact that two different derivations lead to the same equation terms gives one confidence in their validity. The general form of the equations will be the focus of this section, with emphasis on the overall consequences of their form.

The Payne formulation of FREFLO is contained in the following equations:

$$(\partial\rho/\partial t) + (\partial q/\partial x) = S(x,t) \tag{1}$$

$$\partial u/\partial t = [- u(\partial u/\partial x)] + [(1/c)(u_e - u)] - [b(\mu\rho/\partial x)(1/\rho)] \tag{2}$$

$$u_e = u_e(\rho) \tag{3}$$

where

$\rho(x,t)$: density at x and t,
$u(x,t)$: speed at x and t,
$q(x,t) = \rho(x,t) \cdot u(x,t)$: flow rate at x and t,
$S(x,t)$: ramp flow source term at x and t,
$u_e(\rho)$: equilibrium speed-density relation,
c: relaxation term coefficient, and
b: anticipation term coefficient.

Equation 1 represents the conservation of vehicles. The ramp source term is a static approximation of the true behavior of the ramp flows. Ramp capacity and metering rates are accounted for here. Merging effects are treated as a function of the mainline flow in the right-hand lane. The dynamics, or time change of momentum, are contained in Equation 2. The first term on the right-hand side of Equation 2 is the convection term. The second term represents the relaxation of vehicles to an equilibrium speed. This speed, in Equation 3, is experimentally determined. Figure 1 shows an equilibrium speed-density relation for the Harbor and Hollywood freeways in Los Angeles (7). The last term of Equation 2 is called an anticipation term by Payne. It represents driver response to changes in density over space.

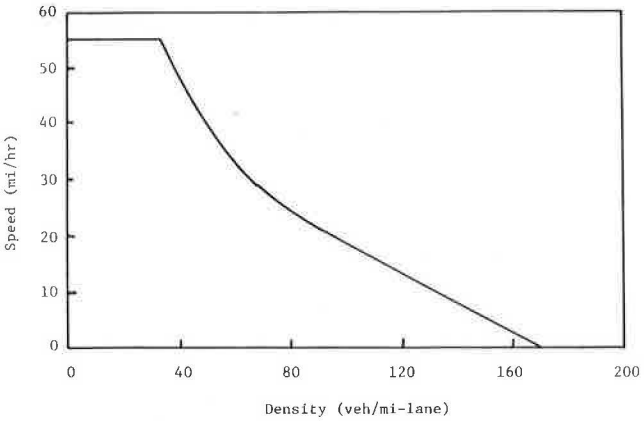


FIGURE 1 Continuous equilibrium speed-density relation.

The coefficient in the relaxation term of Equation 2 is important in two ways. First, it is the inverse of the time constant for relaxation to the equilibrium speed. This can be observed when the traffic speed has been lowered (for example, at an incident) and then suddenly allowed to return to the speed limit. This change in speed must correspond to a realizable vehicle acceleration. Second, the only capacity information in the model is contained in the equilibrium speed-density relation. For example, short time constants force the speed to equilibrium quickly and reduce the excursions from capacity flow when high flow rates are present.

The coefficient of the anticipation term represents driver sensitivity to changes in density. This term becomes important in situations where there is a large spatial gradient in density. For example, as free-flowing vehicles approach a congested region they start to slow before they enter the congestion.

Simulation studies of simple geometries were used to tune the coefficients in the momentum equation. The primary concerns are that the resulting traffic flows show realizable accelerations, have a proper emphasis on capacity information, and behave reasonably during transients. The dynamic equation coefficients found to give the most realistic simulations for simple geometries are

$c = 0.075$ sec and
$b = 1{,}200$ (mph)^2.

These coefficients could have been set only if the model was properly discrete. In the following sections the nature of the discretization and examples of its proper and improper use will be shown.

DISCRETE FORMULATION

Given the continuous model represented in Equations 1 and 2, how can these equations be solved in an efficient manner (i.e., on a computer)? Clearly a discrete model that is equivalent to the continuous one is required. Partial derivatives with respect to time or space would be transformed into changes over discrete steps in time or space. However, information is naturally lost during this transformation. Values previously defined over a continuum become specified only at discrete points or as averages over a region. This loss is a natural consequence of discretization but is acceptable because the resulting formulation is computer compatible. Extreme care must be taken to ensure that the resulting discrete model is a faithful representation of the continuous one and therefore can represent realistic freeway flows.

Discretization is proper when either of the following two statements is true: the discrete model must give solutions that coincide, at discrete locations, with the solutions of the continuous model; or the discrete model's solution should not be dependent on the selected discretization. These rules are used to find the maximum allowable intervals for a given model.

There is also a minimum limit on step sizes. Otherwise, the effects of machine roundoff start to cloud the solution. In the range between these two extremes (the range of proper discretization), altering the discrete steps has only a proportionate effect on the solution. In other words, step size alters only the detail of the solution.

Payne transformed the continuous formulation of FREFLO into a discrete one by defining spatially aggregated variables and then integrating each over time (2,4). Using this procedure on Equations 1 and 2 results in the following model:

$$\rho(x,t+\Delta t) = \rho(x,t) + \Delta t\left(-\{[q(x-\Delta x,t) - q(x,t)]/\Delta x\} + S(x,t)]\right) \quad (4)$$

$$u(x,t+\Delta t) = u(x,t) + \Delta t\Big\{-u(x,t)\,[u(x,t) - u(x-\Delta x,t)/\Delta x]\} + \{(1/c)\,[u_e(\rho) - u(x,t)]\} - \Big(b \cdot [1/\rho(x,t)]\,\{[\rho(x + \Delta x,t) - \rho(x,t)]/\Delta x\}\Big)\Big\} \quad (5)$$

where dx and dt have been replaced by discrete steps in space and time. The state variables are now defined only at discrete locations and times.

This system of equations could also have been derived by the explicit finite-difference method. Because the model is an explicit instead of an implicit finite-difference representation, the selection of the step sizes will affect model stability. There are two restrictions placed on the spatial and temporal step sizes. These are called the stability and accuracy limits. Because the model has two independent variables, their ratio is limited to ensure model stability. Also, both the temporal and spatial intervals must be small enough to follow all significant transitions in time and space, respectively. This limit determines the model's tracking accuracy.

The problem of stability has been discussed in the literature (3,4) and demonstrated on a simple heat transfer problem [Appendix C (1)]. The stability limit maintains causality in the flow of information down the freeway. The established stability limit for the model of Equations 4 and 5 is

$$(\Delta t/\Delta x) < 22 \text{ sec/mile} \quad (6)$$

This stability limit applies only when the discretization intervals also satisfy the accuracy limit. The stability limit is severely reduced if the step sizes violate the accuracy requirements.

The accuracy limitation arises from the definition of a derivative and its approximations. Because the discrete model obtains its next data point by extrapolating the local slope, frequent observations of this slope are needed when it changes abruptly in time or space. Let us examine the anticipation term of Equation 2:

$$b(\partial\rho/\partial x)\,(1/\rho) \quad (7)$$

This term uses information on the local slope of the density with respect to space.

Figure 2 shows a simulated density profile of congestion resulting from a lane drop at x = 0.5 mile. For the discrete model to follow this contour the spatial discretization must be small enough to ensure that the spatial slope of the density has not changed appreciably over each step. Spatial step sizes similar to the extent of the shock front at x = 0.25 are required to ensure that the entire region of congestion is modeled accurately.

The proper value for the stability and accuracy limits can be found experimentally. A properly discrete model is one with simulation results that are independent of the step sizes used. Therefore, the discretization limits can be found by repeated simulations with various step sizes. This process is discussed in the following section.

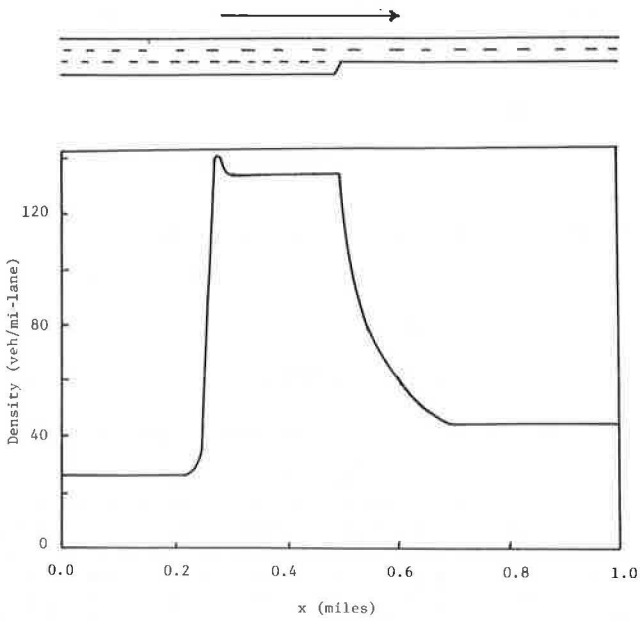


FIGURE 2 Density contour at lane drop.

EFFECTS OF DISCRETIZATION

To show the consequences of proper and improper discretization more concretely, the simple geometry of a lane drop with a constant input flow (Figure 3) is used. In this example the lane capacity is set to 2,000 vehicles per hour and the input demand is 4,500 vehicles per hour. It is expected that this demand level would cause capacity flow in subsections 6 through 10 and queueing on the subsections upstream of subsection 6.

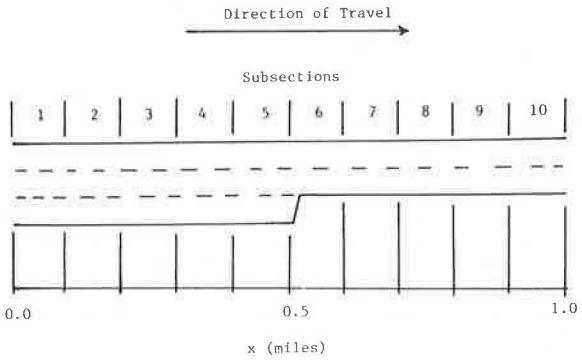


FIGURE 3 Lane drop geometry.

The model represented in Equations 4 and 5 and Figure 1 has been coded into a simple FORTRAN program (Table 1). To maintain simplicity in this program all inputs are assigned statements and the output is printed at each time step. Although this program is used to present simulations of the geometry of Figure 3, the output has been processed here into a more readable format. For the following procedures, the model generates instantaneous speeds and densities averaged over each subsection and the instantaneous flow leaving each subsection.

To demonstrate the effects of various discretization intervals the program is used to simulate a 10-min period of the bottleneck geometry of Figure 3. The spatial steps in the model are arbitrarily set equal to the subsection lengths (0.1 mile) and the temporal steps are set to satisfy the stability criterion (Equation 6). The results (Figure 4 and Table 2) show that the capacity of the bottleneck is exceeded and the highest densities (more than 100 vehicles per mile-lane) are inside the bottleneck. This is similar to a simulation presented in the literature (4).

Clearly this is not the expected behavior for an overloaded bottleneck. The first aspect to be investigated is whether the stability limit is valid for this spatial discretization. Dividing the time steps by two gives the simulation shown in Figure 5 and Table 3. Although this result is quite different from that of the previous simulation, it is no more real because the bottleneck flows are less than 35 percent of capacity. Further reductions of the time step do not alter the result. Therefore, the stability limit for this spatial step is in fact less than the prescribed stability limit.

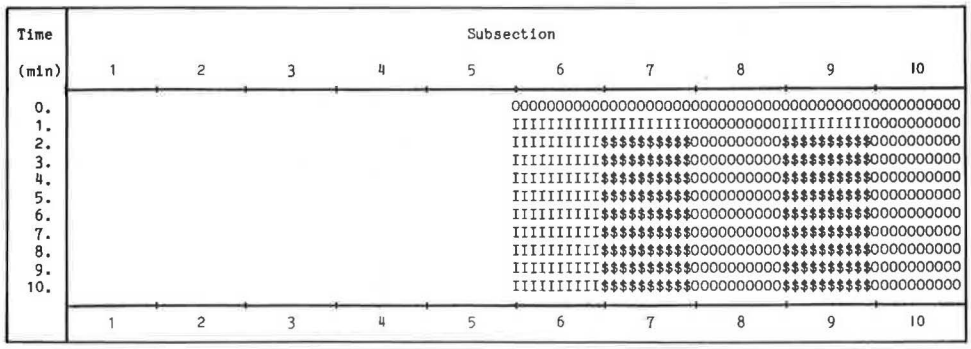
To get a more realistic simulation the equilibrium speed-density relation could be artificially altered as has been suggested in the literature (3,4). It is recommended instead that an investigation of the spatial accuracy limit be performed. In Figures 6-9 and Tables 4-7 results with spatial step sizes set at 0.05, 0.02, 0.01, and 0.005 mile, respectively, are presented. The time steps used were also reduced in each case to satisfy the stability limit. For convenience the model output is still presented as averaged speed and densities over each of the ten original subsections. However, now each of these subsections is comprised of subdivisions (the spatial steps used in the program). The flow rate still represents the flow leaving each subsection.

The solution changes dramatically as each subsection is divided into smaller pieces. This evolution approaches a limit when the subsection size is 0.01 mile (52 ft). Further reduction in the step sizes does not alter the nature of the simulation result, only the detail observed in the transition regions. This experiment may be duplicated by altering the number of subsections (nsect), length of the subsections (x), and the time step (delt) in the FORTRAN program (Table 1).

In the final simulations (Figures 8 and 9 and Tables 6 and 7) the bottleneck capacity is adhered to and the bottleneck densities are typical for capacity flow. There is queueing on subsections upstream of the lane drop and an acceleration region exists in subsection 6. It should be emphasized that the only difference between the simulation of Figure 4 and Table 2 and that of Figures 8 and 9 and Tables 6 and 7 is in the length of the spatial discretization (and the associated temporal steps needed for stability). As the model becomes more properly dis-

TABLE 1 FORTRAN Program for Lane Drop Simulation

```
C       SIMPLE FREEWAY SIMULATION MODEL
        dimension x(200),lanes(200),u(200),rho(200),vol(200)
C       NUMBER OF SUBSECTIONS (FIRST, LAST ARE BOUNDARY CONDITIONS)
        nsect = 12
C       SET INITIAL CONDITIONS AND LANE NUMBERS
        do 10 i=1,nsect/2
         lanes(i) = 3
         rho(i) = 27.
   10    u(i) = 55.
        do 20 i=(nsect/2)+1,nsect
         lanes(i) = 2
         rho(i) = 50.
   20    u(i) = 40.
C       SET LENGTH OF SUBDIVISIONS AND INITIAL FLOWS
        do 30 i=1,nsect
         x(i) = 0.1
   30    vol(i+1) = rho(i)*u(i)
C       SET COEFFICIENTS, INPUT FLOW, TIME STEP, AND STOP TIME
        c = 0.075/3600.
        b = 1200.
        vol(1) = 4500.
        delt = 1.0/3600.
        tstop = 600./3600.
C
   40   continue
C       CALCULATE NEXT SPEED AND FLOW FOR EACH SUBDIVISION
        u(1) = u(2)
        vol(2) = rho(1)*u(1)
        do 50 i=nsect-1,2,-1
         vol(i+1) = rho(i)*u(i)
   50    u(i) = u(i) - delt * ( u(i)*2.*(u(i)-u(i-1))/
     1     (x(i)+x(i-1)) + (u(i)-ue(rho(i),i,nsect))/c +
     2     b*2.*(rho(i+1)-rho(i))/(rho(i)*(x(i+1)+x(i))) )
C       CALCULATE NEXT DENSITY FOR EACH SUBDIVISION
        rho(1) = rho(1) + delt*(vol(1)-lanes(1)*vol(2))/lanes(1)
        do 70 i=2,nsect-1
   70    rho(i) = rho(i) + delt * ( vol(i)*lanes(i-1) -
     1     vol(i+1)*lanes(i))/(x(i)*lanes(i))
        rho(nsect) = rho(nsect-1)
C       UPDATE TIME AND PRINT CURRENT STATES
        time = time + delt
        timsec = time*3600.
        write (6,90) (timsec,i,u(i),rho(i),vol(i+1),i=2,nsect-1)
   90   format (5x,f7.1,i6,3f10.2)
        if (time.lt.tstop) go to 40
        stop
        end
C
C
C       EQUILIBRIUM SPEED-DENSITY RELATION
        function ue(rho,i,nsect)
        d = rho
        if (rho.gt.100) d = 100.
        ue = ((-7.4e-5*d+0.0215)*d-2.31)*d+107.
        if (rho.gt.100.) ue = ue - ((rho-100.)*ue/70.)
C       SET CAPACITY AT 2000.
        ue = ue * (2000./1800.)
C       FOLLOWING LINE USED FOR DISCONTINUOUS RELATION
C       if (rho.gt.60.) ue = ue * 0.8
        if (ue.gt.55.) ue = 55.
        if (ue.lt.0.) ue = 0.
        return
        end
```



Legend For Contour Map

Lower Limit	Upper Limit	Symbol
0.	40.	
40.	50.	.
50.	60.	0
60.	80.	I
80.	120.	$
120.	1000.	#

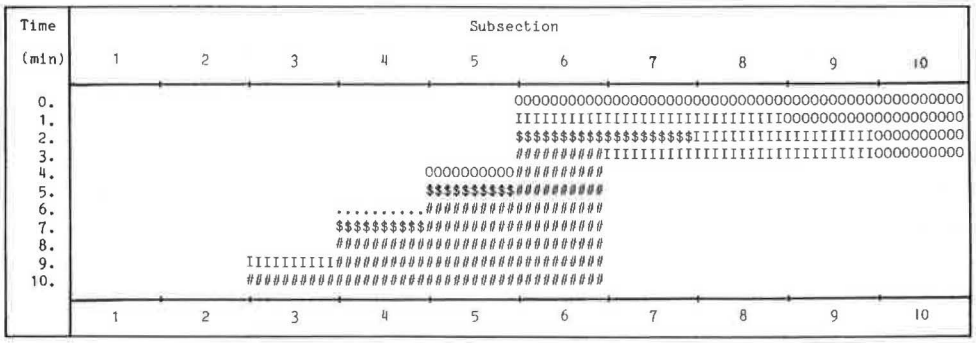
FIGURE 4 Unstable subsection averaged density (vehicles/mile-lane)—continuous equilibrium relation with $\Delta x = 0.1$ mile.

TABLE 2 Unstable Volume Leaving Subsections (Vehicles/Hour)—Continuous Equilibrium Relation with $\Delta x = 0.1$ Mile

Time	Subsection Number									
(min)	1	2	3	4	5	6	7	8	9	10
0	4455	4455	4455	4455	4455	4000	4000	4000	4000	4000
1	4468	4465	4463	4458	4410	4203	3848	3811	3533	3506
2	4487	4486	4484	4482	4479	4472	4448	4408	4313	4314
3	4495	4494	4494	4493	4492	4490	4488	4485	4478	4416
4	4498	4498	4497	4497	4497	4496	4495	4494	4493	4561
5	4499	4499	4499	4499	4499	4499	4498	4498	4497	4427
6	4500	4500	4500	4500	4500	4499	4499	4499	4499	4569
7	4500	4500	4500	4500	4500	4500	4500	4500	4500	4429
8	4500	4500	4500	4500	4500	4500	4500	4500	4500	4570
9	4500	4500	4500	4500	4500	4500	4500	4500	4500	4430
10	4500	4500	4500	4500	4500	4500	4500	4500	4500	4570

TABLE 3 Volume Leaving Subsections (Vehicles/Hour)—Continuous Equilibrium Relation with $\Delta x = 0.1$ Mile

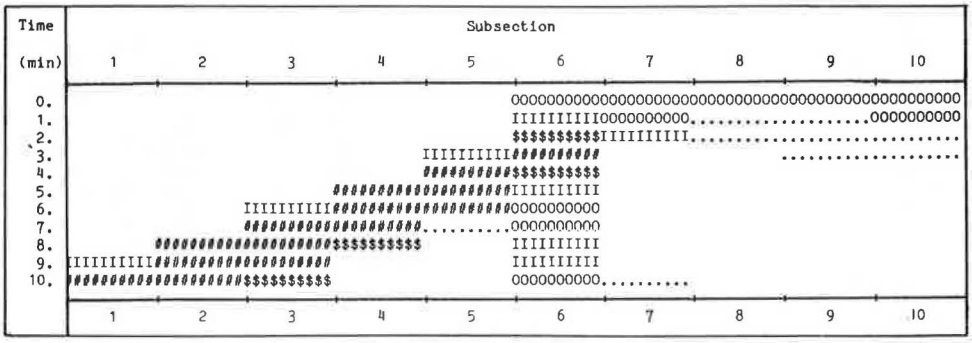
Time	Subsection Number									
(min)	1	2	3	4	5	6	7	8	9	10
0	4455	4455	4455	4455	4455	4000	4000	4000	4000	4000
1	4469	4465	4463	4460	4436	4168	4065	3708	3893	3567
2	4487	4486	4484	4482	4462	4167	3893	4126	3576	3978
3	4495	4494	4494	4492	4454	3640	3884	3828	4148	4002
4	4498	4498	4497	4485	4112	1430	1940	2509	3004	3759
5	4499	4499	4498	4458	3367	2176	2134	2056	2005	1962
6	4500	4499	4493	4301	514	2066	2116	2184	2212	2235
7	4500	4499	4467	3579	755	1492	1548	1580	1610	1699
8	4499	4496	4382	1281	1468	1423	1411	1424	1421	1426
9	4496	4474	3780	59	1112	1379	1400	1396	1407	1413
10	4484	4394	1748	843	1029	1266	1269	1293	1302	1315



Legend For Contour Map

Lower Limit	Upper Limit	Symbol
0.	40.	
40.	50.	.
50.	60.	0
60.	80.	I
80.	120.	$
120.	1000.	#

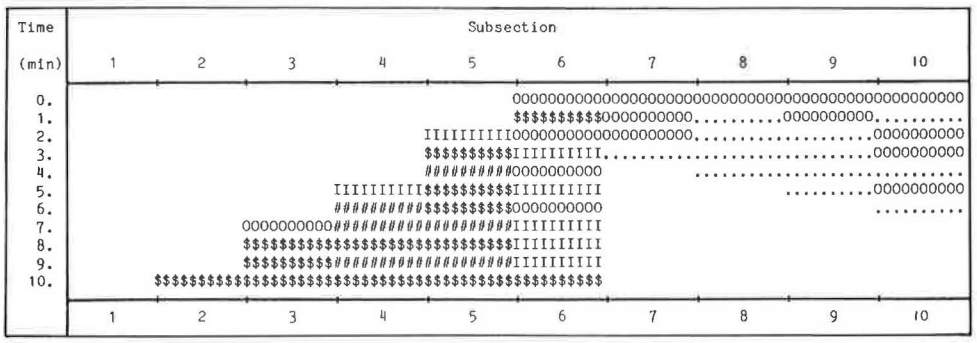
FIGURE 5 Subsection averaged density (vehicles/mile-lane)—continuous equilibrium relation with $\Delta x = 0.1$ mile.



Legend For Contour Map

Lower Limit	Upper Limit	Symbol
0.	40.	
40.	50.	.
50.	60.	0
60.	80.	I
80.	120.	$
120.	1000.	#

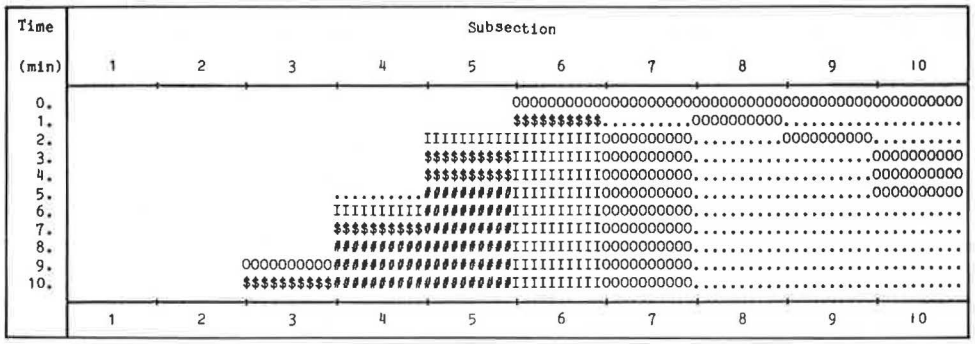
FIGURE 6 Subsection averaged density (vehicles/mile-lane)—continuous equilibrium relation with $\Delta x = 0.05$ mile.



Legend For Contour Map

Lower Limit	Upper Limit	Symbol
0.	40.	
40.	50.	.
50.	60.	0
60.	80.	I
80.	120.	$
120.	1000.	#

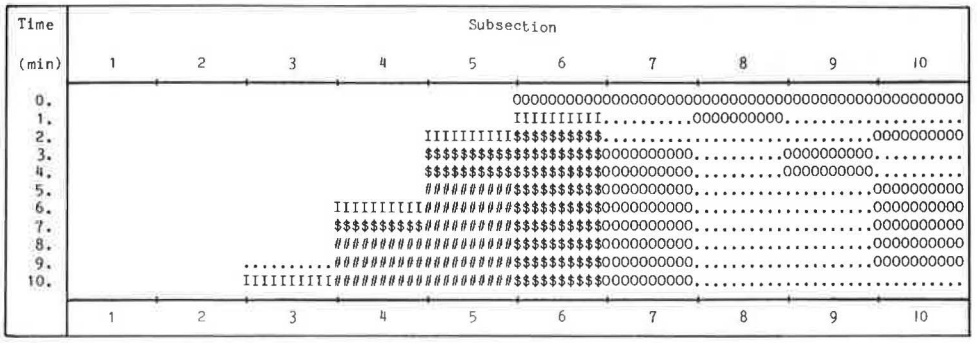
FIGURE 7 Subsection averaged density (vehicles/mile-lane)—continuous equilibrium relation with $\Delta x = 0.02$ mile.



Legend For Contour Map

Lower Limit	Upper Limit	Symbol
0.	40.	
40.	50.	.
50.	60.	0
60.	80.	I
80.	120.	$
120.	1000.	#

FIGURE 8 Subsection averaged density (vehicles/mile-lane)—continuous equilibrium relation with $\Delta x = 0.01$ mile.



Legend For Contour Map

Lower Limit	Upper Limit	Symbol
0.	40.	
40.	50.	.
50.	60.	0
60.	80.	I
80.	120.	$
120.	1000.	#

FIGURE 9 Subsection averaged density (vehicles/mile-lane)—continuous equilibrium relation with $\Delta x = 0.005$ mile.

TABLE 4 Volume Leaving Subsections (Vehicles/Hour)—Continuous Equilibrium Relation with $\Delta x = 0.05$ Mile

Time	Subsection Number									
(min)	1	2	3	4	5	6	7	8	9	10
0	4455	4455	4455	4455	4455	4000	4000	4000	4000	4000
1	4468	4465	4463	4461	4431	4085	3971	3988	4001	4001
2	4487	4486	4484	4483	4456	4056	3950	3953	3994	4009
3	4495	4495	4494	4490	3768	3122	3640	3903	3975	4010
4	4498	4498	4497	4428	1420	2102	2211	2338	2623	2970
5	4499	4499	4486	2189	1436	1860	1761	1714	1712	1733
6	4500	4497	3809	773	2946	3126	3058	2974	2889	2786
7	4499	4381	490	1625	3732	3727	3686	3622	3520	3396
8	4471	1568	898	3624	3863	3791	3767	3748	3731	3725
9	3398	499	1979	3691	3788	3804	3779	3775	3769	3772
10	106	1125	3653	3697	3747	3808	3784	3780	3777	3777

TABLE 5 Volume Leaving Subsections (Vehicles/Hour)—Continuous Equilibrium Relation with $\Delta x = 0.02$ Mile

Time	Subsection Number									
(min)	1	2	3	4	5	6	7	8	9	10
0	4455	4455	4455	4455	4455	4000	4000	4000	4000	4000
1	4468	4465	4463	4461	4431	3995	3978	4005	4000	4000
2	4487	4486	4485	4483	3626	3986	3950	3990	4011	3995
3	4495	4495	4494	4491	4062	3890	4015	3960	4004	3997
4	4498	4498	4498	4402	3761	3874	3949	4021	3975	4019
5	4499	4499	4499	3610	4052	3852	3859	3937	4032	3980
6	4500	4500	4493	3672	3530	3859	3845	3861	3915	4041
7	4500	4500	4014	4227	4067	3836	3862	3843	3858	3903
8	4500	4499	3446	3408	3747	3862	3837	3856	3845	3854
9	4500	4297	4286	4182	3854	3837	3857	3843	3852	3846
10	4500	3387	3450	3679	4069	3853	3846	3852	3845	3851

TABLE 6 Volume Leaving Subsections (Vehicles/Hour)—Continuous Equilibrium Relation with $\Delta x = 0.01$ Mile

Time	Subsection Number									
(min)	1	2	3	4	5	6	7	8	9	10
0	4455	4455	4455	4455	4455	4000	4000	4000	4000	4000
1	4468	4465	4463	4461	4388	3971	4000	4000	3999	4000
2	4487	4486	4485	4483	3942	3986	3953	4023	3991	4001
3	4495	4495	4494	4493	3974	3977	3968	3969	4016	3990
4	4498	4498	4498	4490	3973	3975	3975	3967	4020	3966
5	4499	4499	4499	4218	3974	3975	3975	3974	3979	3985
6	4500	4500	4500	3960	3975	3975	3975	3975	3975	4028
7	4500	4500	4500	3975	3975	3975	3975	3975	3975	4001
8	4500	4500	4459	3975	3975	3975	3975	3975	3974	3977
9	4500	4500	4041	3974	3975	3975	3975	3975	3975	3975
10	4500	4500	3977	3975	3975	3975	3975	3975	3975	3975

TABLE 7 Volume Leaving Subsections (Vehicles/Hour)—Continuous Equilibrium Relation with $\Delta x = 0.005$ Mile

Time	Subsection Number									
(min)	1	2	3	4	5	6	7	8	9	10
0	4455	4455	4455	4455	4455	4000	4000	4000	4000	4000
1	4468	4465	4463	4461	4258	3969	4008	3999	4000	4000
2	4487	4486	4485	4483	4050	3981	3971	3995	4004	4000
3	4495	4495	4494	4493	4012	3988	3955	4026	3987	4003
4	4498	4498	4498	4496	4000	3987	3967	3955	3989	3982
5	4499	4499	4499	4270	3993	3986	3973	3963	4036	3937
6	4500	4500	4500	3996	3989	3985	3977	3970	3991	3959
7	4500	4500	4500	3987	3987	3984	3979	3974	3971	3963
8	4500	4500	4492	3986	3985	3984	3980	3977	3973	3959
9	4500	4500	4135	3984	3984	3983	3981	3978	3976	3969
10	4500	4500	3985	3984	3984	3983	3981	3980	3977	4025

crete in space, the stability limit gradually increases to the value in Equation 6. This property can be confirmed by altering the value of the time step (delt) in the FORTRAN program (Table 1).

It is not surprising that simulation accuracy is dependent on the length of the spatial steps. As discussed earlier, spatial steps on the order of 0.01 mile would be needed to follow the changing slope near the queueing shock front. Use of larger intervals where the slope is changing leads to erroneous solutions that have no relation to the continuous formulation nor bear any resemblance to freeway behavior.

The spatial steps required by the model are a result of the form of the differential terms in the system equations and the model parameters. It has been found that these limits are quite insensitive to the shape of the equilibrium speed-density relation. For example, consider the alteration of the equilibrium speed-density relation shown in Figure 10. Ceder (8) has found that traffic flow may be better approximated by discontinuous, multiregime curves of this type.

Figure 11 and Table 8 show a simulation using this curve and spatial steps of 0.1 mile. The solutions evolve toward the simulation result shown in Figure 12 and Table 9. Here the spatial steps are 0.01 mile. Further reduction of the step sizes does not significantly alter the result. This process can

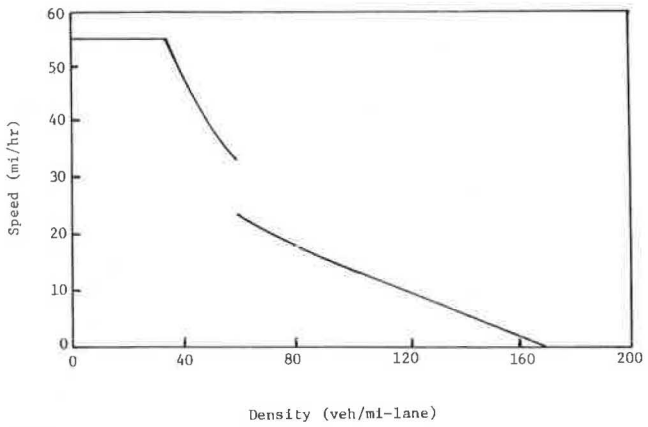
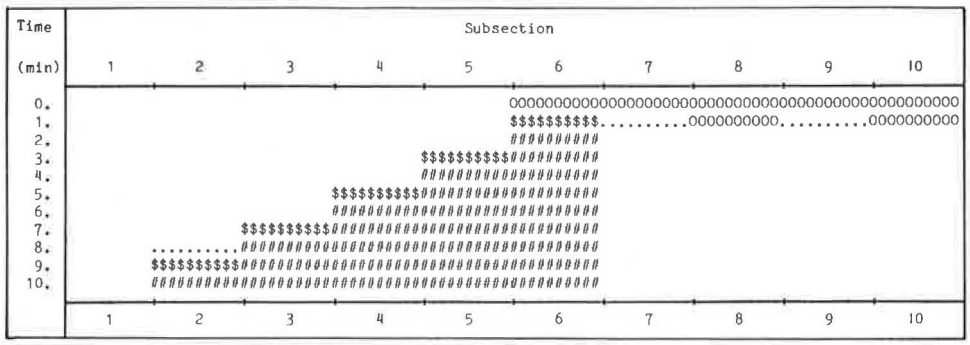


FIGURE 10 Discontinuous equilibrium speed-density relation.



Legend For Contour Map

Lower Limit	Upper Limit	Symbol
0.	40.	
40.	50.	.
50.	60.	0
60.	80.	I
80.	120.	$
120.	1000.	#

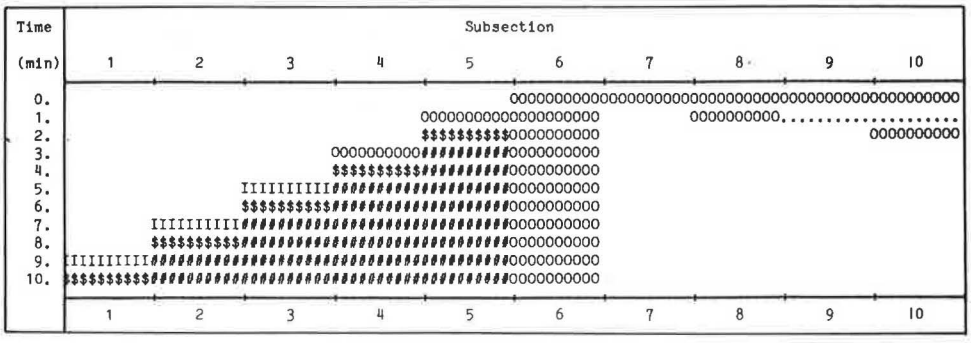
FIGURE 11 Subsection averaged density (vehicles/mile-lane)—discontinuous equilibrium relation with $\Delta x = 0.1$ mile.

TABLE 8 Volume Leaving Subsections (Vehicles/Hour)—Discontinuous Equilibrium Relation with $\Delta x = 0.1$ Mile

Time	Subsection Number									
(min)	1	2	3	4	5	6	7	8	9	10
0	4455	4455	4455	4455	4455	4000	4000	4000	4000	4000
1	4468	4465	4463	4459	4418	3838	3901	3743	3897	3562
2	4487	4486	4484	4478	4352	2201	2521	3137	3410	3829
3	4495	4494	4493	4462	3555	1840	1766	1679	1663	1782
4	4498	4498	4493	4348	828	1890	1921	1969	1981	1986
5	4499	4498	4465	3484	669	1986	1439	1470	1526	1581
6	4499	4494	4329	603	1201	1293	1287	1304	1305	1313
7	4495	4454	3223	270	944	1214	1235	1235	1248	1257
8	4467	4266	292	724	942	1121	1122	1142	1148	1159
9	4300	2882	156	595	920	1061	1076	1072	1081	1087
10	3613	172	488	624	880	1007	1006	1021	1025	1031

TABLE 9 Volume Leaving Subsections (Vehicles/Hour)—Discontinuous Equilibrium Relation with $\Delta x = 0.01$ Mile

Time	Subsection Number									
(min)	1	2	3	4	5	6	7	8	9	10
0	4455	4455	4455	4455	4455	4000	4000	4000	4000	4000
1	4468	4465	4463	4461	3896	3811	4012	3997	4000	4000
2	4487	4486	4485	4481	3576	3581	3589	3814	4011	4002
3	4495	4495	4494	3888	3580	3580	3580	3580	3589	3804
4	4498	4498	4495	3578	3580	3580	3580	3580	3580	3580
5	4499	4499	3879	3580	3580	3580	3580	3580	3580	3580
6	4500	4494	3579	3580	3580	3580	3580	3580	3580	3580
7	4500	3827	3580	3580	3580	3580	3580	3580	3580	3580
8	4490	3580	3580	3580	3580	3580	3580	3580	3580	3580
9	3771	3580	3580	3580	3580	3580	3580	3580	3580	3580
10	3580	3580	3580	3580	3580	3580	3580	3580	3580	3580



Legend For Contour Map

Lower Limit	Upper Limit	Symbol
0.	40.	
40.	50.	.
50.	60.	0
60.	80.	I
80.	120.	$
120.	1000.	#

FIGURE 12 Subsection averaged density (vehicles/mile-lane)—discontinuous equilibrium relation with $\Delta x = 0.01$ mile.

be duplicated by activating the commented "if" statement in the FORTRAN program (Table 1).

One curious result of this simulation is that the bottleneck is not filled to capacity. This is due to the upstream congested region reaching equilibrium on the discontinuous curve of Figure 10. This equilibrium cannot simultaneously satisfy continuity at the lane drop and provide flows high enough to fill the bottleneck. This situation may not occur in real traffic because the equilibrium relation used here was not derived from traffic data. However, the discretization limits remained the same in the face of these very different traffic flows.

With spatial step sizes of 0.01 mile required to follow changes in the density-space slope, the possibility of unmanageable input and output arises. This is avoided by permitting the user to use standard subsections to define the freeway geometry and demands. These subsections are related to physical divisions and have no direct bearing on the required mathematical discretization. The model internally divides each subsection into the necessary subdivisions so that model accuracy is maintained. The results are printed out as subsection values to keep the output readable.

A further difficulty that arises from the use of proper discretization is that of computing time. Consider a simulation of the Santa Monica Freeway in Los Angeles. The site is 7.7 miles long and a 3-hr period is to be simulated. The use of spatial steps of 0.01 mile and their associated temporal steps of 0.2 sec results in 770 subdivisions requiring 54,000 time steps to complete the simulation. This process would use 40 central processing unit (CPU) hours of computing time on a DEC VAX11/750. It would be advantageous to find a way of reducing this computing time while still maintaining the discrete model's integrity.

HEURISTIC SCHEME FOR ADAPTING THE DISCRETIZATION

Although the use of very small spatial steps is required to maintain the model's accuracy, these steps are not required at all times and at all locations during a simulation. This idea motivates the use of an adaptive scheme. The term that most restricts the maximum spatial step needed is the anticipation term (Equation 7). It implies that small steps are required to follow the changing density-space slope. Further, this process will become more critical at higher densities. It is possible to exploit some of the freeway flow properties to develop a heuristic scheme for placing subdivisions within the subsections only where and when they are required to maintain the model's accuracy. This adaptive scheme is at the heart of the FRECON model.

Consider again the contour of density for the congested bottleneck (Figure 2). In the region from x = 0.0 to 0.25, large subdivisions could be used because the density is low and the slope is constant. The shock front at x = 0.25 requires very short steps because the slope changes abruptly. The queued region (x = 0.3 to 0.5) also needs short divisions because the density is large. The acceleration region (x = 0.5 to 0.7) has a changing slope so it too requires small subdivisions. The capacity flow area (x = 0.7 to 1.0) needs only moderate sized subdivisions because of its density level.

Similar examinations can be made of other simple geometries. In general, all freeway situations can be divided into the following classes: lane drop (or heavy on-ramp), lane add (or heavy off-ramp), capacity flow, queued flow, and incident. Each has a pattern of maximum subdivision lengths associated with it. These are required when the flow rises above a given threshold inducing sharp gradients in the density-space profile. Therefore, it is possible to select the subdivision pattern by examining the geometry and the current flow level. When all the patterns for a given time period have been established, they are placed on the freeway to provide the needed spatial discretization.

