TRANSPORTATION RESEARCH RECORD

No. 1283

Planning and Administration

Transportation Systems Planning and Applications 1990

A peer-reviewed publication of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1990

Transportation Research Record 1283
Price: $31.00

Subscriber Category
IA planning and administration

Modes
1 highway transportation
2 public transit

Subject Areas
11 administration
12 planning
13 forecasting
15 socioeconomics
16 user needs
53 vehicle characteristics
54 operations and traffic control
55 traffic flow, capacity and measurements

TRB Publications Staff
Director of Publications: Nancy A. Ackerman
Senior Editor: Naomi C. Kassabian
Associate Editor: Alison G. Tobias
Assistant Editors: Luanne Crayton, Kathleen Solomon, Norman Solomon
Graphics Coordinator: Diane L. Ross
Production Coordinator: Karen S. Waugh
Office Manager: Phyllis D. Barber
Production Assistant: Betty L. Hawkins

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

Transportation systems planning and applications, 1990.
p. cm.—(Transportation research record, ISSN 0361-1981 ; no. 1283)
ISBN 0-309-05069-3
1. Transportation and state—United States.
2. Transportation—United States—Planning. I. National Research Council (U.S.). Transportation Research Board.
II. Series: Transportation research record ; 1283.
TE7.H5 no. 1283
[HE206.2]
388 s—dc20 90-29281
[388'.068] CIP

Sponsorship of Transportation Research Record 1283

GROUP 1—TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Chairman: Ronald F. Kirby, Metro Washington Council of Governments

Transportation Data, Economics, and Forecasting Section
Chairman: Edward Weiner, U.S. Department of Transportation

Committee on Transportation Data and Information Systems
Chairman: Arthur B. Sosslau, Comsis Corporation
Secretary: James J. McDonnell, Federal Highway Administration, U.S. Department of Transportation
Suzanne D. Cook, Robert M. Donnelly, Kenneth J. Dueker, Stanely Gee, Brad H. Johnson, Philip D. Patterson, Alan E. Pisarski, Charles L. Purvis, Phillip A. Salopek, Gordon A. Shunk, Howard J. Simkowitz, Richard P. Steinmann, Ronald W. Tweedie, Marcus Ramsay Wigan, George F. Wiggers.

Committee on Transportation Supply Analysis
Chairman: Hani S. Mahmassani, University of Texas at Austin
Yupo Chan, Carlos F. Daganzo, Mark S. Daskin, Randolph Hall, William C. Jordan, Eric J. Miller, Anna Nagurney, Jossef Perl, Earl R. Ruiter, K. Nabil A. Safwat, Yosef Sheffi, Mark A. Turnquist

Task Force on Transportation Planning Applications
Chairman: Gordon A. Shunk, Director of Transportation and Energy, North Central Texas Council of Governments
Daniel Brand, Frederick W. Ducca, Jerry Faris, Christopher R. Fleet, Larry R. Goode, David T. Hartgen, George T. Lathrop, Eugene J. Lessieu, Richard S. Marshment, Richard H. Pratt, George L. Reed, G. Scott Rutherford, Arthur B. Sosslau, Montie G. Wade, W. T. Watterson, Edward Weiner, Debra Ann Whitmore, George V. Wickstrom

Transportation Systems Planning Section
Chairman: John W. Fuller, University of Iowa

Committee on Transportation Planning Needs and Requirements of Small and Medium-Sized Communities
Chairman: Montie G. Wade, Texas Department of Highways and Public Transportation
Secretary: Marion R. Poole, Planning and Research Branch, North Carolina Department of Transportation
Katheryn Briscoe, Charles E. Dare, David J. Forkenbrock, Steven B. Gayle, Joseph W. Guyton, Antoine G. Hobeika, R. Keith Jones, C. Jotin Khisty, Dennis E. Lebo, John A. Lundell, Keith G. Meyer, Michael M. Moravec, George L. Reed, Richard W. Schmid, George E. Schoener, Arthur B. Sosslau, Vergil G. Stover, C. Michael Walton, Pam Ward, Linda Wilshusen, Eugene M. Wilson

Statewide Modal Planning Section
Chairman: Kevin E. Heanue, Federal Highway Administration, U.S. Department of Transportation

Committee on Statewide Multimodal Transportation Planning
Chairman: Michael D. Meyer, Georgia Institute of Technology
Secretary: Philip I. Hazen, Federal Highway Administration
David Preston Albright, Linda Bohlinger, Kirk P. Brown, Darryl E. Durgin, Johns S. Hassell, Jr., Carl A. Hennum, Thomas F. Humphrey, Colin Ian MacGillivray, Douglas S. McLeod, Lance A. Neumann, Neil J. Pedersen, Harry A. Reed, Larry H. Sams, Roger L. Schrantz, Carl N. Swerdloff, James P. Toohey, Richard A. Torbik

GROUP 5—INTERGROUP RESOURCES AND ISSUES
Chairman: William J. Harris Jr., Texas A&M University

Task Force on Expert Systems
Chairman: Roswell A. Harris, University of Louisville
Domenick J. Billera, William Bowlby, Anselmo Osvaldo Braun, Edmond Chin-Ping Chang, Louis F. Cohn, Michael J. Demetsky, Jon D. Fricker, Geoffrey D. Gosling, Jerry J. Hajek, Chris T. Hendrickson, Mike Kaul, Olin W. Mintzer, Satish Mohan, Peter S. Parsonson, Michel Ray, Stephen G. Ritchie, Stephen Edwin Rowe, Alberto J. Santiago, Jerry B. Schneider, John R. Stone, George F. Swetnam, Hossein B. Takallou, James A. Wentworth, Marcus Ramsay Wigan

Task Force on Advanced Vehicle and Highway Technologies
Chairman: Daniel Brand, Charles River Associates Inc.
David E. Boyce, Richard P. Braun, Peter Davies, Shaun S. Devlin, Robert D. Ervin, Robert L. French, Franz K. Gimmler, Thomas A. Griebel, William J. Harris Jr., Jack L. Kay, Peter G. Koltnow, Bill M. McCall, Robert L. Mitchell, Lawrence G. O'Connell, Robert E. Parsons, Ivars Renga, Stephen Edwin Rowe, Lyle Saxton, Joseph L. Schofer, William M. Spreitzer, Burton W. Stephens, Joseph M. Sussman, John Vostrez, David K. Willis

James Scott, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each paper. The organizational units, officers, and members are as of December 31, 1989.

Transportation Research Record 1283

Contents

Foreword

In his paper, Hartgen reports on a study to identify similarities among the financial, infrastructure, and operational structures of the state highway systems and to group states according to similar systems. The purpose of the study was to classify states according to similar systems and problems rather than according to closeness or geography. Results show that, using 19 variables, the state highway systems can be accurately clustered into five distinct groups. Hartgen concludes that although nearest neighbor geographies may be useful for many political liaisons, state highway agencies should also look at broader similarities with states that are not immediate neighbors.

Poole describes how environmental factors have been incorporated in thoroughfare planning studies in North Carolina. Poole also discusses documentation of environmental inventories, the use of environmental data in development of the transportation plan, and the importance of system level environmental studies in right-of-way protection.

DeCorla-Souza and Fleet, in their analysis of the supply and use of urban highways by functional class, suggest that the potential for shifts in travel share from lower order facilities to freeways and expressways will increase as urban travel increases and urban population centers increase in size, and that the greatest potential shifts will be in the suburbs of the largest urban areas. The authors further suggest that a more favorable mix of highway facility types might be achieved by new capacity investments on freeways and expressways.

Huard-Spencer discusses a three-phase proactive planning process used by the Orange County Transit District for developing transportation demand and system management actions at suburban activity centers and implementing them with assistance from the private sector.

Moore and Muller discuss methodology for developing defensible transportation impact fees, such as quantifying the benefits that are derived from new transportation infrastructure, identifying the recipients of the benefits, and calculating the size of an equitable impact fee. Their methodology employs an equity-based approach utilizing net-present-value techniques and addresses the legal criteria of "rational nexus."

Glass et al. discuss the New York State Department of Transportation's long-range planning efforts to cooperate with the Metropolitan Planning Organizations in identifying the data necessary for travel forecasting in addition to its regular monitoring and special data collection activities. The authors emphasize that the state's approach is to coordinate new travel survey activities to maximize their ability and to rely on the appropriate sharing and reuse of existing data when and where possible.

Bruggeman discusses the potential of a transit corridor to connect Dulles Airport with the rest of the Washington Metropolitan Area. The study examined a variety of transit options and recommended an express bus system be implemented using the high level of service provided by the Dulles Access Road, with station sites identified for interim use for express bus operation and potential conversion to a rail or other high technology when operational and financial conditions warrant such action.

Allen states that the presence of traffic control devices is rarely simulated in regional network modeling. The author demonstrates that an alternative approach is available. The normal BPR-based link capacity restraint function built into the MINUTP model can be modified to provide an approximation of the delay associated with traffic control devices. This estimate of delay is developed on the basis of detailed theoretical models of delay and is not usually consistent among control device types. The method is relatively simple and can be implemented within the existing planning software systems.

Khisty and Rahi evaluate three simplified travel demand models suitable for small urban areas that make use of routinely collected ground counts. Suitable models are discussed and then evaluated by applying them in a common setting to the city of Pullman, Washington. The authors conclude that the models described in this paper will help small urban areas to forecast travel demands using routinely collected traffic ground counts and socioeconomic data.

Faghri presents an analysis of new expert systems and existing simulation, evaluation, and optimization models in traffic and transportation. Different taxonomies and configurations for combining expert systems and other computerized models are shown and advantages and disadvantages of each configuration are stated. An application problem in transportation dealing with disaster evacuation and response planning is presented.

Safwat and El-Araby introduce an expert system to assist practicing planners and engineers in selecting microcomputer packages for the Network-Based Transportation Planning Software Selection Advisor (NETSSA). NETSSA is highly interactive and user friendly. According to the authors, its current knowledge base includes nine software packages and it can be expanded to include additional software packages.

Crainic et al. describe the Strategic Transportation Analysis model (STAN) as an interactive graphic system for national or regional strategic planning of multimode, multiproduct transportation systems. The authors describe the general concepts underlying the system, the data base structure, the assignment procedure, and the results. Data and results from an actual application are used to illustrate these concepts.

Baaj and Mahmassani describe TRUST as a program to analyze and evaluate a given set of bus transit routes and associated frequencies in terms of several descriptions, including measures with user costs and service quality and operation resources. The application of the program to the transit network of the city of Austin, Tex., urban is presented, illustrating the program capabilities and computational performance.

Crainic and Mondou explore a number of issues related to the development of systems for the tactical and strategic planning of intercity freight transportation and distribution systems, with emphasis on the specification of interactive graphic interfaces. The authors analyze the requirements and possible ways of graphically representing the various components of the transportation system and present a number of solutions. The discussion is illustrated with computer images obtained from an actual rail application.

Shladover reviews the history of roadway automation in light of current transportation needs and technological progress. The main body of the paper outlines the technical questions that need to be answered in order to make roadway automation a reality, based on nine locations with an automated roadway system discussed in the paper.