This process is repeated at regular, user-set intervals (a typical value is every minute) so that the freeway is always properly, but not excessively, subdivided. In cases where the input demands vary at a higher frequency, the adaptation of the subdivisions can be performed at a comparable frequency or a worst-case can be assumed and the adaptation can take place at a lower frequency.

Each time the subdivision patterns are laid down on the freeway subsections, care must be taken that the original geometry is not altered. Also, in regions where the subdivision patterns overlap, the limiting step sizes (i.e., the smallest of the steps) must be maintained. Each time new patterns replace old patterns the new ones are initialized to maintain the spatial speed and density profile.

When there are subdivisions of varying sizes present, it becomes advantageous to integrate each one at its maximum allowed time step. This step is set by the stability limit (Equation 6). Such integration is accomplished through nonsequential calculations of the subdivisions. In this scheme the time argument of each state variable remains as near to the others as possible. The time arguments of all state variables are identical at user-set intervals so that valid outputs can be printed.

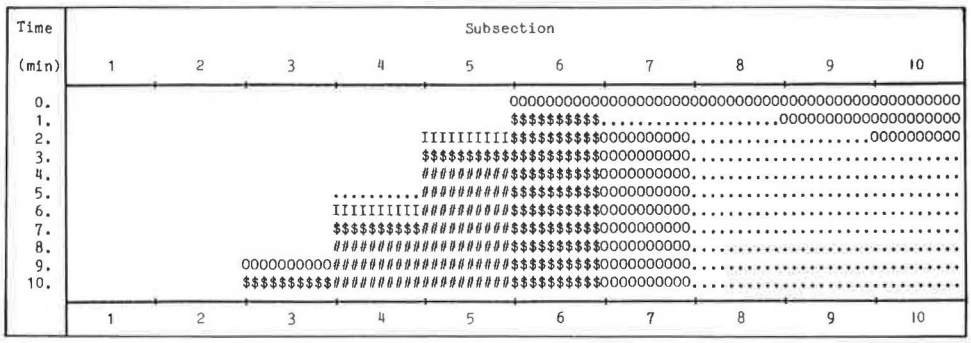
For example, consider a region of length 0.01 mile between regions of 0.1 and 0.2 mile. At the start each region would be integrated with its respective time step. Then the 0.01-mile region would be stepped forward in time by nine of its steps. At this point both the 0.01- and the 0.1-mile regions would be calculated. Then the 0.01-mile region would be integrated nine more times. Finally, all three regions would be calculated.

These asynchronous calculations are valid only if the subdivisions are properly discrete. In this case the only subdivisions that change quickly are the short ones. The stability limit ensures that these regions are integrated at a rate that maintains causality. The validity of this scheme can be shown by comparing results of nonadapted simulations with those of simulations where the adaptation is used.

Return to the lane drop simulation of Figure 3. In the acceptable simulation of Figure 8 and Table 6 the freeway was divided into 100 subdivisions each with a length of 0.01 mile. The stability limit restricts the time steps for these divisions to 0.2 sec. Therefore, the 10-min simulation requires 100 subdivisions to be calculated 3,000 times each. These 300,000 calculations of the state equations took 4.3 min of CPU time on a DEC VAX11/750.

Figure 13 and Table 10 give the simulation results for the same situation using the adaptive scheme. The spatial and temporal steps were adapted at each minute during the simulation. The total number of calculations of the state equations was 3,700. The simulation result remained virtually unchanged from the nonadaptive one, and the running time was reduced to 50 sec.

In general, the use of this spatial and temporal step size adaptation scheme results in large savings of computer time and does not jeopardize the model's accuracy. The savings is even greater in simulations of real freeways than in the lane drop example. This is because there are often large regions where there is no congestion; longer subdivisions can be used in these areas. For example, the 3-hr simulation of the



Legend For Contour Map

Lower Limit	Upper Limit	Symbol
0.	40.	
40.	50.	.
50.	60.	0
60.	80.	I
80.	120.	$
120.	1000.	#

FIGURE 13 Subsection averaged density (vehicles/mile-lane)—continuous equilibrium relation with Δx = adaptive.

Santa Monica Freeway (February 17, 1981) now would require 1 hr 50 min of CPU time on a DEC VAX11/750. This is a reduction by a factor of 21 from the non-adaptive simulation. Of this CPU time, only 3 percent is used in setting the adaptive spatial and temporal steps. The run time is not a fixed value because the number of subdivisions needed with the adaptive module depends on the congestion level.

There are some interesting benefits to using the adaptive scheme included in FRECON. First, it is possible to place special subdivisions at detector locations. They give true local information as would a real roadway detector. The ability to include realistic detectors in a macroscopic model greatly influences the results of tests with on-ramp controllers that use detector information. Second, because each subdivision acts as a miniature subsection within the model, a selected subdivision can be used to duplicate capacity loss during an incident. In this way the incident is localized instead of spread over an entire subsection. These alterations are handled within the adaptive module and require only conventional user inputs.

This adaptive scheme has been successful, but the adaptation relies on presumed consequences of given geometries and flows. Therefore, it must by nature be conservative. If there is any doubt, shorter subdivisions are applied. This ensures that the next flows examined by the adaptive module are, in fact, valid. The following adaptive scheme avoids this overcautiousness and results in a more elegant and natural discretization.

A MORE NATURAL ADAPTATION

To avoid the conservative subdividing present in the heuristic adaptation scheme and to provide a clearer and more natural adaptation, the model has been reformulated into a Lagrangian reference frame (9). Here all changes are measured relative to a particle in motion with the local flow. This differs from the stationary (Eulerian) reference frame used in many fluid flow problems. In the Lagrangian frame the discretization can become a natural part of the state equations, causing many of the adaptive problems to disappear.

Start by transforming the model of Equations 1 and 2 into the Lagrangian frame:

TABLE 10 Volume Leaving Subsections (Vehicles/Hour)—Continuous Equilibrium Relation with Δx = Adaptive

Time	Subsection Number									
(min)	1	2	3	4	5	6	7	8	9	10
0	4455	4455	4455	4455	4455	4000	4000	4000	4000	4000
1	4468	4465	4463	4461	4368	3963	3975	4003	4001	4000
2	4487	4486	4484	4483	3938	3980	3941	3972	4011	4003
3	4495	4494	4494	4493	3969	3971	3965	3957	3985	4020
4	4498	4498	4498	4489	3968	3970	3969	3966	3970	4004
5	4499	4499	4499	4202	3969	3970	3970	3969	3970	3980
6	4500	4500	4500	3957	3969	3970	3970	3970	3970	3972
7	4500	4500	4499	3970	3969	3970	3970	3970	3970	3970
8	4500	4500	4435	3970	3969	3970	3970	3970	3970	3970
9	4500	4500	4048	3969	3969	3970	3970	3970	3970	3970
10	4500	4500	3970	3969	3969	3969	3970	3970	3970	3970

$$(\partial\rho/\partial t) + [\rho(\partial u/\partial x)] = S(x,t) \quad (8)$$

$$(\partial u/\partial t) = [(1/c)u_e(\rho) - u] - [b(1/\rho)(\partial\rho/\partial x)] \quad (9)$$

In this form the convective terms are no longer present because the observer moves with the flow.

Before the discretization of the model, consider a local region of constant mass (vehicles) but changeable length and density:

$$V = \rho \cdot \delta x \quad (10)$$

where

V: number of vehicles in region δx and
δx: length of region.

If the problems of on- and off-ramps are ignored for the moment (which would alter the fixed number of vehicles in the region), it is possible to transform the conservation relation (Equation 8) into a form based on the length of the constant vehicle region:

$$[\partial(\delta x)/\partial t] = (\delta x) \cdot (\partial u/\partial x) \quad (11)$$

Further, the length of the region can become the model's discrete spatial step. These discrete regions will be referred to as a "box" to emphasize their constant vehicle number (when not near a ramp). In this case the discrete model is

$$\delta x(x,t+\Delta t) = \delta x(x,t) + \Delta t\,[u(x+\Delta x,t) - u(x,t)] \quad (12)$$

$$u(x,t+\Delta t) = u(x,t) + \Delta t \left((1/c)[u_e(\rho) - u(x,t)] - b \cdot [1/\rho(x,t)]\,\{[\rho(x+\Delta x,t) - \rho(x,t)]/\Delta x\}\right) \quad (13)$$

where

$$\rho(x,t) = V/\delta x(x,t)$$

Note that the discrete regions (boxes) are no longer fixed on the freeway; they now travel with the flow and lengthen or shorten as the local density changes. It is this natural altering of the box lengths that provides the spatial adaptation.

In this particular model, special care must be taken when each box passes a ramp (the number of vehicles in the box changes) and when it passes a lane drop or add (the box's density changes). In addition to solving Equations 12 and 13, each box's position must be updated at each time step. The critical length of the boxes is still 0.01 mile (at their highest density) and the stability limit remains the same as in the heuristic scheme.

The use of temporal adaptation with this form of the model also requires care. Previously, all subdivisions maintained their size for a set period of time. In this scheme, the boxes vary their lengths at each time step. Still it is possible to find time periods (5 to 10 sec) over which the boxes do not change their sizes significantly. A temporal adaptive scheme like the one used in the heuristic method can be used if it is updated at intervals of 5 to 10 sec.

In implementing this adaptive scheme, a few points must be considered. The upstream boundary condition is now provided by the creation of boxes waiting to approach the freeway section. The downstream boundary condition is met by allowing boxes to completely exit the freeway before they are no longer considered part of the simulation. The number of vehicles in the boxes is set by the queueing density, the number of lanes, and the maximum allowed spatial step during congestion (0.01 mile for this model). As the boxes pass ramps the number of vehicles they contain changes. These changes must be monitored at regular intervals so the discretization limit will not be violated. When a limit violation occurs, the boxes on the freeway are reformed and initialized so that the simulation progresses smoothly and the model's accuracy is maintained.

Natural adaptation of the spatial steps permits the use of realistic detectors and localized incidents as in the heuristic scheme. The inherent nature of the spatial adaptation and the accuracy that results reduce the computing times for this scheme. However, the temporal steps must be set more often because the spatial steps vary continually. The net effect is that simulation tests have shown this method to give accurate results with savings of 10 to 20 percent over the heuristic adaptive scheme. In cases where the demands have a high frequency component, this natural adaptation is better suited and gives savings as high as 50 percent in the run times.

CALIBRATION AND VALIDATION OF FRECON

The preceding discussion of how to properly discretize a freeway model is of interest only if the resulting model can duplicate the behavior of a real-world freeway. The Santa Monica Freeway (I-10) in Los Angeles was selected as the test site for the calibration and validation of the FRECON model. This selection was greatly influenced by the desire to use a site that would also be compatible with later comparative studies of control strategies (1).

The study site is a 7.7-mile, eastbound section from the San Diego Freeway (I-405) heading toward downtown (Figure 14). The hours of operation examined were the "counterpeak" from 3 p.m. to 6 p.m. This section is part of the Los Angeles 42-mile-loop surveillance system so main-line and ramp-detector information was available. Data sets were provided for 5 weekdays during the winter of 1981 that were believed to be normal and incident free. Each contained main-line input and on- and off-ramp flows at 5-min intervals. Also, a set of 35 mph contours from the main-line detectors was provided for each day. The first day of data was to be used to calibrate the model's capacities and the remaining 4 were for validating the model's performance. The roadway geometry and detector locations used in the modeling were obtained from detailed maps and field visits.

Using the first day of data (February 17, 1981), the model's capacities were tuned. The calibration was performed by comparing the field detector's 35-mph contour with that of the model's detectors. Figure 15 shows the result of this comparison for the calibrated capacities. The 35-mph contour is generally a reliable indicator of the extent of main-line queueing. In general, the model and field contours show a similar extent of main-line queueing.

The unusual field contour at detectors 9 and 10 (subsection 11) is due to an unusual field geometry. Subsection 11 is bypassed by a collector-distributor road that also carries a high percentage of through traffic. This causes subsection 11 to be a bottleneck. The field configuration is modeled with the collector-distributor through traffic being carried on the model's main line. The unusual congestion contour is caused by merging patterns. This results in a model capacity for subsection 11 that is 16 percent lower than that of neighboring subsections. This modeling approximation cannot duplicate all of the field behavior around detectors 9 and 10.

When the model capacities were once calibrated on the first day's data, they remained fixed for the validation study. Figure 16 shows a comparison of the model and field 35-mph contours for one of the validation days (February 23, 1981). As with the calibration day, the two contours show similar patterns for the upstream bottleneck (detectors 1-4). Also, there is difficulty matching the field pattern around detectors 9 and 10 because of the modeling of the field geometry. The other three validation days also showed field-model contour compatibility as shown in Figure 16.

The five simulations of the Santa Monica Freeway resulted in a range of total travel times from 3,321 to 4,528 vehicle-hours and total services from 163,000 to 176,000 vehicle-miles. For all five simulations the model was able to duplicate the overall field congestion pattern. This was possible over this range of traffic demands because of two properties of the model. First, it is properly discrete. Hence, it can follow the density-space profile that is present during congestion. Second, the model uses localized point detectors. It is possible to duplicate the sudden changes in detector state as the main-line queue passes over only if the detector has a restricted, local range.

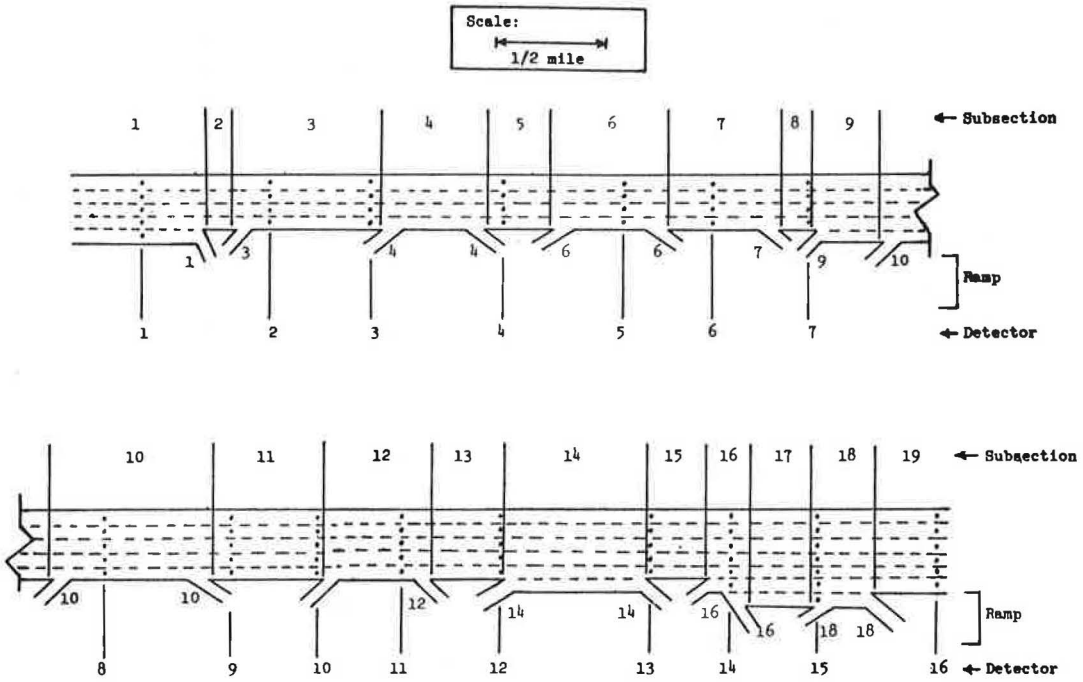


FIGURE 14 Santa Monica Freeway geometry.

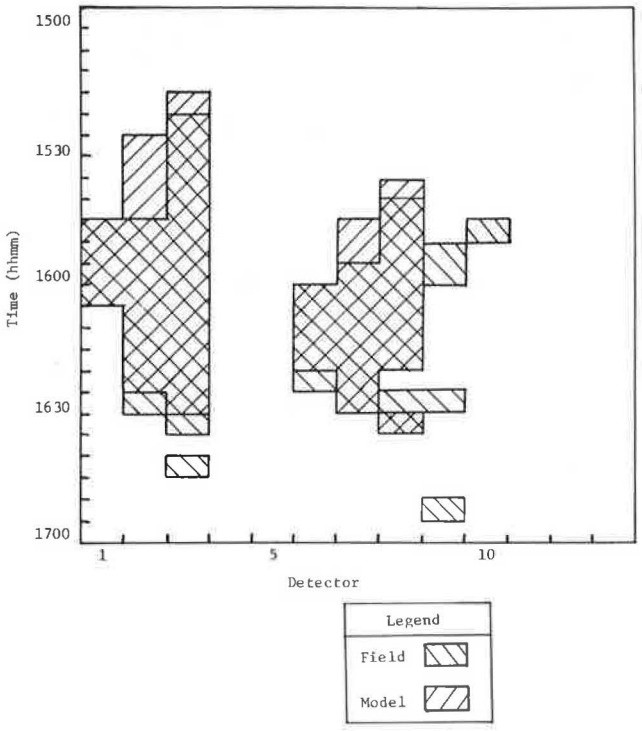


FIGURE 15 Santa Monica Freeway detectors < 35 mph—February 17, 1981.

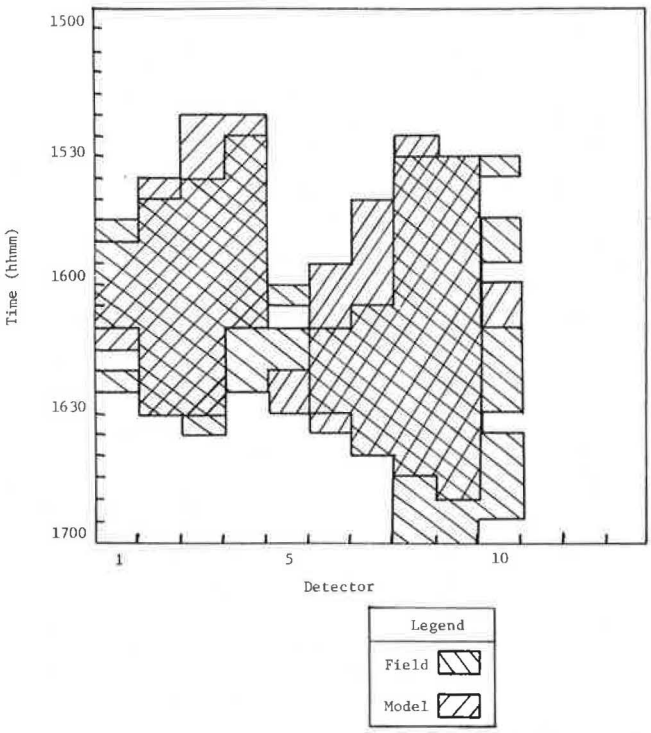


FIGURE 16 Santa Monica Freeway detectors < 35 mph—February 23, 1981.

SUMMARY

All discrete models have limits on the maximum allowable step sizes. This limit must be adhered to if the model is to be a faithful representation of the original continuous model or the real world. If the intervals are larger than the permitted maximum, the simulation results are irrelevant and merely an artifact of improper discretization. Proper discretization creates a model the results of which are an accurate approximation of the continuous solution. The model must be made properly discrete before its parameters can be tuned and the resulting model calibrated.

When freeway models have terms like those in FREFLO, the proper discretization is not the traditional subsection. Although the concept of a subsection (the region between geometric features on the

roadway) is useful for model input and output, it is usually an invalid mathematical discretization. In particular, the anticipation term of the model limits the size of the spatial steps permitted so that changes in the density-space gradient can be followed. To ensure that shock fronts at the boundary of main-line queueing are modeled properly, spatial step sizes must be less than or equal to 0.01 mile. This subdividing is made necessary by mathematical limitations. It can be performed internally in the model so that the user can continue to interact with the model using traditional geometric subsections.

The direct application of this proper spatial discretization and the proper temporal step sizes needed to maintain the model's stability result in excessive computer run times. Because the discretization is limited by changes in the density-space slope, it is possible to find circumstances where the spatial step sizes may be lengthened without an adverse effect on the simulation results. These occur when the density is low and the density-space gradient is constant. When this situation is identified, it is possible to adjust the spatial and temporal steps the model uses, thereby increasing its efficiency.

These ideas motivated the development of the adaptive modules used in the FRECON model. The first is a heuristic scheme where flow levels and geometry are examined and compared with a library of required subdivision patterns. In this way the spatial steps are adjusted at regular intervals so that the maximum allowed steps are always used. When these spatial steps have been established, the corresponding maximum temporal steps that ensure stability are known. Therefore, by using these asynchronous time steps, it is possible to reduce the number of times each subdivision state needs to be integrated.

The second adaptive scheme is more natural because it is built into the mathematics. By transforming the model from a stationary reference frame to a moving one, it is possible to define regions that move with the flow of traffic. Each region is discretized into a "box" that contains a constant number of vehicles and a length that is inversely proportional to its density. This natural variation of length with density provides the spatial adaptation. A scheme similar to the heuristic adaptation is used to reduce the frequency of the integration steps required.

Both of these adaptation schemes have proven useful in reducing the computing time required for a simulation. The first scheme, used in the FRECON model, reduced the CPU time for a typical simulation by a factor of 21. Further, the presence of an adaptive scheme permits the use of realistic local detectors and localized incidents in the model.

The FRECON model was calibrated on the Santa Monica Freeway for a 3-hr peak period. The calibration of model capacities was performed by comparing field and model detector 35-mph contours. These capacities, and the model's performance, were then validated using four more peak periods of data from the same freeway. With the exception of a complex collector-distributor geometry, which could not be accurately modeled, the simulation results generally agreed with the field data.

The performance of the FRECON model on the 4 days of validation indicated that the model can duplicate freeway behavior over a wide range of demand levels and can realistically model the response of localized detectors. These properties have proved helpful in studying the use and design of freeway control strategies.

ACKNOWLEDGMENT

This research was supported in part by the California Department of Transportation and the FHWA under grant F80-T003.

REFERENCES

1. P.S. Babcock, A.D. May, D.M. Auslander, and M. Tomizuka. Freeway Simulation and Control. Research Report UCB-ITS-RR-82-13. Institute of Transportation Studies, University of California, Berkeley, 1982.
2. H.J. Payne. Models of Freeway Traffic and Control. Mathematics of Public Systems, Vol. 1, No. 1, 1971, pp. 51-61.
3. H.J. Payne. FREFLO: A Macroscopic Simulation Model of Freeway Traffic--A User's Guide. ESSCOR, San Diego, Calif., 1978.
4. H.J. Payne. FREFLO: A Macroscopic Simulation Model of Freeway Traffic. In Transportation Research Record 722, TRB, National Research Council, Washington, D.C., 1979, pp. 68-77.
5. W.F. Phillips. A Kinetic Model for Traffic Flow with Continuum Implications. Transportation Planning and Technology, Vol. 5, 1979.
6. W.F. Phillips. A New Continuum Model for Traffic Flow. Final Report DOT-RC-82018. Utah State University, Logan, 1979.
7. H.J. Payne, W.A. Thompson, and L. Isaksen. Design of a Traffic Responsive Control System for an L.A. Freeway. Transactions on Systems, Man, and Cybernetics, Vol. SMC-3, No. 3, May 1973, pp. 213-224.
8. A. Ceder. A Deterministic Traffic Flow Model for the Two-Regime Approach. In Transportation Research Record 567, TRB, National Research Council, Washington, D.C., 1976, pp. 16-30.
9. K.R. Symon. Mechanics. Addison-Wesley, Reading, Mass., 1971.

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

Simulation Study of Guidelines for Rural Road Improvements

C. J. HOBAN

ABSTRACT

The rural traffic simulation model TRARR was used to evaluate a range of options for improving traffic performance on two-lane rural roads in Australia. The options include auxiliary lanes, widening to four lanes, and reconstruction on an improved alignment. Simulated travel times, bunching rates, and overtaking rates are presented for these options for a range of terrains and traffic volumes. Benefit-cost analysis was used to derive a volume warrant for each option, based on construction costs and reductions in travel times and accidents. Minimum volume warrants of 385 to 6,800 vehicles per day were found, depending on the terrain, the existing road standard, and the type of road improvement considered. Practical warrants are likely to be higher than these minimum values, but the relative rankings appear to be robust.

The rural traffic simulation model TRARR (1) was developed at the Australian Road Research Board and is now being used by a number of organizations in Australia and New Zealand. In recent years the model has been applied to several case studies of specific two-lane road sections to evaluate various road improvement options.

The case studies have considered widely different terrains and design standards, with different traffic volumes and compositions. The improvement options have included reconstruction on improved alignment, the provision of auxiliary lanes, and upgrading to four-lane divided-road standard (2-4). A structured simulation experiment that was undertaken to develop a broad set of guidelines for rural road improvements using a common basis is described.

TRAFFIC SIMULATION

Detailed descriptions of the TRARR simulation model are given elsewhere (5,6). The model has been calibrated to some extent, and its output is reasonable and consistent. A thorough validation is currently in progress, and Figure 1 shows a typical comparison of observed and simulated traffic performance. The figure shows observed traffic bunching over 15-min periods, compared with the mean and range of simulated bunching results for the same road and traffic conditions. A vehicle was considered to be following in a bunch if its headway to the preceding vehicle was less than 3.5 sec.

Overall results to date suggest that the model tends to overpredict traffic bunching on two-lane roads and underpredict that on four-lane roads. The model may therefore overestimate delays on an existing road and the expected benefits of various improvement options. The errors, however, do not appear to be large compared with other uncertainties in the evaluation of road projects.

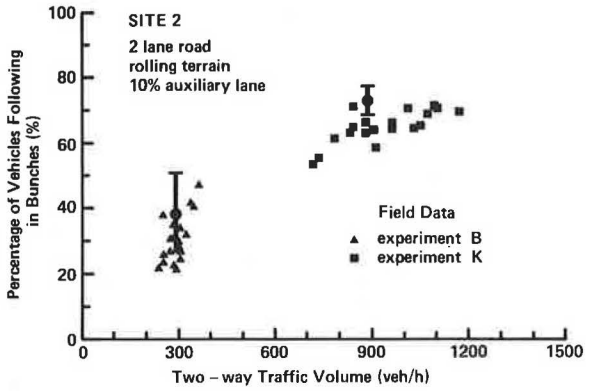


FIGURE 1 Comparison of observed and simulated traffic bunching: site 2.

SIMULATION STUDY FRAMEWORK

Road Alignment and Terrain

Three 12-km road segments were chosen:

1. Straight level road (SLR): an idealized perfect road with no grades, curves, barrier lines, or restrictions on sight distance or speed.
2. Maroondah Highway, Victoria (MHV): a segment of rural highway in rolling terrain with 110 km/hr design speed. Seventy-eight percent of the road has sight distance less than 450 m, and 26 percent has grades in excess of 4 percent.
3. Bass Highway, Tasmania (BHT): a segment of rural highway with fairly low geometric standards and few opportunities for overtaking. This segment has an average highway speed of 71 km/hr and 98 percent of sight distance less than 450 m; 46 percent of its length has grades over 4 percent.

Road Improvement Options

The first option to be simulated on each road was the existing condition. Seven improvement options were then considered, as shown in Figure 2. The first four of these are 500-m auxiliary lanes at various spacings. The remaining three provide for widening to four lanes over part or all of the road length. The effects of improved road alignment may be investigated by comparing the three different road segments chosen for analysis. The investigation of auxiliary lane and partial four-laning options was strongly recommended in previous studies (3,4).

Traffic Conditions

Traffic volumes of 200, 800, and 1,600 vehicles per hour were simulated, with 2:1 directional split and

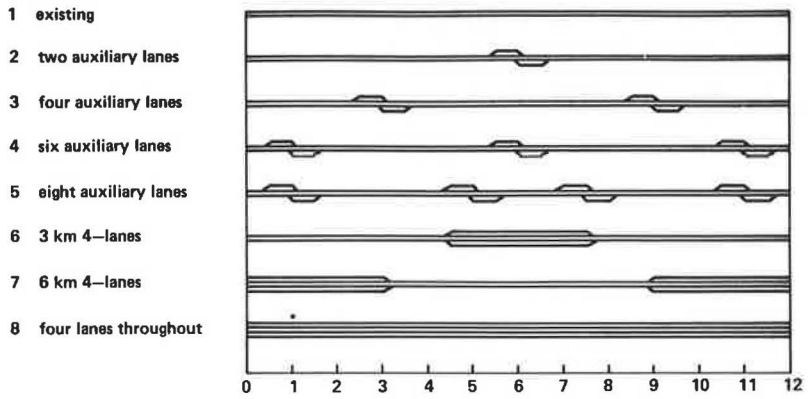


FIGURE 2 Road improvement options.

10 percent trucks in the traffic stream. The simulation time and settling down times were varied so as to observe approximately the same 3,000 vehicles for each terrain, road option, and traffic volume.

SIMULATED TRAFFIC PERFORMANCE

Travel Time

Figure 3 shows the travel time over 12 km for each simulated case. On the existing road (option 1), travel times increase with traffic volume and with decreasing standard of road alignment. Travel times are then reduced in varying degrees by road improvement options 2-8. Several points may be noted.

1. The road improvements yield the greatest travel time savings where the existing road alignment is poorest;
2. Travel time savings generally increase with traffic volume;
3. Four-lane roads show no increase in travel times with increasing traffic volumes up to 1,600 vehicles per hour;
4. Improvement options 2-7 offer intermediate levels of traffic performance between two-lane and four-lane roads.
5. The options of improved road alignment are shown in Figure 4. The existing Maroondah Highway, for example, offers lower travel times than the Bass Highway with closely spaced auxiliary lanes.

Time Spent Following in Bunches

The mean percentage of journey time spent following (at headways less than 3.5 sec) is shown for each case in Figure 4. Time spent following reflects the degree of traffic bunching along the road and is a useful measure of the quality of service as perceived by the driver.

At 200 vehicles per hour, bunching on the straight level road is generally below 10 percent. This rises to 15-30 percent on the Maroondah Highway, and 25-55 percent on the Bass Highway. Even at this low traffic volume, it would appear that bunching can be quite substantial on roads with poor geometric standards. In all cases, bunching rises substantially with increasing traffic volume, although the extent of bunching is still quite dependent on road alignment. As might be expected, simulated bunching on four-lane roads is quite low over the range of traffic volumes considered.

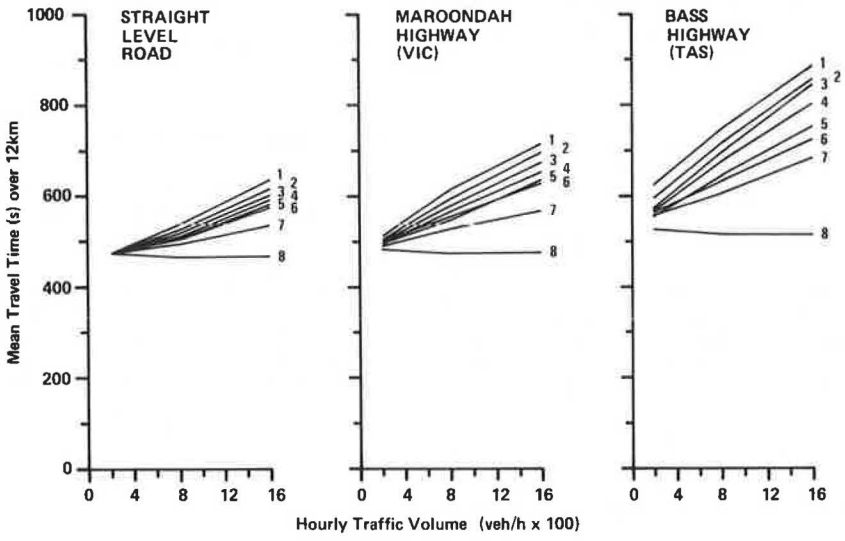


FIGURE 3 Simulated travel times.

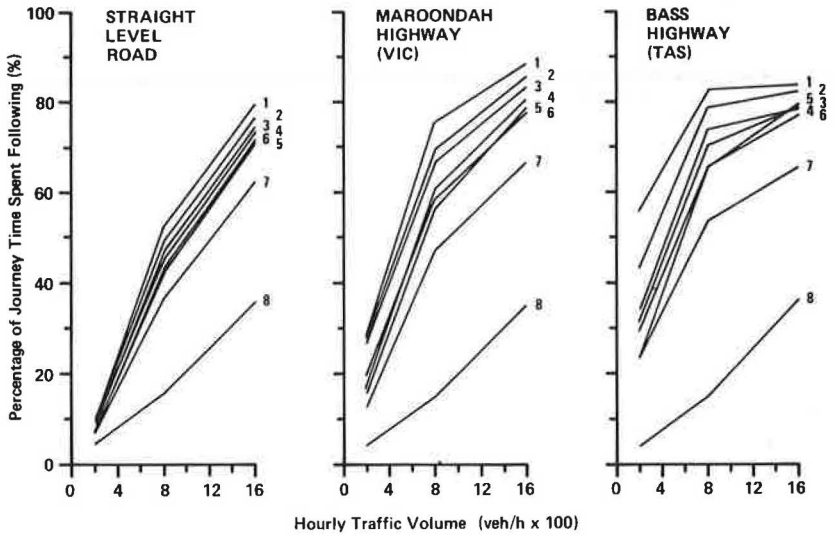


FIGURE 4 Simulated percentage of journey time spent following.

BENEFIT-COST ANALYSIS

The benefit-cost analysis procedure is described in some detail elsewhere (7). Benefit-cost (B/C) ratios were calculated for 36 road improvement options for a range of average annual daily traffic (AADT) volumes. The ratios ranged from 0.0 to 19.0, depending on terrain, improvement type, design standard, and traffic volume.

Table 1 gives, for selected options, the minimum volume warrant, Qw, which is defined as the initial-year AADT volume for which the benefit-cost ratio is 1.0. The values of Qw in Table 1 vary widely, from 385 to 6,800 vehicles per day, and terrain has a strong influence on these results. The minimum warrant for a pair of auxiliary lanes, for example, is 385 vehicles per day on the Bass Highway, 1,335 vehicles per day on the Maroondah Highway, and 2,570 vehicles per day on the straight level road.

TABLE 1 Estimated Minimum Volume Warrants for Selected Options

Option	Initial-Year AADT (veh/day)		
	Bass Highway, Tasmania	Maroondah Highway, Victoria	Straight Level Road
8 percent auxiliary lanes	385	1,335	2,570
25 percent auxiliary lanes	625	1,670	3,200
25 percent four lanes	1,670	3,500	5,430
Four lanes throughout	3,670	5,000	6,800
Major realignment	3,330	6,500	

DISCUSSION

The B/C analysis is based on a number of uncertain assumptions regarding construction costs, accident reductions, traffic growth, and other parameters. Sensitivity analysis (7) showed that the values of B/C and Qw are sensitive to changes in these assumptions, but the relative rankings of various options are robust. The uncertainties in evaluation assumptions appear to be larger than any likely errors in the simulation model predictions, so that the TRARR model is quite suitable for use in this type of analysis.

The model, however, appears at present to have a bias that leads to overestimation of the benefits of improvements on two-lane roads. The minimum volume warrants in Table 1, therefore, probably underestimate true minimum warrants. Practical minimum warrants are probably higher still, because these are usually based on B/C ratios greater than 1.0.

Despite these reservations, the results in Table 1 indicate that short auxiliary lanes offer a low-cost road improvement option that can be warranted at quite low traffic volumes. The use of auxiliary lanes and four-laning on existing alignments may in many cases be preferable to realignment on substandard two-lane roads.

ACKNOWLEDGMENT

The author would like to thank the Executive Director of the Australian Road Research Board, M.G. Lay, for permission to publish this paper.

REFERENCES

1. G.K. Robinson. A Model For Simulating Traffic on Rural Roads. Technical Manual ATM 10. Australian Road Research Board, Numawading, Victoria, 1980.
2. G.K. Robinson. Use of a Traffic Simulation Model to Formulate a Climbing Lane Warrant. Internal Report AIR 290-1. Australian Road Research Board, Numawading, Victoria, 1980.
3. C.J. Hoban. The Two and a Half Lane Rural Road. Proc., 11th Australian Road Research Board Conference, 1982, pp. 59-70.
4. C.J. Hoban. Simulation of Rural Road Improvement Alternatives. Proc., New Zealand Roading Symposium, Wellington, Aug. 1983.
5. Program and Papers. Workshop on Rural Traffic simulation, Australian Road Research Board, Vermont South, June 1983.
6. C.J. Hoban and J.R. McLean. Progress With Rural Traffic Simulation. Proc., 11th Australian Road Research Board Conference, 1982, pp. 22-33.
7. C.J. Hoban. Guidelines for Rural Road Improvements--A Simulation Study. Internal Report AIR 359-10. Australian Road Research Board, Numawading, Victoria, 1983.

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

Signalized Intersection Delay Models– A Primer for the Uninitiated

V. F. HURDLE

ABSTRACT

Delay is being used increasingly as the primary indicator of level of service at signalized intersections, but for many traffic engineers delay estimation is a new task. Some of the currently available estimation techniques are introduced, and the assumptions on which they are based are examined. In general, these assumptions are unrealistic; so accurate delay estimates are not really possible. The difficulties are particularly acute when the arrival flows approach capacity. Some of the procedures avoid the worst of the problems at high flows by methods that, though they involve considerable mathematics, are based on modeling that is essentially qualitative rather than quantitative. Such methods abandon the quest for accuracy in favor of reasonableness: rather than attempting to provide right answers, they try to avoid answers that are terribly wrong. These models would seem to be useful as long as users do not expect too much from them, and the models can probably be somewhat improved. If accuracy is desired, however, a new generation of models that take more account of variations in travel demand over time is needed. The use of such models would require more information about traffic patterns than users are accustomed to providing.

There is a worldwide trend toward the use of estimated delay as the principal measure of the level of service at signalized intersections. The reasons for this trend are clear: Delay can be measured; it has obvious economic worth; and it is easily understood by both technical and nontechnical people. The price that must be paid for these advantages is that capacity analysis procedures must include methods for estimating delay. This necessarily complicates the procedures, so that they involve more computation and are likely to be somewhat harder to understand. In addition, if delay is to be a useful measure of level of service, it is clearly necessary that the methods of estimating it be reasonably accurate and that analysts have a sense of the degree of accuracy to be expected in various situations.

This review paper is intended as a primer for traffic engineers who are familiar with capacity estimation techniques but have not made much use of delay equations. Because this is a primer rather than a textbook or handbook, the emphasis will not be on the mathematical derivation or detailed use of delay equations. Instead, the discussion will be concentrated on the assumptions underlying the equations and the limitations that stem from these assumptions. Although the paper is addressed to beginners, the author hopes that engineers more experienced with delay equations will find food for thought.

BASIC MODEL

The mathematical models used to estimate intersection delay are queueing models. The modeler views the traffic on each approach as a stream of customers seeking service from a server, a system similar to the lineups at grocery store checkout counters except that the server is the intersection and the "service" provided consists simply of letting the "customer" enter the intersection. The server also differs from most grocery store clerks in that it serves very quickly when it serves at all, but every minute or so it gets a red face and insists on resting.

Assume that the customers arrive at some average rate v (vehicles per unit time) and can enter the intersection at a maximum rate s when the light is green; s is called the saturation flow and, in North American literature, v is usually called the volume but will be called the arrival flow in this paper. If the cycle length is C and the length of the green interval is G, it follows that the approach can handle a maximum of c = (G/C)s vehicles per unit time.

The fraction of time the signal is green, G/C, is one of the important parameters in all delay equations. Usually G will not appear alone but only as a part of the G/C ratio.

A second important parameter is the approach capacity,

$$c = (G/C)s$$

but it will also usually appear in a ratio: v/c, the ratio of the average arrival flow to the capacity. This ratio is often called the degree of saturation and in British and Australian literature is usually denoted x. In queueing theory it is usually denoted ρ and called the traffic intensity.

The operation of the intersection approach can be modeled (1) as shown in Figure 1. The curve labeled A(t) shows the cumulative number of arrivals by time t, and D(t) the cumulative number of departures. The arrival curve is an abstraction because it does not indicate the number of actual arrivals at the stop line, but the number of vehicles that would have arrived had the light always remained green. Such a curve could be obtained by counting some distance upstream, plotting the counts, then moving the curve to the right a distance equal to the time required to drive from the count station to the stop line when there is no queue of traffic waiting to get through the signal.