Prashker presents a transit assignment algorithm that takes into account the actual capacity of transit lines and assigns trips to more than a single path when the shortest path reaches its capacity. This procedure produces a practical Multipath Capacity Limited Transit Assignment (McLAT). The procedure was implemented on an IBM mainframe computer using UMTA's standard UTPS package with the addition of only one FORTRAN program.

Hendrickson et al. describe network representatives of performance analysis parameters and relationships. The intersection performance analysis procedures implicit in the *Highway Capacity Manual* are formalized as an example of network representation. These representations make use of semantic net and frame representations originally developed in the field of artificial intelligence.

Similarities Among the State Highway Systems

DAVID T. HARTGEN

Results are reported for a recent study to identify similarities among the financial, infrastructure, and operational structures of the state highway systems, and to group states according to these similarities. The purpose of the study was to classify states according to similar systems and problems, rather than according to closeness or geography. An additional goal was to determine the degree to which states that are widely separated may have similar problems, thereby suggesting political alliances for various issues. A structure for classifying state highway agency financing is first proposed and described. Then, a data base containing 61 data items for each state is developed using cross-sectional data from 1984, 1985, and 1986. The source of these data items is primarily *Highway Statistics*, supported by organization studies in the transportation literature. Data consist of measures of size, road and bridge condition, taxes, revenues, disbursements, and agency characteristics. After a review of descriptive rankings of states on a number of key variables, the data are then factor-analyzed using varimax rotation. This procedure yields a smaller number of data items found to most clearly separate these states. States are then clustered according to their ratings on these variables. Clusters based on the full data set containing 61 variables are compared with clusters based on reduced data sets containing 19, 13, and 7 variables. Results show that with 19 variables, the state highway systems can be quite accurately clustered into five distinct groups as follows: (1) Alaska; (2) the far western states, along with Michigan, Hawaii, Florida, West Virginia, and northern New England; (3) southern New England and seaboard Middle-Atlantic; (4) midwestern and southern states, including Washington; and (5) very large states (California, Texas, New York, and Pennsylvania). Within each of these groups, a number of subgroupings identify strong regional and content coalitions. The group structure appears to be reasonably robust under a variety of assumptions, with regional subgroupings particularly strong for New England, the southeastern seaboard, the far west, and big states. It is concluded that while nearest neighbor geographies may be useful for many political liaisons, state highway agencies should also look at broader similarities with states that are not immediate or nearest neighbors.

It is well understood that the problems associated with planning, financing, building, and maintaining state highway systems are extensive and complex. Each of the state highway agencies is faced with many problems involving each of these activities. These problems are generally similar from state to state, but their details vary immensely because each state is unique in its location, context, and capabilities. Thus, under the general guidance of federal law, state policy, and goal structure, each of the states operates more or less independently in managing its state highway system.

In spite of these differences, there also are many similarities in the highway management process: each state operates within the general guidance of federal law; each state has organized and maintains a system of state-owned highways and an agency responsible for them; each state has developed and relies upon its own funding sources for revenues necessary to maintain its system. In theory, therefore, while many differences in procedures and problems separate the states, we should also recognize that many similarities bind them as they deal with these complex issues.

Department of Geography and Earth Sciences, University of North Carolina at Charlotte, Charlotte, N.C. 28223.

A fundamental hypothesis worthy of testing is whether the state highway agencies have problems and solutions similar to their closest neighbors, or whether the similarities extend beyond the borders of immediate states to those further away. One might expect that geographical, historical, political, climatic, and development conditions would lead to the greatest similarities being shared by the states closest to each other. Conversely, as the Interstate System has fostered a larger proportion of regional traffic, and as state highway agencies share technology and procedures through technical as well as political processes, one might also expect that these geographically based similarities are declining over time in favor of similarities of problem and solution. If this is so, then state liaisons based solely upon geography are likely to be declining relative to those based upon system similarity. In a nutshell, the purpose of this paper is to review the extent to which geographical or structural similarities among the state highway agencies can be utilized to identify such coalitions. Our purpose is to determine whether the nature of these coalitions is primarily geographic or structural.

THEORY AND METHOD

While each of the state highway agencies is responsible for planning, constructing, and maintaining its highway system, in addition to other duties, each state is largely independent in developing the organizational and fiscal structure necessary to achieve those goals. In addition, the states may be expected to vary widely along the key dimensions relating to these activities. These key dimensions are as follows:

- geography, including soil conditions, climate, freeze-thaw cycles, and weather;
- size and extent, including size of the highway system for which the agency is responsible, measures of area and population size for the state itself, and measures of traffic;
- system condition, specifically congestion, bridge and highway condition, and performance measures;
- sources of revenue, particularly tax rates for different fuels, vehicles, licenses, and other sources;

- agency characteristics, particularly measures of agency size, focus, and structure;
- expenditures, particularly the magnitude and density of expenditures per mile; and
- network access and measures of network accessibility relative to the area and population size.

Figure 1 shows a simple model that describes how most states deal with the problem of financing highway systems (*1–3*). Within the context of a general goal direction and a geographic environment, the state sets tax rates to generate revenue needed to deal with the highway problems or conditions. Expenditures then result in highway improvements that, when compared with goals, allow for revisions of revenue streams. Within legislative, managerial, and operational environments, the process goes along more or less continuously. That is, periodic readings of highway conditions and needs are used to identify requirements, which in turn lead to revenue generation through taxing actions. Over time, system improvements result in changes in condition, which modify needs requirements.