The departure curve D(t) shows the actual departures from the stop line. When the light is red, there are no departures, so the curve is horizontal. In reality it would begin to curve upward as cars began to move after the start of green, then after a few seconds it would become nearly straight with a slope equal to the saturation flow s.

In Figure 1 the D(t) curve is somewhat simpler than this because the figure is a simplified mathematical model. The major simplification is that the outflow does not gradually increase to s but changes instantly from zero to s at the effective start of

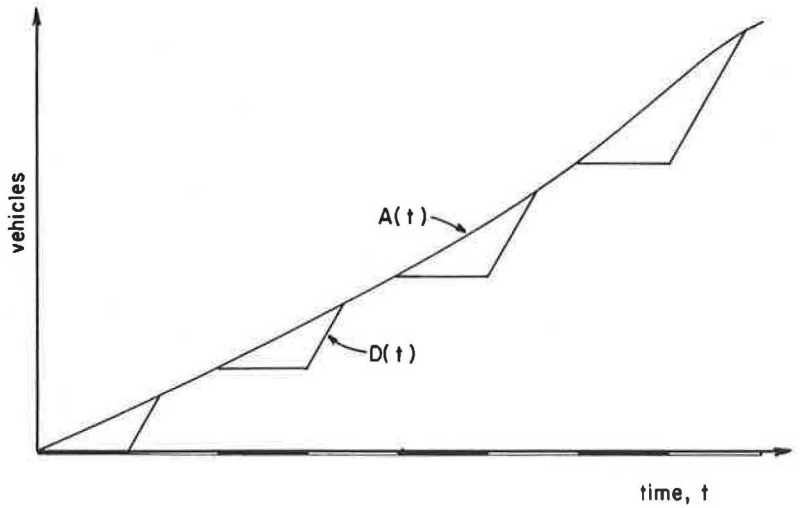


FIGURE 1 Arrival and departure curves for a signalized intersection approach.

green. To keep this simplification from being troublesome, it is, of course, important to be somewhat careful in distinguishing between the real green interval and the effective green interval used in the model, or at least to make a conscious decision that "effective green equals real green" is a good enough assumption for one's purposes. In what follows, the green interval length G will always be the effective length. (Readers wanting to know more about effective green times can find discussions in most traffic engineering textbooks.)

Figure 1 is handy because from it the queue length at any time,

$$Q(t) = A(t) - D(t)$$

can easily be read (Figure 2). Note that, because A(t) is a shifted count curve, Q(t) is an abstract queue length equal to the difference between the number of vehicles that would have crossed the stop line if the signal were not there and the number that actually did cross. Another way of thinking about the model is to say that, in the model, vehicles do not line up along the street but form a vertical stack at the stop line. The real queue is always somewhat longer than the model predicts because the queue engulfs some vehicles that the model assumes are still driving to the vertical stack at the stop line.

If the system works on at least approximately a first in-first out (FIFO) basis, one can also easily determine the waiting time of any individual vehicle as shown in Figure 2 for customer or vehicle number i, which waits w_i between arrival and departure. This is vehicle i's delay, the extra time it has to spend because the signal is there. Unlike the queue length, the abstract delay in the model's vertical stack is equal to the real delay. Vehicle i actually spends more time than w_i in queue, but the extra time, if not spent in queue, would have to be spent driving from the end of the real queue to the stop line.

If one considers vehicle i's delay as the area of a band w_i wide and one vehicle high, it is easy to see that the area between the A(t) and D(t) curves is equal to the total delay suffered by all of the customers who use the intersection approach. With a few more assumptions, a formula for the average delay can be derived by evaluating this area. If the slope of the A(t) curve is always v, the area between the A(t) and D(t) curves is made up of a

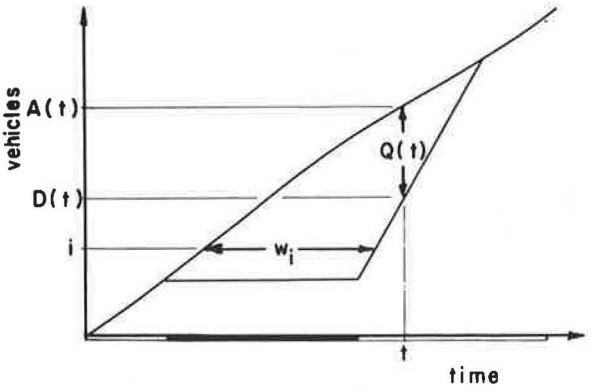


FIGURE 2 Queue length and waiting time.

series of identical triangles and the total delay per cycle can be estimated as the area of any single triangle. Dividing this area by the number of arrivals per cycle, vC, yields the average delay. This takes some algebraic manipulation because one must first locate the upper right-hand corner of the triangle. The answer is

$$UD = \{C[1-(G/C)]^2\}/\{2[1-(v/s)]\} \quad (1)$$

where

UD = average uniform delay per vehicle,
C = cycle length,
G/C = fraction of time the signal is (effectively) green, and
v/s = flow ratio (2,3) = (G/C)(v/c).

This average delay is denoted UD, which stands for uniform delay, because it was derived under the assumption that vehicles arrive at a uniform rate v throughout the cycle. In making this assumption, any random effects and any pattern imposed on the arrival stream by upstream intersections are ignored.

EFFECTS OF RANDOMNESS

The problem with Equation 1, of course, is that vehicles do not arrive at an isolated intersection uniformly, but in a random manner. At low flow

levels, this is not important, but when the arrival flow v gets close to the approach capacity, c = (G/C)s, the actual delay will be considerably larger than the uniform delay predicted by Equation 1. The reason is obvious in Figure 3, where the number of arrivals is the same during every cycle except the second, when some extra vehicles happen to arrive. Because the approach is very nearly saturated (i.e., very near capacity), it takes a long time for the extra queue that builds up in the second cycle to dissipate, and the area between the arrival and departure curves is a good deal larger than if the extra arrivals had not occurred. This extra delay is often called random delay or overflow delay. The latter term reflects the fact that the main effect of random arrivals is to cause occasional overflows of traffic from one cycle to the next (4).

To predict the random or overflow delay, it is necessary to construct a stochastic model using the methods of probability theory or computer simulation. The usual stochastic model depends on four basic assumptions:

1. The number of arrivals in a given time interval has a known distribution, often Poisson, and this distribution does not change with the time of day (i.e., the distribution is stationary) or because of the number of arrivals in any other time interval;
2. The headways between departures from the stop line either have some known distribution, again with a constant mean, or are all the same;
3. $v < c = (G/C)s$ (i.e., the system is not saturated); and
4. The system has been running long enough to have settled into a steady state.

Under these assumptions and some further assumptions about how the system operates, the queue length and waiting time should have a probability distribution that, once a steady state has been reached, does not vary with time but remains fixed. This steady-state waiting time distribution, or at least its mean and possibly its variance, can be determined in simple cases by analytic modeling and in more complicated cases by computer simulation.

Such models are known as steady-state queueing models because of the very important assumption that the system has operated for a sufficiently long time with the same average values of v and s to have settled into a steady state. For small v/c ratios, up to about 0.5, what happens in each cycle is nearly independent of what happened in the previous cycle so the time to reach a steady state is very short. As the v/c ratio increases, however, this relaxation time increases rapidly and approaches infinity as the v/c ratio approaches unity. The reason for this will be explained hereafter.

There are many steady-state models of varying complexity available; Allsop (4) gives an excellent summary, and Katsuhisa (5) gives numerical comparisons. By far the best known of these is the one constructed by Webster (6,7) in the late 1950s, assuming random (Poisson) arrivals and uniform departure headways. Webster's formula for the average delay, d, is

$$d = \{C[1-(G/C)]^2/2[1-(v/s)]\} + [(v/c)^2/2v(1-v/c)] - 0.65(C/v^2)^{1/3}(v/c)^{(2+5G/C)} \quad (2)$$

The first term of this expression is the uniform delay of Equation 1, the second can be predicted theoretically (4), and the third is a correction term obtained empirically from simulation results (6). This third term reduces the estimate by 5 to 15 percent, so a 10 percent reduction can be used as a rough approximation (7). In this latter approximation, any consistent time units may be used (i.e., if d is to be in minutes, then C must be in minutes and v and c in vehicles per minute), but the third term works properly only if all time units are seconds and all flows are in vehicles per second.

Figure 4 shows the delay predicted by Equations 1 and 2 as a function of the v/c ratio for a 60-sec cycle, G/C ratio of 0.5, and saturation flow of 3,600 vehicles/hr. The random delay component is small compared with the uniform delay when the degree of saturation is small, but it increases very rapidly at the higher v/c ratios and approaches infinity at the right edge of the figure. More complicated models with different assumptions would

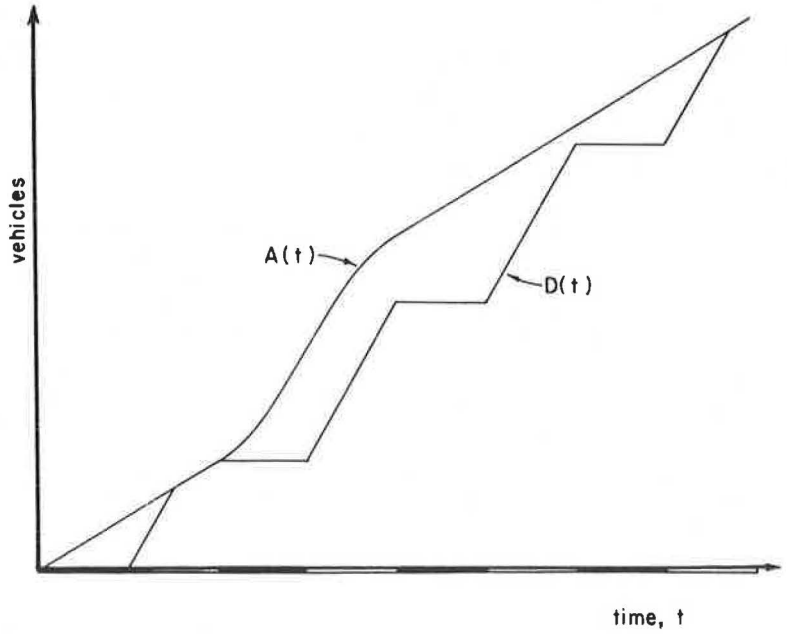


FIGURE 3 Overflow due to random variations in arrival flow.

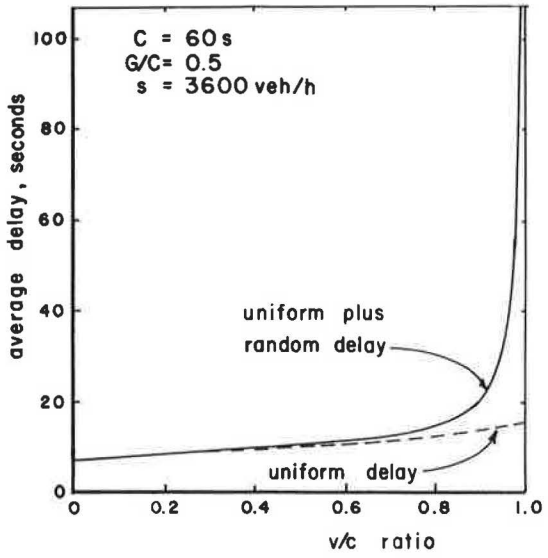


FIGURE 4 Delay predicted by a steady-state, stochastic model.

produce slightly different curves, but all steady-state stochastic models give curves with the general shape shown in Figure 4. In particular, any steady-state model that does not assume completely uniform arrivals will predict that the queue length, and therefore the delay, approach infinity as the v/c ratio approaches unity.

This is, of course, the reason that systems with a high v/c ratio take a long time to settle into a steady state; it simply takes a long time for such long queues to form, particularly since vehicles keep leaking away through the signal. As a result, one seldom sees real delays as large as those predicted for high v/c ratios by Equation 2. This discrepancy is not a result of faulty mathematics but of the unrealistic assumption that the system is in a steady state. If vehicles continued to arrive at a rate v nearly equal to the capacity c, the giant queues really would form, but in reality the peak period ends and v decreases long before a steady state is reached.

As a result, steady-state models are useful for predicting delays only at lightly loaded intersections. Some alternative models are available for the analysis of heavily loaded intersections, but they also have their shortcomings. The next model examined is at the opposite extreme from the steady-state models; it works well only if the intersection is solidly oversaturated (i.e., only if the arrival flow v is considerably greater than the capacity c for a significant length of time).

OVERSATURATION

Oversaturated systems are very easy to model (1). Suppose that a time varying arrival flow v(t) is either known or estimated. Then the cumulative arrival curve,

$$A(t) = \int_0^t v(s)ds$$

can be calculated and plotted as shown in Figure 5. [Traffic counters give A(t) directly, not v.] The dashed line in Figure 5 has slope (G/C)s and has been drawn tangent to A(t) at time zero, the time when the approach first becomes saturated. The departure curve D(t) has been drawn below, so that the dashed line divides the total area representing delay into two components: uniform delay corresponding to Equation 1 below the dashed line and overflow delay above.

The average uniform delay in this case is just half of the red interval, (1-G/C)C/2, as can be seen either by examination of Figure 5 or by substituting v = c = (g/C)s into equation 1. (In this substitution, v is not the arrival flow but the slope of the tops of the triangles.)

To obtain the average overflow delay, simply measure the area between A(t) and the dashed line--with a planimeter, by square-counting, or by any convenient method--to obtain the total delay, then divide by [A(T)-A(0)] to find the average. In practice, it is not necessary to draw the little triangles and it is usual to think of the dashed line as the departure curve, D(t), even though the real D(t) is the stairstep curve outlining the triangles.

This model ignores the effects of random variations. That is no problem when the queue is very large, because the error due to random effects will be small compared with the estimated delay, but this error can be a problem if the intersection is only slightly oversaturated. Thus, there is one group of models, the steady-state queueing models, that work well when v/c is considerably less than one and another type, the deterministic queueing model of Figure 5, that works well when v/c is considerably more than one. In between, there are problems.

Unfortunately, "in-between" is an area where good estimates could be very useful. At the extremes, it may be sufficient to say that operation is excellent or dreadful; in between, some numbers would be nice. (In truth, the ability to assign numerical values to different degrees of dreadfulness is important, but that is not the subject of this paper.)

IN-BETWEEN

As a first step toward the understanding of one model often used for v/c near unity (4,8-11), consider the special case of oversaturation delay shown in Figure 6. Figure 6a shows a very simple (and highly unrealistic) peak-period arrival pattern. Before time zero, the arrival flow is small enough that the overflow delay is negligible; during the peak from time zero to time T, it is equal to v; and after time T, there is some smaller, off-peak flow. Figure 6b shows the cumulative arrival and departure curves. The departure curve shown is not the true, saw-toothed departure curve but a straight line with slope c = (G/C)s like the dashed line in Figure 5; so the area of the triangle is the overflow delay not the total delay.

This area could be calculated with no great difficulty, but there is an easier way to obtain the average overflow delay. Begin by noting that a vehicle that arrives at time zero or τ experiences no overflow delay, but one that arrives at T must wait a time equal to the length of the horizontal dashed line near the center of Figure 6b. Between time zero and time T, the horizontal distance between the A(t) and D(t) curves varies linearly, so the average wait for those that arrive between zero and T is clearly half the length of the dashed line. The same is true for those that arrive between T and τ, so the average delay is half the length of the dashed line for both groups, hence for everyone who is caught in the overflow queue that exists from time zero until τ.

To calculate this average delay, first note that the length of the vertical dashed line is vT-cT =

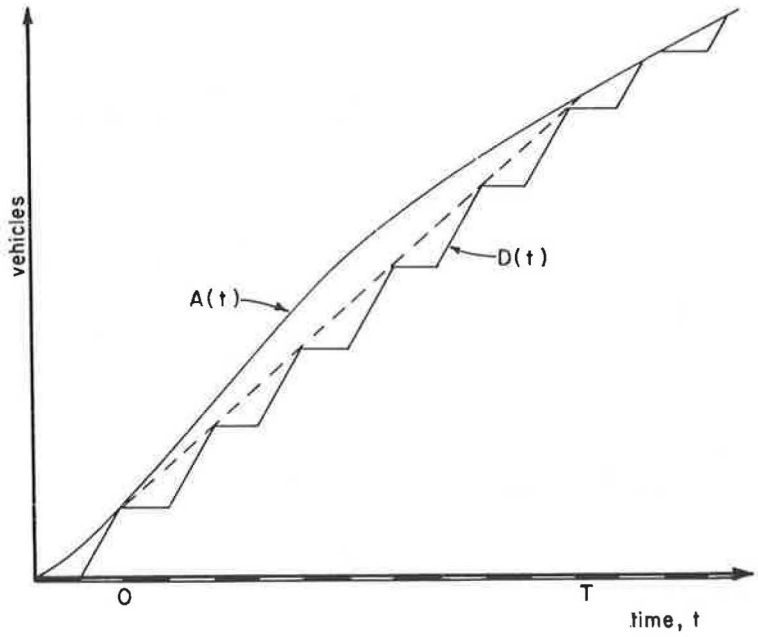


FIGURE 5 Arrival and departure curves for an oversaturated approach.

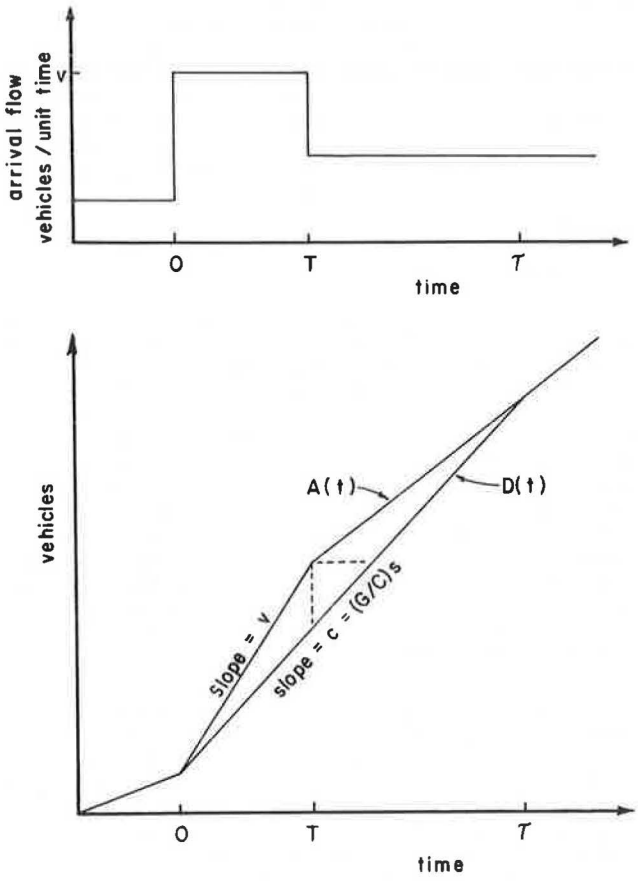


FIGURE 6 Overflow delay in a highly idealized peak period.

T(v-c). Then, because the length of the vertical dashed line is equal to the length of the horizontal dashed line times the slope of the D(t) curve, c, the horizontal dashed line must be T(v-c)/c long and the average overflow delay is

$$OD = (T/2)[(v-c)/c] = (T/2)[(v/c) - 1] \quad (3)$$

Note that the overflow delay is directly proportional to T, the length of time during which the arrival flow exceeds the capacity, and to [(v/c)-1], which is a measure of oversaturation.

Kimber and Hollis (10) and Akcelik (12) have used diagrams like Figure 7 to summarize what has been discussed so far. The solid curve on the left side, for v/c < 1, is the average overflow delay after a steady state is reached, as predicted by Equation 2. Because steady-state models predict infinitely long queues for saturated systems, this curve is asymptotic to the vertical line v/c = 1. As already noted, the enormous queues that cause these large delays do not develop at real intersections.

The solid curve on the right-hand side shows the average delay during time interval $(0,\tau)$ as predicted by the deterministic model of Figure 6 when T = 1 hr. Equation 3 indicates that this curve is a straight line with slope directly proportional to T. As was already noted, the average delay experienced by people who arrive during time interval (0,T) is also given by Equation 3 and the curve in Figure 7, but people who arrive right at time T have a delay equal to double this average.

It is reasonable to expect that both of the solid curves in Figure 7 would be good predictors of the overflow delay in the proper circumstances: the right-side, oversaturation model if the traffic really arrived as shown in Figure 6a with T of moderate length and v substantially larger than c; the left-side, steady-state model if v were substantially smaller than c or if T were very long. It is also reasonable to expect both models to deteriorate as the v/c ratio approaches unity.

Figure 7 shows clearly, however, that the problem when the v/c ratio is near unity is not just that neither model is at its best in this range, nor that one model predicts the average delay over a time interval and the other the average delay at the end of the interval. The problem is much more serious than that: the two models are utterly incompatible for v/c ratios near unity. One predicts that the overflow delay approaches infinity as v/c approaches

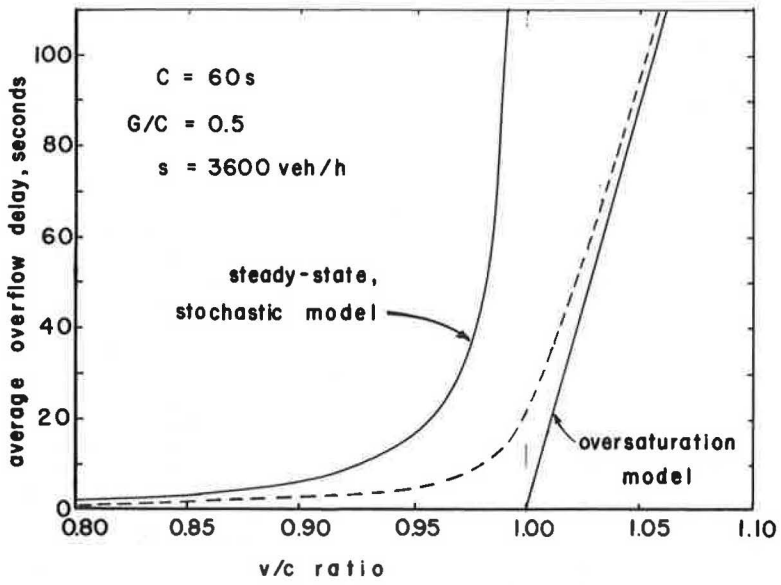


FIGURE 7 Comparison of steady-state and oversaturation models.

1 from below, but the other predicts that the delay approaches zero as v/c approaches 1 from above. A greater incompatibility is hard to imagine.

The problem, of course, lies with the assumptions. It is clear that the delay at v/c ratios slightly less than 1 will always be less than the steady-state models predict because the peak period never lasts long enough for the enormous steady-state queues to develop.

For v/c ratios greater than unity, it is not really possible to say that the real queues will always be longer than the deterministic model predicts, but surely they almost always will be. In this case, the reason is that real arrival flows do not suddenly jump from a low value to v as shown in Figure 6a but increase gradually. Therefore, because of random variations in the arrival flow and capacity, there will usually already be an overflow queue (i.e., a queue remaining at the end of green) at time zero (when the approach first becomes saturated). If this overflow queues is Q_o long at time zero, then the queue length in Figure 6b will be Q_o longer throughout time interval $(0,_T)$ and the actual average delay will be about Q_o/c greater than the deterministic model predicts. For large v/c ratios, this error is a small fraction of the average delay, so it can be ignored, but for v/c ratios near 1, this error can be important.

Thus, the real delays are likely to be more as shown by the dashed curve in Figure 7, which agrees with the two solid curves at the edges of the figure but not in the middle. It is tempting to think that one could produce such a curve by fixing T and varying v in Figure 6a, but the assumptions of the two solid-line models are too incompatible for this to work. There would be a discontinuity at v/c = 1 because the two models average the delay in different ways. This problem could, of course, be overcome by changing the assumptions, but the effort would scarcely be worthwhile unless the shape of Figure 6a were also changed to something more realistic. It is better not to think of the dashed curve as a model of a queueing system but simply as a curve sketched quickly to illustrate some intuitive ideas.

In fact, however, the curve was not just sketched; it was calculated by the method developed by P.D. Whiting for version 6 of the Transport and Road Research Laboratory's popular computer program TRANSYT (8). This copyrighted computation method is perhaps better described as an algorithm, rather than a formula, because it requires only seven quite simple FORTRAN statements but spreads over most of a page when written as an equation. According to Robertson (9), the formula can be approximated by

$$OD = \{15T/c\}\ \{(v-c) + [(v-c)^2 + 240\ v/T]^{1/2}\} \qquad (4)$$

where OD is in seconds, T in minutes, and the flows v and c in vehicles per hour. The actual TRANSYT formula (10) is of the same general form but much more complicated and yields average delay per unit time rather than per vehicle. The curve in Figure 7 was obtained by dividing this result by the average number of vehicles served per unit time [either v or c = (G/C)s, whichever is smaller], but a formula for direct estimation of delay per vehicle is given by Kimber and Hollis (10).

Another much used version of this formula appears in the capacity analysis procedures developed in Australia by Akcelik (11):

$$OD = \begin{cases} (T/4)\left\{(x-1) + \left[(x-1)^2 + 12(x-x_o)/cT\right]^{1/2}\right\} & \text{if } x > x_o \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

where

$x = v/c$

and

$$x_o = 0.67 + sG/600 \qquad (6)$$

can be thought of as the smallest v/c ratio for which the random or overflow delay is large enough to be worth the effort required to calculate it. Unlike equation 4, Equation 5 can be used with any convenient time units, as long as they are consistent throughout the equation. (If the times are in seconds, then s and c should be in vehicles per second.)

The theoretical basis (12) for Equation 5 is similar to that of TRANSYT's random delay algorithm, so they can be regarded as essentially the same for

the purposes of this paper. As can be seen in Figure 8, the TRANSYT version generally seems to bend more sharply, hence produce lower delay estimates. The difference, however, is not always as large as that shown in Figure 8. Another minor, but obvious, difference between the two formulas is that the overflow delay predicted by TRANSYT increases smoothly over the entire range of v/c ratios, but Equation 5 jumps abruptly from zero to some small value when $v/c = x_o$. This both saves some unnecessary calculations and avoids the embarrassment of discovering the hard way that the upper branch of Equation 5 predicts negative overflow delays when v/c is small. (Note that x-1 is a negative number unless the intersection approach is oversaturated.)

A derivation of the TRANSYT random delay equation can be found elsewhere (10). The basic idea is really nothing more than what has already been said with respect to Figure 7, that the two solid curves should be good estimators when the v/c ratio is not close to unity, but that the real delay must be less than the steady-state models predict for v/c slightly less than 1 and more than the oversaturation model predicts for v/c slightly greater than 1. This line of intuitive reasoning leads to imagining a smooth curve that looks like the solid curve near the left side of Figure 7, then somehow cuts across the middle of the diagram and is asymptotic to the right-side, solid curve.

This is exactly what the dashed curve does. It is, in fact, exactly what it was designed to do. Whiting took a steady-state model developed by Robertson and bent it around by means of algebraic manipulation in such a way that the curve is no longer asymptotic to the vertical line, v/c = 1, but instead to the line predicted by the deterministic model of Figure 6b and Equation 3. Details of the transformation can be found elsewhere (10). The interest here, however, is not in the details but in the idea. Equations 4 and 5 and the series of statements in the TRANSYT program are simply mathematical expressions that accomplish the purpose of the dashed curve in Figure 7, provision of a smooth transition between the steady-state and oversaturation models in the region in which the two are inconsistent and likely to be unrealistic.

It is important to understand that the exact form of this transition is not the result of any detailed analysis of queue behavior but of the simple ideas expressed previously. The equations have one and only one justification: they provide a smooth transition in a way that satisfies intuitive ideas of what ought to happen. Kimber and Hollis describe the reliability of such formulas as follows: "In limiting cases . . . their results are correct, and in the intermediate regions, their functional behaviour is sensible" (10,p.11). They key word here is sensible. No claim is made that the formulas are correct; rather they yield answers that do not violate elementary logic in the troublesome region of v/c near unity where neither the steady-state nor the oversaturation models can be expected to yield reasonable results.

IS THERE A BETTER WAY?

For a practitioner needing an answer, Equation 5 or its TRANSYT equivalent obviously offers a better

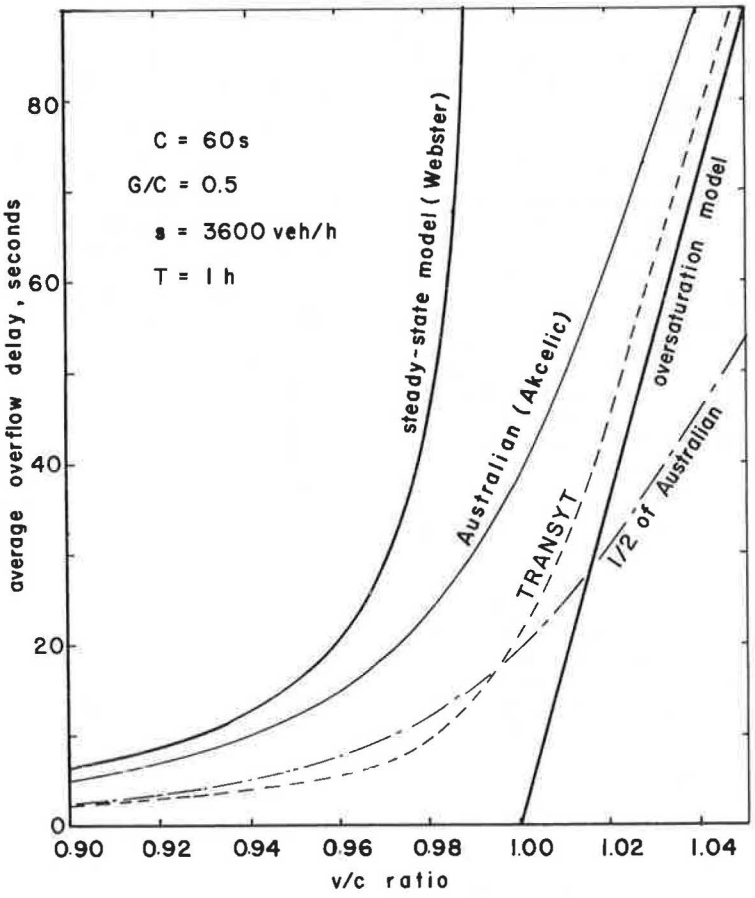


FIGURE 8 Comparison of the overflow delays predicted by five models.

solution for situations where the v/c ratio is near unity than any of the steady-state models. It will not necessarily give a good answer, but the competition offers only unreasonable answers. The choice is easy, but one really would like to be on safer ground.

The road to safer ground has two branches. The first, and obviously easier, starts by noting that the exact form of transition provided by Equations 4 and 5 is arbitrary; the dashed curve must agree with the solid curves at the edges of Figure 7, but it is clear that it could have many shapes in the middle. Therefore, an obvious thing to do is to measure delays in the field and compare then with the predictions. If there is a consistent direction of error, then maybe another transition curve can be found that is equally arbitrary but works better.

This is the approach followed by W.R. Reilly et al. (2,3). Their measurements of delays on undersaturated intersection approaches indicated that Equation 5 consistently overestimated the overflow delay, particularly at v/c ratios approaching unity (3). Therefore, they recommended a new, but very similar, overflow delay equation:

$$OD = 346.2\ T\ \{(x-1) + [(x-1)^2 + 12(x-OF)/cT]^{1/2}\} \quad (7)$$

where

$$OF = 0.67 + Cc/2{,}160{,}000 \quad (8)$$

This differs from Equation 5 in three ways, only one of which is important. The first difference is one of notation and units: OF is what has been called x_o here, and a factor of 3,600 has been introduced in both Equations 7 and 8 so that the delay and the cycle length C can be in seconds, but T is in hours and s is in vehicles per hour.

The second difference is that the equation has been divided by 1.3 to yield stopped delay. Theoretical and simulation models give estimates of the total time lost due to the presence of the signal, which is greater than the amount of time spent actually stopped. Field studies, on the other hand, often yield stopped delay, because it is easier to measure. According to Reilly et al. (13), the total delay can be estimated by multiplying the measured stopped delay by 1.3; so Equation 5 had to be divided by 1.3 to yield stopped delay.

The third difference is that the delay has also been divided by 2 to give results more nearly in agreement with the field measurements. Thus, the 346.2 T in Equation 7 comes from the T/4 of Equation 5 as follows:

$$(T/4) \times (1/1.3) \times (3{,}600\ \text{sec/hr}) \times (1/2) = 346.2\ T$$

The only important difference is the factor of 2.

The effect of this halving is to produce a new curve--shown by a broken line in Figure 8--that is lower than the one based on Equation 5 but no longer asymptotic to the straight line predicted by the oversaturation model. Reilly et al. (2) recommend that Equation 7 be used only for $v/c \leq 1.10$ and the oversaturation model for larger v/c ratios. Thus, there is a discontinuity in the delay prediction at v/c = 1.10 and a tendency to underestimate the delay when $1.00 < v/c < 1.10$, but outside of this range the overall pattern of prediction is likely to be rather similar to that of the TRANSYT model.

The hope is that Equation 7 gives better predictions than Equation 5. The price paid for this increase in accuracy includes the discontinuity, the somewhat illogical behavior for $1.00 < v/c < 1.10$ where the broken curve of Equation 7 may lie below the lower bound set by the oversaturation model, and the loss of the theoretical basis for the form of the equation--a form dictated by the abandoned requirement that the prediction equation be asymptotic to the solid line on the right side of Figure 8.

As an alternative to dividing Equation 5 by 2 or some other factor, one could try to change its shape to improve the predictions. There would seem to be two obvious goals for such an attempt: either to find a new curve that is once more asymptotic to the solid line or, if this goal is abandoned, to find a curve with a simpler formula. Researchers taking the latter path, however, will have to grapple with the problem of finding a new functional form that responds to changes in v/c, c, and T in a reasonable way. For those who would prefer to stay with the basic logic that led to Equation 5 but somehow change the curve's shape, two approaches offer hope of quick and easy results. One of these is to return to the TRANSYT formulation, which seems to agree fairly well with Equation 7 for $v/c \leq 1$. A second is to change the value of T; Equation 5 with T = 15 min gives delay estimates quite similar to Equation 7 with T = 1 hr. For $v/c < 1$, these two estimates are very similar to the TRANSYT estimate, but for v/c much greater than unity, they predict only about half as much delay, a worrisome difference. As will be discussed later, T is a very important parameter, but it is difficult to evaluate.

QUEUEING THEORY APPROACH

The second road to improvement is a modeling road; rather than simply trying to produce a reasonable curve, one can try to model actual queue behavior. This road offers more potential for good answers, but it also presents many difficulties. To be an improvement over what is now available, the model cannot assume a steady-state situation but must account for the way the queue length varies over time--its transient behavior.

Unfortunately, the modeling of transient queue behavior is not a very well-developed science and the models that are available tend to be complicated. On the other hand, more can be done than usually is. Two possible approaches are discussed by Kimber and Hollis (10) and in the last few chapters of Newell's Applications of Queueing Theory (1). Many readers of this paper will probably find both tough going, but they really can be read by graduate engineers armed with only one or two probability courses--a statement that cannot be made about most of the available literature on this subject.

Kimber and Hollis (10) propose a model in which fairly short time intervals are considered sequentially, the results from one forming the starting condition for the next. Newell (1), on the other hand, develops partial differential equations to describe the approximate behavior of the queue length distribution as it changes with time. What both modeling approaches make very clear is that the development of the queue is very dependent on the details of the arrival pattern. For example, if the numbers of arrivals in each 15 min of the peak hour at a saturated or nearly saturated approach are 200, 500, 300, and 200, the peak-hour volume and peak-hour factor will be just the same as with volumes of 200, 300, 500, and 200, but the average delay will be different because of the dynamics of queue growth and decay. If this difference is to be reflected in a model, it is clear that more information about arrival patterns must be provided than is now customary.

Because this paper is directed more to users of models than to builders of models, this is the really important lesson to be learned from the examination of non-steady-state models. If good estimates of signalized intersection delay are desired, better input to the models must be provided. Until that is done and until practical models are developed to use that input, it is not reasonable to expect delay estimation to be a very accurate process.

DIFFICULTIES IN APPLICATION

Returning to Equations 5 and 7, it should by now be clear that the most difficult problem facing an analyst willing to accept their reasonableness is the choice of the two variables describing the traffic arrival pattern, v and T. The definitions of these are given in Figure 6a: vehicles arrive at flow rate v during a period exactly T long. This, of course, is not the way real traffic arrives, and it is not true that all intervals T long in which vT vehicles arrive will have the same average delay. In fact, the delay will depend very much on the flow variations within the time interval. How, then, does one choose v and T?

The available literature seems to indicate that T should usually be 1 hr and v either the average flow during that hour (6,11) or the average within the busiest 15 min of that hour (2). Nobody, however, gives much in the way of reasons. [A partial exception is Reilly et al. (2). Equation 7 is based in part on data and its originators suggest, logically enough, that users select v and T the same way they did when they chose the equation.] Reading between the lines, it seems that nobody is very sure about this matter.

This is not surprising. Theoretical considerations indicate that there is no good answer, so those who must advise are quite correct--if not very helpful--in sounding none too confident about their advice. Nonetheless, the users are stuck with the fact that even if they cannot find a good answer, it may be possible to find a particularly bad one.

SIGNAL COORDINATION

Everything said so far has been based on the implicit assumption that the signal approach is isolated, so that the arrival pattern is at least more or less uniform. In many cases, however, the arriving vehicles are the output stream from an upstream signal. In that case, the cumulative arrival curve will not be smooth, as in Figure 1, but saw-toothed like the departure curve. In such cases, the area between the arrival and departure curves will obviously be very different if the signals are well coordinated--in which case, the area should almost vanish--than if each platoon of arrivals neatly hits a red light.

Two of the procedures discussed previously try to account for this, but in different ways. The TRANSYT model's main purpose is to achieve coordination, so it takes a very detailed approach in which the actual arrival curve is predicted. The estimate of the nonrandom delay is based on this predicted arrival curve rather than on Equation 1. The random or overflow delay estimate does not depend on the coordination.

Reilly et al. (2), on the other hand, recommend that the sum of the uniform and overflow delays predicted by Equations 1 and 7 be multiplied by an adjustment factor based on the degree of coordination (5 categories from poor to very good) and on whether the signal is pretimed, semiactuated, or fully actuated. Adjustment factors range from 0.65 to 1.50.

Obviously the detailed analysis in TRANSYT is the better approach, and just as obviously it is not always a practical one. Its advantages are likely to be realized only when good data are available, and it is really oriented to signal timing and large system design rather than to the usual design situation in which only one or a few intersections are to be designed and the upstream intersections may not even be planned, let alone timed. It is also a computer model and its methods are not always practical for hand calculation. Another difference worthy of note is that the two methods make very different assumptions about the effect of coordination on overflow delay.

SUMMARY

The primary purpose of this paper has been to introduce new users of delay models to the background of some of the models and the assumptions on which they are based. This seemed necessary because the information is widely scattered, some of it in material not readily available to the average traffic engineer. Along the way, it has been noted that the methods currently used either ignore the way in which delays vary with time or try to cope with the variation in ways that are more mathematical applications of common sense than mathematical models of traffic signal systems. The methods that ignore the variation are nearly useless for the most interesting traffic engineering problems, those where the signal system operates close to capacity, but the others seem reasonable enough to be useful as long as one remembers their limitations.

None of the models examined here can be expected to give really consistent and accurate results. To obtain such results, one would need not just better models but better information about traffic patterns. For some purposes it may be unrealistic to suppose that such information will be available. For detailed operational studies, however, it could be obtained. If estimated delay is to be a good indicator of level of service in such studies, the information will have to be obtained and the models developed to use it. This is not an impossible task.

ACKNOWLEDGMENT

This work was supported in part by the National Science and Engineering Research Council of Canada.