The structure described in Figure 1 can be represented as a series of simultaneous equations in which data for a number of years is used as the basis for model development. For instance, let:

g = geographic attributes,
G_y = goals vector, year y,
T_y = tax rates,
S_y = size measures,
Rev_y = revenue,
$Needs_y$ = needs,
$Cond_y$ = conditions,
$Disb_y$ = disbursements,
$Maint_y$ = maintenance expenditures, and
Cap_y = capital expenditures.

Then, a system of simultaneous equations can be developed to describe this process. For instance, the relationship in Figure 1 can be expressed as:

$$T_y = f(G_{y-1}, N_{y-1}),$$
$$Rev_y = f(T_{y-1}, Size_y, Disb_{y-1}),$$
$$Cond_y = f(Size_y, Cap_{y-1}, Maint_{y-1}),$$
$$Needs_y = f(Cond_{y-1}),$$
$$Disb_y = f(Rev_{y-1}, Cond_{y-1}),$$
$$Cap_y = f(Disb_y), \text{ and}$$
$$Goals_y = f(Cap_y, Maint_y, Needs_y).$$

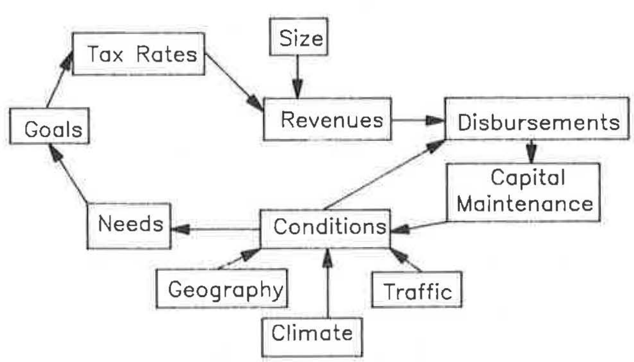


FIGURE 1 Simplified state highway financing model.

The specific functional form of these models would be determined by two-stage least squares calibration. Generally, linear models are used.

Models of this structure are particularly difficult to develop and calibrate because they require data on systems expenditures and performance for a number of time periods. Recognizing the limitations of a simpler design, we have chosen to use primarily cross-sectional data for model estimation purposes.

Numerous studies contain comparative cross-sectional (and occasionally time series) data on highway financing. The most extensive and comprehensive is *Highway Statistics*, published since the 1920s. This series contains aggregate data, by state, on many aspects of financing, road conditions, and traffic. Its primary shortcomings are in agency employment, which it does not contain, and in "quality" measures of system performance (such measures as congestion and pavement condition were only recently added) (*4*). These reports can be analyzed over time, but the Federal Highway Administration (FHWA) did not (until recently) keep its data in that fashion, so comparisons are difficult. The summary document, *Highway Statistics: Summary to 1985*, contains five- and ten-year trend statistics for key indicators, by state; and a few data items are kept annually (*5*). An earlier document contains state-level trend data for the years 1957–1975, by modal programs (*6*). The series *Highway Taxes and Fees* (1981, 1984, 1987) describes procedures needed by the states to collect and disburse taxes; a companion document, *Financing Federal Aid Highways*, describes the federal process (*7,8*). Other documents contain data on bridge statistics and agency size (*9,10*).

Reports prepared for fiscal reviews by individual state highway departments also typically contain general comparisons with other states deemed "similar," but generally no detailed comparative analysis of states are undertaken. Apparently, comparative reviews of state highway funding practices are not as common a subject of investigation as they once were. Of 24 studies reviewed for this paper, most were conducted before 1970. An NCHRP Project (1970) reviewed state level budgeting practices, relying on states' reports from the field. In one of the very few comparative analytical studies of state highway financing, Phelps (*1*) developed a stock-adjustment model based on economic theory to describe the timing of state and local highway capital outlays; data were time series for 1951–66, primarily from *Highway Statistics*. An even earlier study (*3*) compared the collection and distribution practices of motor vehicle revenues in 34 states, and concluded that the primary uses (71 percent) were for state highway construction and maintenance. This detailed paper, written in 1927, contains 1925 data, perhaps commenting indirectly on the speed of today's "modern" data delivery systems. Rao (*2*) compared the funding practices and procedures of the states, focusing on bonding and other revenue-building approaches, but did not analyze the similarities of states statistically. None of the documents reviewed used such tools as factor analysis or cluster analysis. The author is not aware of any such applications to highway data, although at least one study (*11*) attempted a path analysis of transit property performance statistics. In general, the subject is unresearched.

One way to determine the similarities among state financing and organization structure is to calibrate a model such as the above, using time series data for each state. A comparison of the coefficients of the models for each state would then lead to rejection or acceptance of the null hypothesis that the states are homogeneous in their financing structures. However, such a methodology would be extremely consuming in both time and effort. Therefore, a short approach relying primarily on cross-sectional data is used instead. This approach relies heavily on the cross-sectional data available for each of the state highway agencies, primarily reported in *Highway Statistics*, supplemented by data on transportation agency employment (*9*) and data on bridge conditions (*10*).

Data representing primarily 1984, but also some 1980, 1983, 1985, and 1986 data, were consolidated from these sources into a state level data base using spreadsheet (EXCEL) and SAS environments. The specific variables used in the analysis consist of 61 data items drawn from the above sources and grouped into several major classifications. These data items, extracted in Table 1, contain information on the following dimensions: (1) system size and state size measures, (2) measures of condition, (3) fiscal sources and revenues, (4) expenditures, (5) agency size, (6) expenditure efficiency.

TABLE 1 MEASURES OF STATE HIGHWAY AGENCY STRUCTURE

Category	Subclass	Variable	Source	Name
Size	Mileage	• Rural miles under State control	(1)	RUSC
		• Urban miles under State control	(1)	UUSC
		• Total miles under State control	(1)	TOTUSC
		• Total state mileage	(1)	TRUM
		• Rural Interstate mileage	(4)	RITOT
		• Rural principal arterial mileage	(4)	RPATOT
		• Urban Interstate miles	(4)	UVCTOT
		• Urban principal arterial mileage	(4)	UPATOT
	Traffic	• Annual VMT (million)	(2)	VMT
	Population	• Resident population, 1983	(8)	RESPOP
	Drivers	• Licensed drivers/1000 pop.,1983	(8)	LDRP
	Vehicles	• Registered motor vehicles/ 1000 pop., 1983	(8)	RVMPOP
		• Licensed drivers/reg. motor veh.	(8)	LDRMV
	Bridges	• Federal aid number of bridges	(9)	FANOBR
		• Non-Fed. aid number of bridges	(9)	NFANOBR
		• Total number of bridges	(9)	NOBRIDGE
Condition	Congestion	• Miles of rural Interstate with V/C.7-.95	(4)	RVC1
		• Miles of rural Interstate with V/C >.95	(4)	RVC2
		• Miles of rural principal arterial with V/C.7-.95	(4)	RPAVC1
		• Miles of rural principal arterial with VC>95	(4)	RPAVC2
		• Miles of Urban Interstate with V/C.7-.95	(4)	UVC1
		• Miles of Urban Interstate with V/C>.95	(4)	UVC2
		• Miles of Urban Principal Arterial with V/C .77-95	(4)	UPAVC1
		• Miles of Urban Principal Arterial with V/C>.95	(4)	UPAVC2
		• Interstate + Arterial Networks % of roads with V/C>.77		PVC
	Pavement	• Rural Interstate miles PSR<2.5	(5)	RIPSRMI
		• Urban Interstate miles PSR<2.5	(5)	RIPSRMI
		• Interstate % PSR<2.5	(5)	PSR
	Bridges	• Number of FA deficient bridges	(9)	FADEF
		• Number of non-FA deficient bridges	(9)	NFADEF
		• Percent of deficient bridges (FA & NFA)	(9)	DEFBR
		• Percent of FA deficient bridges		PDFABR
		• Percent of non-FA deficient bridges		PDNFABR
Taxes	Gasoline	• Gasoline tax rate cents/gal	(3)	GASTAX
	Diesel	• Diesel tax rate	(3)	DIESLTAX
	Others	• Special fuels rate	(3)	OTHERTAX
Revenues	Fuels	• Receipts for State hwy.-motor fuels	(6)	FUELTAX
	Vehicles	• Receipts for State hwy.-vehicle/carriers	(6)	MVMCTAX
	Total	• Total receipts		TOTREV

TABLE 1 (*continued on next page*)

TABLE 1 (*continued*)

Category	Subclass	Variable	Source	Name
Disbursements	Capital	• Capital outlay for roads & bridges	(7)	RBCORB
	Maintenance	• Outlay for maintenance & traffic services	(7)	SERVMTS
	Administrative	• Outlay for administrative & misc.	(7)	MISCADM
	Total	• Total disbursements	(7)	TOTDISB
Employees	Number	• No. of engineers in management, 1984	(10)	MGMTENG
		• Total no. of engineers, 1984	(10)	TOTENG
		• Total no. of planners	(10)	PLANRS
		• Total no. of computer employees	(10)	CMPTR
		• Total employees, 1980	(10)	TOTEMP80
		• Total employees, 1984	(10)	TOTEMP84
		• % change in employees 1980-84		CHGEMP
	Salaries	• District Engineer's salary, 1984, $	(10)	DE
		• Project Engineer's salary, 1984, $		PE
		• CE graduate's salary, 1984, $		BSCE
Rates				
	Staff density	• Miles of road under State control per employee		MIPEREMP
	Expenditure density	• Capital outlay for roads and bridges per mile		CORBPUSC
		• Maintenance & traffic services expenses per mile		MTSPUSC
		• Administrative expenses per mile		ADMPUSC
		• Total disbursements per mile		DISBPUSC
		• Total disbursements per population		DISPOP
		• Total disbursements per motor vehicle		DISBPRMV
	Expenditure efficiency	• Capital outlay for road & bridges per employee		CORPEMP
		• Maintenance & traffic services expenses per employee		MTSPEMP
		• Administrative expenses per employee		ADMPEMP
	Engineering focus	• Engineers per employee		ENGPEMP
		• Mgt. engineers per employee		MGTEPEMP

Notes
(1) Highway Statistics, 1984, Table HM10
(2) Highway Statistics, 1984, Table VM2
(3) Highway Statistics, 1984, Table MF1
(4) Highway Statistics, 1984, Table HM61
(5) Highway Statistics, 1984, Table HM62
(6) Highway Statistics, 1984, Table SF-3
(7) Highway Statistics, 1984, Table SF-4
(8) Highway Statistics, 1984, Table SF-17
(9) Sixth Annual Bridge Report, 1984 data, Table 10
(10) TRB Special Report No. 207, 1984 data, Table 2-16.

The methodology for analyzing this data is relatively straightforward. It consists of three basic steps:

1. Descriptive statistics were prepared for key measures of state systems. These ranked data streams were used to develop a preliminary overview of relative state groupings.
2. Using factor analysis techniques and varimax rotation, the descriptive statistics were factor analyzed to develop a reduced set of variables representative of each factor group.
3. Using these variables, the states were clustered along similarity lines using Ward "nearest neighbor" clustering algorithms. Reduced variable structures were identified for 7, 13, 19, and a total of 61 variables. The SAS procedure CLUSTER was used for this approach. Separate cluster analyses were then prepared for large and small states, to determine whether state size is critical in the grouping process. Characteristics of each cluster were then prepared.