REFERENCES

1. G.F. Newell. Applications of Queueing Theory, 2nd ed. Chapman and Hall, London, England, 1982.
2. JHK & Associates and the Traffic Institute, Northwestern University. Urban Signalized Intersection Capacity. Draft Final Report, NCHRP Project 3-28(2). TRB, National Research Council, Washington, D.C., Feb. 1983.
3. JHK & Associates. Urban Signalized Intersection Capacity. Uncorrected Draft, NCHRP Project 3-83(2). TRB, National Research Council, Washington, D.C., 1982.
4. R.E. Allsop. Delay at a Fixed Time Traffic Signal--I: Theoretical Analysis. Transportation Science, Vol. 6, 1972, pp. 260-285.
5. K. Ohno. Computational Algorithm for a Fixed Cycle Traffic Signal and New Approximate Ex-

pressions for Average Delay. Transportation Science, Vol. 12, 1978, pp. 29-47.
6. F.V. Webster. Traffic Signal Settings. Road Research Technical Paper 39. Her Majesty's Stationary Office, London, England, 1958.
7. Road Research Laboratory. Research on Road Traffic. Her Majesty's Stationary Office, London, England, 1965.
8. D.I. Robertson and P. Gower. User Guide to TRANSYT Version 6. Supplementary Report 255. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1977.
9. D.I. Robertson. Traffic Models and Optimum Strategies of Control--A Review. Proc., International Symposium on Traffic Control Systems, edited by W.S. Homburger and L. Steinman, Vol. 1, Berkeley, Calif., 1979, pp. 262-288.
10. R.M. Kimber and E.M. Hollis. Traffic Queues and Delays at Road Junctions. Laboratory Report 909. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1979.
11. R. Akcelik. Traffic Signals: Capacity and Timing Analysis. Research Report 123. Australian Road Research Board, Victoria, 1981.
12. R. Akcelik. Time-Dependent Expressions for Delay, Stop Rate and Queue Length at Traffic Signals. Internal Report AIR 367-1. Australian Road Research Board, Numawading, Victoria, 1980.
13. W.R. Reilly, C.C. Gardner, and J.H. Kell. A Technique for Measurement of Delay at Intersections. Report FHWA-RD-76-137. FHWA, U.S. Department of Transportation, 1976.

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

Automated Collection of Vehicular Delay Data at Intersections

JAY F. LEGERE and A. ESSAM RADWAN

ABSTRACT

Most current methods used to estimate vehicular delay at intersections involve some form of manual data collection. These methods rely on statistical techniques (such as multiple linear regression) to improve the accuracy of the delay estimates. In addition, most require significant data collection and reduction efforts. The theory, design, operation, and evaluation of a microprocessor-based system for the collection of vehicular delay data at intersections is presented. The principle of the automated system, including definitions of pertinent variables and equations, is discussed. An overview of the system design, including a description of the vehicle detection scheme and the microprocessor's recognition of vehicle arrivals and departures, is presented. There is a discussion of the system software as well as a description of the data collection and reduction processes. The system performance was evaluated both in the laboratory and through analysis of data collected in the field. Recommendations for further development of the device are presented.

Since the Arab oil embargo of 1973, the United States has become increasingly concerned with its energy supply and with ways in which that supply can be conserved. One area of particular interest is the conservation of energy within the transportation sector. About 40 percent of this nation's petroleum consumption is attributable to passenger travel by automobile.

Much traffic engineering research has been done on delay and fuel consumption at signalized intersections simply because they are considered the locations where most delay and excess fuel consumption occur. Unfortunately, the most accurate methods of data collection and analysis have proven to be extremely time consuming and costly.

Estimates of intersection delay are used in numerous applications, some of which are validation and calibration of computer simulation models, estimates of road-user costs, before-and-after studies, comparisons of the efficiency of various types of intersection control, and comparison of specific signal timing and phasing.

The theory, design, implementation, and evaluation of a microprocessor-based system for the collection of vehicular delay data at intersections are presented here. The primary application of the system is for the collection of data at intersections that are under some form of signalized control. The system is also applicable to any intersection or, in general, to any section of highway for which values of average travel time and delay are desired. Details of the hardware specifications, the software routines, and the assembly language program for this application are fully documented elsewhere (1).

PRINCIPLE OF AUTOMATED DATA COLLECTION METHOD

A report prepared for the FHWA by JHK & Associates (2) presents the results of a research project on defining and measuring delays at intersections. Four basic methodologies were identified for use in estimating delay: point sample, input-output, path trace, and modeling. The input-output method has proven to be convenient and reliable when used to measure approach delay. However, the data reduction process is extremely tedious and time consuming. If the input-output method could be automated by the use of a microprocessor-based system, the problem of data reduction would be eliminated and the accuracy of the data collected would be improved.

The input-output method requires

1. Definition of an approach delay section. The downstream end or exit point of the section is located just beyond the stop line. The entrance to the section is located at an arbitrary point upstream such that the length of the section includes all delay associated with the signal.
2. Determination of a sample interval for the data collection process. The boundaries of the approach delay section were defined by two sets of detectors and a sample interval of 1 sec was chosen to increase the accuracy of the data collected.

Figure 1 shows a simplified diagram of the system setup. Each set of detectors is controlled by its own microprocessor. For each sample interval in the data collection period, each microprocessor counts and records the number of wheels that pass over its corresponding set of detectors. The method of wheel counting is a new concept that resulted as a by-product of the detection scheme chosen. It was possible that a new source of error would be introduced with this method. However, it was believed that the accuracy gained by decreasing the sample interval would result in a net improvement in the accuracy of the data collected.

The number of wheels crossing the upstream detectors for any sample interval i is denoted by NWI_i (number of wheels in for sample interval i). Likewise, the number of wheels crossing the downstream detectors is denoted by NWO_i (number of wheels out for sample interval i). The number of wheels on the approach delay section for any sample interval i, $NWADS_i$, is given by

$$NWADS_i = NWADS_{i-1} + NWI_i - NWO_i$$

where

$NWADS_i$ = number of wheels on the approach delay section for the previous sample interval.

For the first sample interval (i = 1) the number of wheels on the approach delay section for the previous interval must be determined by a manual count.

Total travel time for all wheels traversing the approach delay section in a 15-min data collection period is given by

$$TTT = \sum_{i=1}^{n} (NWADS_i)(\Delta t)$$

where

$NWADS_i$ = number of wheels on the approach delay section for sample interval i,
Δt = length of the sample interval (seconds), and
n = number of sample intervals in the data collection period.

Dividing the total travel time (in wheel-seconds as a result of the multiplication) by the number of wheels entering the system yields the average travel time per wheel (in seconds).

To estimate average approach delay, an estimate of the approach free flow travel time is required. This estimate may be obtained by two methods: (a) averaging a sample of travel times for unimpeded vehicles over the approach delay section or, (b) for types of control other than traffic signals, calculating a free flow travel time based on upstream or downstream free flow travel speeds. Subtracting the estimated free flow travel time from the average travel time yields an estimate of the average approach delay.

Also shown in Figure 1 is a communications line connecting the two processors. This line is used to pass control information from the "master" processor

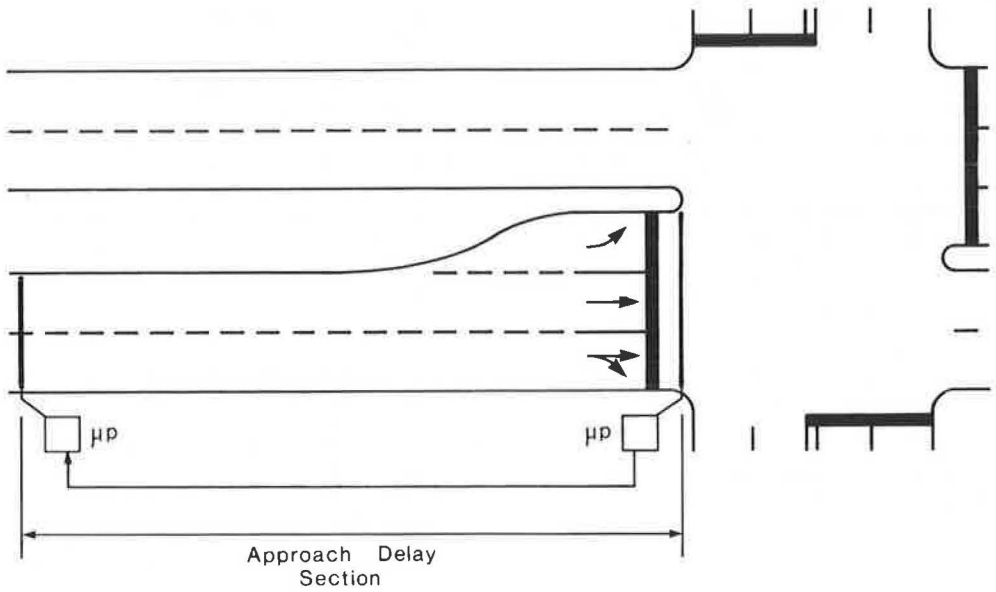


FIGURE 1 Microprocessor system setup.

at the stop line to the "slave" processor. The control information is used to synchronize the two processors at the beginning of a data collection period.

MICROPROCESSOR SYSTEM DESIGN OVERVIEW

Vehicle Detection Scheme

Two forms of vehicle detection were considered for this application. The most common form, the loop detector, was examined as well as a more temporary form of detection, the electrical tapeswitch. Both forms of detection were analyzed in several configurations to determine which would most nearly provide the accuracy required by this application.

Loop Detectors

The principle of loop detection (disturbance of a magnetic field by a heavy metal object) makes it difficult to accurately define the detection area of the loop. For this reason a standard detector configuration presents two possible sources of error that are shown in Figure 2.

In Figure 2(a), vehicles A and B pass over the loop detector with a headway sufficiently small to cause a continuous disturbance within the loop. This situation would probably not occur at the upstream detectors. However, at the stop line, where speeds can be low due to departure from a queue or execution of a turning movement, this case could occur frequently.

In Figure 2(b), a vehicle performs a lane change between two loop detectors. The question here is whether the vehicle generates a single count, two counts (one on each detector), or no count at all. This case may not occur at the stop line, but it might occur at the upstream detectors. In addition to the inaccuracy considerations, the cost of loop detection made it an extremely unattractive alternative for this application.

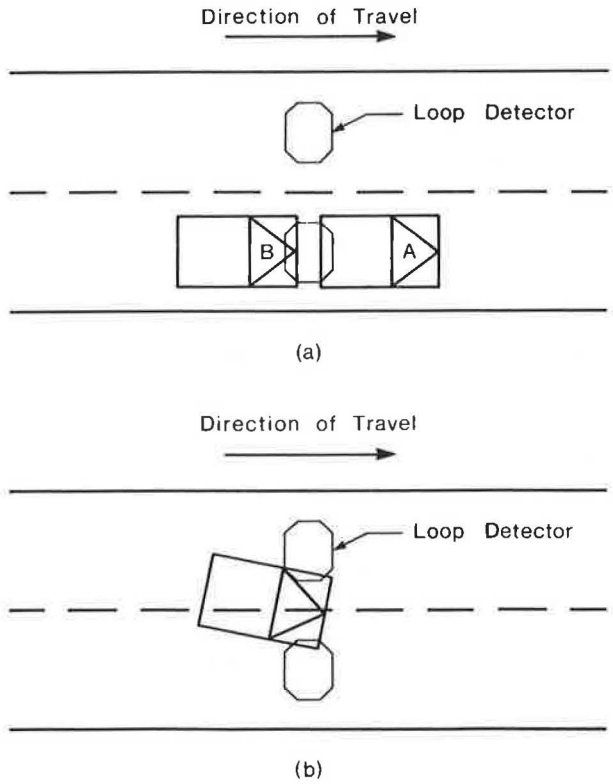


FIGURE 2 Sources of error in loop detection.

Electrical Tapeswitches

Electrical tapeswitches were preferred to pneumatic pressure tubes because the interface between the microprocessor and the tapeswitches could be more easily developed.

The first tapeswitch configuration scheme analyzed is shown in Figure 3. The figure shows two tapeswitches, one switch covering both lanes of traffic and the other covering only one lane. The possible sources of error for this configuration include (a) axle counting as opposed to vehicle counting (multiaxled vehicles); (b) for angled vehicles, four counts (one for each wheel) instead of two counts; and (c) coincident closure of a switch by two adjacent vehicles.

To eliminate the first two sources of error, an attempt was made to find an angle at which a tapeswitch could be mounted to ensure a count for each wheel. This second configuration is shown in Figure 4. However, the problem of coincident closure would remain if adjacent vehicles were slightly staggered as shown in the figure.

It became obvious that, to alleviate that third source of error, shorter tapeswitches would be necessary. A configuration using 2-ft tapeswitches placed end-to-end was investigated. This configuration is shown in Figure 5. With the shorter switches, each closure would consistently represent a vehicle's wheel rather than its axle. Because of this, the counting of angled vehicles would present no problem. Also, it would be impossible for two adjacent vehicles to actuate the same switch. After considering all possible sizes, speeds, and combinations of vehicles, it was decided that the most cost-effective and accurate detection scheme would be that shown in Figure 5.

Switch Scanning and Closure Detection

After the switch configuration was chosen, it was necessary to determine how the switches were to be

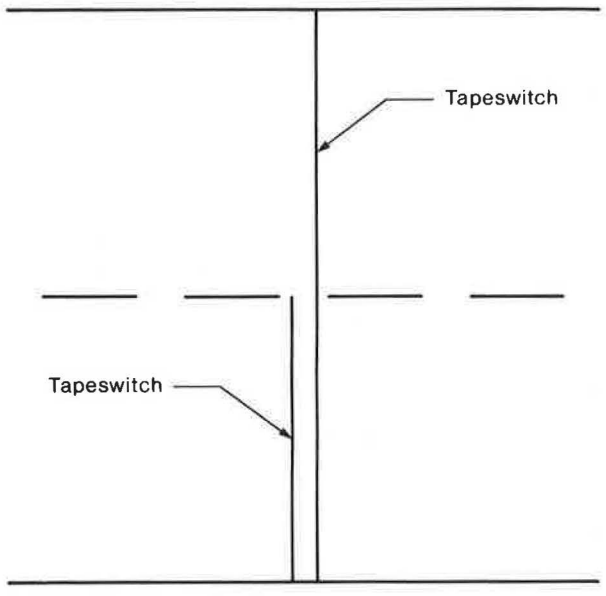


FIGURE 3 First electrical tapeswitch configuration.

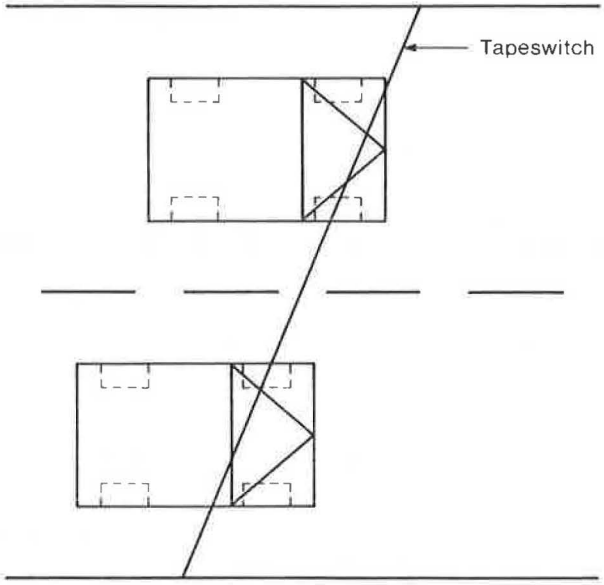


FIGURE 4 Second electrical tapeswitch configuration.

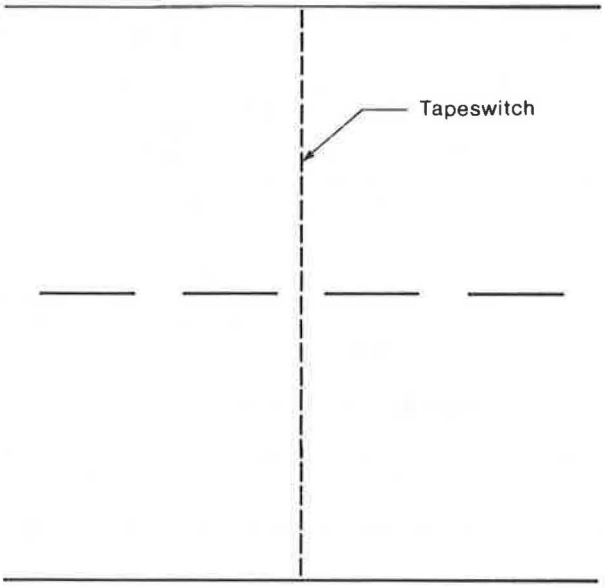


FIGURE 5 Series of 2-ft tapeswitches.

scanned by the microprocessor so as not to miss the passage of any vehicle wheels. A tapeswitch interface circuit was designed to provide a logic 0 level to the microprocessor when the switch was open and a logic 1 level when closed. With this interface, a typical logic-level diagram for the passage of a vehicle's wheel over a tapeswitch would be as shown in Figure 6(a), and a vehicle passage (two wheels) would be as shown in Figure 6(b). The time lapse depicted in the figure may be only a fraction of a second. Because of this, the microprocessor must scan the switches quickly enough to detect the switch closures caused by both the front and rear wheels.

Switch-Status Flags

Each tapeswitch has two switch-status flags assigned in the microprocessor's main memory. The first flag is set when the switch is hit (logic 0 to 1 transition shown in Figure 6) by a vehicle's wheel and the second flag is set when the switch has been cleared (logic 1 to 0 transition) by a wheel passing completely over it. The logical process required to set these flags is as follows: (a) Determine if the switch has previously been hit. (b) If the switch has not been hit, determine the current status of the switch (open or closed). If the switch is still open, do nothing. If the switch is now closed, set the switch hit flag. (c) If the switch has been hit, determine the current status of the switch. If the switch is still closed, do nothing. If the switch is now open, set the switch cleared flag. (d) Repeat the process for each switch.

The switch-status flags are used by another software routine that logs or counts the passage of each vehicle's wheels and resets the flags when the wheel has passed completely over the switches.

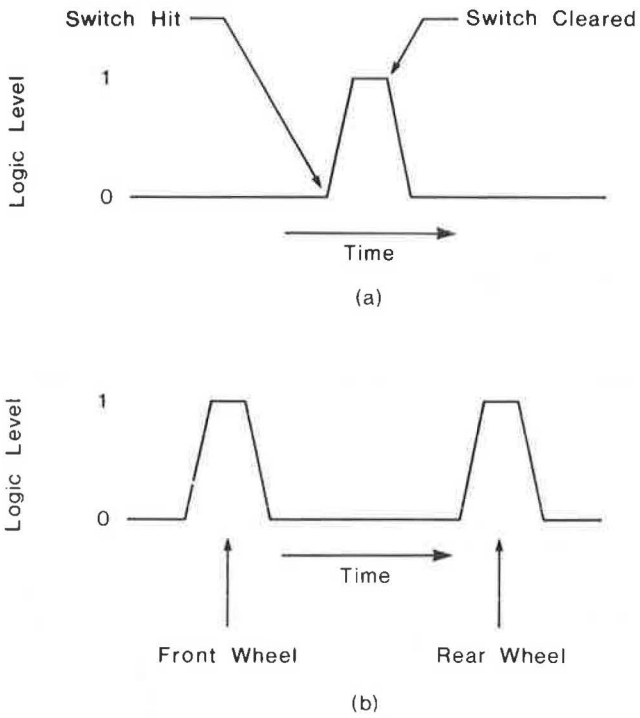


FIGURE 6 Tapeswitch logic-level diagrams.

Flag-Check Interval

The switch-scanning routine described in the previous section is executed repeatedly by the microprocessor. However, it was also necessary to check the switch-status flags at regular intervals to determine the presence of any vehicle wheels to be counted. A flag-check interval was chosen based on calculations of front and rear wheel passage time. Assuming a minimum of 5 ft between a vehicle's front and rear wheels, maximum vehicle passage speeds were calculated for various flag-check intervals. Some of the intervals and their corresponding maximum speeds are

Flag-Check Interval (msec)	Maximum Speed (mph)
10	341
20	170
25	136
30	114
40	85
50	68

The 25-msec flag-check interval was selected because it provided a safe (not likely to be exceeded) maximum speed and because 1 sec is evenly divisible by this interval. A flag-check interval of 25 msec implies that the tapeswitch status flags are examined 40 times per second. Each time the flags are examined, the wheel count is incremented if it is determined that a wheel has passed over the tapeswitch. This wheel count is stored in memory and reset to zero every second.

Switch Patterns and Wheel Counting

After the flag-check interval was selected, it was necessary to develop a method for recognizing the possible switch closure patterns and a method for counting vehicle wheels. It was recognized that, with the detection scheme used, it was likely that a vehicle's wheels would not always pass directly over the middle of a switch. In fact, four possible switch closure patterns were identified. These patterns are shown in Figure 7 and are (a) A wheel passes directly over the middle of a switch. (b) A wheel passes directly between two adjacent switches. (c) A vehicle with a wheelbase of less than 4 ft has its left wheel pass directly between two adjacent switches and its right wheel over a third switch. (d) A vehicle with a wheelbase approximately equal to 4 ft has its left wheel pass directly between two adjacent switches and its right wheel pass directly between another pair of adjacent switches.

These patterns result in the closure of one, two, three, or four switches in a row and each possible pattern was recognized by a software routine and handled as follows: Scan the array of switches sequentially from one side of the roadway to the other, and, when a switch pattern is recognized, check the pattern to be sure that each switch in the pattern has been cleared. If not, ignore the pattern until all switches have been cleared. If all switches in the pattern have been cleared, increment the wheel count by 1 for one or two switches in a row, or by 2 for three or four switches in a row, then reset the appropriate switch-status flags.

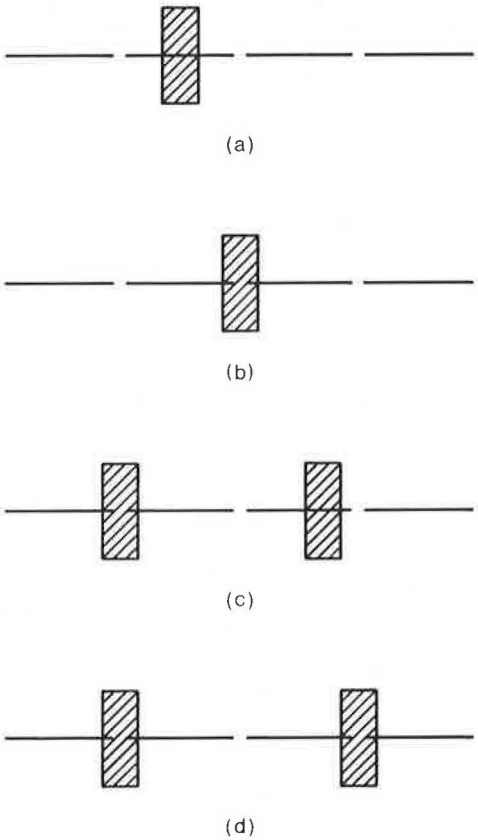


FIGURE 7 Possible switch closure patterns.

SYSTEM HARDWARE AND SOFTWARE

The automated data collection system is based on the Intel 8085 microprocessor. The foundation of the system is Intel's SDK-85 system design kit. This kit provides all the necessary components to build a complete 8085-based microcomputer system (3,4).

The electrical tapeswitches used with the data collection system were connected to the microprocessor input ports through the tapeswitch interface circuitry. Each tapeswitch is made up of two metal contacts separated by thin plastic spacers. At one end of the switch, two lead-in wires are spot soldered, one to each metal contact. To protect the sensitive ends of the tapeswitches from the impact of vehicle wheel passages, small pieces of U-shaped steel channel were used to bridge the wheels over switches. Each tapeswitch is actuated by applying approximately 20 pounds pressure at any point along its length. The switches are catalog number 170-IS Temporary Roadway Instrumentation Switches manufactured by the Tapeswitch Corporation of America (5).

The software that controls the data collection system performs four basic functions:

1. Initialization;
2. System synchronization and start-up;
3. Switch scanning and status-flag update; and
4. Status-flag check, wheel counting, and data storage.

The initialization process is essentially the same for the two microprocessors. The direction of data flow for all input-output (I/O) ports is defined and all memory locations, flags, and pointers are set up.

The master processor and slave processor operate independently during initialization and data collection. However, synchronization of the two processors is required before data collection begins to ensure that both operate together as a system. After its initialization routine, the slave processor remains in a waiting loop until a key is pressed on the keypad of the master processor. When this key is pressed, both processors enter a short start-up routine after which data collection begins.

The switch-scanning routine interacts directly (through the I/O ports) with the electrical tapeswitches. The routine scans each switch being hit (logic 0 to 1 transition) and cleared (logic 1 to 0 transition) and updates the switch-status flags accordingly. The data update routine is executed once every 25 msec (the flag-check interval). This routine uses the switch-status flags to determine the passage of vehicle wheels over a switch or set of switches. When a wheel has completely passed over a switch, the wheel count is incremented. This count is stored in memory and is reset every second.

SYSTEM EVALUATION

Data collected from nine 15-min test periods were used to evaluate the performance of the microprocessor system. This evaluation involved three steps:

1. For each data collection period, true average travel time valves were determined as follows: (a) Each microprocessor was used to display elapsed time on a light-emitting diode (LED) display. Each time a vehicle crossed the upstream or downstream set of detectors, its arrival or departure time was recorded. (b) For each vehicle, travel time was determined by subtracting that vehicle's arrival time from its departure time. (c) These travel time values were averaged to obtain the true average travel time.
2. Data collected by the microprocessor system were used to calculate average travel time estimates. The percent difference (or percent error) was calculated between the microprocessor system values and the true values.
3. The arrival and departure times for each vehicle in each data collection period were used to simulate the conditions that would be seen by a field observer. The input-output method was simulated using a 15-sec sample interval, and the average travel time values obtained were compared with the true average travel time values. After the accuracy of the field-observer method had been evaluated, it was possible to compare both field data collection methods to determine which was the more accurate.

A FORTRAN program was developed to accept the microprocessor-collected data and to perform the calculations necessary to transform this data into values of

1. Average travel time (true) in seconds,
2. Average travel time (microprocessor system) in seconds,
3. Average travel time (field observer) in seconds,
4. Percent error between the true values and the microprocessor system values, and
5. Percent error between the true values and the field-observer values.

In addition, average approach delay values were calculated using a free flow travel time that was taken as the average travel time of 50 unimpeded vehicles.

To compare the accuracy of the microprocessor system with that of a field observer, it was necessary to simulate the operation of a field observer with the data analysis program. A simple input-output method was simulated using a 15-sec sample interval. For every sample interval, the observer determined the number of vehicles crossing the upstream and downstream detectors. The number of vehicles on the approach delay section was then calculated for each sample interval. The summation of these values was used to obtain the total travel time in vehicle-seconds and the average travel time was determined by dividing the total travel time by the number of vehicles arriving during the data collection period. The microprocessor data were reduced in a similar manner except that wheel counts, rather than vehicle counts, were used. Also, a sample interval of 1 sec was used to provide accuracy greater than that of the field-observer method.

The next step was to determine the accuracy of the microprocessor system and the field-observer method with respect to the true travel time values. This was done by calculating the percent error of each method. A positive error indicated that the measured value was greater than the true value and a negative error indicated that the measured value was less than the true value.

Finally, the desired system output, average approach delay, was calculated for each data period by subtracting the free flow travel time from the average travel time.

Average Travel Time and Delay

Data were collected on the southbound approach to the intersection of Progress Street and Giles Road in Blacksburg, Virginia. This intersection handles very low traffic volumes and is under pretimed signal control. The low volumes at this location were desirable for system testing. Data were collected in 15-min periods beginning at 11:00 a.m. and ending at 6:00 p.m. In this time period, nine sets of data were collected.

Table 1 gives the true average travel time values for each data collection period as well as the values measured by the two field measurement techniques. A free flow travel time for the study approach was determined by averaging the travel times of 50 vehicles that passed through the intersection without stopping. This free flow travel time was subtracted from the average travel time values to obtain estimates of average approach delay. The average approach delay values are given in Table 2.

Percent Error and Sources of Error

Field-Observer Method

Table 3 gives the true average delay values, the field observer values, and the corresponding percent error values for each data collection period. The errors for this method range from -1.52 to 14.68 percent. In addition, the errors are both positive and negative indicating both overestimation and underestimation of the true average values.

The primary source of error with the field-observer method is the length of the sample interval. This interval must be long enough to accommo-

TABLE 1 Average Travel Time Values

Period	Travel Time (sec)		
	True	Microprocessor	Field-Observer
1	33.0	31.5	32.5
2	19.9	19.9	21.4
3	30.6	30.4	27.0
4	33.6	33.2	31.9
5	27.3	26.9	31.3
6	18.9	18.7	20.0
7	20.4	20.3	19.5
8	22.3	21.7	20.4
9	13.8	13.5	15.0

TABLE 2 Average Delay Values

Period	Delay (sec)		
	True Average Approach	Microprocessor Average Approach	Field-Observer Average Approach
1	23.1	21.6	22.6
2	9.9	9.9	11.5
3	20.7	20.5	17.1
4	23.7	23.3	22.0
5	17.3	17.0	21.3
6	9.0	8.8	10.1
7	10.5	10.3	9.6
8	12.4	11.8	10.4
9	3.9	3.6	5.1

TABLE 3 Percent Error Comparison, Field-Observer Method

Period	No. of Vehicles	True Travel Time (sec)	Field-Observer Travel Time (sec)	Error (%)
1	6	33.0	32.5	-1.52
2	7	19.9	21.4	7.91
3	10	30.6	27.0	-11.76
4	8	33.6	31.9	-5.20
5	12	27.3	31.3	14.68
6	18	18.9	20.0	5.88
7	10	20.4	19.5	-4.41
8	14	22.3	20.4	-8.65
9	15	13.8	15.0	8.70

TABLE 4 Percent Error Comparison, Microprocessor Method

Period	No. of Vehicles	True Travel Time (sec)	Microprocessor Travel Time (sec)	Error (%)
1	6	33.0	31.5	-4.42
2	7	19.9	19.9	0.00
3	10	30.6	30.4	-0.65
4	8	33.6	33.2	-1.30
5	12	27.3	26.9	-1.22
6	18	18.9	18.7	-1.10
7	10	20.4	20.3	-0.74
8	14	22.3	21.7	-2.52
9	15	13.8	13.5	-2.29

date the limitations of a human observer. However, as the length of the interval increases, the amount of error associated with the data also increases. It should be noted that the percent error for this method will decrease as the observed travel times increase. However, the error can be either positive or negative, which makes it difficult to apply a correction factor to the results.

Microprocessor System

Table 4 gives the true average delay values, the microprocessor system values, and the corresponding percent error values for each data collection period. The automated system errors range from -4.42 to 0.00 percent. The improvement in accuracy due to shortening of the sample interval is obvious. However, a new source of error has been introduced with the microprocessor system. This new error is due to wheel counting.

The error due to wheel counting is significant only when the travel time of a vehicle's front wheels differs greatly from the travel time of its rear wheels. When a red signal is encountered by a vehicle, it is possible for the front wheels of the vehicle to pass over the detectors before the vehicle comes to a complete stop. In this case, the measured travel time of the vehicle's front wheels will be less than the travel time of the vehicle causing the average travel time for the four wheels to be less than the true travel time of the vehicle. This explains the negative percent error values in Table 4. The fact that the percent error values for the microprocessor system are always either zero or negatives suggests that a correction factor could easily be applied to the these results.

CONCLUSIONS AND RECOMMENDATIONS

The automated data collection system developed in this research was proven to be theoretically sound. However, a problem was encountered whenever the device was taken into the field for data collection. Close examination of the software, including a detailed analysis of the time-dependent routines, revealed no problems. Similarly, in extensive laboratory testing using a function generator to simulate switch closures, the system displayed correct, predictable results.

For these reasons, it was concluded that the performance of the tapeswitches was probably not as "clean" as presumed in the design. A detailed analysis of the mechanical and electrical characteristics of the tapeswitches was beyond the scope of this research and is therefore recommended for further research.

As designed, the system cannot tolerate so-called "ghost" vehicles that may appear or disappear from entry or exit points between the two sets of detectors. It would be possible to develop additional hardware and software capable of handling intermediate entry and exit points within the approach delay section.

The possibility of applying a correction factor to the microprocessor data was alluded to previously. A large quantity of data would be required to identify a factor that could be correlated to the error introduced by wheel counting.

Finally, the feasibility of developing a more convenient detection scheme, or an improved version employing electrical tapeswitches, should be investigated. A single strip containing the separate 2-ft tapeswitches would significantly reduce the cost of setup and removal.

REFERENCES

1. J.F. Legere. Automated Collection of Vehicular Delay Data at Intersections. M.S. thesis. Virginia Polytechnic Institute and State University, Blacksburg, 1983.
2. C.C. Gardner, J.H. Kell, and W.R. Reilly. A Technique for Measurement of Delay at Intersections, Vol. 1. JHK & Associates, Alexandria, Va., Sept. 1976.
3. SDK-85 System Design Kit User's Manual. Intel Corporation, Santa Clara, Calif., 1978.
4. MCS-80/85 Family User's Manual. Intel Corporation, Santa Clara, Calif., Oct. 1979.
5. Ribbon Switches. Industrial Catalog C-8. Tapeswitch Corporation of America, Farmingdale, N.Y., 1981.

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

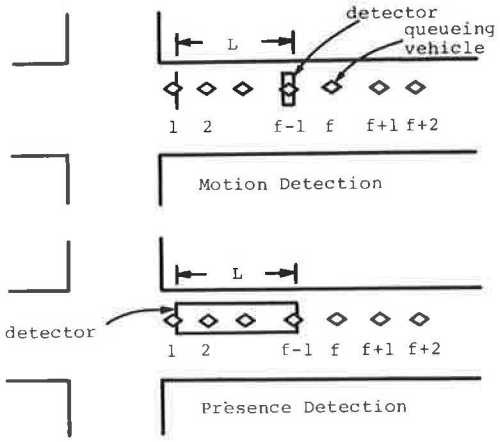


FIGURE 1 Detector placements for traffic-actuated control.

cles on an exclusive left-turn lane on Erie Boulevard were also included in the data collection.

At the three sites, parking was prohibited and pedestrian influences on the vehicular flows were virtually nonexistent. The approach speeds of the observed vehicles were generally between 30 and 35 mph. The flows in both selected left-turn lanes were characterized by a significant number of U-turns around curbed medians.

To facilitate data collection, hypothetical placements of detectors corresponding to detector setbacks or detector lengths of 30, 50, 80, and 120 ft were used. No distinctions were made between motion detection and presence detection. The reason for this is that data can be collected with respect to the same variable, L, as shown in Figure 1, in either case.

For each detector placement, two types of data concerning the movement of a queueing vehicle were recorded. One was the elapsed time from the onset of a green duration to the moment the front bumper of a queueing vehicle reached the upstream edge of the detector. This elapsed time will be referred to as arrival time for the sake of convenience. The other type of data was the elapsed time from the onset of the same green duration to the moment the rear bumper of the same vehicle crossed the stop line (for motion detection) or the downstream edge of the detector (for presence detection). This elapsed time will be defined as departure time. The arrival times and the departure times of the queueing vehicles were recorded simultaneously by two observers as audio signals on tape recorders. The signals were then decoded with a strip chart recorder.

In addition to the data specific to each queueing vehicle, the queueing position of the first vehicle upstream of a detector at the onset of a green duration was recorded for each observed queue. This information serves as a basis for correlating the arrival time of a vehicle with the departure time of another vehicle. The number of queueing vehicles ahead of the first queueing vehicle upstream of the detector varies from one queue to another. It can affect arrival times and, therefore, was included in the data collection.

QUEUE DISCHARGE HEADWAYS

The operations of both pulse control and presence control are influenced by the departure times of queueing vehicles after a green duration begins. The departure time, D_i, of a vehicle can be represented by the sum of a series of queue discharge headways:

$$D_i = \sum_{K=1}^{i} H_K \quad (1)$$

where H_K is the queue discharge headway of a vehicle in the Kth queueing position.

As given in Table 1 the average queue discharge headway (m) of the straight-through vehicles on Almond Street was 3.1 sec for vehicles in the first queueing position. This was based on a sample size (N) of 256 queues. The average headways fell to 2.5 sec for vehicles in the second queueing position and rapidly leveled off to about 2.1 sec. The standard deviations (S) of the queue discharge headways were in the range of 0.5 to 0.9 sec.

The left-turn vehicles on Erie Boulevard had average queue discharge headways 0.1 to 0.3 sec longer than the averages for the straight-through vehicles on Almond Street. However, the standard deviations of their discharge headways were also in the range of 0.5 to 0.9 sec. The average headways of the left-turn vehicles on Almond Street were up to 0.2 sec longer than those observed on Erie Boulevard. This is probably due to more restricted geometric design features (e.g., 8-ft lanes) encountered on Almond Street.

The cumulative distributions of the discharge headways of the vehicles in the first queueing positions were distinctly different from those of the others. Significant variations were also present among the distributions of the headways of those vehicles not in the first queueing positions. These variations were largely associated with headways of between 2 and 3 sec. The ranges of the observed headways decrease from about 6 seconds for vehicles in the first queueing position to about 4 seconds for those in the eight and ninth positions.

For each of the queueing positions given in Table 1, the cumulative distribution of the discharge headways can be replaced by a cumulative distribution of the discharge headways as percentage of the mean headway. This requires the headways, t, in the table to be converted into the percentages of the mean, m. With this simple manipulation, an interesting characteristic of the discharge headways emerges. As shown in Figure 2, the headway distributions for the various queueing positions can be realistically represented by a single normalized distribution for both straight-through and left-turn movements.

This normalized distribution can be used in several ways for simulating the discharge headways. Let

p = percentage of the mean headway for a given queueing position and
w = probability of a headway being less than or equal to p.

Then, one convenient way is to represent the distribution in terms of the following five linear equations:

$$p = 40 + 300\,w \quad \text{if } w \leq 0.1 \quad (2a)$$
$$p = 64 + 63\,w \quad \text{if } 0.1 < w \leq 0.5 \quad (2b)$$
$$p = 53 + 83\,w \quad \text{if } 0.5 < w \leq 0.8 \quad (2c)$$
$$p = 13 + 134\,w \quad \text{if } 0.8 < w \leq 0.95 \quad (2d)$$
$$p = -1{,}380 + 1{,}600\,w \quad \text{if } w > 0.95 \quad (2e)$$

In a simulation process w can be generated as a random number with a value between 0 and 1. The p value associated with this w can then be determined from one of the above equations. Multiplying this p by the mean headway for the corresponding queueing position results in a headway that belongs to the observed distribution.

Vehicle-Detector Interactions and Analysis of Traffic-Actuated Signal Controls

FENG-BOR LIN and MARTIN C. PERCY

ABSTRACT

In modeling the queueing behavior of vehicles for analyzing traffic-actuated signal controls, existing simulation studies have treated only queue discharge headways explicitly. Other vehicle-detector interactions, which govern the initiation, extension, and termination of a green duration, are largely embedded in obscure algorithms for processing vehicles. This negligence has led to misleading and unrealistic findings. Based on data collected at three intersections in Syracuse, New York, some basic characteristics of vehicle-detector interactions that should be accounted for in simulating traffic-actuated controls are described. Further studies are needed to establish baseline information for examining the role of such interactions in shaping control efficiency and for calibrating simulation models.

Traffic-actuated signal controls commonly used today include semiactuated control, full-actuated control, and volume-density control. These controls are operated either on pulse mode with small-area, motion detectors or on presence mode with large-area detectors. One of the major problems in employing these controls lies in the determination of ways of achieving high control efficiencies. This problem is often dealt with through computer simulation analysis.

Existing simulation analyses of traffic-actuated signal controls appear to be very weak in modeling the interactions between queueing vehicles and detectors. These vehicle-detector interactions can be perceived to begin when a vehicle enters a detection area and to end when the vehicle moves downstream into an intersection. In a simulation analysis, usually only queue discharge headways have been treated explicitly. The way queueing vehicles actuate detectors is invariably buried in an embedded algorithm for processing vehicles. This lack of attention to such interactions has quite possibly led to incompatible or even unrealistic conclusions regarding the operating characteristics of traffic-actuated controls. Several existing analyses serve to underline this possibility.

In an analysis of pulse mode full-actuated control, Morris and Pak-poy (1) point out that, to minimize delays, vehicle interval should be reduced from about 8 seconds at an approach volume of 200 vehicles per hour (vph) to about 2 seconds at 800 vph. A study by Michalopoulos et al. (2) based on a UTCS-1 model (3) revealed a different operating characteristic for such a control. It showed that vehicle intervals of 2 sec resulted in better control efficiencies than longer ones for traffic volumes ranging from 200 to 700 vph per lane. Using the NETSIM model (4), Tarnoff and Parsonson (5) further found that vehicle intervals of 1 sec produced shortest delays for traffic volumes ranging from 200 to 1,500 vph per approach. In the same study they also examined presence mode full-actuated controls for detector lengths of 30, 60, and 90 ft. These lengths corresponded to extension intervals of 0.6, 1.2 and 1.8 sec under the simulated conditions. The analysis showed that delays decreased as detector length was shortened.