RESULTS

Size Measures

Table 2 shows various measures of size for these states. California, New York, and Texas lead the list in population, vehicles, drivers, and VMT, with Pennsylvania, Illinois, Ohio, and Florida close behind. With respect to highway system extent, Georgia, North Carolina, Texas, and Virginia have the largest state highway systems, all over 50,000 miles.

Measures of Condition

Network condition can be measured according to a number of criteria. Data readily available include the percent of mileage with volume-to-capacity ratios greater than .77, the per-

TABLE 2 MEASURES OF SIZE

OBS	STATE	POP	VEHICLES	DRIVERS	VMT	MILES	BRIDGES
1	CA	25044.6	17781.7	16714.8	196537	18213	22260
2	NY	17713.1	8421.0	9599.9	87268	16394	17419
3	TX	15758.2	11734.8	11500.1	137737	71448	44036
4	PA	11896.0	6843.6	7459.6	74297	44000	21719
5	IL	11484.8	7507.1	6981.6	69910	17609	25058
6	OH	10725.9	7778.3	7389.4	74895	20221	28969
7	FL	10686.9	8812.0	8371.4	85475	11536	10005
8	MI	9109.4	6300.6	6363.7	63470	9510	10401
9	NJ	7473.7	4939.4	5433.3	52312	3167	5732
10	NC	6094.6	4615.0	3968.9	48182	76920	15712
11	MA	5786.7	3835.7	3682.3	38537	3613	4776
12	GA	5713.3	4206.5	3743.8	50486	86655	14191
13	VA	5535.7	3890.0	3695.5	44527	54782	12493
14	IN	5506.3	3837.7	3530.7	41074	11344	17682
15	MO	4949.9	3433.4	3330.4	38535	32317	23726
16	WI	4752.4	3212.9	3084.4	35367	12519	12822
17	TN	4672.2	3548.0	2944.9	36523	11171	18110
18	LA	4418.7	2872.1	2757.3	31588	16419	14215
19	MD	4304.5	3015.2	2804.1	31702	5239	4059
20	WA	4304.3	3339.2	2871.7	34248	18525	6796
21	MN	4150.4	3288.6	2367.8	31826	13443	12906
22	AL	3953.2	3147.0	2391.7	32961	11688	15513
23	KY	3710.9	2621.8	2202.3	27951	25120	12484
24	OK	3304.2	2775.6	2192.7	30981	13056	22021
25	SC	3263.6	2052.9	2011.8	25971	40338	8890
26	CO	3151.2	2641.4	2218.8	24588	9301	7147
27	CT	3131.4	2308.8	2262.6	21076	3896	3724
28	AZ	2955.5	2294.8	2180.1	20613	5786	5140
29	IA	2910.6	2474.9	1930.4	20497	10160	26112
30	OR	2655.2	2117.2	1905.5	20943	10856	6869
31	MS	2587.6	1561.3	1811.1	18442	10324	16728
32	KS	2429.0	2050.2	1681.1	18717	10692	25656
33	AR	2322.2	1444.9	1647.2	16621	16111	14336
34	WV	1967.5	1295.2	1411.8	12671	31356	6608
35	UT	1622.5	1075.6	925.0	11661	5584	2371
36	NE	1594.7	1232.7	1097.1	11968	10385	16197
37	NM	1401.2	1238.3	767.7	12432	12406	3420
38	ME	1148.0	766.9	774.6	9345	7999	2592
39	HI	1026.7	617.0	573.8	6505	1059	1038
40	ID	987.6	877.9	649.6	7768	5085	3622
41	NH	958.7	801.8	697.6	7294	4398	2551
42	RI	952.7	598.7	604.7	5300	1952	689
43	NV	890.1	730.6	679.4	7332	5183	1011
44	MT	816.2	828.3	488.7	7386	7830	4777
45	SD	698.6	629.3	484.5	6401	7896	7061
46	ND	689.2	675.5	439.1	5377	7304	5475
47	DE	605-5	427.0	431.2	5138	4616	691
48	VT	523.6	367.6	360.3	4403	2787	2654
49	WY	514.3	502.4	391.9	5127	6622	2851
50	AK	478.6	350.6	287.5	3589	11426	835
51	DC	.	.	.	3214	1102	236

cent of roads in poor condition, and the percentage of deficient bridges. Table 3 shows these statistics, and indicates that the District of Columbia, Missouri, Connecticut, Kansas, and West Virginia lead these lists. The District is, of course, particularly high on congestion.

Fiscal Sources

A simple measure of state gasoline tax rates, gas tax and diesel tax, shows that Washington has the highest gasoline and diesel tax combination, followed by Minnesota, Louisiana, Montana, and Wisconsin (Table 4).

Expenditures

Table 5 shows that with respect to disbursements, a number of southern states, particularly North Carolina, South Carolina, Georgia, and Virginia, have the highest mileage per dollar expenditures, while several eastern states, including Maryland, Connecticut, Massachusetts, and New Jersey, have the lowest rates.

Overall Status

A useful measure of overall status is the "misery index." This is constructed as a combination of condition (percentage of deficient bridges, deficient pavement, and congested roads), gasoline and diesel taxes, and miles of road per dollar available. High ratings on the "misery index" mean that a state has many problems (per mile) and not much money (per mile) to deal with them. Table 5 also rates states according to the "misery index" and shows that a group of southern states scored particularly high on this index, primarily because of their very high mileage per dollar expended. These states include North Carolina, South Carolina, Georgia, West Vir-

TABLE 3 STATES RANKED BY ROAD CONDITION

OBS	STATE	OVERALL CONDITION	PERCENT V/C	POOR CONDITION	PERCENT DEFICIENT BRIDGES
1	DC	1.41	0.88	0.33	0.19
2	MO	1.01	0.06	0.27	0.68
3	CT	0.91	0.31	0.00	0.60
4	KS	0.86	0.01	0.28	0.56
5	WV	0.84	0.12	0.14	0.58
6	MD	0.82	0.31	0.15	0.36
7	MT	0.82	0.00	0.10	0.70
8	MS	0.80	0.01	0.17	0.61
9	NY	0.79	0.12	0.00	0.66
10	NC	0.76	0.07	0.03	0.65
11	OK	0.75	0.04	0.15	0.54
12	RI	0.73	0.21	0.31	0.20
13	NJ	0.72	0.37	0.01	0.34
14	WI	0.68	0.02	0.16	0.49
15	MI	0.68	0.10	0.23	0.33
16	NH	0.67	0.14	0.00	0.52
17	CO	0.67	0.08	0.16	0.42
18	LA	0.66	0.09	0.04	0.52
19	HI	0.65	0.38	0.00	0.27
20	PA	0.65	0.19	0.10	0.34
21	AR	0.63	0.01	0.01	0.60
22	NE	0.62	0.00	0.00	0.62
23	CA	0.62	0.26	0.08	0.27
24	IN	0.62	0.02	0.00	0.58
25	VT	0.61	0.04	0.04	0.52
26	TN	0.60	0.07	0.01	0.51
27	KY	0.59	0.12	0.05	0.41
28	MA	0.58	0.25	0.00	0.33
29	AL	0.58	0.03	0.00	0.54
30	ND	0.58	0.00	0.00	0.57
31	IA	0.57	0.00	0.02	0.54
32	GA	0.56	0.10	0.00	0.45
33	TX	0.54	0.10	0.04	0.40
34	AK	0.53	0.04	0.29	0.19
35	ID	0.53	0.00	0.18	0.34
36	WA	0.51	0.24	0.00	0.26
37	NM	0.49	0.00	0.28	0.20
38	VA	0.49	0.09	0.05	0.34
39	IL	0.47	0.06	0.07	0.32
40	SD	0.46	0.01	0.00	0.45
41	OH	0.40	0.10	0.08	0.21
42	MN	0.39	0.04	0.03	0.32
43	ME	0.39	0.03	0.04	0.31
44	DE	0.37	0.21	0.00	0.16
45	FL	0.32	0.03	0.00	0.29
46	WY	0.30	0.00	0.03	0.26
47	SC	0.30	0.04	0.01	0.23
48	OR	0.29	0.06	0.00	0.22
49	UT	0.28	0.07	0.02	0.18
50	NV	0.19	0.01	0.01	0.16
51	AZ	0.14	0.02	0.04	0.07

*Condition = Percent v/c > .77 + Percent PSR < 2.0 + Percent Deficient Bridges

ginia, Missouri, and Virginia. At the bottom of the list, states with good conditions, low tax rates, and high expenditures per mile of system, include Nevada, Utah, Wyoming, Arizona, and Florida.

Factor Structure

Tables 6 and 7 show results of the factor analysis of these 61 variables. Results show that approximately six factors are sufficient to describe the data structure. These factors are described as follows: (1) size and congestion, (2) expenditure per mile of system, (3) total mileage and fuel revenues, (4) number of bridges, (5) deficient bridges, and (6) tax rates, pavement conditions, and expenditures. Consequently, the model structure suggested here assumes that size, tax rates, bridge deficiencies, and expenditures sufficiently distinguish between the state fiscal patterns.

Cluster Analysis

To undertake the cluster analysis, variables were selected from each of the factor groups described above in decreasing numbers. The full cluster analysis contained all 61 data items, while reduced cluster analyses contained 19, 12, and ultimately 7 data items, respectively. The specific items selected for each cluster analysis were chosen based on factor loadings, internal independence, and representativeness for each factor group. Table 6 shows the specific variables used for each cluster run. States were grouped into large and small classifications on the basis of the total number of miles under state

TABLE 4 STATES RANKED BY TOTAL FUEL TAX RATE (1984)