Something is obviously amiss in the results of these analyses. Queueing vehicles in a traffic lane often require more than 2 sec to reach and actuate a detector successively. When vehicle intervals of 1 or 2 sec are used, queueing vehicles may not have ample opportunities to actuate and extend green durations. Consequently, premature termination of a green duration may take place. Such an event is more likely to degrade the efficiency of a control than to improve it. Realizing this, Tarnoff and Parsonson cautioned that the simulation results should not be taken for granted.

Their caution deserves attention. Simulation is perhaps the only practical means of providing a comprehensive analysis of a signal control. Yet simulation based on unrealistic modeling of vehicle-detector interactions may result in misleading information. Grave consequences may follow if such information is unknowingly used as a basis for operating existing signal controls or for developing future generations of control.

In light of these possible implications, there is a need to identify the vehicle-detector interactions under traffic-actuated controls. In response to this need, a preliminary effort was made to collect relevant field data at three intersections in Syracuse, New York. The observed vehicle-detector interactions are described, and several basic requirements for simulating such interactions are identified.

DATA COLLECTION

When motion detectors are used, the operation of a control is dictated by the arrivals of vehicles at the detectors. These arrivals can be influenced by the detector setback that is denoted as L in Figure 1. If presence detectors are used instead, the arrivals of vehicles at the upstream edges of the detectors as well as the subsequent departures of the same vehicles from the downstream edges can affect the operation of the control. Therefore, the movement of a vehicle in relation to the detector length becomes an important factor. This detector length is also denoted as L in Figure 1.

Three sites in Syracuse, New York, were selected for data collection. These sites were the intersections between Almond Street and Harris Street, Almond Street and Adam Street, and Erie Boulevard and Kinne Avenue. At the first site, the queueing vehicles in three straight lanes on Almond Street were the subject of the data collection. At the second site, the movements of queueing vehicles in an exclusive left-turn lane on Almond Street were recorded. And, at the last site, the queueing vehi-

TABLE 1 Cumulative Distribution of Queue Discharge Headways (H_K)

Headway t (sec)	Probability of $H_K \leq t$ Queueing Position 1	2	3	4	5	6	7	8	9
	a) Straight-Through Vehicles on Almond								
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.5	0.01	0.03	0.07	0.07	0.12	0.10	0.10	0.14	0.12
2.0	0.09	0.25	0.35	0.40	0.47	0.43	0.38	0.52	0.51
2.5	0.26	0.53	0.71	0.72	0.75	0.75	0.71	0.89	0.78
3.0	0.49	0.79	0.87	0.89	0.91	0.92	0.87	0.94	0.97
3.5	0.75	0.93	0.96	0.97	0.98	0.97	0.98	0.98	1.00
4.0	0.86	0.97	0.98	0.99	1.00	0.99	1.00	0.99	1.00
4.5	0.92	0.98	1.00	0.99	1.00	0.99	1.00	1.00	1.00
5.0	0.95	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5.5	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
m	3.1	2.5	2.3	2.2	2.1	2.2	2.2	2.0	2.1
S	0.9	0.9	0.7	0.6	0.6	0.6	0.6	0.6	0.5
N	256	256	256	241	206	157	115	86	59
	b) Left-Turn Vehicles on Erie								
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.5	0.01	0.03	0.05	0.04	0.08	0.05	0.08	0.04	0.04
2.0	0.05	0.23	0.26	0.33	0.31	0.33	0.36	0.28	0.48
2.5	0.19	0.53	0.55	0.64	0.66	0.58	0.58	0.67	0.78
3.0	0.45	0.77	0.83	0.87	0.88	0.77	0.80	0.89	0.93
3.5	0.64	0.90	0.93	0.95	0.98	0.92	0.92	0.98	0.98
4.0	0.79	0.95	0.99	0.98	0.98	0.96	0.96	1.00	1.00
4.5	0.89	0.99	1.00	0.99	0.99	0.99	0.99	1.00	1.00
5.0	0.94	0.99	1.00	1.00	0.99	1.00	1.00	1.00	1.00
5.5	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
m	3.2	2.6	2.5	2.4	2.3	2.5	2.4	2.3	2.2
S	0.9	0.7	0.7	0.6	0.6	0.8	0.7	0.5	0.5
N	175	146	152	141	131	120	85	57	46

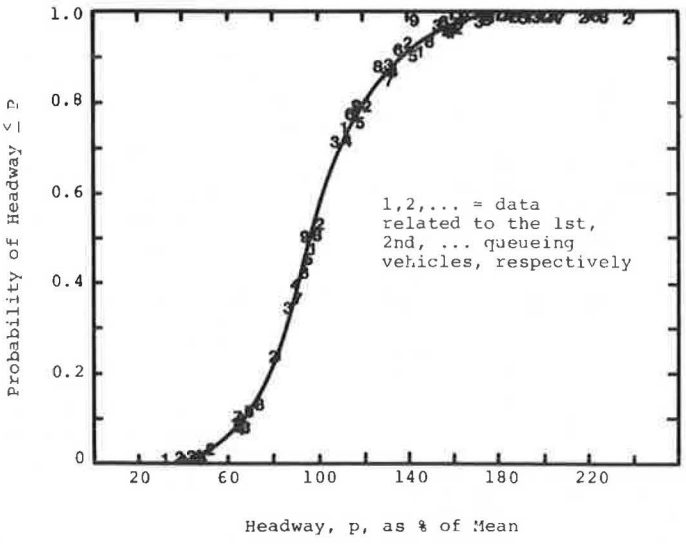


FIGURE 2 Normalized cumulative distribution of departure headways.

ARRIVAL CHARACTERISTICS

Let

f = queueing position of the first vehicle upstream of a detector at the onset of a green duration (Figure 1),
A_f = arrival time of the fth queueing vehicle, and
T_K = arrival headways of queueing vehicles behind the first upstream queueing vehicle (K = f + 1, f + 2, ...).

Then the arrival time, A_i, of a queueing vehicle in the ith queueing position can be determined as

$$A_i = A_f + \sum_{K=f+1}^{i} T_K \quad i > f \tag{3}$$

The value of A_f can be affected by the movement of the vehicle in the (f-1)th queueing position. This vehicle is the one just ahead of the first queueing vehicle upstream of a detector at the onset of a green duration. It is also the last queueing vehicle that has reached the detector by the time the green

light is turned on (Figure 1). Let the departure time of this leading vehicle be denoted as D_{f-1}. The field data show that the following general relationship exists between A_f and D_{f-1}:

$$A_f = V(D_{f-1}) \pm R \quad (4)$$

where $V(D_{f-1})$ is a deterministic component of A_f expressed as a function of D_{f-1} and R is a probabilistic component representing the variations of A_f from $V(D_{f-1})$.

A typical relationship between A_f and D_{f-1} is shown in Figure 3. Table 2 gives a summary of the specific forms of Equation 4 as determined from the field data. The variations of A_i from $V(D_{f-1})$ were found to be mostly less than 1.5 sec. This implies that the values of R range from 0 to 1.5 sec. Existing data, however, are not sufficient to determine the distribution of R at a given level of D_{f-1}.

The use of the equations in Table 2 requires the determination of D_{f-1} first. D_{f-1} depends very much on the position of the (f-1)th queueing vehicle. This position corresponds to the number of queueing vehicles that are at least partly inside the area defined by L (Figure 1) at the onset of a green duration. The data in Table 3 indicate that, for a given detector placement, this number cannot be reasonably assumed to be a constant. Therefore, a simulation analysis should also take this characteristic into consideration. Otherwise, systematic underestimates or overestimates of A_f will result.

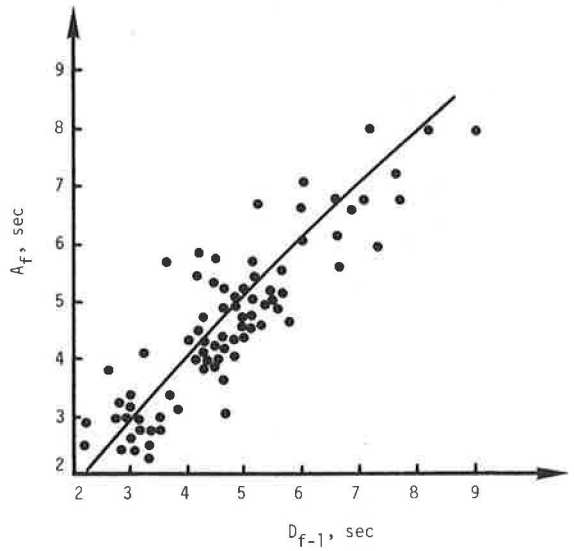


FIGURE 3 A_f - D_{f-1} relationship of straight-through queueing vehicles on Almond Street (L = 30 ft).

In contrast to A_f, the arrival headway T_K in Equation 3 can be conveniently determined from an observed distribution of arrival headways. The cumulative distributions of the arrival headways observed at the study sites are given in Tables 4 and 5. The means, standard deviations, and ranges of these distributions follow a pattern similar to those of the queue discharge headways.

Again, each of the distributions given in Tables 4 and 5 can be normalized by expressing the headways as percentages of the respective means. It can be shown that the resultant distributions are virtually the same as the one shown in Figure 2. Thus, the departure headways and the arrival headways may be governed by the same natural law regardless of the queueing position of a vehicle. If this is true, the input requirements for achieving a realistic simulation of the vehicle-detector interactions can be significantly reduced.

PREMATURE TERMINATION OF GREEN

Presence Control

Under a presence control a vehicle can demand right-of-way and hold the green by staying in a detection area. After a vehicle departs from the detection area, the green is continued for a period equal to a specified extension interval. If no detector actuations take place during this extension interval, the green could be terminated. This condition for holding the green can be represented by

$$D_{i-1} - \beta_{i-1} + E \geq A_i + \alpha_i \quad (5a)$$

or

$$A_i - D_{i-1} \leq E - \alpha_i - \beta_{i-1} \quad (5b)$$

where A_i and D_{i-1} are, respectively, the arrival time of the ith queueing vehicle and the departure time of the (i-1)th queueing vehicle; E is the ex-

TABLE 2 A_f - D_{f-1} Relationships

L(ft)	Equation
a) Straight-Through Flows on Almond	
30	$A_f = -0.900 + 1.450\ D_{f-1} - 0.050\ D_{f-1}^2 \pm R$
50	$A_f = -1.875 + 1.300\ D_{f-1} - 0.025\ D_{f-1}^2 \pm R$
80	$A_f = -1.615 + 1.029\ D_{f-1} - 0.018\ D_{f-1}^2 \pm R$
120	$A_f = 0.332 + 0.701\ D_{f-1} - 0.008\ D_{f-1}^2 \pm R$
b) Left-Turn Flows on Almond and Erie	
30	$A_f = -0.195 + 1.164\ D_{f-1} - 0.033\ D_{f-1}^2 \pm R$
50	$A_f = -1.525 + 1.050\ D_{f-1} - 0.017\ D_{f-1}^2 \pm R$
80	$A_f = -4.369 + 1.617\ D_{f-1} - 0.050\ D_{f-1}^2 \pm R$

TABLE 3 Percent of Queues with M Vehicles Fully or Partially in the Area Defined by L

M	Straight-Through on Almond L(ft)				Left Turns on Almond and Erie L(ft)			
	30	50	80	120	30	50	50[1/]	80
1	48.1	-	-	-	34.5	-	-	-
2	51.9	57.1	1.9	-	55.2	66.7	38.0	-
3	-	42.9	38.9	-	10.3	33.3	62.0	50.0
4	-	-	59.2	12.5	-	-	-	50.0
5	-	-	-	53.1	-	-	-	-
6	-	-	-	34.4	-	-	-	-

1/ On Almond Street

TABLE 4 Cumulative Distributions of Arrival Headways (T_K) of Straight-Through Vehicles on Almond Street.

Headway t (sec)	Probability of $T_K \leq t$ Position f+1	f+2	f+3	f+4	f+5	Position f+1	f+2	f+3	f+4	f+5
	a) L = 30 ft					b) L = 50 ft				
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.5	0.00	0.01	0.00	0.00	0.06	0.00	0.01	0.09	0.07	0.15
2.0	0.06	0.22	0.27	0.24	0.35	0.09	0.29	0.40	0.56	0.45
2.5	0.33	0.61	0.75	0.60	0.71	0.34	0.69	0.69	0.80	0.90
3.0	0.59	0.84	0.93	0.88	0.82	0.60	0.84	0.93	0.97	1.00
3.5	0.82	0.97	0.95	1.00	1.00	0.82	0.97	1.00	1.00	1.00
4.0	0.95	0.99	1.00	1.00	1.00	0.91	1.00	1.00	1.00	1.00
4.5	0.99	0.99	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
5.0	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
m	2.9	2.4	2.3	2.5	2.1	2.9	2.3	2.2	2.1	2.0
S	0.6	0.5	0.5	0.5	0.5	0.8	0.6	0.5	0.5	0.4
N	84	66	44	25	17	95	68	45	30	20
	c) L = 80 ft					d) L = 120 ft				
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1.5	0.00	0.02	0.07	0.16	0.07	0.00	0.00	0.08	0.10	0.00
2.0	0.10	0.25	0.36	0.42	0.53	0.03	0.09	0.33	0.19	0.22
2.5	0.23	0.58	0.64	0.74	0.67	0.17	0.59	0.69	0.51	0.67
3.0	0.48	0.88	0.89	0.95	0.93	0.59	0.83	0.92	1.00	0.94
3.5	0.65	0.98	0.96	1.00	1.00	0.79	0.96	0.92	1.00	1.00
4.0	0.83	0.98	1.00	1.00	1.00	0.93	1.00	0.96	1.00	1.00
4.5	0.92	0.98	1.00	1.00	1.00	0.93	1.00	1.00	1.00	1.00
5.0	0.98	0.98	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00
5.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
m	3.2	2.5	2.3	2.1	2.2	3.1	2.4	2.3	2.4	2.4
S	0.9	0.7	0.6	0.6	0.5	0.7	0.4	0.8	0.5	0.4
N	48	40	28	19	15	29	27	26	21	18

tension interval; β_{i-1} represents the departure response time (i.e., the time required for a detector to recognize the departure of a vehicle before the rear bumper of the vehicle crosses the downstream edge of a detector); and α_i is the arrival response time (i.e., the time required for a detector to recognize the arrival of a vehicle after the front bumper of the vehicle crosses the upstream edge of the detector).

The arrival response time, α_i, and the departure response time, β_i, of a loop presence detector depend in part on the sensitivity of the detector and the type of vehicle. They are also a function of individual vehicle speed. In a field test using a 1981 Reliant K over a 62-ft-long detector, it was found that the front bumper has to be 2.5 to 3 ft inside the detector to actuate the detector. The same test also showed that the rear bumper should be at least about 2 ft inside the downstream end of the detector to hold the green. Therefore, a vehicle traveling at a constant speed of 10 mph would have an α_i of just under 0.2 sec and a β_i of approximately 0.13 sec. For a given detector length, queueing vehicles further upstream from the detector would have lower α_i and β_i than those closer to it because of their higher speeds.

The right-hand side of Equation 5b can be referred to as adjusted extension interval. To prevent premature termination of a green duration, the value of A_i minus D_{i-1} cannot exceed the corresponding adjusted extension interval. For a given adjusted extension interval, the chance of premature termination of a green duration increases as the length of the detection area is shortened. This characteristic is shown in Figure 4 in terms of the observed rela-

TABLE 5 Cumulative Distributions of Arrival Headways (T_K) of Left-Turn Vehicles on Almond and Erie streets.

Headway t (sec)	Probability of $T_K \leq t$ Position f+1	f+2	f+3	f+4	f+5	Position f+1	f+2	f+3	f+4	f+5
	a) L = 30 ft (Erie)					b) L = 50 ft (Erie)				
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-	-
1.5	0.00	0.04	0.07	0.08	0.07	0.00	0.05	0.00	-	-
2.0	0.14	0.46	0.35	0.58	0.40	0.09	0.17	0.33	-	-
2.5	0.41	0.88	0.57	0.89	0.87	0.26	0.62	0.80	-	-
3.0	0.79	0.96	0.83	0.89	0.93	0.59	0.90	0.87	-	-
3.5	0.93	0.96	1.00	1.00	0.93	0.85	1.00	1.00	-	-
4.0	0.93	1.00	1.00	1.00	1.00	0.85	1.00	1.00	-	-
4.5	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	-	-
5.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-	-
m	2.7	2.1	2.4	2.0	2.3	2.9	2.4	2.3	-	-
S	0.6	0.5	0.6	0.5	0.8	0.7	0.5	0.5	-	-
N	29	24	23	19	15	27	21	15	-	-
	c) L = 80 ft (Erie)					d) L = 50 ft (Almond)				
1.0	0.00	0.00	0.00	0.00	-	0.00	0.00	0.00	0.00	0.00
1.5	0.00	0.04	0.04	0.17	-	0.00	0.00	0.03	0.00	0.04
2.0	0.12	0.46	0.36	0.52	-	0.06	0.12	0.21	0.21	0.22
2.5	0.33	0.75	0.75	0.78	-	0.23	0.47	0.67	0.59	0.63
3.0	0.64	0.89	0.93	0.96	-	0.57	0.81	0.92	0.82	0.89
3.5	0.85	0.96	0.96	0.96	-	0.87	0.95	1.00	0.94	1.00
4.0	1.00	1.00	1.00	1.00	-	0.89	1.00	1.00	1.00	1.00
4.5	1.00	1.00	1.00	1.00		0.94	1.00	1.00	1.00	1.00
5.0	1.00	1.00	1.00	1.00		0.96	1.00	1.00	1.00	1.00
5.5	1.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00	1.00
m	2.9	2.2	2.3	2.2		3.0	2.6	2.4	2.5	2.4
S	0.6	0.6	0.6	0.7		0.8	0.5	0.4	0.6	0.5
N	33	28	28	23		47	43	39	33	27

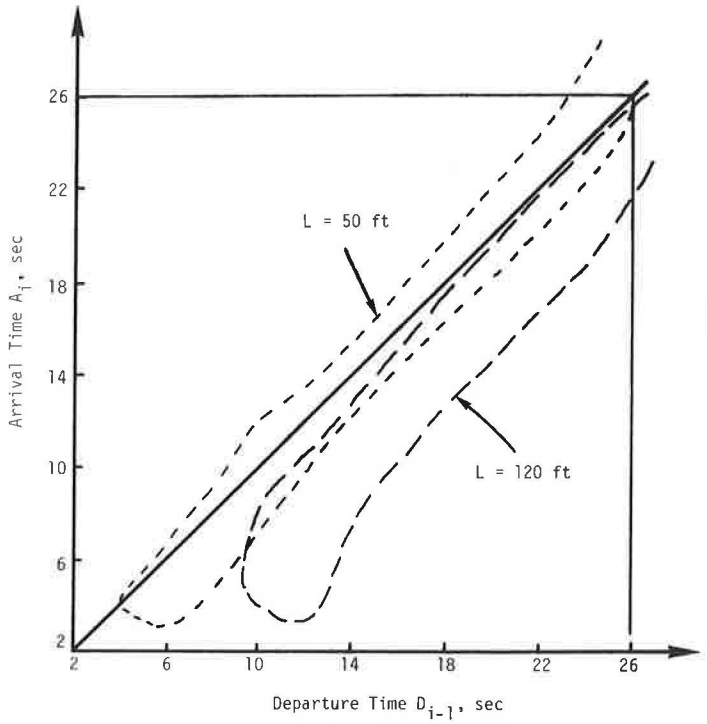


FIGURE 4 Domains of observed correlations between A_i and D_{i-1} of straight-through queueing vehicles on Almond Street.

tionship between A_i and D_{i-1} for the straight-through queueing vehicles on Almond Street. With a 120-ft detector, D_{i-1} exceeded A_i in all the observations. This implies that a 120-ft detector allows a queueing vehicle to actuate the detector before the vehicle ahead leaves the detection area. Consequently, premature termination of a green duration is an unlikely event. In contrast, a large number of the observed queueing vehicles had $A_i > D_{i-1}$ for a detector 50 ft long. In such a

case, a short adjusted extension interval can lead to frequent premature termination of green durations.

Figure 5 shows the probabilities of premature termination of a green duration for detector lengths of 30 and 50 ft. These probabilities were determined from the field data by considering only one traffic lane. The corresponding probabilities for detector lengths greater than 80 ft are not shown because they are negligibly small. It is evident from Figure 5 that the use of detectors shorter than 50 ft will require a careful selection of an extension interval to avoid premature termination of a green duration.

The probabilities shown in Figure 5 can be reduced if more than one lane is associated with a signal phase. To what extent the probabilities can be reduced, however, depends on the actual traffic and signal control conditions.

Premature termination of a green duration is usually undesirable. To prevent it or to lessen the chance of its occurring, long detectors can be used. This alternative, however, would increase the dwell time of a vehicle in a detection area (Figure 6). The dwell time is the time a vehicle spends in a detection area after a green duration begins. A long dwell time of a vehicle results in sluggish transfers of right-of-way and, thus, may increase delays. Another alternative to ease the problem is to use short detectors in conjunction with long extension intervals. It is still not clear, however, what combinations of detector length and extension interval would result in high control efficiencies.

Motion Control

When a motion control is employed, premature termination of a green duration is governed primarily by the formation of queues and by the settings of initial interval and vehicle interval. Following actuation by a vehicle, a motion detector requires a short time to produce a pulse. This time interval varies with the make and the sensitivity setting of the detector and ranges from about 4 milliseconds to about 200 milliseconds. The pulse duration is usually 75 to 150 milliseconds per vehicle unless a vehicle stalls in the detection area. The time required for a detector to sense the arrival of a vehicle is also very short. Moving a 1973 Valiant over a detector 6 ft long, it was found that the detector can be actuated when the front bumper is less than 1.5 ft inside the detector. This is equivalent to an arrival response time of about 0.1 sec at a vehicle speed of 10 mph.

The elapsed time between two successive actuations of a detector is $A_{i+1} + \beta_{i+1} - A_i - \beta_i$ or $T_{i+1} + \beta_{i+1} - \beta_i$. This elapsed time should be shorter than the vehicle interval in order to extend a green duration. The difference between $\beta_{i+1} - \beta_i$ is likely very small. Therefore, the headway T_{i+1} should be shorter than the vehicle interval in order to extend a green duration.

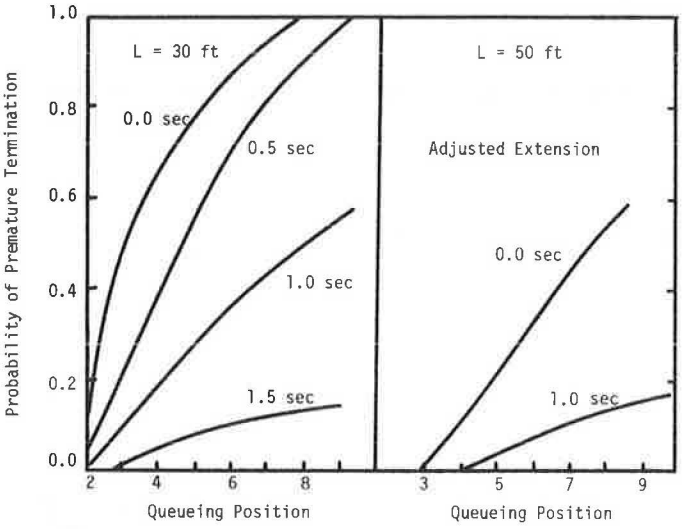


FIGURE 5 Probabilities of premature termination of green under a presence control (single lane flow).

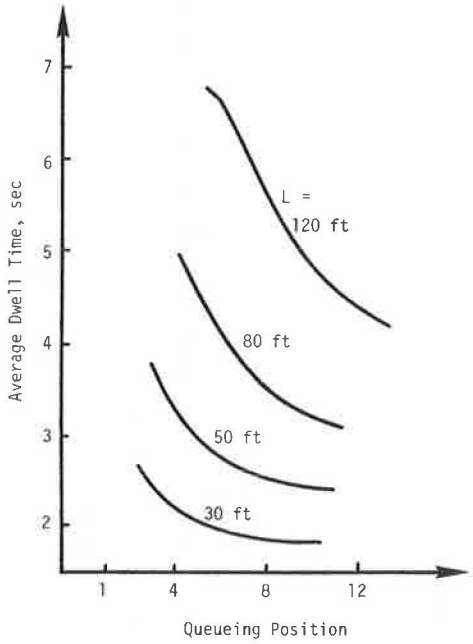


FIGURE 6 Average dwell time as a function of L and queueing position.

It can be seen from Tables 4 and 5 that arrival headways of less than 1.0 sec did not exist at the study sites. Furthermore, the probability of an arriving headway being less than 2 sec was very low for the queueing vehicles just upstream of a detector. Therefore, if a queue extends upstream of a detector, the use of a vehicle interval of 2 sec can easily result in premature termination of a green duration. This can be illustrated with a simple example.

Consider a motion control with a detector setback of 80 ft. At the beginning of a green duration there are six queueing vehicles upstream of the detector. Assume that the first three vehicles (f, f+1, and F+2) can either cross the detector during the initial interval or extend the green duration. The vehicle interval is set at 2 sec. The problem is to determine the probability of premature termination of the green for the remaining three vehicles.

Based on the data in Table 4 for L = 80 ft, the approximate probabilities that vehicles f+3, f+4, and f+5 will extend a green duration are, respectively, 0.36, 0.42, and 0.53. Therefore, the probability that at least one of these three vehicles will be unable to extend the green duration is 1 - 0.36 x 0.42 x 0.53 = 0.92. Even with a vehicle interval of 2.5 sec, the corresponding probability of premature termination of the green duration is still 0.73. For this reason, it is unlikely that vehicle intervals as short as 1 or 2 sec can improve the performance of a motion control under a wide range of traffic flow conditions.

IMPLICATIONS

The vehicle-detector interactions described previously could create several problems in the analysis of a control. One major problem is the danger of generating misleading information when such interactions are not properly accounted for. This danger is illustrated with two examples shown in Figure 7 for a full-actuated motion control and in Figure 8 for a presence control. In both examples a two-phase operation with two lanes in each phase was examined through computer simulation. The analysis relied on

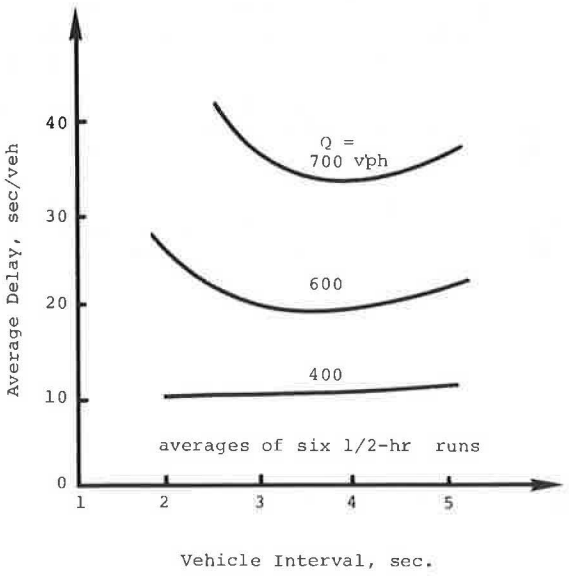


FIGURE 7 Average vehicle delay as a function of vehicle interval and flow rate for L = 120 ft (pulse control).

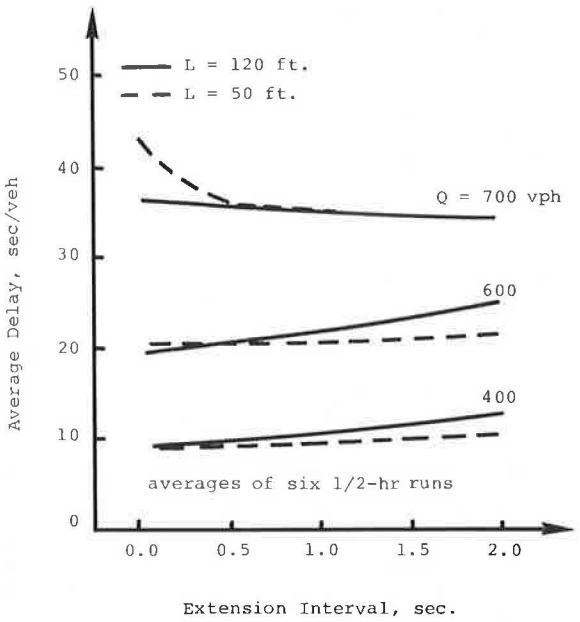


FIGURE 8 Average delay as a function of detector length, extension interval, and flow rate (presence control).

a simulation model (RAPID) that was developed at Clarkson College of Technology. In the analysis, the traffic volume, Q, was assumed to be the same in all lanes. Only straight-through vehicles were considered. And the observed vehicle-detector interactions shown in Tables 1-4 for such vehicles were used as input in the simulation.

For the motion control, the initial interval, maximum green, and clearance interval were, respectively, 10, 50, and 4 sec for either phase. The initial interval was followed automatically by one vehicle interval. The detector setback was 120 ft. And the same maximum green and clearance intervals were used for the presence control.

Contrary to the findings of other studies, Figure 7 shows that, for a motion control, shortening the vehicle interval will not necessarily reduce the delays. When traffic flows are light (e.g., Q = 400 vph/lane), the chance for queues to extend beyond a detector with 120 ft in setback is slim. Under this condition premature termination of a green duration is unlikely to occur. As a result, the average delays become insensitive to the vehicle interval. With heavier flows long queues may exist in most of the cycles. Consequently, the use of short vehicle intervals makes it difficult for many queueing vehicles to extend green durations to satisfy their needs. This explains why in Figure 7 vehicle intervals of less than 3 sec increase delays substantially for flows of 600 and 700 vph/lane.

Similarly, Figure 8 shows that, under a presence control, a short detector length (L = 50 ft) is not always better than a long one (L = 120 ft). This is a characteristic not identified in other studies. The use of 120-ft-long detectors virtually eliminates premature termination of a green duration, but it produces long dwell times. As expected, delays increase with the extension interval under light flow conditions. When 50-ft-long detectors are used, the interplays between premature termination, dwell time, and extension interval could give rise to inefficient controls. For example, an extension interval of 0.0 sec allows the operation of a control to be dominated by premature termination of a green duration. This, in turn, brings about excessive delays under heavy flow conditions.

Obviously misleading or unrealistic simulation results would only raise doubts in the minds of informed engineers about the reliability of a simulation model. More serious consequences may emerge if such results are accepted by unwary users. Errors in simulation results caused by improper modeling of flow characteristics are usually difficult to detect. Therefore, it is advisable that simulation models be calibrated with observed vehicle-detector interactions.

Another problem in dealing with the vehicle-detector interactions is related to the sensitivity of signal operations to such interactions. A high sensitivity would require accurate input data or demand more simulation runs. A low sensitivity could tolerate greater errors in the measurement of the interactions. Of particular concern in this regard are the A_f-D_{f-1} relationship, the values of the detector response times α_i and β_i, and the distributions of the queue discharge headways and the arrival headways.

The measured relationship between A_f and D_{f-1} may be biased by the discrepancies in observers' reactions to the beginning of a green duration. The experience gained in this study indicates that the discrepancies can be confined to a value of less than 0.3 sec. Allowing the value of A_f to increase or decrease by 0.3 sec from those estimated from the equations given in Table 3, the resultant simulated average delays differ by less than 0.4 sec per vehicle. For all practical purposes, such differences are not significant. Varitions in α_i and β_i of up to 0.3 sec are also found to have negligible impact on control efficiency.

On the other hand, variations in the queue discharge headways and the arrival headways could have significant effects. For example, let the distributions of the straight-through headways given in Tables 1 and 4 be shifted by 0.15 sec to shorter headways. This shift results in substantial reductions in the average delays presented in Figures 7 and 8 for Q = 600 and 700 vph/lane. The reductions reach as high as 20 to 30 percent, particularly when the problem of premature termination of a green duration exists. Therefore, the accurate representation of the distributions of queue discharge headways and arrival headways is of foremost concern in the analysis of a traffic-actuated control.

CONCLUSIONS

The vehicle-detector interactions under a traffic-actuated control can be characterized by (a) queue discharge headway, (b) arrival headway, (c) the A_f-D_{f-1} relationship, (d) the number of queueing vehicles in the area defined by L (Figure 1) at the onset of a green duration, and (e) dwell time. These flow characteristics are probabilistic entities and should be treated as such. For simulation analysis of a traffic-actuated control, a model should be calibrated in terms of the observed characteristics of such parameters and relationships.

For both straight-through and left-turn queueing vehicles, the distributions of the observed queue discharge headways and arrival headways were found to follow similar trends. When the headways are expressed as percentages of the means for respective queueing positions, these distributions can be replaced by a single normalized distribution.

The difficulty observers have starting measurement simultaneously in response to the beginning of a green duration can bias the observed A_f-D_{f-1} relationships. The biases, however, are small and do not significantly affect simulated vehicle delays. The arrival response time and the departure response time of a detector are not fully understood. Nevertheless, they appear to be rather short, corresponding to the time required for a vehicle to travel a distance of 2 to 3 ft. Their effects on vehicle delays also seem to be negligible. In contrast, vehicle delays are rather sensitive to the distribution of queue discharge headways.

Under a presence control, the chance for premature termination of a green duration increases when detector lengths are shortened. A detector length of longer than 80 ft can effectively eliminate the premature termination. Using long detectors, however, results in long dwell times and may thus reduce control efficiency. The tradeoffs between long detectors and short detectors are not well understood at present.

Arrival headways of less than 2 sec constitute only a small portion of the observed headways. Therefore, when a vehicle interval of 1 or 2 sec is used, a motion control cannot be expected to achieve a high control efficiency under a wide range of traffic conditions. A simulation model should be capable of revealing such a characteristic.

This study demonstrates the need to assess existing simulation models that are being used for the analysis of traffic-actuated control. Before a model is used and a conclusion drawn, one should make sure that the model has an adequate representation of the vehicle-detector interactions. Otherwise, misleading information may be generated.

REFERENCES

1. R.W.T. Morris and P.G. Pak-poy. Intersection Control by Vehicle-Actuated Signals. Traffic Engineering and Control, Oct. 1967, pp. 288-293.
2. P.G. Michalopoulos, B. Papapanou, and E.B. Binseel. Performance Evaluation of Traffic Actuated Signals. Journal of the Transportation Engineering Division of ASCE, Vol. 104, No. TE5, Sept. 1978, pp. 621-636.
3. R.D. Worral and E. Lieberman. Network Flow Simulation for Urban Traffic Control Systems, Vol. 1-5. FHWA, U.S. Department of Transportation, 1973.
4. Traffic Network Analyses with NETSIM-A User Guide. Implementation Package FHWA-IP-80-3. FHWA, U.S. Department of Transportation, Jan. 1980.
5. P.J. Tarnoff and P.S. Parsonson. Selecting Traffic Signal Control at Individual Intersections. NCHRP Report 233. TRB, National Research Council, Washington, D.C., June 1981, 133 pp.

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

Investigation of Network-Level Traffic Flow Relationships: Some Simulation Results

HANI MAHMASSANI, JAMES C. WILLIAMS, and ROBERT HERMAN

ABSTRACT

Results of an exploratory study of network-level relationships in an isolated network with a fixed number of vehicles circulating according to the microscopic rules embedded in the NETSIM traffic simulation model are presented. The primary concern was to assess the usefulness of such simulation-based approaches in the investigation of macroscopic network-level traffic relationships. Three specific objectives were addressed: (a) identification of network-level descriptors that are related in operationally useful and simple ways, (b) exploration of some aspects of the two-fluid theory of town traffic, and (c) examination of the traffic flow distribution over the network components. A number of simulations were conducted on the same network under two different sets of control schemes and traffic characteristics at widely varying vehicle concentrations. The results were analyzed with respect to the three objectives, yielding useful insights into network-level traffic phenomena and suggesting promising avenues for further research.

The characterization of traffic in city networks by a small number of network-level descriptors and the quantification of their interrelationships constitute an important problem that has not received sufficient attention to date in traffic research. This problem is important because there is a need (a) to evaluate and compare the overall resulting quality of service of traffic control strategies and designs at the network level, (b) to make comparisons among cities or of the same city over a number of time periods, and (c) to identify possible deficiencies in current operations that result in overall conditions that are below the norm.

A theory of town traffic, based on a two-fluid representation of traffic in a city network, was proposed by Herman and Prigogine (1). This theory leads to a relationship between average time stopped and average time running on a network. Empirical validation of the theory and related phenomena has since been undertaken (2-4). These efforts have relied primarily on data obtained using chase-car techniques, supplemented by aerial photography, and thus have required considerable time and other data collection and analysis resources.

A complementary approach to gaining further insights into the relationships between network-level descriptors and the effect of various network features and system controls is the use of computer simulation. For example, it seems plausible that microscopic traffic simulation models can yield potentially useful insights into a wide variety of situations and conditions that would otherwise require enormous resources to investigate. Computer simulation provides the capability of looking at a network under controlled conditions at relatively reasonable costs. Simulation experiments can guide further theoretical development as well as assist in the judicious selection and design of observational studies to validate key assumptions and results.

One such exploratory study, in which a set of simulation experiments was performed using the NETSIM traffic simulation package, is described. The primary objective of the study was to assess the usefulness of a simulation-based approach to the investigation of macroscopic network-level traffic relationships. More specific technical objectives included attempts (a) to identify relevant network-level descriptors that may be related in operationally useful and simple ways, including an initial determination of these relationships; (b) to explore some aspects of the two-fluid model of town traffic under controlled conditions, including the sensitivity of its parameters to network features and controls; and (c) to examine the distribution of flow on various components of a closed (i.e., isolated) network under a variety of operating conditions and the implication for overall traffic service quality.

The next section briefly introduces the essential conceptual background. The details of the experimental set-up, including the features and characteristics of the test network, are described in the third section. This is followed by a presentation and critical discussion of the principal results obtained in the effort to date. Concluding remarks include suggestions for further research in this highly promising area.

CONCEPTUAL BACKGROUND

The key concepts and theories underlying the research objectives addressed in these simulation experiments can be grouped in three categories: (a) network traffic flow theory, (b) the two-fluid theory of town traffic, and (c) flow distribution on network components. A complete theoretical presentation will not be given here because such presentations can be found in the literature. This introduction is limited to only those concepts necessary for an understanding of the new results presented.

Network Traffic Flow Theory

The concern here was with a macroscopic theory of traffic flow in a network, not with traffic flow on roadways and arterials or isolated intersections, which have been the subject of nearly three decades of traffic research. Three major tasks of this effort were (a) to identify appropriate fundamental macroscopic traffic flow variables, (b) to determine how these variables are interrelated, and (c) to understand how these relationships are affected by the characteristics of the network, including its topology and its control system.

In the first two tasks, it is natural to be guided by the extensive body of knowledge accumulated over years of studying traffic relationships

for arterials. Three fundamental variables are thus of concern: speed, concentration, and flow. To the extent that networkwide averages of these quantities are of interest, the definition, measurement, and computation of the appropriate averages, particularly for the flow variable, may be problematic. For a given period of observation, τ, average speed, V, will be taken as the ratio of total vehicle-miles to total vehicle-hours over the network during period τ (yielding an average speed in miles per hour). This average is taken both over time and over all vehicles in the network.

The average concentration, K, for the same period, τ, is the time average of the number of vehicles per unit lane length. Letting N(t) denote the number of vehicles at time t, and L the lane-miles of roadway, the average concentration, K, can be expressed as

$$K = (1/\tau) \int_{t_o}^{t_o+\tau} [N(t)/L]\, dt$$

where t_o is the beginning of the observation period. Effectively, K can be obtained by dividing total vehicle-hours by the product τL. In these experiments, to avoid averaging over widely fluctuating traffic conditions, concentration has been maintained constant throughout a given observation period.

The definition of an average flow variable, Q, is somewhat less obvious; the one pursued here considers average network flow to be the average number of vehicles per unit time that passes through an "average" point of the network. Letting q_i and ℓ_i, respectively, denote the average flow over the observation period and the length of link i, i = 1, ..., M, where M denotes the total number of links,

$$Q = \left(\sum_{i=i}^{M} \ell_i q_i\right) / \left(\sum_{i=1}^{M} \ell_i\right)$$

Two principal relationships between these three variables will be investigated. The first concerns the joint variation of speed, V, and concentration, K, at the macroscopic networkwide level. A number of studies have addressed this question for arterials, and although several functional forms have been suggested, the underlying trend of a decreasing average speed as a function of increasing concentration is well accepted (5,6). At the network level, although this same general trend can be expected to hold, it cannot be analytically derived from the single roadway relationships because of the added complexity of the interconnections in a network. Furthermore, virtually no empirical evidence is available on this matter; this is undoubtedly due to the considerable resources required for gathering the needed data. The simulations described thus provide an opportunity to examine how K and V may be related at the network level.

The second relationship examined is one that is fundamental in traffic flow theory, namely that Q = KV. This relationship is well established for arterials, but its validity when Q, K, and V are defined for the whole network remains to be verified. Preliminary indications of its validity are offered. The importance of this type of investigation and its eminent relevance to engineering practice are perhaps best underscored by the significant role that traffic stream models for arterials have played in the development of the Highway Capacity Manual (7) and in everyday applications.

Two-Fluid Theory of Town Traffic

The principal concept underlying Herman and Prigogine's two-fluid theory of town traffic (1) initially appeared in their kinetic theory of multilane highway traffic (8) when the transition to the so-called collective flow regime was made at sufficiently high vehicular concentrations. Essentially, traffic in an urban network is viewed as consisting of two traffic "fluids": one composed of the moving vehicles and the other of vehicles that are stopped as a result of congestion, traffic control devices, and other obstructions. Parked vehicles are not considered part of the "stopped" fluid but as a part of the geometrics.

A basic postulate of the two-fluid theory states that V_r, the average speed of the moving vehicles, is related to f_s, the fraction of vehicles stopped, in the following way:

$$V_r = V_m(1-f_s)^n \quad (1)$$

where V_m is the average maximum running speed in the network and n is a parameter that was subsequently found to be a useful indicator of the quality of traffic service in a network (3,4).

That postulate leads to a relationship between three principal variables: T, T_r, and T_s, respectively average travel time, average running (or moving) time, and average time stopped, all per unit distance (note that $T = T_r + T_s$) of the following form (1):

$$T_s = T - T_m^{(1/n+1)}\, T^{(n/n+1)} \quad (2)$$

where T_m is a parameter of the model, equal to V_m^{-1} and thus reflecting the minimum travel time per unit distance in the network under free-flow conditions.

The calibration of the model parameters T_m and n, which were found to be robust characteristics of a given network, for a particular network has to date been conducted using observations on stopped time and moving time gathered by one or more test vehicles circulating in the network and passively following randomly selected vehicles (chase-car technique). The details of the measurement procedures and subsequent data processing and analysis are extensively described elsewhere (3,4).

An important result used in the derivation of the two-fluid model is that the mean fraction of time stopped, $E(T_s/T)$, (taken over the population of vehicles circulating in the network) is equal to f_s, the average of the fraction of vehicles stopped taken over a given observation period. A proof of this result can be found elsewhere (3,4). This result is actually at the root of the measurement and calibration procedure, where it is assumed that the fraction of time stopped for one or more test vehicles sampling the network passively would provide estimates of $E(T_s/T)$ and thus of f_s.

Observational studies to validate the model have been quite successful (3,4) though limited in nature by the obvious difficulty of obtaining the required data on a whole population of vehicles at the network level. Aerial photography is currently being used to measure f_s, but it is rather costly and time consuming, especially at the interpretation and reduction stages, which are still in progress (4).

The results presented will be analyzed from the perspective of the two-fluid representation to illustrate the use of simulation to further explore some aspects of this theory. Moreover, insights into the sensitivity of the model parameters T_m and n to physical and operating features of the network

could be gained under controlled conditions on a scale that generally precludes real-world experimentation.

In addition, the simulations will address a proposed extension (1) of the two-fluid model that postulates a relation between the fraction of time stopped, f_s, and the network concentration, K, of the form

$$f_s = (K/_m)^p \quad (3)$$

where K_m is the average maximum concentration level at which the network jams (in a sense, a network-level parallel to the jam concentration of an arterial) and p is a network-specific parameter reflecting the sensitivity of f_s to increasing levels of use. Combining their expression with the previous equations of the two-fluid model yields (1)

$$K = K_m[1-(V/V_m)^{(1/n+1)}]^{1/p} \quad (4)$$

The relationship between K and V ties the two-fluid concept to the earlier discussion of a network traffic flow theory.

Flow Distribution on Network Components

In addition to the network performance models discussed previously, the distribution of traffic on various components of the network will be examined. The importance of this question is primarily methodological in that it relates to the design of simulation experiments to examine the circulation of vehicles in a network under controlled conditions.

As will be seen in the next section, an attempt has been made to maintain a certain degree of homogeneity and uniformity within each experiment (at least in these early stages) to avoid unnecessary complications that would make insights more difficult to obtain. A question arises about whether there are long-term trends of traffic distributing itself in some parts of the network to the exclusion of others. For instance, for a given (closed) test network configuration and control system and a given set of parameters governing the movement of vehicles on that network (i.e., the inputs to the simulation model), particular circulatory patterns involving only those links that define the boundary of the study network might be observed. This issue will become clearer in the next section in the context of the description of the test network and of the other details of the simulation experiments.

EXPERIMENTAL DESIGN AND SETUP

In all simulations, interest was in the network-level properties of a fixed number of vehicles circulating in a closed system. Thereby a constant concentration of vehicles in the network throughout the simulation period was maintained. Although the same network configuration was used in all the experiments, the traffic characteristics (including usage patterns) as well as the control strategy were varied. Two groups of simulation runs were performed, each group corresponding to a different combination of traffic characteristics and control strategy. Within each of the two groups, two parameters were varied across runs: (a) the concentration level (ranging from 1 to 75 vehicles per lane-mile) and (b) the mean "desired" speed, specified as either 25 or 35 mph (the desired speed of a given driver is the speed at which he would travel in the absence of other vehicles and traffic controls) (9).

Further details on these elements of the simulation experiments are presented hereafter, following a brief overview of the NETSIM model, which, as was noted earlier, was used to perform all of these experiments.

The NETSIM Model

NETSIM is a fixed-step, microscopic, network traffic simulation model, which was developed primarily as a tool for the analysis of alternative urban street network control and traffic management strategies. It is generally accepted as a well-validated model and has been used extensively by traffic researchers and engineers.

Each vehicle in the system is treated separately during the simulation. Vehicle behavior is governed by a set of microscopic car-following, queue-discharge and lane-switching rules. An array of performance characteristics is stochastically assigned to each vehicle as it enters the network. All vehicles are processed once every second and their time-space trajectory recorded to a resolution of 0.1 sec. [Further details are given elsewhere (9,10).] The NETSIM model was selected primarily for its high level of detail in the representation of microscopic traffic phenomena, its sensitivity to the factors of interest in the investigation, and its rather well-developed user-related features such as types of output and summaries.

Network Configuration and Geometric Features

As noted in the previous section, a degree of regularity and uniformity in the test network used at this stage of the investigation was sought. This network consists of 25 nodes, arranged in a 5 node by 5 node square, connected by two-way, four-lane streets forming a regular, central business district-like grid. Because only directed links should be used in representing the network in NETSIM (i.e., all links are one-way), there are 80 one-way, two-lane links, as shown in Figure 1. Each link (block) is 400 ft long, with no right- or left-turn bays, and all grades are zero.

Vehicles are injected onto the network via 12 entry links placed around the perimeter, three to a side, with each entry link connecting a source node (source nodes are labeled 801 to 812 in Figure 1) to

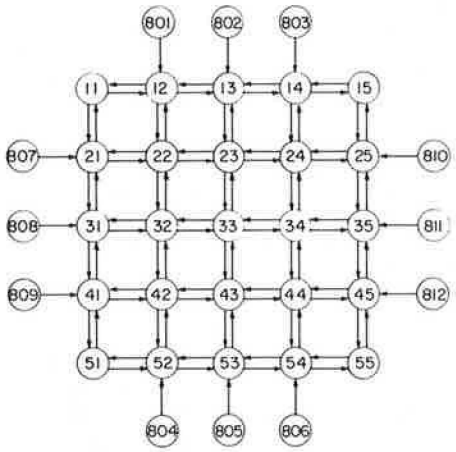


FIGURE 1 Network configuration (nodes 801-812 are source nodes and not part of the circulation network).

a noncorner boundary node. No sink nodes have been designated, because a closed system where vehicles remain in the network once they have entered it is under consideration.

Traffic Characteristics and Control Strategies

As mentioned previously, the simulations reported here belong to one of two groups. In the first group, vehicle turning movements were uniform throughout; at the interior nodes, one-third of all incoming vehicles (on any of the approaches incident to a given node) turned left, one-third turned right, and one-third continued straight through. At the boundary nodes, incoming traffic split equally between the two available options. Fixed-time traffic signals were placed at all 25 nodes, each with a 50-sec cycle length and a 50/50 split (i.e., green time equally divided between the intersecting streets), with no protected turning movements. All 25 signals were operated simultaneously, so that all the north and south approaches were green together, followed by all the east and west approaches.

In the second group, vehicle movements at the interior nodes were changed to 10 percent left, 15 percent right, and 75 percent through; they were not changed (relative to group 1) at the boundary nodes. A phase was added to the signals along the boundary to provide a protected left turn for vehicles reentering the interior of the system. The 50-sec cycle length was apportioned equally among the three approaches (phases); however, splits at the interior intersections were not changed. The four signals at the corner nodes were removed, thus allowing traffic to move unhindered through the corner nodes. Signals were operated according to a single alternative system whereby offsets between adjacent signals were all 50 percent of the cycle length so that, at any given time, every other signal along a street was green and the intervening signals were red.

There were no pedestrians, and right-turn-on-red was allowed at all intersections in both groups. Additional details of individual runs are presented next.

Individual Runs

A 5-min start-up period was used in all runs during which vehicles were generated uniformly on the 12 entry links. The vehicles were injected directly into the network interior by not allowing turns onto the boundary from the entry links. The vehicles were then allowed to circulate in the network for the desired simulation period (15 min for most of cases, although longer runs were made). Intermediate output was printed every minute, providing a "snapshot" of each link's condition at that time along with some cumulative information for each link. Network-level cumulative information as well as additional cumulative link data were printed every 3 min during the simulation period.

As mentioned earlier, two quantities were allowed to vary within each of the two groups. The mean desired speed in the network had an assumed value of either 25 or 35 mph in a given simulation. Vehicle concentration, which is a key quantity in these experiments, varied across a wide range of up to 75 vehicles per lane-mile. The results of these simulations are analyzed and critically discussed in the next section.

ANALYSIS OF SIMULATION RESULTS

The results of the simulation experiments conducted to date are analyzed with respect to those aspects of network traffic behavior identified earlier. The discussion is organized in the same three categories that were used in the conceptual background description, namely, flow distribution on network components, network traffic flow theory, and the two-fluid theory of town traffic.

Flow Distribution on Network Components

This aspect is discussed first because of its implications for the remainder of the analysis. In particular, the determination of the appropriate observation period, over which time averages of the network variables of interest were defined, was predicated on the results of the analysis of the long-run circulatory patterns developing in the test network. Two phenomena are investigated: the relative distribution of actual flow in the boundary links versus the interior links and the respective fractions of vehicles on boundary links and on interior links. In both cases, the chief concern is with the time-varying patterns of the quantities under consideration. Note that boundary links are defined as those one-way links, shown in Figure 1, with both end nodes belonging to the set of nodes {11, 12, 13, 14, 15, 21, 25, 31, 35, 41, 45, 51, 52, 53, 54, 55}. All other links, with the exception of the entry links, in the circulation network of Figure 1 are interior links.

To observe the dynamic behavior of the traffic measures pertaining to the relevant questions, a number of runs were conducted, each with 30 min of simulation beyond the initial 5 min required to load the network. The relative distribution of flow in boundary versus interior links is addressed first. Link flow can be calculated from the cumulative discharges given for each link in each intermediate output of the NETSIM model. The arithmetic averages of these minute-by-minute flows were taken separately over the boundary and interior links. These minute-by-minute boundary-average and interior-average flows are shown in Figure 2 for a representative simulation run. This run is from group 1 (groups 1 and 2 are used hereafter to denote the two combinations of traffic characteristics and control strategies described earlier).

As expected, average flows for both boundary and interior links start by increasing monotonically during and shortly after the network loading period. Subsequently, fluctuations can be observed, and there is a clear tendency for these fluctuations to occur around an apparently stable average value. The

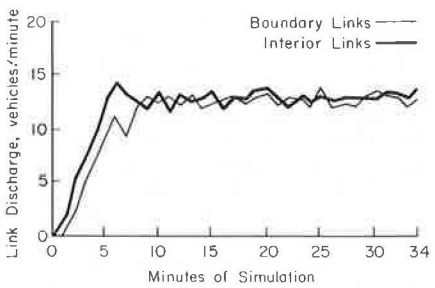


FIGURE 2 Link discharge for interior and boundary links.

system is considered stable when the link flows stabilize about an average value particular to each simulation. It can be seen in Figure 2 that this condition is achieved after approximately 8 min of total simulation time (i.e., 3 min after all the vehicles have been emitted at the source nodes). Furthermore, Figure 2 shows that the average flows for the boundary and interior links, respectively, have approximately the same value. The same conclusions could be reached for the other runs as well. Accordingly, it was decided to start the observation period, for the purpose of these analyses, at the 8th minute in all cases. In most runs analyzed, the observation period was ended at the 20th minute.

The second perspective from which to examine the general stability of the network is based on the respective fractions of vehicles on the boundary links and the interior links. In all cases, vehicles were injected directly into the interior of the network. Thereafter, given the rules governing turning movements at the intersections, vehicles eventually reached the boundary links, and, after some period of time, the fractions of vehicles on the boundary and interior links, respectively, seemed to stabilize about some constant value that was different for each of the two link classes. As is the case with average link flows, the network appears to be generally stable after about 8 min.

The time variation of the respective fractions of vehicles on the boundary and interior links for a representative 35-min run (also from group 1) are shown in Figure 3. These fractions were calculated on a minute-by-minute basis from tallies of link occupancies (separately for boundary and interior links) from the intermediate output of NETSIM. Figure 3 shows that, although the vehicle fractions undergo fluctuations, these appear to stabilize about a mean value of approximately 0.440 for boundary links and 0.560 for interior links for the simulation run under consideration. These values are actually extremely close to the arithmetic averages taken over all group 1 simulations. In this network, boundary links comprise 40 percent of the total network length, and the interior links comprise the remaining 60 percent. As the fractions here indicate, the boundary links have a higher average vehicle concentration than do the interior links. The results from the other group 1 runs are given in Table 1.

Vehicle fractions on the boundary and interior links for a simulation from group 2 are shown in Figure 4. Similar results were found for the other runs in group 2 and are given in Table 2. As in the previous cases, the network appears to stabilize after about 8 min of simulation, but the vehicle fractions on the links are different: 0.402 on the boundary links and 0.598 on the interior links. This is much closer to the ratio of total length of the boundary and interior links than is evident in the group 1 runs. The specified turning movements between groups 1 and 2 changed only for the interior intersections, and the primary cause of this change is probably the modified traffic signal phasing at the boundary nodes. The addition of a protected left-turn phase to allow traffic to more readily re-enter the network interior has apparently led to a reduction in vehicle concentrations on the boundary.

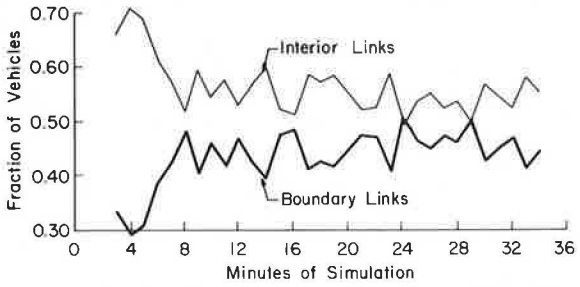


FIGURE 3 Fraction of vehicles on boundary and interior links, group 1.

TABLE 1 Average Fractions of Vehicles on Boundary and Interior Links (f_b and f_i), Group 1

	f_b	f_i
	0.440	0.560
	0.453	0.547
	0.460	0.540
	0.426	0.574
	0.424	0.576
Overall average	0.441	0.559

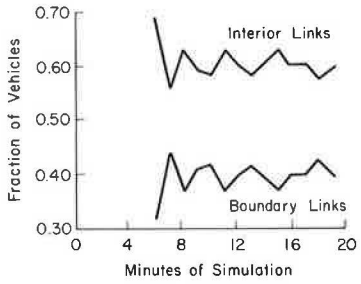


FIGURE 4 Fraction of vehicles on boundary and interior links, group 2.

TABLE 2 Average Fractions of Vehicles on Boundary and Interior Links (f_b and f_i), Group 2

	f_b	f_i
	0.395	0.605
	0.396	0.604
	0.394	0.606
	0.391	0.609
	0.400	0.600
	0.401	0.599
	0.416	0.584
	0.407	0.593
	0.398	0.602
	0.405	0.595
	0.425	0.575
	0.389	0.611
Overall average	0.402	0.598

Network Traffic Flow Theory

As discussed previously, the relationship between network concentration, K, and average speed, V, (both taken over the same observation period) is of prime interest. This was explored by varying concentration levels, K, while keeping all other features of the network, including the control system and the traffic characteristics, constant. For each

run, the average speed was found by dividing the total amount of time spent in the system by the vehicle-miles traveled. Both quantities were aggregated over the population of vehicles in the network between the 8th and 20th minutes of simulation, which defines the observation period for the reasons explained earlier in this section.

The average speeds thus obtained for a number of group 2 simulation runs (with a 35-mph mean desired speed) at different concentration levels of up to 75 vehicles per lane-mile are shown in Figure 5. Note

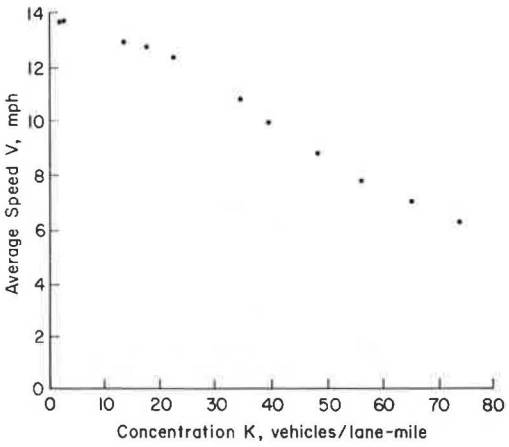


FIGURE 5 Average speed versus network concentration, group 2 runs (35 mph mean desired speed).

here that each point corresponds to an individual simulation run at a given concentration level and not to an average over a number of replicated runs [i.e., using different seeds for the random number generator underlying the stochastic features of NETSIM (9,10)] at the same concentration. Replications were not needed for the exploratory study because of the very large number of vehicles and the long observation periods over which the average quantities of interest were calculated. This was confirmed by a number of replications that were performed using different seeds for a variety of conditions, all of which yielded practically identical results for the averages under investigation (greater sensitivity to this aspect could, however, be problematic at very low concentration levels where stochastic effects are more pronounced). It can be seen in Figure 5 that the average network speed clearly decreases as a function of increasing network concentration. This is not unlike the K-V relationship that prevails on arterials. Particular care must be exercised in interpreting the extreme values (low K and high K); at the lower end of the concentration spectrum, the aforementioned stochastic effects might be prevalent, whereas at the higher end instabilities, which may not be adequately captured in the simulation, might arise. The results shown are nevertheless insightful because they reveal a clear qualitative trend between K and V. Note also that the rate of decrease of V slowly decreases at the higher concentration values shown in Figure 5.

To highlight the effect of the operational control system and associated traffic characteristics on the relationship between V and K, Figure 6 shows the average speeds obtained for a wider selection of simulation conditions under varying concentration

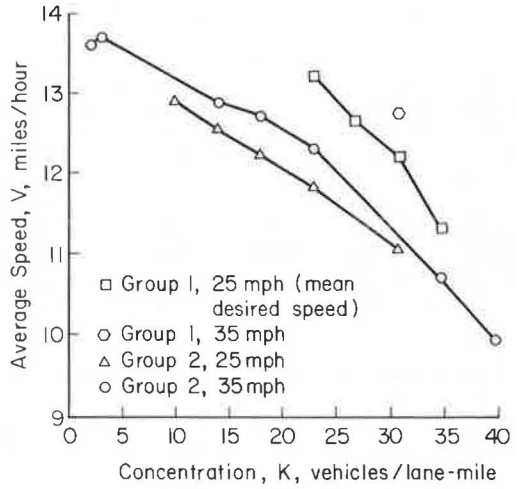


FIGURE 6 Comparison of average speed-network concentration relationship for various simulation groups.

levels but over a more restricted range. Thus both groups 1 and 2 are represented in Figure 6 at both mean desired speeds (25 and 35 mph). For each desired speed within each group, a separate K-V trend can be distinguished over the range of speeds and concentrations under consideration. Although there is only one reading for a group 1 simulation at 35-mph mean desired speed, it is included in Figure 6 because it clearly lies off the other three lines. It is apparent in Figure 6 that, within a given group, simulations conducted with a 35-mph mean desired speed displayed a higher average speed, for the same concentration level, than a corresponding run with 25-mph mean desired speed. Furthermore, the group 1 combination of traffic controls and operating characteristics resulted in better performance, in terms of average speed at a given concentration level, than did the group 2 combination at both mean desired speed levels. The main conclusion is that although average speed in a network shows a clear decreasing trend as network concentration increases, the trend itself seems to vary as a function of the operational characteristics and controls governing the use of the network configuration. However, considerable further probing is required before more generalizable conclusions can be reached.

Next the Q = KV relationship, discussed earlier, is investigated. The concentration, K, remains constant over the observation period during each simulation, and the speed is calculated as described previously. The average link flow, also defined earlier and used in determining network stability, is averaged over all links over the period between the 8th to 20th minutes of simulation. Because the link discharges represent the flow in each link and each link is identical (400 ft long, two-lane, and one-way), the average of the link discharges over the period of time in question represents the average flow seen by an average point of the network. The results for a selected number of simulations (corresponding to a variety of conditions) are given in Table 3. The value of the product of concentration and speed is quite close, in all cases, to the observed flow found directly from the link discharge information. The fact that the latter quantity was calculated by averaging over a series of network "snapshots" taken at 1-min intervals whereas the average speed was determined by parameters accumulated semicontinuously over the same period may account for some of the minor discrepancies apparent in this table.

TABLE 3 Comparison of Average Network Flow with Product of Average Speed and Concentration for Selected Runs

V[a]	K[b]	KV[c]	Q[d]	Difference[e] (%)
12.21	61.38	749	762	1.7
12.74	61.38	782	792	1.3
11.32	69.30	784	810	3.3
12.64	53.46	676	690	2.1
13.20	45.54	601	618	2.8
12.54	27.72	348	354	1.7
12.87	27.72	357	360	0.8
12.24	35.64	436	438	0.5
12.71	35.64	453	456	0.7
11.19	61.38	687	696	1.3
12.90	19.80	255	258	1.2
11.84	45.54	539	546	1.3
12.32	45.54	561	570	1.6
10.70	69.30	742	750	1.1
9.93	79.20	786	798	1.5
11 30	53.46	604	600	-0.7

[a] Average speed in mph, calculated by dividing total vehicle-miles traveled by total time spent in the system by all vehicles.
[b] Concentration in vehicles per link-mile calculated by dividing the constant number of vehicles on the network by total link-miles. Link-miles are used here instead of lane-miles because all links are identical (2 lanes).
[c] Product of V and K in vehicles per hour.
[d] Average flow in vehicles per hour found from the vehicle discharges for each link. (Compare with column 3.)
[e] Percentage difference between KV and Q.

Two-Fluid Theory of Town Traffic

The results of the simulations are analyzed here from the perspective of the two-fluid representation of traffic in a network. As described earlier, traffic is divided into moving vehicles and stopped vehicles. Thus, how various properties aggregated separately over stopped and moving vehicles vary as a function of increasing network concentration and with respect to each other is examined in the context of idealized simulations.

It should first be noted that the relation given by Equation 1 between the average speed of the moving fluid and the average fraction of vehicles running is predicated on the interactions that occur among moving vehicles. In particular, as the concentration in the network increases, and the average speed decreases (meaning that T, the average trip time per unit distance, increases), the average moving time, T_r (per unit distance), has been observed to increase in real urban street networks. This results from perturbations including short stops to pick up or drop off passengers or goods, pedestrian infringement on the right-of-way, illegal and double parking maneuvers, and other incidents that induce sudden braking or forced lane switching. These sources of turbulence occur primarily along the links rather than at the nodes where most controlled stopping takes place.

In the test network, these sources of intralink friction are virtually absent. For instance, no driveways, parking garages, unsignalized minor streets, buses, or pedestrians have been specified. Furthermore, it is not clear that adequate analytic models of the microscopic aspects of these interactions have been developed to date. This deficiency was acknowledged in earlier studies of NETSIM (11, pp.278-280) where intralink "rare events" were believed to be the primary cause of some discrepancy between simulated results and field observations for a few links in the Washington, D.C., central business district. (This led the model developers to introduce the short and long rare event and blockage features in NETSIM to increase its ability to represent realistic features of urban street networks.) In view of the this, the effect of intralink interactions, which are essential in the description offered by the two-fluid theory, was anticipated to be insignificant in idealized simulations; this is reflected in the results presented later in this section.

A key identity invoked in the derivation of the relationship between T and T_s (Equation 2) in the two-fluid theory is that the mean fraction of time stopped is equal to f_s, the average fraction of vehicles stopped. This identity, which can be proved mathematically for a constant number of vehicles circulating in a closed network (3), could be verified. Over a given observation period (following the stabilization period discussed earlier), the mean fraction of time stopped can be readily obtained from the networkwide cumulative statistics generated by the simulation model. The estimation of f_s is not as straightforward; the arithmetic average of a number of snapshots of the network taken at 1-min intervals during the observation period was used. For each snapshot, vehicles stopped in queue were tallied over all links and divided by the total number of vehicles on the network. Twelve snapshots (between the 8th and 20th min of simulation) were thus taken for each run, with the exception of the longer runs where more snapshots could be taken, and averaged to yield an estimate $\hat{f}_s$ of f_s.

Table 4 gives $\hat{f}_s$ alongside the corresponding mean fraction of time stopped, $E(T/T_s)$, for a selected number of runs from both simulation groups (and mean desired speeds) and for varying concentration levels. The percent difference between $E(T/T_s)$ and $\hat{f}_s$, also shown for each run in Table 4, clearly indicates the close correspondence between the two values in nearly all cases. The observed discrepancy stems of course from the coarse discretization on which $\hat{f}_s$ is based. It is actually rather remarkable that as few as 12 snapshots yield such close estimates of f_s,

$$(1/\tau)\int_{t_o}^{t_o+\tau} f_s(t)dt$$

where $f_s(t)$ is the time-varying fraction of vehicles stopped, and τ the duration of the observation period. Insights into the dynamic behavior of $f_s(t)$ can be obtained by looking at successive

TABLE 4 Comparison of Fraction of Total Time Stopped with Average Fraction of Vehicles Stopped[a] for Selected Runs

Mean Fraction of Time Stopped	$\hat{f}_s$[b]	Difference (%)
0.291	0.298	2.4
0 288	0.295	2.4
0.349	0.352	0.9
0.258	0.269	4.3
0.240	0.242	0.8
0 279	0.297	6.5
0.284	0.274	-3.5
0.292	0.298	2.1
0.295	0.307	4.1
0.351	0.357	1.7
0.353	0.362	2.5
0.260	0.261	0.4
0.313	0.310	-1.0
0.311	0.292	-6.1
0.394	0.405	2.7
0 437	0.443	1.4
0.341	0.350	2.6

[a] Estimated from discrete snapshots.
[b] Based on 1-min snapshots.

snapshots, obtained at 1-min intervals, for one or more runs. Figure 7 shows such snapshots for two selected runs. The bottom line corresponds to a group 1 run (35-mph mean desired speed) with a network concentration of 30.69, and the top line corresponds to a group 2 run (35-mph mean desired speed) and concentration of 39.60. For both runs, the variation is rather substantial, which is typical of all the simulations performed. Variation of average fraction of vehicles stopped as a fraction of network concentration is discussed later in this section.

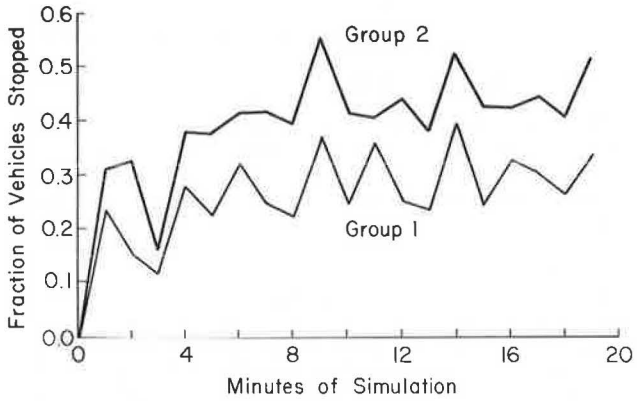


FIGURE 7 Instantaneous fraction of vehicles stopped evaluated at successive network snapshots for two selected runs.

The variation of T, T_s, and T_r as a function of network concentration, K, while network control features and traffic characteristics are kept fixed, is examined next. Group 2 simulation runs with mean desired speed of 35 mph are considered for this purpose. Table 5 gives a summary of the values of T, T_r, and T_s that correspond to concentrations of up to 75 vehicles per lane-mile; also given are the ratios T_r/T and T_s/T. As expected (and seen earlier in the discussion of the K-V relationship), the average trip time per unit distance, T, increases with increasing concentration. It can also be seen from the values of T_s that this increase in T is due overwhelmingly to increasing stopped time; the average running time (per unit distance), T_r, is seen to remain at approximately the same level throughout. This behavior, which does not correspond to that observed in real urban networks, confirms the concerns about the inadequate representation of perturbations and interactions that are an integral aspect of urban street traffic. The almost constant T_r obtained in these runs corresponds to a value of the parameter n (of the two-fluid model), which is not significantly different from zero, indicating the relative lack of sensitivity of T_r to changes in the fraction of vehicles running in the range of concentrations considered in the idealized system. It is, however, conceivable that greater realism could be achieved by specification of random "rare events" (allowed by NETSIM) or other sources of interference; an investigation of these possibilities was beyond the scope of the present exploratory study. Further thoughts on this matter are offered in the next section.

Table 5 also allows examination of the variation of f_s (which is identical to T_s/T) as a function of network concentration. An analytical relationship between f_s and K, proposed by Herman and Prigogine (1) as an extension of the two-fluid model, has been presented (Equation 3). The general trend predicted by Equation 3 is present in the results of the simulations whereby f_s increases nonlinearly with increasing concentration; this trend is shown in Figure 8. It can also be seen in Figure 8 that the fraction of vehicles stopped tends to level off and not fall below a minimum threshold at lower concentrations. A plausible explanation is that, over any meaningful observation period, it is inevitable that a number of stoppages will occur due to the nature of the traffic control system governing the network used in the simulations. In other words, it is virtually impossible to circulate in a network of this type for 12 min within some fraction of the vehicles stopping at some of the frequently encountered signals.

A least-squares estimation of the parameters p and K_m in Equation 3 yielded the values 0.589 and 156.0 vehicles per lane-mile, respectively. Note that these estimates are for K in the range of 15 to 75 vehicles per lane-mile. The corresponding R^2 was equal to 0.988. It is interesting to note here that the magnitude of K_m is consistent with its interpretation as the concentration at which the network "jams." A partial simulation run with an intended K of 150 vehicles per lane-mile showed extreme congestion levels with prevailing spillbacks at most intersections, which effectively delayed the complete loading of the vehicles onto the network by as much as 20 min after the last vehicle was emitted from the source nodes. (Computational cost considerations prevented the pursuit of the matter further at this stage with such unrealistically high concentration levels.)

TABLE 5 Average Trip Time, Stop Time, and Running Time Characteristics for Group 2 Runs[a] Under Varying Network Concentration Levels

Concentration (veh/lane-mile)	Trip Time[b] (T)	Running Time[b] (T_r)	Stop Time[b] (T_s)	T_r/T	T_s/T
1.98	4.41	3.34	1.07	0.758	0.242
2.97	4.38	3.30	1.08	0.754	0.246
13.86	4.66	3.34	1.33	0.716	0.284
17.82	4.72	3.33	1.39	0.705	0.295
22.77	4.87	3.35	1.52	0.689	0.311
34.65	5.61	3.40	2.21	0.606	0.394
39.60	6.04	3.40	2.64	0.563	0.437
48.51	6.86	3.41	3.44	0.498	0.502
56.43	7.80	3.41	4.39	0.437	0.563
65.34	8.62	3.39	5.24	0.393	0.607
74.25	9.68	3.37	6.31	0.349	0.651

[a] 35 mph mean desired speed.
[b] T, T_r, and T_s are averages (over time and over vehicles) expessed in minutes per mile.

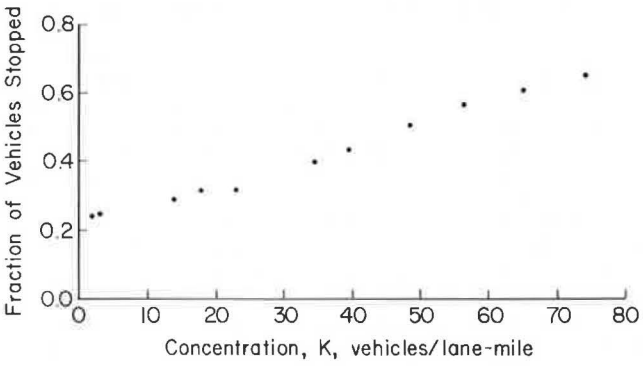


FIGURE 8 Average fraction of vehicles stopped versus network concentration, group 2 runs (35 mph mean desired speed).

SUMMARY AND ASSESSMENT

An exploratory study designed to assess the usefulness of a simulation-based approach to support the investigation of network-level macroscopic traffic flow relationships has been discussed. The NETSIM microscopic network traffic simulation model was used in conjunction with an idealized system consisting of an isolated CBD-like rectangular grid network operating under various combinations of traffic control schemes and characteristics. The results were analyzed with respect to (a) network traffic flow theory, (b) the two-fluid theory of town traffic, and (c) flow distribution on network components.

The analysis of flow distribution on network components was concerned primarily with the dynamic behavior of the relative concentrations and flows on boundary links and interior links, respectively. Useful insights were gained regarding the general stabilization of the respective fractions of vehicles and flows on boundary and interior links after an initial period. This provided the basis for defining an appropriate observation period over which average traffic descriptors, used in other parts of this study, were calculated.

Two principal network traffic flow theoretic relationships were addressed. Average network speed was found to decrease as a function of increasing network concentration, as anticipated, with an overall trend that is not unlike that observed for arterials. This trend was also explored under varying signal timing schemes and traffic characteristics. The second relationship addressed was $Q = KV$, which is fundamental for arterials but unverified at the network level. Simulations seemed to indicate that, for properly defined averages of the three quantities, $Q = KV$ could be expected to hold.

The analysis of the results from the perspective of the two-fluid theory of town traffic verified the identity between the average fraction of vehicles stopped and the mean fraction of time stopped for a fixed number of vehicles circulating in a closed network over the same observation period. The dynamic behavior of the instantaneous fraction of vehicles stopped was also examined. In addition, the variation of the average fraction of vehicles stopped as a function of network concentration was studied and seemed to indicate that a proposed extension of the two-fluid model (1) holds rather well over a middle range of concentrations. On the other hand, the idealized simulation conditions apparently did not generate the interactions among moving vehicles that constitute an essential feature of real urban traffic systems operations and that, as such, are critical to the description encapsulated by the two-fluid theory.

It is essential to re-emphasize here the exploratory nature of the research described. It is evident that a number of probably severe limitations are present in the preliminary results. The list of these limitations includes the highly idealized nature of the system and its operating conditions, the well-known fact that a simulation model is only an abstraction of the real world, and many others ranging from specifics of the execution to the restricted scope of the results and associated conclusions. It is nevertheless believed that such exploratory studies yield useful insights into the behavior of traffic systems under a variety of conditions that cannot be easily achieved in the real world. Such studies can provide a useful complement to observational studies of network-level traffic phenomena, which remain the cornerstone of a scientific approach to the development of a network traffic flow theory. By raising questions and probing some possible answers, simulation-based studies can suggest to the analyst important directions for both theoretical and observationally based advances.

A number of avenues exist for further effort in this general area of investigation. Some of the more immediate ones include the assessment of how NETSIM could be used to adequately describe the interactions among moving vehicles or intralink phenomena that are fundamental in real urban street systems. It was mentioned earlier that the introduction of short- and long-term rare events and blockages, in addition to heavy vehicles, pedestrian interference, driveways, and parking maneuvers, is likely to improve the realism of this representation. However, more fundamental modifications in the car-following and lane-switching procedures embedded in NETSIM may be required. A related possibility exists for using some of the recent empirical results obtained in conjunction with the two-fluid theory whereby one or two test vehicles circulating in a network can yield sufficient information for characterizing the quality of traffic service in an urban network (3,4). This relatively easily acquired information could provide the basis for calibrating a simulation model like NETSIM, particularly with respect to difficult-to-model intralink interactions.

ACKNOWLEDGMENTS

Partial support for the research on which this paper is based was provided by a grant from the Bureau of Engineering Research at the University of Texas (to H. Mahmassani), as well as by U.S. Department of Transportation grant DOT DTRS 5781-P-81539. The assistance of Patricia A. Best, undergraduate research assistant at the University of Texas at Austin, in the processing of the tons of computer output and in the graphic presentation is particularly appreciated. The authors remain, of course, solely responsible for the contents of this paper.

REFERENCES

1. R. Herman and I. Prigogine. A Two-Fluid Approach to Town Traffic. Science, Vol. 204, 1979, pp. 148-151.
2. M.-F. Chang and R. Herman. Trip Time Versus Stop Time and Fuel Consumption Characteristics in Cities. Transportation Science, Vol. 15, 1981, pp. 183-209.
3. S. Ardekani and R. Herman. Quality of Traffic Service. Research Report 304-1. Center for Transportation Research, University of Texas, Austin, 1982.
4. R. Herman and S. Ardekani. Characterizing Traffic Conditions in Urban Areas. Smeed Memorial Lecture. University College, London, England, May 1983.
5. D.L. Gerlough and M.J. Huber. Traffic Flow Theory: A Monograph. TRB Special Report 165. TRB, National Research Council, Washington, D.C., 1975.
6. D. Gazis, ed. Traffic Science. Wiley, New York, 1974.
7. Highway Capacity Manual 1965. HRB Special Report 87, HRB, National Research Council, Washington D.C., 1965, 397 pp.
8. I. Prigogine and R. Herman. Kinetic Theory of Vehicular Traffic. American Elsevier, New York, 1971.
9. Traffic Network Analysis with NETSIM--A User

Guide. Implementation Package FHWA-IP-80-3. FHWA, U.S. Department of Transportation, 1980.
10. Peak, Marwick, Mitchell, and Co. Network Flow Simulation for Urban Traffic Control System: Phase II, Volumes 1-5. FHWA, U.S. Department of Transportation, 1973.
11. Peat, Marwick, Mitchell, and Co. and General Applied Science Laboratories, Inc. Network Flow Simulation for Urban Traffic Control System. Final Report. FHWA, U.S. Department of Transportation, 1971.

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

Another Look at Storage Requirements for Bank Drive-In Facilities

JOHN L. BALLARD, JOHN G. GOBLE, RICHARD J. HADEN, and PATRICK T. McCOY

ABSTRACT

Observations of the operation and performance of bank drive-in facilities in Lincoln, Nebraska, indicated that current storage requirements for these facilities were excessive. The objective of this research was to determine why these theoretically and empirically developed requirements were excessive and to develop more reasonable storage requirements. Arrival and service-time data collected at bank drive-in facilities were analyzed. It was determined that the arrivals were Poisson. But, contrary to the usually employed queuing theory assumptions of negative exponential serving times, which had been used to develop previous storage requirements, the service-time distributions were found to be gamma distributions with shape parameters between 2.75 and 5.00. Because of the intractability of using queuing theory with gamma service-time distributions, simulation models of single-queue and multiple-queue, multiple-channel queuing systems typical of bank drive-in facilities were developed and validated. The models were then used to determine more appropriate storage requirements.

Before April 1981 the storage requirements for bank drive-in facilities imposed by the city of Lincoln, Nebraska, were those given in Table 1. These requirements were developed from a review of the literature, primarily papers written by Woods and Messer (1) and Scifres (2), and the results of field studies conducted by the city in 1974, which in general confirmed the findings presented in the literature. These requirements were generally accepted as reasonable for several years. Beginning in 1980 they were challenged for requiring too much storage, and the need for updated studies became apparent.

TABLE 1 City of Lincoln, Nebraska, Drive-in Bank Storage Requirements Before April 1981

No. of Windows	Minimum Storage Required[a,b] (vehicles)
1	7
2	14
3	21
4	28
5	30
6	30

[a] In addition to the service position.
[b] 22 ft per vehicle required in storage lanes.

The need for updated studies resulted primarily from major changes in the banking industry in Lincoln. Among these changes were

1. A sharp increase in the number of drive-in facilities available, spreading the business around and reducing peaking at any one facility.
2. The introduction of 24-hour electronic teller machines at sales points such as grocery stores provided a new convenience for customers. This raised the customers' expectations and reduced their tolerance of delay.
3. There was an increased trend toward staggered payrolls among major employers, which reduced peaking characteristics for deposits and withdrawals.

Consequently, in early 1981, the city conducted studies of traffic operations at drive-in banking facilities to determine the reasonableness of its storage requirements. A total of 1,142 transactions were observed during which the average traffic intensity was 0.89. However, the maximum queue length observed in any one storage lane was only five vehicles, and it existed for only 18 sec. Otherwise, the maximum queue length was four vehicles. The average transaction time observed was 2.12 min, which is considerably lower than the average service times often assumed in design guidelines (1,2). In addi-

tion, the average time a vehicle spent in the system (i.e., waiting time plus service time) was 3.55 min, which is also much less than the waiting time normally used in the development of design guidelines (1,2).

On the basis of the results of these studies, it was concluded that the current standards were unreasonable in that they did require too much storage. Therefore, revision of the storage requirements for bank drive-in facilities was recommended, and revised requirements were adopted by the Lincoln City Council in April 1981. The revised standards require storage for four vehicles per drive-in window and storage for 20 vehicles maximum for an entire facility.

OBJECTIVE

Although the storage requirements were revised out of practical necessity, questions remained about why these empirical standards were so different from the city's previous standards and the guidelines recommended in the literature (1,2). Therefore, the objective of the research reported in this paper was to determine the reasons for these discrepancies and develop storage requirement guidelines for bank drive-in facilities.

PROCEDURE

A bank drive-in facility is a queuing system. Depending on its configuration, it may be classified as a single-queue, multiple-channel system or a multiple-queue, multiple-channel system. In a single-queue, multiple-channel system all vehicles wait for service in one line. If the distribution of arrivals is Poisson and the distribution of service times is negative exponential, the operation and performance of this type of system can be evaluated using queuing theory as was done by Woods and Messer (1). Otherwise, evaluation using queuing theory may be intractable. In a multiple-queue, multiple-channel system vehicles wait in queues in front of each drive-in window. When the queue storage of all windows is full, vehicles wait in a single queue. Regardless of the nature of the arrival and service-time distributions, the evaluation of the operation and performance of this type of system using queuing theory is intractable.

Because of the limitations of queuing theory, simulation was used in this research. Simulation models were developed and validated for both types of bank drive-in window queuing systems. The data collected by the city in early 1981 were analyzed to determine the nature of the observed arrival and service-time distributions. The results of this analysis and the simulation models were then used to determine storage requirements of bank drive-in facilities.

ARRIVAL AND SERVICE-TIME DISTRIBUTIONS

The arrival and service-time data collected by the city during its studies of bank drive-in facilities in early 1981 were analyzed to determine the nature of their distributions. The chi-square goodness-of-fit test was applied at the 0.01 level of significance to determine the arrival and service-time distributions in each of the 12 peak hours studied. As as result of these tests, it was found that all 12 of the arrival distributions were Poisson, a finding consistent with the common assumptions of queuing theory and the observations of previous studies of drive-in banking facilities (1,3).

The results of the analysis of the service-time distributions were not consistent with the usual assumption of queuing theory that service times are distributed negative exponentially. If fact, none of the 12 service-time distributions was found to be negative exponential. Instead, all but one of them were found to fit a gamma distribution with a shape parameter between 2.75 and 5.00. This finding is consistent with that of Thurgood (4), who found that the service-time distributions of drive-in banking facilities in the Chicago area were definitely not negative exponential.

SIMULATION MODELS

Two simulation models of bank drive-in facilities were developed. One was a model of the single-queue, multiple-channel system, and the other was a model of the multiple-queue, multiple-channel system. Both of these models were written in the GPSS/H simulation language (5), a discrete simulation language commonly used to simulate queuing systems.

Using the validation procedure outlined by Law and Kelton (6), the outputs of the simulation models were first compared with the theoretical results of queuing theory. Fifty replications of 100 hr each were simulated for worst-case conditions (7). In all cases, the model output deviated from the theoretical results by less than 10 percent. Also, no significant differences were found at the 0.05 level of significance between the model and the theoretical mean values of number in the system, number in the queue, and time in the system.

STORAGE REQUIREMENTS

Previous Studies

A search of the literature published in the past 15 years, during the tremendous increase in the number of drive-in banking facilities, revealed little information on storage requirements for these facilities. One exception was a paper by Woods and Messer (1) published in 1970, which provided guidelines for the design of drive-in banking facilities. Based on field observations of 227 service times, they concluded that the "usually assumed" average service time of 1.5 min or 40 vehicles per hour per window was "reasonably valid" and that the service times tended to be "negative exponential in nature." Scifres (2) also developed guidelines for the planning and design of drive-in financial institutions.

Model Results

In general, the theoretically based guidelines developed by Woods and Messer require less storage than do the empirically based guidelines developed by Scifres. The traffic intensity used by Woods and Messer was 0.875 and the average traffic intensity observed by the city was 0.89. The expected average time in the system used by Woods and Messer was 5 min and the average time in the system observed by the city was 3.55 min. Because of this consistency, the simulation models were used to develop storage requirements for conditions similar to those used by Woods and Messer.

Therefore, a traffic intensity of 0.875 was used in conducting the simulation runs, and storage requirements were determined for 5 and 15 percent probabilities of the queue exceeding the storage provided. Also, Poisson arrivals were used in all cases. However, unlike Woods and Messer, gamma service-time distributions were used instead of the

negative exponential service-time distribution. Storage requirements were determined using gamma service-time distribution with shape parameters of 2.75 and 5.00, which was the range of gamma service-time distributions observed by the city of Lincoln.

The storage requirements determined using the simulation models and the gamma service-time distributions are given in Tables 2 and 3. The storage requirements for single-queue, multiple-channel facilities are given in Tables 2, and those for multiple-queue, multiple-channel facilities are given in Table 3. Also given in these tables are the storage requirements determined by Woods and Messer using the negative exponential service-time distribution.

TABLE 2 Storage Requirements for Single-Queue, Multiple-Channel Bank Drive-in Facilities[a,b]

No. of Drive-in Windows	Minimum Storage (P = 15%)[c] (vehicles/window)			Desirable Design Storage (P = 5%)[c] (vehicles/window)		
	Negative Exponential[d]	Gamma		Negative Exponential[d]	Gamma	
		2.75[e]	5.00[f]		2.75[e]	5.00[f]
2	12	6	5	20	12	11
3	12	6	5	20	12	11
4	11	6	5	20	12	10
5	11	6	5	19	12	10
7	10	5	4	19	12	10
9	10	5	4	18	11	9

[a] In addition to the service positions.
[b] Traffic intensity = 0.875 and Poisson arrivals.
[c] P = percent of time that queue length would be greater than the storage provided.
[d] Negative exponential distribution of service times.
[e] Gamma distribution of service times with shape parameter = 2.75.
[f] Gamma distribution of service times with shape parameter = 5.00.

TABLE 3 Storage Requirements for Multiple-Queue, Multiple-Channel Bank Drive-in Facilities[a,b]

No. of Drive-in Windows	Minimum Storage (P = 15%)[c] (vehicles/window)			Desirable Design Storage (P = 5%)[c] (vehicles/window)		
	Negative Exponential[d]	Gamma		Negative Exponential[d]	Gamma	
		2.75[e]	5.00[f]		2.75[e]	5.00[f]
2	7	4	3	10	6	6
3	4	3	2	7	4	4
4	3	2	1	5	4	4
5	3	1		4	3	2
7	2			3	2	1
9	2			2		

[a] In addition to the service positions.
[b] Traffic intensity = 0.875 and Poisson arrivals.
[c] P = percent of time that queue length would be greater than the storage provided.
[d] Negative exponential distribution of service times.
[e] Gamma distribution of service times with shape parameter = 2.75.
[f] Gamma distribution of service times with shape parameter = 5.00.

Examination of these tables reveals that the storage requirements determined using the gamma service-time distributions are in all cases less, and in many cases substantially less, than those determined using the negative exponential service-time distribution. Of course, this is due to the fact that the gamma distributions have variances that are lower than those of the negative exponential distribution, and, as the shape parameter is increased, the variance is reduced. Consequently, the storage requirements determined using the gamma service-time distribution with a shape parameter of 5.00 are the lowest in every case.

Comparison of the data in Tables 2 and 3 indicates that in general less total storage is required in single-queue, multiple-channel systems than in multiple-queue, multiple-channel systems. This is to be expected because of the pooling of service provided in the single queue, multiple-channel system.

CONCLUSION

According to the findings of this research, the usual queuing theory assumptions of Poisson arrivals and negative exponential service times are not valid for the determination of storage requirements for bank drive-in facilities. Although the distribution of arrivals at bank drive-in facilities was found to be Poisson, the distribution of service times at these facilities was not found to be negative exponential. Instead, the distribution of service times was found to be a gamma distribution with a shape parameter ranging from 2.75 to 5.00. Consequently, storage requirements developed using queuing theory with the usual assumptions, as was done in previous studies (1,4), are excessive. Although such requirements may be considered conservative by a traffic engineer, they are considered extravagant by the financial institutions.

Therefore, it was concluded that the storage requirements given in Tables 2 and 3, which were determined in this research using simulation models and gamma service-time distributions, should be used as guidelines in the planning and design of bank drive-in facilities. It should be noted that these requirements were developed using a traffic intensity of 0.875. This was the same traffic intensity used by Woods and Messer (1), and it was found to be consistent with observations of peak-hour operations at bank drive-in facilities in Lincoln, Nebraska.

REFERENCES

1. D.L. Woods and C.J. Messer. Design Criteria for Drive-In Banking Facilities. Traffic Engineering, Dec. 1970, pp. 30-37.
2. P.N. Scifres. Traffic Planning for Drive-In Financial Institutions. Traffic Engineering, Sept. 1975, pp. 21-24.
3. G. Kelly. Banks Say They Need to Branch. Sunday Lincoln Journal Star, Lincoln, Neb., April 5, 1981.
4. G.S. Thurgood. The Application of Stochastic Queuing Theory in the Development and Evaluation of Suggested Traffic Design Guidelines for Drive-In Service Facilities. M.S. thesis. Texas A&M University, College Station, 1975.
5. J.O. Henriksen. GPSS/H User's Manual. Wolverine Software, Falls Church, Va., 1978.
6. A.M. Law and W.D. Kelton. Simulation Modeling and Analysis. McGraw-Hill, New York, 1982.
7. J.G. Goble. Estimating the Performance of Drive-In Service Systems with Varying Operating Characteristics. M.S. thesis. University of Nebraska-Lincoln, 1983.

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

Simulation of Traffic Performance, Vehicle Emissions, and Fuel Consumption at Intersections: the TEXAS-II Model

CLYDE E. LEE and FONG-PING LEE

ABSTRACT

TEXAS-II is a computer simulation model that teams the TEXAS Model for Intersection Traffic, the Environmental Protection Agency Modal Analysis Model for light-duty vehicles, and a new heavy-duty vehicle model to make quantitative estimates of several traffic performance parameters, emissions, and fuel consumption that will result from mixed traffic using any practical intersection configuration operating under signalized, signed, or unsigned traffic control. The quantitative information produced by the model is useful for identifying existing intersection problems and for evaluating feasible alternative solutions. TEXAS-II has been run more than 300 times in an extensive series of experiments designed to study the relative effects of specified intersection environments on traffic delay, queue lengths, emissions, and fuel consumption.

Vehicle emissions and fuel consumption at or near street intersections are usually higher than on other street segments because the intersection frequently causes vehicles to slow, stand, and accelerate. Pollutants emitted from vehicles in the vicinity of intersections can sometimes accumulate, and the resulting concentrations in the air can be potentially dangerous to human health. Excessive fuel consumption at intersections is also a major concern in traffic engineering and in transportation economics because it relates to the use and conservation of energy.

A practical means of estimating both vehicle pollutants and fuel consumption near intersections in quantitative terms is needed. Existing and potential locations with excessive emissions and fuel consumption need to be identified so that appropriate remedial and preventive measures can be programmed.

Among the various emitted pollutants, carbon monoxide (CO), hydrocarbons (HC), and nitrogen oxides (NOx) are of most concern. Carbon monoxide is so toxic that it can cause death within minutes in high concentrations. Hydrocarbons, in the gaseous form, combine with nitrogen oxides in the presence of sunlight to form photochemical smog. Smog frequently causes watering and burning of the eyes and adversely affects the human respiratory system, especially of those persons in marginal physical condition. Nitrogen oxides tend to combine with the hemoglobin in the blood and react with moisture in the lungs to form dilute nitric acid. Even when the amounts of NOx are minute, the effect on the human body is cumulative (1). HC and NOx, which sometimes react in the atmosphere, can form oxidants and thus are difficult, if not impossible, to monitor accurately with existing equipment and sampling methods. Only CO concentrations can be measured practically by field techniques at this time.

To predict the vehicle-generated pollutant concentrations that might exist at any selected location on or adjacent to a roadway, the source of emissions must first be estimated. Vehicle source emissions can be characterized by a time-dependent instantaneous rate with respect to location along the roadway. The type and the amount of pollutants emitted from any vehicle traveling along the roadway actually depend on the vehicle type, its condition, and the performance of traffic at the location. Vehicle emissions are displaced almost immediately from the instantaneous point of deposit by movement of the air around the vehicles traveling on the roadway, wind, and thermal convection. For certain modeling purposes, however, the emissions deposited along a highway lane or on a set of intersection approach lanes in a short time period, before being dispersed into the air or modified by reaction with other constituents in the air, might be viewed collectively as a line source of pollutants in relation to the overall intersection space (2). The pollutants in this line source may be further dispersed into the air, quickly or slowly, depending on the localized meteorological conditions.

In developing the TEXAS-II model, estimating the source of vehicle emissions in quantitative terms was the major objective. Mixing and dispersion of pollutants is the subject of related ongoing research that is using field measurements of pollutants as the basis for developing improved models of pollutant concentrations in or near road intersections. Fuel consumption, which likewise varies with respect to time and location along the roadway, was also addressed because the estimation techniques are somewhat similar and the subject is one of continuing interest and concern.

STRUCTURE OF THE TEXAS-II MODEL

To quantify the effects of intersection geometry, traffic control, and traffic flow on air pollution and fuel consumption, the TEXAS-II simulation model was developed to compute estimates of vehicle emissions and fuel consumption on a microscopic basis. TEXAS-II is a modified and extended version of the TEXAS Model for Intersection Traffic (3-7). The basic information flow process for the TEXAS-II model is shown in Figure 1.

The TEXAS Model for Intersection Traffic includes three data processors: GEOPRO (geometry), DVPRO (driver-vehicle), and SIMPRO (simulation) for describing, respectively, the geometric configurations, the stochastically arriving traffic, and the behavior of traffic in response to the applicable traffic controls. SIMPRO integrates all the defined elements and computes deterministically the response of each driver-vehicle unit.

The TEXAS Model for Intersection Traffic, which is suitable for a single, multileg, multilane, mixed-traffic intersection operating under any conventional type of traffic control, thus simulates

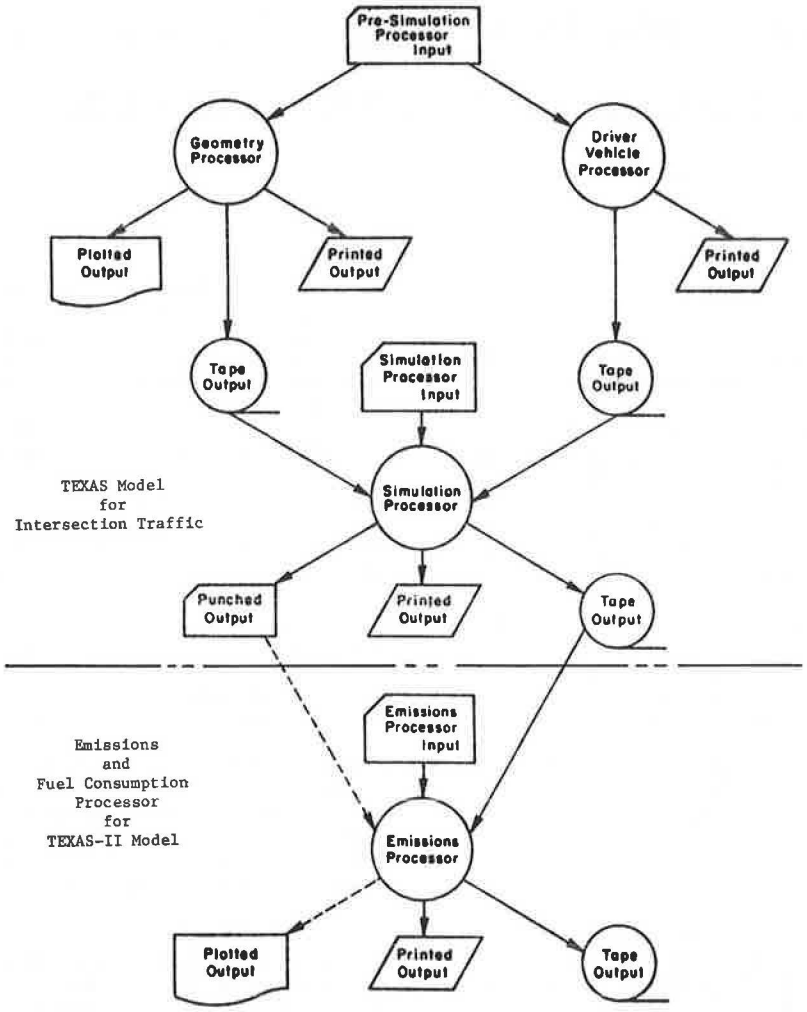


FIGURE 1 Information flow process for the TEXAS-II model.

the instantaneous behavior of each individually characterized driver-vehicle unit as it approaches, passes through, and departs from an intersection. The premise is that each simulated driver will attempt to maintain safety and comfort while sustaining desired speed and obeying traffic laws. At any time, a unit may maintain or change speed and retain or change lanes depending on the relative positions and movements of neighboring units and the effects of applicable traffic control devices. The instantaneous traffic behavior of each unit including speed, location, and time are written on a tape by the TEXAS model for subsequent use in the emission processor (EMPRO). Statistics about delays and queue lengths are also gathered by the TEXAS model to evaluate the performance of traffic at the intersection.

The new emissions processor, EMPRO, incorporates models to predict the instantaneous vehicle emissions of carbon monoxide (CO), hydrocarbons (HC), nitrogen oxides (NOx), and fuel flow (FF) for both light-duty vehicles and heavy-duty vehicles. The existing Environmental Protection Agency (EPA) models for 1975 light-duty vehicles operating at low altitude (8) were used, but the models that estimate the behavior of heavy-duty vehicles were developed recently (9-11) with experimental data supplied by Southwestern Research Institute (12-19).

EMPRO uses information from SIMPRO about the instantaneous speed and acceleration of each vehicle to compute instantaneous vehicle emissions and fuel consumption at points along the vehicle path. For evaluation purposes, each lane on each approach is partitioned into a series of buckets, and the emissions and fuel flow are accumulated on a bucket basis to show the spatial variation of emissions and fuel consumption with respect to time. The intersection proper is treated as one bucket, which collects the emissions and fuel consumption values generated by vehicles crossing it from all approaches. The length of buckets on each inbound or outbound lane can be specified by the user as input data to EMPRO. The bucket is set as a 100-ft section of a lane; therefore, each inbound or outbound lane is partitioned into eight buckets because all approaches are 800 ft long. The EPA emission and fuel consumption models for light-duty vehicles require input directly in terms of instantaneous vehicle performance (speed and acceleration), but the emission and fuel consumption models for heavy-duty vehicles described herein use functions of engine performance (engine torque and engine speed). EMPRO incorporates a subprogram that relates vehicle performance to engine performance for heavy-duty vehicles to estimate emissions and fuel consumption.

Emissions and Fuel Consumption Models for Light-Duty Vehicles

The emission models for CO, HC, NOx and CO_2 developed by EPA for light-duty vehicles and referred to as the Modal Analysis Model (8), are presented in quadratic form of speeds for a steady state of vehicle motion, and in quadratic form of speeds and accelerations for transient states. The fuel consumption model is expressed as a linear function of the amounts of HC, CO, and CO_2 emitted. The emission models are formulated as follows:

Steady state

$$L(V) = S_1 + S_2V + S_3V^2$$

where

L = instantaneous emission rate (grams/second),
V = speed (mph), and
S = coefficients (given in Table 1).

Transient state

$$L(V, A) = B_1 + B_2V + B_3A + B_4AV + B_5V^2 + B_6A^2 + B_7V^2A + B_8A^2V + B_9A^2V^2$$

where

A = acceleration or deceleration (mph/second) and
B = coefficients (given in Table 1).

In the Modal Analysis Model, vehicles are classified in 18 groups by model year from 1957 to 1975 and by low or high operating altitude. The 1975 low-altitude group, which provides the most current information that is generally available and matches the terrain situations of many American cities, was selected for use in the TEXAS-II model. The models and coefficients for estimating the emissions of CO, HC, NOx, and CO_2 for the 1975 low-altitude group of automobiles are given in Table 1. These coefficients can be modified by the user if necessary. An evaluation of the models indicates that in steady-state driving the emissions of CO and HC decrease with speed while the emissions of NOx and CO_2 increase with speed. The fuel consumption rate in steady-state driving stays almost constant for speeds up to 10 mph and then increases with speed. In transient-state driving, acceleration increases emissions and fuel consumption. The effect of acceleration is higher when speed is higher. Use of the coefficients for the 1975 low-altitude group of vehicles can produce negative values of emissions and fuel flow. In the TEXAS-II model, all such negative values are automatically set to zero.

Emissions and Fuel Consumption Models for Heavy-Duty Vehicles

A series of models (see Table 2) for estimating instantaneous values of emissions and fuel consumption for heavy-duty vehicles powered by gasoline or diesel engines was developed for incorporation into the data postprocessor called EMPRO in the TEXAS-II model. This development involved combining rational approximations of vehicle dimensions and operating characteristics with empirical data on engine performance to produce the models.

Experiments conducted at Southwest Research Institute (12-19) indicated that emission rates from heavy-duty vehicles are functions of speed, acceleration, engine make, type of pollution control devices, maintenance condition, and engine operating temperature. These experiments involved testing a representative sample of heavy-duty vehicles (HDVs) and measuring the produced emissions for comparison with proposed regulatory standards. A 13-mode steady-state test schedule was used for diesel en-

TABLE 1 Instantaneous Emission and Fuel Consumption Models for Passenger Cars (8)

INSTANTANEOUS EMISSION MODELS

Steady State Model: $L(V) = S_1 + S_2V + S_3V^2$

L = Instantaneous Emissions Rate (gram/second)

V = Speed (mph)

Transient State Model: $L(V,A) = B_1 + B_2V + B_3A + B_4VA + B_5V^2 + B_6A^2 + B_7V^2A + B_8VA^2 + B_9V^2A^2$

A = Acceleration or Deceleration (M/H^2)

COEFFICIENTS FOR EMISSION MODELS

State		CO	HC	NO_x	CO_2
Steady	S1	1.16557780E - 01	5.38159910E - 03	1.46895690E + 00	2.65079990E - 03
	S2	-4.62989880E - 03	-1.45500000E - 04	7.06690180E - 03	-3.53700020E - 04
	S3	6.98999940E - 05	1.99999980E - 06	1.61370010E - 03	2.34000040E - 05
Transient	B1	2.15785210E - 01	8.06840140E - 03	2.28404900E + 00	1.08160000E - 02
	B2	-1.25777980E - 02	-4.00200020E - 04	-2.62799000E - 02	-1.22500000E - 03
	B3	5.14772980E - 02	9.00400100E - 04	6.55900840E - 02	-7.35400010E - 04
	B4	-2.34259990E - 03	6.50000000E - 05	5.39221990E - 02	5.39399920E - 04
	B5	1.67800000E - 04	6.60000020E - 06	2.12890000E - 03	4.44000030E - 05
	B6	-1.57559990E - 03	-7.35699900E - 04	-1.65571990E - 01	-3.29720000E - 03
	B7	2.82299940E - 04	8.98000000E - 05	3.02321020E - 02	5.26600050E - 04
	B8	1.25299990E - 04	-3.00000010E - 07	-9.01000020E - 05	3.11999970E - 06
	B9	4.85000060E - 05	-6.00000020E - 07	-4.12700000E - 04	-8.40000030E - 06

INSTANTANEOUS FUEL CONSUMPTION MODEL

$$FF = 0.866 * HC + 0.429 * CO + 0.273 * CO_2$$

TABLE 2 Instantaneous Emissions and Fuel Consumption Models for Gasoline and Diesel Trucks

EMISSION AND FUEL CONSUMPTION MODELS FOR GASOLINE TRUCKS[1]	
HC	= 6.526E - 03 + 1.088E - 08 * ABS(TRQ) * RPM + 4.153E - 11 * TRQ * TRQ * TRQ * TRQ - 5.46E - 09 * ABS(TRQ) * TRQ * TRQ
CO	= 10.0**(-2.636 + 3.190E - 05 * TRQ * TRQ + 4.257E - 02 * SQRT(RPM) - 2.205E - 06 * ABS(TRQ) * RPM + 1.659E - 10 * TRQ * TRQ * TRQ * TRQ)
NO	= 10.0**(-1.702 + 2.505E - 02 * SQRT(ABS(TRQ)) - 8.991E + 02/RPM - 3.815E - 10 * TRQ * TRQ * TRQ * TRQ + 8.504E - 03 * ABS(TRQ))
FF	= -1.301 + 7.409E - 06 * ABS(TRQ) * RPM + 7.105E - 02 * SQRT(RPM) + 3.555E - 10 * TRQ * TRQ * TRQ * TRQ
EMISSION AND FUEL CONSUMPTION MODELS FOR DIESEL TRUCKS[1]	
HC	= -1.183E - 02 + 3.459E - 05 * RPM - 7.560E - 06 * ABS(TRQ) - 4.833E - 09 * RPM * RPM
CO	= 3.069E - 02 - 1.107E - 03 * ABS(TRQ) + 2.212E - 07 * ABS(TRQ) * RPM + 1.103E - 05 * TRQ * TRQ
NO	= 2.602E - 02 - 2.035E - 04 * ABS(TRQ) + 4.024E - 07 * ABS(TRQ) * RPM + 6.591E - 04 * SQRT(ABS(TRQ))
FF	= -2.898E - 02 + 3.726E - 03 * ABS(TRQ) + 8.097E - 06 * ABS(TRQ) * RPM + 8.467E - 04 * (ABS(TRQ) + RPM) - 1.180E - 01 * SQRT(ABS TRQ))

[1] Units = grams/second

Where: TRQ = Engine torque in foot-pounds
RPM = Engine speed in revolutions per minute

gine dynamometer tests (20) and a 23-mode schedule was used for gasoline engines (18) in steady-state operation. For each of these test procedures, a test engine was placed on a dynamometer and run through each mode in the prescribed sequence. For the duration of each mode, HC, CO, and NOx exhaust emissions were accumulated in a container (bag) for subsequent weighing while the engine operated at a specified number of revolutions per minute (RPM) and resistive torque (TORQ).

In developing the heavy-duty vehicle emissions and fuel-consumption models for EMPRO, a regression analysis technique was used to relate a series of independent variables, including primarily RPM and TORQ, to the bag values of emissions produced by 30 diesel engines of various makes and types (11). Fifteen emissions models (three pollutant models for each of five engine makes) resulted from this process. In order to have only one emissions model for each pollutant, weighting factors, which represented the percentages of each make of engine included in the total sample, were used. A summary of the resulting three emissions models, along with the fuel consumption model, is given in Table 2. An underlying assumption in this model-building process is that the accumulation of all steady-state emission or fuel consumption predicted for a series of small time increments closely approximates the integral of an instantaneous emission or fuel consumption function over the same time period.

A similar model-building process was used to relate engine performance parameters (RPM and TORQ) for gasoline heavy-duty engines to emissions and fuel consumption. The number of sample engines (two 1975 engines each with 350-in.^{3} displacement) was small, but the predictive models for gasoline-powered HDVs that resulted are given in Table 2. These engines are used in TEXAS-II to power heavy-duty vehicles with appropriate characteristics to represent this overall class of vehicles.

To convert vehicle speed and acceleration values to corresponding engine speed and torque values for use in the predictive emissions and fuel consumption models, a motion equation for a heavy-duty vehicle was derived. Four main resistive forces act on a moving vehicle: (a) rolling resistance, (b) air resistance, (c) resistance due to steepness of the road grade, and (d) drive train resistance (21). The motive force of the vehicle must equal the sum of these resistive forces to maintain a given velocity or produce an acceleration. The power output of the engine must act through the transmission and other drive train components as well as the tires to effect the resulting instantaneous state of vehicle motion. The required engine torque (TORQ) is expressed mathematically as

$$TORQ = 11.106(GRO)\{[0.0076 + 0.00006136(V)]W + [0.09941(V^2)] + W(dh/ds) + [40 + 0.45(V)] + (W/32.2 + nI/R^2)a\}$$

where

GRO = overall gear ratio including axle ratio, tire revolutions per mile, and transmission gear ratio;
V = vehicle speed, ft/sec = (RPM)(GRO);
dh/ds = gradient, ft/ft;
n = number of tires;
I = moment of inertia of wheel, lb/ft/sec^2;
R = loaded wheel radius, ft; and
a = acceleration of vehicle, ft/sec^2.

This expression is adopted to estimate, at selected time intervals, the required torque and RPM

of a HDV engine given instantaneous velocities and accelerations, which are generated by the simulation processor in the TEXAS Model for Intersection Traffic (3).

It was necessary to define a typical HDV transmission. A typical transmission found in HDVs is the nine-gear manual transmission, in which the gear ratios range from 12.5:1 (first gear) to 1:1 (ninth gear). Based on empirical observations the following criteria were used to determine the most appropriate gear ratio at each instant of the simulation process.

Starting with the transmission in first gear, the overall gear ratio and the RPM for a given vehicle speed are calculated. If the RPM exceed the specified RPM that produce maximum torque, the transmission is shifted to the next higher gear. The lower the gear, the higher the torque. This criterion is enforced until the transmission is in the highest gear. Beyond this point, the engine RPM can exceed the specified RPM that produce maximum torque and can reach the manufacturer's specified maximum RPM.

An algorithm for the emissions simulation process for heavy-duty vehicles is shown by the flowchart in Figure 2. Typical vehicle and engine specifications are provided in TEXAS-II. For each time increment of the simulation, a velocity and an acceleration are generated from the TEXAS Model for Intersection Traffic simulation processor, as mentioned previously. From the velocity and acceleration, the operating mode of the vehicle can be determined: acceleration, deceleration, cruise, or idle. During the next step, torque and RPM are calculated by applying the criterion for gear shifting. When engine torque and RPM are known, emissions and fuel consumption rates are estimated.

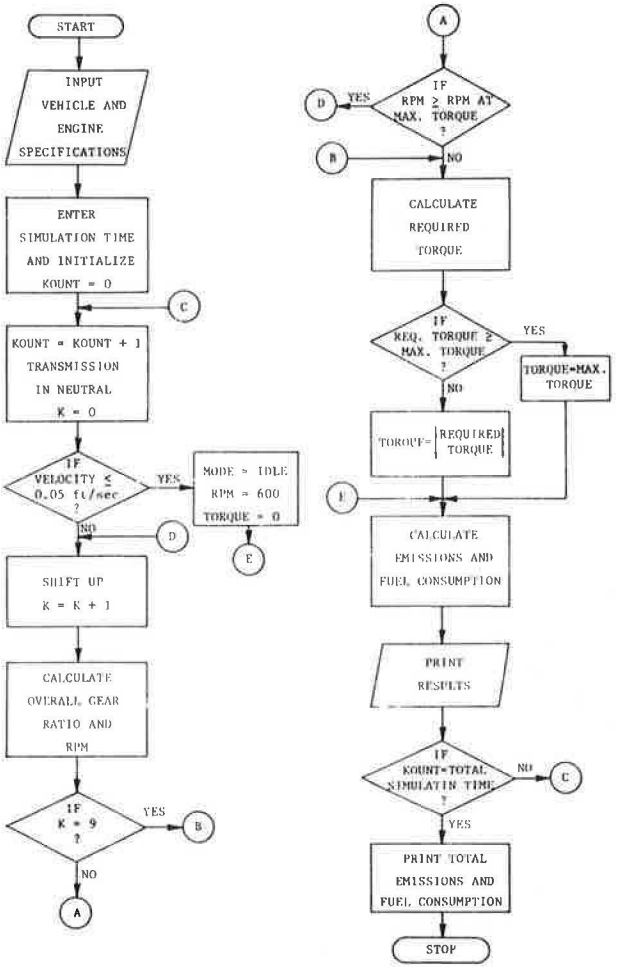


FIGURE 2 Emissions simulation process for heavy-duty vehicles

Summary Statistics from TEXAS-II

A wide range of traffic performance and traffic control device statistics is calculated and tabulated by the TEXAS Model for Intersection Traffic (3), which is incorporated in TEXAS-II. Of these, such factors as speed, delay, queue lengths, and vehicle miles of travel are of particular interest in analyzing the cause-and-effect relationship of traffic behavior at an intersection.

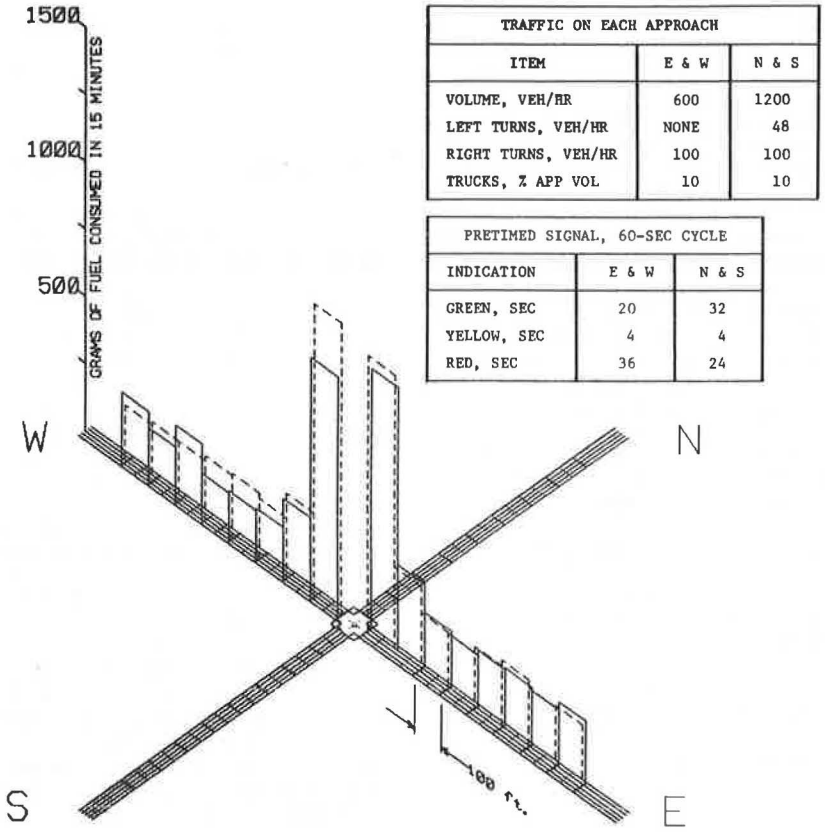
EMPRO, the new data postprocessor in TEXAS-II, computes quantitative estimates of CO, HC, NOx, and fuel consumption and accumulates them in the form of summary statistics. The statistics are tabulated according to bucket, lane, leg, total intersection system, and vehicle class for any user-selected time interval. Small buckets and short time intervals can be chosen to help minimize the effects of displacement, dispersion, or reaction of pollutant sources when modeling concentrations at selected locations in or near the intersection system. Bucket statistics show the longitudinal variation of emissions and fuel consumption along each inbound and outbound lane. Lane statistics are the sum of all buckets along a lane and show the transverse variation in emissions and fuel consumption on each intersection leg. Approach statistics are the sum of all lane statistics, regardless of direction, on each leg. Total intersection system statistics are summed about all approaches and at the intersection proper. Both approach statistics and intersection statistics can be used to analyze the significant effects of selected intersection environmental factors on emissions and fuel consumption.

Figures 3-5 show the pattern of fuel consumption statistics that are produced by EMPRO on the inbound lanes and in the intersection proper where two four-lane streets intersect. Figure 3 shows the grams of fuel consumed by mixed traffic passing through each 100-ft bucket in each inbound lane on the east-west street in a 15-min period for the geometry, traffic, and traffic control conditions shown. Figure 4 shows similar statistics for the north-south street. Figure 5 shows a summary of fuel consumption on all inbound lanes and the amount of fuel consumed by all vehicles crossing the intersection proper.

Application of TEXAS-II

The TEXAS-II model is currently operational on mainframe CDC and IBM computers that support FORTRAN 66 and FORTRAN 77 languages. The package is being converted to FORTRAN 77 and adapted to a user-friendly, interactive format of input and output that will probably incorporate interactive computer graphics. This enhancement, which will be completed in a few months, also includes adaptation to run on superminicomputers such as the VAX 11/780.

TEXAS-II has been run more than 300 times recently in a series of experiments designed to obtain quantitative estimates of the effects of various traffic and intersection factors on emissions, fuel consumption, traffic delays, and queue lengths (22). The resulting data have been used to build predictive models of emissions and fuel consumption at intersections. The factors that were used for simulating the intersection environment were (a) intersection size, (b) presence or absence of a special



TRAFFIC ON EACH APPROACH		
ITEM	E & W	N & S
VOLUME, VEH/HR	600	1200
LEFT TURNS, VEH/HR	NONE	48
RIGHT TURNS, VEH/HR	100	100
TRUCKS, % APP VOL	10	10

PRETIMED SIGNAL, 60-SEC CYCLE		
INDICATION	E & W	N & S
GREEN, SEC	20	32
YELLOW, SEC	4	4
RED, SEC	36	24

FIGURE 3 Fuel consumption estimates on inbound E-W lanes.

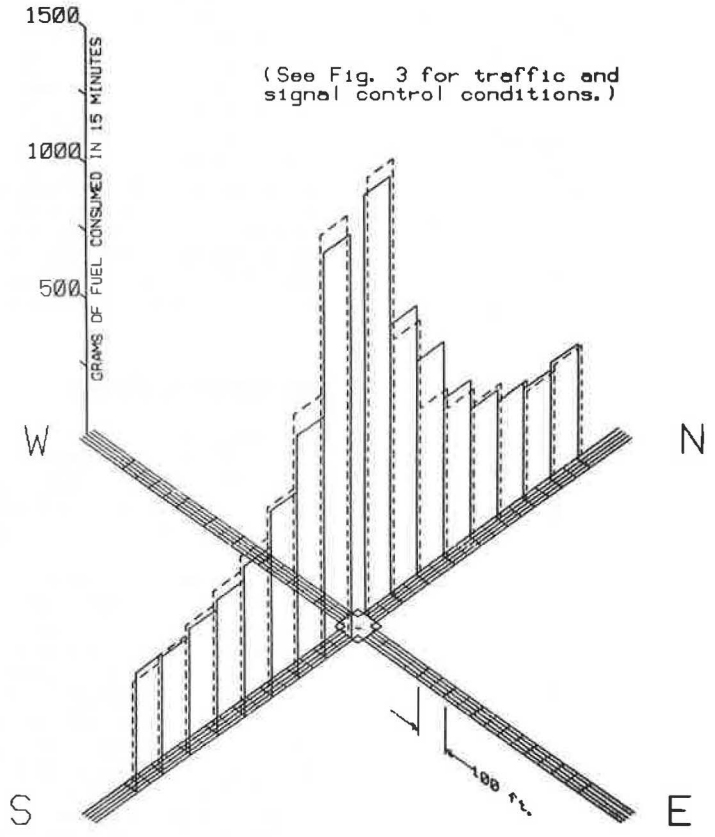


FIGURE 4 Fuel consumption estimates on inbound N-S lanes.

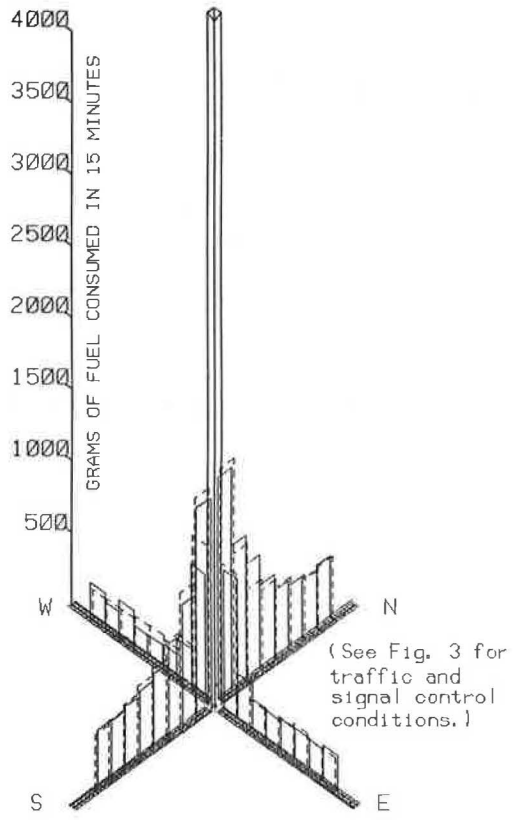


FIGURE 5 Fuel consumption estimates on all inbound lanes and within the intersection proper.

left-turn lane, (c) pretimed signal control, (d) fully actuated signal control, (e) all-way stop-sign control, (f) traffic volume, (g) left turns, and (h) heavy-duty vehicles. A typical four-way intersection, with moderate traffic, run for 5 min start-up time and 15 min simulation time took about 300 TM seconds on a Cyber 170/750 system for the simulation processor. The emissions processor took an additional 100 TM seconds to calculate the emissions and to place them in the appropriate buckets along each lane and in the intersection proper.

The results of this study (22) can be used in one of three ways. First, the predictive models can be applied to calculate the expected source of emissions, fuel consumption, and traffic performance parameters for any selected intersection situation that was included in the range of simulated conditions. Second, a series of tables can be used for convenient lookup of these values. Finally, the TEXAS-II computer simulation program can be run to obtain detailed data concerning any specific intersection environment of practical interest. The values thus obtained can serve as a basis for further emission dispersion studies or for direct comparison of the effects of various intersection features on emission sources, fuel consumption, vehicular delay, and queue lengths.

REFERENCES

1. G.S. Springer and D.J. Patterson. Engine Emissions: Pollutant Formation and Measurement. Plenum, New York, 1973.
2. L.E. Haefner, D.E. Lang, R.W. Meyer, J.L. Hutchins, and B. Yarjani. Line Source Emissions Modeling. In Transportation Research Record 648, TRB, National Research Council, Washington, D.C., 1977, pp. 71-73.
3. C.E. Lee, T.W. Rioux, and C.R. Copeland, Jr. The TEXAS Model for Intersection Traffic--Development. Research Report 184-1. Center for Highway Research, University of Texas at Austin, Dec. 1977.
4. C.E. Lee, T.W. Rioux, V.S. Savur, and C.R. Copeland, Jr. The TEXAS Model for Intersection Traffic--Programmer's Guide. Research Report 184-2. Center for Highway Research, University of Texas at Austin, Dec. 1977.
5. C.E. Lee, G.E. Grayson, C.R. Copeland, Jr., J.W. Miller, T.W. Rioux, and V.S. Savur. The TEXAS Model for Intersection Traffic--User's Guide. Research Report 184-3. Center for Highway Research, University of Texas at Austin, July 1977.
6. C.E. Lee, V.S. Savur, and G.E. Grayson. Application of the TEXAS Model for Analysis of Intersection Capacity and Evaluation of Traffic Control Warrants. Research Report 184-4F. Center for Highway Research, University of Texas at Austin, July 1978.
7. C.R. Copeland, Jr. Modifications of the TEXAS Model to Include Estimates of Vehicular Emissions. M.S. thesis. University of Texas at Austin, 1980.
8. H.T. McAdams, P. Kunselman, C.J. Domke, and M.E. Williams. Automobile Exhaust Emission Modal Analysis Model. EPA-460/3-74-005. Office of Air and Water Programs, Environmental Protection Agency, Jan. 1974.
9. H.H. Wu. Modeling Heavy-Duty Gasoline Vehicle Emissions and Fuel Consumption. M.S. thesis. University of Texas at Austin, 1980.
10. P. Athalye. Modeling Heavy-Duty Diesel Vehicles Emissions and Fuel Consumption. M.S. thesis. University of Texas at Austin, 1980.
11. C. Simeonidis. Emissions and Fuel Consumption as Functions of Vehicle Speed and Acceleration. M.S. thesis. University of Texas at Austin, 1981.
12. C.M. Urban and K.J. Springer. Study of Emissions From Heavy-Duty Vehicles. Southwestern Research Institute, San Antonio, Texas, May 1976.
13. M.N. Ingalls and K.J. Springer. Summary and Comparison of Mass Emissions--All Gasoline and Diesel Powered Trucks Tested on the San Antonio Road Route 1971-1975. Southwestern Research Institute, San Antonio, Texas, March 1975.
14. K.J. Springer and C.D. Tyree. Exhaust Emissions From Gasoline-Powered Vehicles Above 6,000 lb Gross Vehicle Weight. Southwestern Research Institute, San Antonio, Texas, 1972.
15. M.N. Ingalls and K.J. Springer. In-Use Heavy-Duty Gasoline Truck Emissions: Part I, Mass Emissions From Trucks Operated Over a Road Course. Southwestern Research Institute, San Antonio, Texas, Feb. 1973.
16. C.M. Urban and K.J. Springer. 13-Mode Diesel Emissions Test Results on 12 Trucks. Southwestern Research Institute, San Antonio, Texas, Feb. 1976.
17. M.N. Ingalls and R.L. Mason. Heavy-Duty Fuel Economy Program: Phase I, Specific Analysis of Certain Existing Data. EPA-460/3-77-001. Southwestern Research Institute, San Antonio, Texas, Jan. 1977.
18. C.M. Urban and K.J. Springer. Heavy-Duty Fuel Economy Program: Phase II, Evaluation of Emission Control Technology Approaches. EPA-460/3-

77-010. Southwestern Research Institute, San Antonio, Texas, July 1977.
19. C.M. Urban. Heavy-Duty Fuel Economy Program: Phase III, Transient Cycle Evaluations of the Advanced Emissions Control Technology Engineer. EPA-460/3/78-005. Southwestern Research Institute, San Antonio, Texas, May 1978.
20. Federal Register, Vol. 42, No. 174, Sept. 8, 1977.
21. F.T. Buckley, Jr., C.H. Marks, and H.W. Walston, Jr. Analysis of Coast-Down Data to Access Aerodynamic Drag Reduction on Full-Scale Tractor-Trailer Trucks in Windy Environments. Report 760850. Society of Automotive Engineers Transactions, Mechanical Engineering Department, University of Maryland, College Park, 1976.
22. F.-P. Lee, C.E. Lee, R.B. Machemehl, and C.R. Copeland, Jr. Vehicle Emissions at Intersections. Research Report 250-1. Center for Transportation Research, Bureau of Engineering Research, University of Texas at Austin, Aug. 1983.

The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the FHWA and the Texas State Department of Highways and Public Transportation. This paper does not constitute a standard, specification, or regulation.

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

Discontinuity in Equilibrium Freeway Traffic Flow

HAROLD J. PAYNE

ABSTRACT

Analysis of freeway traffic flow data reveals a discontinuity in the equilibrium relationship between speed and density, supporting the dual-mode or two-regime theories of traffic flow. This work was done during the development of an appropriate equilibrium speed-density relationship for a dynamic macroscopic freeway simulation model, FREFLO. This context assisted in clarifying the distinction between equilibrium and nonequilibrium conditions in traffic data and as a consequence made the discontinuity in equilibrium conditions perfectly evident. Use of the new relationship greatly improved the quality of FREFLO model predictions. The corresponding discontinuity in the volume that can be maintained appears to have great significance for the design of freeway control systems.

Traffic speeds bear a generally consistent relationship to traffic density (i.e., mean spacing). A multitude of theories has been developed to represent both static, average relationships at a macroscopic level and dynamic relationships at both microscopic and macroscopic levels. These theories are connected by the fact that the possible steady-state conditions of the dynamic models, expressed as a set of speed-density pairs, can be viewed as a macroscopic, equilibrium speed-density relationship.

The earliest work proposed fairly simple speed-density relationships (e.g., speed linearly decreasing with increasing density). However, Edie (1) first observed a difference in character between two regimes of traffic, roughly characterized as uncongested and congested, and proposed a more sophisticated two-regime model for the speed-density relationship. Empirical analyses (2-5) tended to support this view and others investigated this concept (6).

Further support for the existence of two modes of behavior, and in particular for a discontinuity in the speed-density relationship, is provided. This work was done during the development of an appropriate equilibrium speed-density relationship for a dynamic macroscopic freeway simulation model, FREFLO (7,8). This context assisted in clarifying the distinction between equilibrium and nonequilibrium conditions in traffic data and as a consequence made the discontinuity in equilibrium conditions per-

fectly evident. Use of the new relationship greatly improved the quality of FREFLO model predictions.

This discontinuity also reveals itself as a discontinuity in the volume-density relationship. In this form, implications for freeway control are evident. This was first recognized by Bullen (9), and was elaborated by him and his colleagues (10,11).

The remainder of this paper is organized as follows. The next section describes FREFLO and the role of the equilibrium speed-density relationship in it. Subsequent sections describe the acquisition, preparation, and selection of data; data analysis procedures; and regression results. The final section addresses implications for freeway control.

FREFLO

The TRAFLO simulation code, recently completed under FHWA sponsorship, consists of three street network component models and a macroscopic freeway component model, TRAFLO. FREFLO is the macroscopic freeway component of the TRAFLO simulation code (12). FREFLO, as described in several reports and papers (12-14), is based on an aggregate variable model first developed for simulation purposes by Payne (7).

FREFLO is based on a representation of traffic flow in terms of three aggregate variables:

ρ: traffic density (units: vehicles/lane-mile),
u: (space-mean) speed (units: miles/hour), and
q: volume (units: vehicles/lane-hour).

The freeway itself is represented by a network of freeway sections. Traffic flow is assumed to be homogeneous within a section. The two aggregate variables, ρ and u, describe conditions over a section at an instant in time. Volume, on the other hand, describes the rate of movement of vehicles past section boundaries.

To make FREFLO specific to a site, traffic demand data, geometric data, and traffic flow behavior data must be specified. The traffic flow behavior parameters were the object of the calibration effort described elsewhere (15).

The parameters calibrated were

1. Speed-density relationship parameters, including nominal capacity for each section of roadway, free-flow speed, and several more coefficients that define the shape of the speed-density relationship and
2. Dynamic interaction parameters, including the reaction time coefficient k_T and the anticipation coefficient k_v.

Attention is restricted to the speed-density relationship. Results pertinent to the dynamic interaction parameters are available elsewhere (15). Freeway traffic surveillance data gathered from the freeway systems in Los Angeles (16,17) were acquired for use in this study.

ACQUISITION, PREPARATION, AND SELECTION OF DATA

Automated surveillance equipment for acquiring speed-density data is available in the Los Angeles area. Occupancy and counts over 20-, 30-, or 60-sec intervals are generated. Speed traps were not in place, so direct measurement of individual speeds was not provided.

With automated surveillance data, the occupancy data can be scaled to obtain density. The appropriate scale factor depends on the traffic mix (distribution of vehicle lengths), the size of the presence detector loop, and the selection of thresholds in the associated electronics. Hence, the appropriate scale factor will be site specific and can vary significantly from one location to another.

Counts yield volume measurements by application of the appropriate factor. For example, 60-sec counts yield hourly volumes according to

Hourly volume = 60 x (60-sec count).

Finally, given any two of the three (desired) measurements, density, speed and volumes, the third can be derived via the relationship

Volume = speed x density.

In Table 1, surveillance data, occupancy, and speed (derived from volume and occupancy) measurements from the Santa Monica Freeway in Los Angeles are given. (Density values were computed from occupancy by applying the scale factor 2.72, a value appropriate for the Los Angeles freeways from which data were obtained for this research.) As suggested in Table 1, attention is focused on station 18.

The dynamic speed-density relationship includes the equilibrium speed-density relationship. It also includes additional terms that are intended to entirely account for spatial and temporal variations in traffic conditions of the FREFLO model (15). To obtain data suitable for calibrating the speed-density relationship, it is necessary to exclude data at times and places where substantial spatial and temporal variations exist. Therefore, in examining data pertaining to a particular location (i.e., a surveillance station) it is necessary to consider data from adjacent stations to properly identify equilibrium data.

Data at a particular station are not suitable if values (of density or speed) at adjacent stations are substantially different, indicative of substantial spatial variations. It is also necessary to observe variations over time within a station; data values are not suitable if adjacent values (in time) are substantially different, indicative of substantial temporal variations.

In Table 1, three segments of data pertaining to station 18 have been distinguished. The first segment contains high-speed, low-density, equilibrium data. The second segment involves nonequilibrium data, so judged because the substantial spatial variations in speed, or occupancy, or both; these data are not used for calibration. The third segment involves low-speed, high-density equilibrium data. Such data will entail some temporal and spatial variations, but there is no persistent difference, as there is in the second segment.

Analysis of Equilibrium Speed-Density Data

The analysis of speed-density data to produce the desired equilibrium speed-density relationship involved two major steps: (a) careful selection of equilibrium data and (b) application of regression software to extract coefficients and measures of fit.

Data Selection

Selection of data for developing the equilibrium speed-density relationship requires great care. Figures 1 through 4 show the differences in the data that arise depending on the degree of care that is taken in selecting data. Figure 1 shows composite data from several adjacent stations on the Santa Monica Freeway. Table 1 gives a portion of that data. It will be observed that there appears to be a continuous reduction of speed with density. However,

TABLE 1 Occupancy and Speed Data from the Santa Monica Freeway in Los Angeles

	OCCUPANCY (%) STATION NUMBER							SPEED (MI/HR) STATION NUMBER							
TIME	15	16	17	18	19	20	21	15	16	17	18	19	20	21	
	14	13	13	12	14	14	10	49	53	49	47	44	44	49	HIGH SPEED, LOW
	14	13	12	15	17	14	14	50	54	48	44	45	43	46	DENSITY EQUILIBRIUM
	14	15	14	13	18	18	15	47	50	51	36	41	44	44	DATA
	11	12	15	16	17	18	14	46	49	46	37	40	41	47	
	15	11	10	20	18	15	16	49	50	44	40	39	42	46	
	17	17	15	14	18	19	15	49	53	47	31	40	39	45	
	13	14	18	16	17	16	15	49	52	47	26	41	40	44	
	15	13	14	21	16	17	15	48	53	50	24	40	42	44	
	14	13	15	23	17	19	15	50	51	44	28	40	43	46	NONEQUILIBRIUM DATA:
	14	13	13	25	16	16	16	48	53	40	27	39	41	46	SUBSTANTIAL SPATIAL
	13	12	15	21	16	14	14	44	50	23	20	40	38	46	AND/OR TEMPORAL
	16	14	13	24	19	17	13	42	49	22	22	39	38	42	VARIATIONS IN
	18	14	27	28	18	19	15	43	40	20	21	40	39	42	OCCUPANCY AND/OR
	17	14	25	27	16	17	18	46	24	21	21	38	38	44	SPEED ARE PRESENT
	17	18	29	28	17	16	14	45	21	9	24	40	38	43	
	18	28	32	29	16	17	16	34	28	19	24	36	41	45	
	18	32	49	25	17	18	16	30	15	22	22	37	39	44	
	21	26	31	24	18	14	15	33	13	23	22	37	38	44	
	24	33	30	27	17	17	15	14	21	19	24	32	38	43	
	22	36	26	31	18	20	15	20	27	21	21	37	34	44	
	38	29	31	26	23	18	16	32	19	18	24	37	38	45	
	29	24	28	26	17	21	15	17	21	17	25	35	38	41	
	22	30	30	26	19	19	16	22	21	12	24	34	37	41	
	35	29	32	24	19	18	18	28	20	20	22	36	34	36	
	28	28	42	27	20	16	16	26	8	23	24	34	33	32	
	24	27	28	27	19	20	17	16	16	22	22	37	35	31	
	28	47	27	25	20	23	22	14	24	21	24	37	32	15	
	36	38	28	26	17	20	22	23	28	19	24	35	31	28	
	35	27	30	25	19	20	36	29	28	8	25	36	12	30	
	25	22	35	26	18	23	22	32	26	19	24	32	20	32	
	22	24	51	24	20	40	21	33	16	15	26	26	24	35	
	22	24	32	24	18	32	21	35	18	16	26	16	27	32	LOW SPEED, HIGH
	21	31	35	23	25	26	20	25	23	19	26	21	29	24	DENSITY EQUILIBRIUM
	20	31	32	23	35	24	20	28	12	17	21	18	31	28	DATA
	29	26	33	24	31	23	26	31	15	21	18	15	25	21	
	25	39	34	28	33	21	22	28	20	19	18	16	32	19	

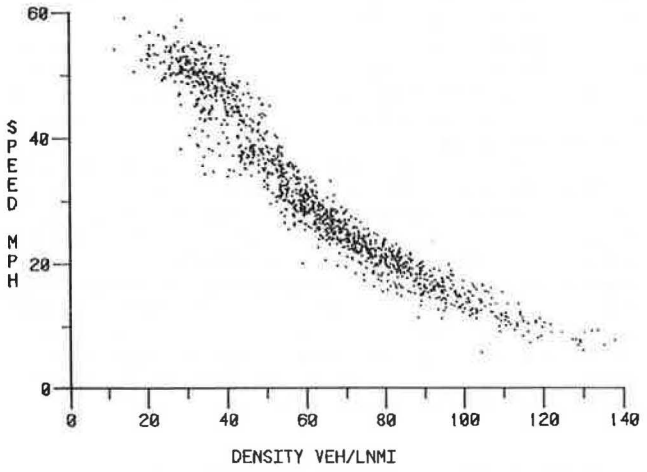


FIGURE 1 Composite scatterplot for several Santa Monica Expressway stations.

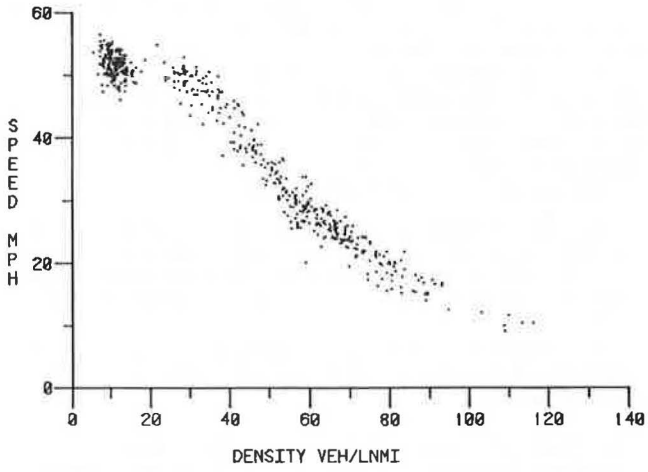


FIGURE 2 Scatterplot for Santa Monica Expressway station 18.

one must recognize that data from different stations involve different electronics so that different scale factors may be required to accurately estimate density from occupancy at the several stations. In the composite data displayed, a single scale factor was applied. In addition, differences in that data for distinct stations can be expected because of local geometrics and resultant differences in traffic behavior. Finally, the data of Figure 1 include data in nonequilibrium conditions.

Figure 2 shows data pertinent to a single station, SME 18. (A portion, but not all, of the data in Figure 2 is present in Figure 1.) Compared with the data in Figure 1, there is clearly a great deal less scatter. These data, however, also contain data in nonequilibrium conditions. Figures 3 and 4 clearly show the distinctive differences among several traffic regimes. In Figure 3, only nonequilibrium data are shown. It will be noted that speeds are generally in the 30- to 40-mph range, with densities in the 40- to 60-vehicles per lane-mile range. Equilibrium data are shown in Figure 4. These data clearly fall into two regimes, corresponding to high-speed flow and congested flow. Most striking, however, is the clear discontinuity in speeds: Equilibrium speeds in the 30- to 40-mph range are rare. Thus, the appearance of continuity of speed as a function of density is seen to be derived from mixing nonequilibrium data with equilibrium data.

In the analyses that were undertaken, data were

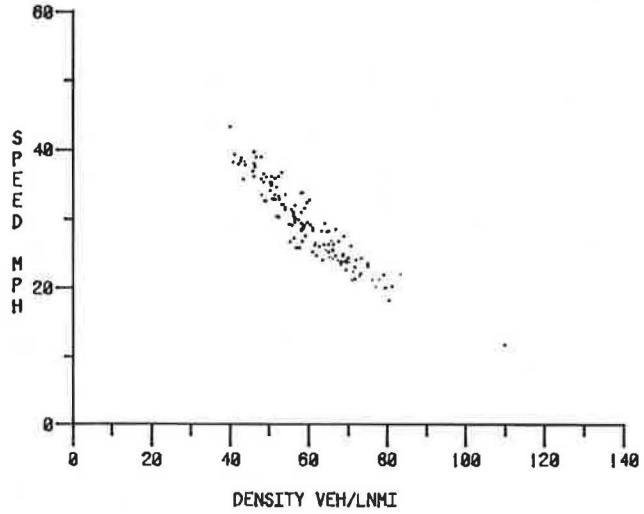


FIGURE 3 Nonequilibrium data for Santa Monica Expressway station 18.

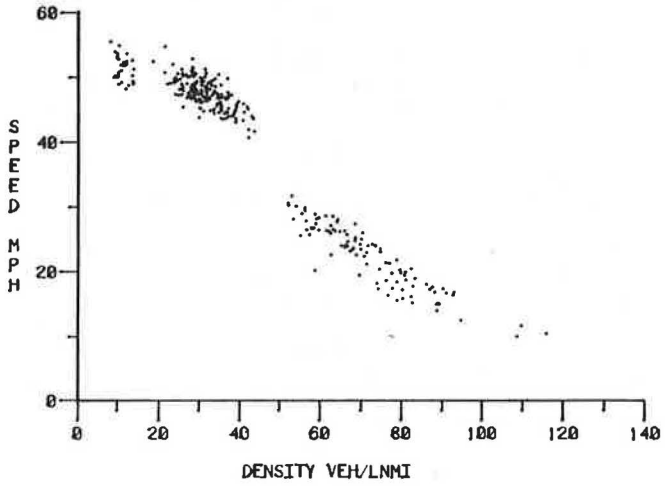


FIGURE 4 Equilibrium data for Santa Monica Expressway station 18.

carefully selected to remove nonequilibrium data. Analyses were first conducted on data from a single station. When a proper understanding of the form of the speed-density relationship was obtained, analysis of a composite data set, formed by aggregating all the station-specific data, was conducted to establish a universal form for the speed-density relationship. The data selected for analysis are given in Table 2.

DATA ANALYSIS PROCEDURES

The steps adopted for determination of the equilibrium speed-density relationship were

1. Selection of data,
2. Selection of functional form,
3. Selection of density breakpoints to define flow regimes,
4. Application of regression technique in each flow regime, and
5. Qualitative and quantitative assessment of alternatives.

TABLE 2 Data Selected for Equilibrium Speed-Density Analysis

STATION	TOTAL NUMBER OF EQUILIBRIUM DATA PAIRS	DATA SET*	NUMBER OF POINTS
SME12	491	740904-62	141
		740904-63	175
		740423-53	175
SME18	301	740904-62	110
		740904-63	30
		740423-53	61
		740423-56	100
HAS13	404	740904-60	130
		740322-01	64
		740325-02	72
		740426-05	138
HAS18	429	740904-60	155
		740322-01	95
		740325-02	88
		740426-05	91
HAS24	170	740904-60	22
		740325-02	88
		740426-05	60

*See Payne *(16)*.

This procedure is of course iterative. Selection of data required careful attention to temporal and spatial variation, as discussed previously.

Functional Forms

Several functional forms were examined, including the two-regime polynomial and inverse polynomial forms and the three-regime forms. The first two functional forms involve one density breakpoint that divides the free-flow from the impeded and constrained flow regime. The third functional form also involves a second density breakpoint, dividing the impeded and constrained flow regimes.

These functional forms were selected by visual inspection of data to select data that provide adequate conformance to the trends and degrees of freedom in matching the data. Data in each regime were treated separately and manipulated as necessary so that linear regression techniques could be applied. Finally, assessment of the alternative results was made to arrive at a standard form for the equilibrium speed-density relationship that is believed appropriate for implementation of FREFLO. The three-regime form is shown in Figure 5.

The method adopted for estimating coefficients from data was to apply linear regression to the data in each of the two- or three-flow regimes separately. In certain circumstances, this required a preliminary manipulation of the data so that a linear regression technique could be applied.

In all instances, the free-flow regime was represented by a constant free-flow speed. The value of free-flow speed was simply obtained as the average of speed values where the corresponding density was below the density breakpoint, ρ_1, defining the free-flow regime.

Three functional forms were considered for the impeded and constrained flow regime:

1. Polynomial,
2. Inverse polynomial, and
3. Shifted inverse polynomial.

The polynomial form,

$$u_e(\rho) = \alpha_0 + \alpha_1\rho + \alpha_2\rho^2 + \ldots + \alpha_n\rho^n$$

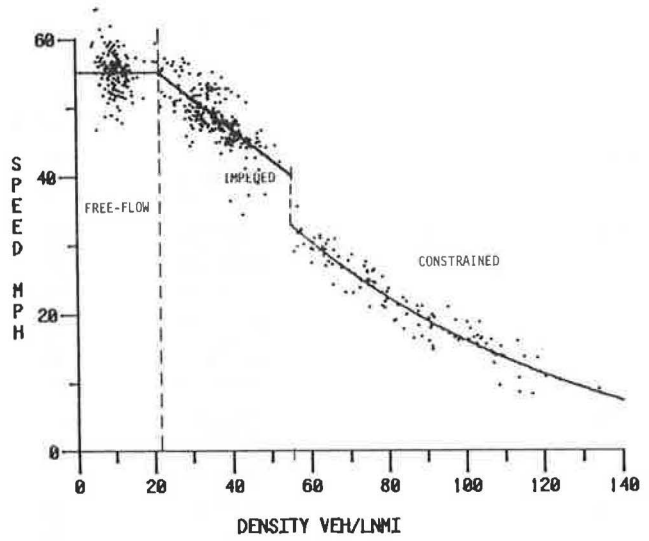


FIGURE 5 Three-regime, discontinuous speed-density relationship.

can be directly treated by standard linear regression techniques. IBM's Scientific Subroutine Package was in fact used for this purpose. The inverse polynomial form

$$u_e(\rho) = 1/(\beta_0 + \beta_1\rho + \beta_2\rho^2 + \ldots + \beta_n\rho^n)$$

can be rewritten as

$$1/u_e(\rho) = \beta_0 + \beta_1\rho + \beta_2\rho^2 + \ldots + \beta_n\rho^n$$

Hence, if each data pair (ρ,u) is replaced by $(\rho,1/u)$, the same standard linear regression technique can be applied. The third form, the shifted inverse polynomial,

$$u_e(\rho) = -u_0 + 1/(\beta_0+\beta_1\rho + \beta_2\rho^2 + \ldots + \beta_n\rho^n)$$

can similarly be written as

$$1/[u_0 + u_e(\rho)] = \beta_0 + \beta_1\rho + \beta_2\rho^2 + \ldots + \beta_n\rho^n$$

so that the data transformation form (ρu) to $[\rho,1/(u_0+u)]$ again allows application of the same technique.

Note that the selection of the density breakpoints and the speed offset (u_0) are not a part of the linear regression technique and must be done before curve fitting. The method adopted was to

1. Select density breakpoints (and speed offset, U_0, in the third form),
2. Apply linear regression in each regime, and
3. Combine results (compute total sum of squared deviations).

Variations in density breakpoints, functional form, polynomial order, and speed offsets produced alternative results that were then assessed quantitatively, using the total sum of squared deviations, and qualitatively, considering the resultant volume-density relationship and the consistency of the selection of breakpoint with actual match of separate regression at the density breakpoint. This method is suboptimal in the sense that an optimal, nonlinear regression technique embodying selection of polynomial coefficients, density breakpoints, and speed offset could have been defined. This method was selected primarily because of its simplicity and economy, which depend on available subroutines.

REGRESSION RESULTS

The methods described earlier were applied to the data sets given in Table 2--initially to each station-specific data set and then to the composite data set. Initial analysis focused on the determination of an appropriate functional form. This effort resulted in selection of the three-regime, discontinuous form. More specifically, it was found that the first-order polynomial form in the impeded regime and the inverse first-order polynomial form in the congested regime were satisfactory.

Results of the regression analysis are given in Table 3. Shown are full calibration results and results for calibration to the standard form. The former is discussed first. Each station-specific data set was subjected to a series of regression analyses, with the functional form fixed as indicated and with variations in density breakpoints and speed offset. Table 3 gives results of these analyses in the form of regression coefficients, selected breakpoints, capacity, and fit. The relative consistency of these results suggested that a universal form might be obtained from the composite data set. Results of that analysis are also given in Table 3.

The results obtained from this analysis of the composite data defined the standard form for the speed-density relationship. This relationship and the corresponding volume-density relationship are shown in Figures 6 and 7, respectively.

The next set of analyses involved application of a second calibration method (15) to each of the station-specific data sets. Each such analysis established a free-flow speed and a scale factor that determined the nominal capacity. The method of arriving at the two parameters was twofold: (a) linear regression to establish free-flow speed and (b) an iterative, mean-square error technique to establish the scale factor. Details appear elsewhere (15). These results are given in Table 3 under the heading "Calibration to Standard Form." The free-flow speed is the same as obtained in the full calibrations. Fits were generally only slightly degraded from the full calibration results.

It is interesting to note that these results imply a corresponding discontinuity in volume at the second breakpoint of significant size. For the standard relationship identified, volume decreased from 2,136 to 1,723 veh/lane-hr at the breakpoint value of 55 veh/lane-mile.

IMPLICATIONS FOR CONTROL

Bullen (9) first called attention to the importance of the dual-mode behavior of traffic flow as manifested by the distinction in the character of traffic flow in the uncongested and congested regimes, and particularly as a consequence of the discontinuity in volume. His observations generally accord with the views of many freeway operations traffic engineers who recognize the instability of traffic near peak volumes. The general principle that is derived from this understanding is a need to operate the freeways at volumes less than peak volumes to avoid a transition to the congested regime with an associated substantial reduction in volume.

A full understanding of the discontinuity is not yet available. One expects that the basis for the discontinuity should be traceable to specific characteristics of driver-vehicle behavior. Some suggestions along these lines were presented by Ceder (5), but no definitive studies have been done. Recently, however, Cohen, according to a private communication, has made controlled simulation runs using the microscopic INTRAS model (18,19), which have also

TABLE 3 Summary of Regression Results

	FULL CALIBRATION										CALIBRATION TO STANDARD FORM		
				IMPEDED REGIME			CONSTRAINED REGIME						
DATA SET	CAPACITY VEH/LN-HR	FREE SPEED MI/HR	FIRST BREAKPOINT VEH/LN-MI	u_1 MI/HR	a	SECOND BREAKPOINT VEH/LN-MI	OFFSET MI/HR	β_0	β_1	FIT MI/HR	*** SCALE FACTOR	CAPACITY* VEH/LN-HR	*** FIT MI/HR
SME12	2177	55.1	20	63.8	-.423	52	28	$.867\times10^{-2}$	$.139\times10^{-3}$	1.90	1.0706	2287	2.55
SME18	2069	51.4	20	57.8	-.328	50	28	$.903\times10^{-2}$	$.153\times10^{-3}$	1.95	.9962	2128	2.14
HAS13	2096	49.0	27	57.4	-.310	50	NO DATA AVAILABLE			1.53	1.0163	2171	1.54
HAS18	2147	50.0	12	53.5	-.295	60	28	$.975\times10^{-2}$	$.135\times10^{-3}$	1.17	.9457	2020	1.81
HAS24	2237	47.0	33	56.3	-.284	55	28	$.700\times10^{-2}$	$.167\times10^{-3}$	1.76	1.0245	2188	1.56
COMPOSITE	2136*	58.1	0	58.1	-.351	50/60**	28	$.912\times10^{-2}$	$.141\times10^{-3}$	2.68			

*Extrapolated to volume at density of 55 veh/ln-mi.

**Data with densities in the range 50-60 veh/ln-mi were excluded from the regression analysis.

***Root-mean-squared-error.

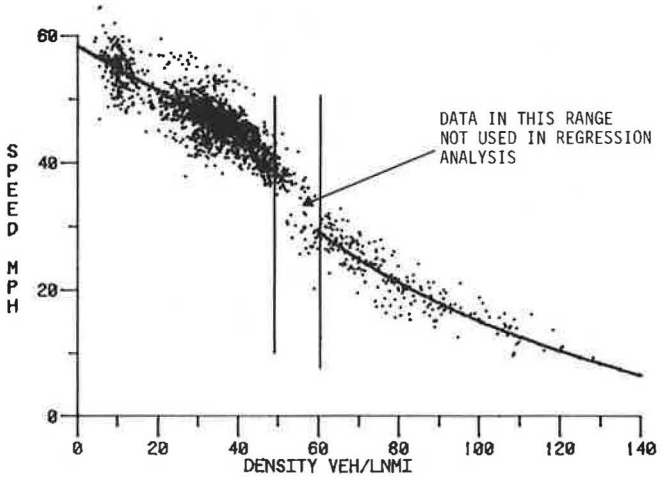


FIGURE 6 Regression for composite data set (speed-density).

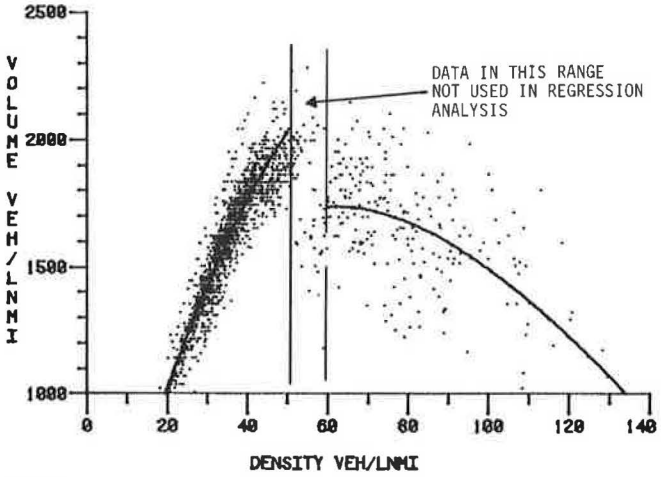


FIGURE 7 Regression for composite data set (volume-density).

exhibited a reduction of substantial volume after breakdown in a lane-drop situation (INTRAS models individual car behavior).

Figure 8 shows such a lane-drop situation with associated ramps. Uncontrolled, traffic can enter the upstream on-ramp in such volume as to exceed the capacity, leading to breakdown (i.e., transition to the congested conditions). In a long section of two

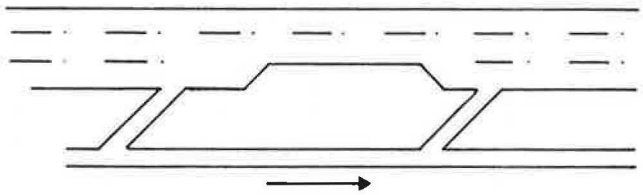


FIGURE 8 Freeway bottleneck.

lanes, the capacity will as a result be substantially reduced, contributing further to congestion and delay. On-ramp control, either by storing vehicles at the ramp or by diverting vehicles to an alternate route, can prevent breakdown and thereby maintain a higher service rate (i.e., volume along the two-lane section).

Traffic dynamics are substantially more complex than this simple analysis might suggest. The FREFLO model, by representing effects due to spatial variations in density, for example, admits increased volumes above the congested capacity in transition areas (e.g., at the head of a queue). This effect is also observed in real traffic. As a result, the effect of the discontinuity can be overcome in small areas. Development of control strategies to account for the discontinuity in volume must therefore be fairly sophisticated and take into account dynamic as well as the underlying static equilibrium relationships.

ACKNOWLEDGMENT

This research was conducted under Contract DTFH61-81-C-00038, sponsored by the Federal Highway Administration, Steven Cohen and Paul Ross, contract managers.

REFERENCES

1. L.C. Edie. Car-following and Steady State Theory for Noncongested Traffic. Operations Research, Vol. 9, No. 1, 1961, pp. 66-76.
2. P. Athol. Interdependence of Certain Operational Characteristics Within a Moving Traffic Stream. In Highway Research Record 72, HRB, National Research Council, Washington, D.C., 1965, pp. 58-87.
3. J.S. Drake, J.L. Schoefer, and A.D. May, Jr. A Statistical Analysis of Speed Density Hypotheses. In Highway Research Record 154, HRB,

National Research Council, Washington, D.C., 1967, pp. 53-87.
4. A.D. May, Jr. and H.E.M. Keller. Evaluation of Single- and Two-Regime Traffic Flow Models. Proc., 4th International Symposium on the Theory of Traffic Flow, Karlsruhe, 1968.
5. A. Ceder. Investigation of Two-Regime Traffic Flow Models at the Micro- and Macroscopic Levels. Ph.D. dissertation. University of California, Berkeley, 1975.
6. H.S. Mika, J.B. Kreer, and L.S. Yuan. Dual-Mode Behavior of Freeway Traffic. In Highway Research Record 279, HRB, National Research Council, Washington, D.C., 1969, pp. 1-12.
7. H.J. Payne. Models of Freeway Traffic and Control. Mathematics of Public Systems, Vol. 1, 1971, pp. 51-61.
8. E. Lieberman, B. Andrews, M. Davila, and M. Yedlin. Macroscopic Simulation for Urban Traffic Management, 3 volumes. Report FHWA-RD-80-113, 114, and 115. Office of Research and Development, FHWA, U.S. Department of Transportation, 1982.
9. A.G.R. Bullen. Strategies for Practical Expressway Control. Proc., ASCE 1972 and Transportation Engineering Journal, Aug. 1972.
10. P. Athol and A.G.R. Bullen. Multiple Ramp Control for a freeway Bottleneck. In Highway Research Record 456, HRB, National Research Council, Washington, D.C., 1973, pp. 50-54.
11. D.C. Haughton. A Mackov Model for Expressway Control. Presented at Operation Research Society of America National Meeting, Las Vegas, Nevada, 1975.
12. H.J. Payne and C. Wolfe. Macroscopic Simulation for Urban Traffic Management, The TRAFLO Model, General Technical Specifications, Supplement: Macroscopic Freeway Model. Report ES-R-B001-2. ESSCOR, San Diego, Calif., Oct. 1978.
13. H.J. Payne. FREFLO: A Macroscopic Simulation Model of Freeway Traffic (Version I), User's Guide. ESSCOR Report ES-R-78-1. July 1978.
14. H.J. Payne. A Critical Review of a Macroscopic Freeway Model. In Computer Control of Urban Traffic Systems, Pacific Grove, Calif., 1979, pp. 251-265.
15. H.J. Payne. Calibration and Validation of FREFLO, Vol. I: User Calibration Procedures and Supporting Data Analysis. VERAC Report R-001-82. VERAC, Inc.; FHWA, U.S. Department of Transportation, Dec. 1981.
16. H.J. Payne and E.D. Helfenbein. Freeway Surveillance Data. Report FHWA-RD-76-74. FHWA, U.S. Department of Transportation, April 1976.
17. Data tapes PD 263 644/AS and PD 263 543/AS. National Technical Information Service.
18. D.A. Wicks and E.B. Lieberman. Development and Testing of INTRAS, a Microscopic Freeway Simulation Model, Vol. 1: Program Design and Parameter Calibration. Report KLD TR-40. KLD Associates; FHWA, U.S. Department of Transportation, Jan. 1976.
19. D.A. Wicks and B.J. Andrews. Development and Testing of INTRAS, a Microscopic Freeway Simulation Model, Vol. 2: User's Manual. Report FHWA-RD-076-77. FHWA, U.S. Department of Transportation, June 1977.

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