OBS	STATE	TAXES	GAS TAX	DIESEL TAX	OTHER TAX
1	WA	36.0	18.0	18.0	.
2	MN	34.0	17.0	17.0	.
3	LA	32.0	16.0	16.0	.
4	MT	32.0	15.0	17.0	.
5	WI	32.0	16.0	16.0	.
6	DC	31.0	15.5	15.5	.
7	CT	30.0	15.0	15.0	.
8	MI	30.0	15.0	15.0	.
9	NE	29.4	14.7	14.7	.
10	ID	29.0	14.5	14.5	.
11	IA	28.5	13.0	15.5	13
12	ME	28.0	14.0	14.0	.
13	NH	28.0	14.0	14.0	.
14	UT	28.0	14.0	14.0	.
15	AL	27.0	13.0	14.0	.
16	MD	27.0	13.5	13.5	.
17	VT	27.0	13.0	14.0	.
18	IL	26.5	12.0	14.5	12
19	AZ	26.0	13.0	13.0	.
20	ND	26.0	13.0	13.0	.
21	RI	26.0	13.0	13.0	.
22	SC	26.0	13.0	13.0	.
23	SD	26.0	13.0	13.0	11.0
24	CO	25.0	12.0	13.0	.
25	NC	24.0	12.0	12.0	.
26	NV	24.0	12.0	12.0	.
27	OH	24.0	12.0	12.0	.
28	PA	24.0	12.0	12.0	.
29	TN	23.0	10.0	13.0	9.0
30	IN	22.2	11.1	11.1	.
31	DE	22.0	11.0	11.0	.
32	KS	22.0	11.0	11.0	10.0
33	MA	22.0	11.0	11.0	6.9
34	NM	22.0	11.0	11.0	.
35	VA	22.0	11.0	11.0	.
36	WV	21.0	10.5	10.5	.
37	AR	20.0	9.5	10.5	7.5
38	KY	20.0	10.0	10.0	.
39	TX	20.0	10.0	10.0	.
40	FL	19.4	9.7	9.7	.
41	MS	19.0	9.0	10.0	8.0
42	CA	18.0	9.0	9.0	6.0
43	NY	18	8.0	10.0	8
44	OK	18	9.0	9.0	.
45	OR	18	9.0	9.0	.
46	HI	17	8.5	8.5	6
47	AK	16	8.0	8.0	.
48	NJ	16	8.0	8.0	4
49	GA	15	7.5	7.5	.
50	MO	14	7.0	7.0	.
51	WY	8	8.0	0.0	.

control, plus the number of federal aid bridges, with 15,000 (miles plus bridges) being the dividing line between large and small states. This results in 29 states being included in the "large" category and 22 states in the "small" category. Measures of system size were used, rather than area, to better relate the data to fiscal expenditures.

A series of cluster analyses was then run on each of these structures, using Ward's minimum distance cluster analysis. The Ward's model is a straightforward minimum distance model of the following form:

$$D_{KL} = [X_K - X_L]^2/(1/N_K + 1/N_L)$$

where

X = sample mean vector, and
N = number of observations.

In this method, the distance between two clusters is the ANOVA sum of squares between the two clusters, added up over all variables. As observations are grouped into clusters, within-cluster sum of squares is minimized over all partitions obtainable by merging the two clusters from the previous generation. The sums of squares are easier to interpret when divided by the total sum of squares to give proportions of variance. As noted in the SAS procedures manual (*12*), the method tends to join clusters with a small number of observations, and is strongly biased toward producing clusters with roughly the same number of observations. It is also quite sensitive to outliers.

The resulting cluster structure can be viewed as a tree, in which clustering occurs at higher levels as one moves up the tree. Corresponding to these different cluster levels is an associated number of clusters. The choice of a particular number of clusters with which to identify the result is essentially arbi-

TABLE 5 STATES RANKED BY "MISERY INDEX"

OBS	STATE	MISERY	CONDITION	TAXES	MONEY
1	NC	3.26	0.76	0.24	2.25
2	SC	2.46	0.30	0.26	1.90
3	GA	2.44	0.56	0.15	1.73
4	WV	2.21	0.84	0.21	1.15
5	MO	2.16	1.01	0.14	1.01
6	VA	2.06	0.49	0.22	1.35
7	AR	1.88	0.63	0.20	1.04
8	NH	1.67	0.67	0.28	0.72
9	ME	1.65	0.39	0.28	0.97
10	MT	1.64	0.82	0.32	0.50
11	ND	1.61	0.58	0.26	0.77
12	NE	1.59	0.62	0.29	0.66
13	VT	1.59	0.61	0.27	0.70
14	KS	1.57	0.86	0.22	0.49
15	MS	1.52	0.80	0.19	0.53
16	OK	1.48	0.75	0.18	0.55
17	WI	1.46	0.68	0.32	0.45
18	SD	1.46	0.46	0.26	0.73
19	PA	1.45	0.65	0.24	0.56
20	TX	1.42	0.54	0.20	0.68
22	KY	1.40	0.59	0.20	0.60
23	CT	1.39	0.91	0.30	0.17
24	AK	1.38	0.53	0.16	0.68
25	CO	1.35	0.67	0.25	0.43
26	WA	1.32	0.51	0.36	0.44
27	LA	1.31	0.66	0.32	0.32
28	NM	1.27	0.49	0.22	0.55
29	IN	1.25	0.62	0.22	0.41
30	MI	1.24	0.68	0.30	0.26
31	RI	1.23	0.73	0.26	0.24
32	MD	1.23	0.82	0.27	0.13
33	IA	1.21	0.57	0.28	0.35
34	TN	1.21	0.60	0.23	0.37
35	AL	1.20	0.58	0.27	0.35
36	NY	1.17	0.79	0.18	0.20
37	MN	1.07	0.39	0.34	0.34
38	OR	1.06	0.29	0.18	0.59
39	DE	1.05	0.37	0.22	0.45
40	OH	1.02	0.40	0.24	0.37
41	CA	1.02	0.62	0.18	0.21
42	MA	0.97	0.58	0.22	0.16
43	NJ	0.96	0.72	0.16	0.08
44	HI	0.95	0.65	0.17	0.12
45	IL	0.92	0.47	0.26	0.18
46	NV	0.87	0.19	0.24	0.44
47	UT	0.84	0.28	0.28	0.28
48	WY	0.73	0.30	0.08	0.34
49	FL	0.70	0.32	0.19	0.17
50	AZ	0.64	0.14	0.26	0.23
51	DC		1.41	0.31	

Money = 10,000 x Miles of Road/Disbursements
Misery = Condition + Taxes + Money

trary, so for purposes of this analysis we selected approximate breakpoints that clearly identify four, five, or six relatively clean clusters. The selection of this breakpoint was guided by technical data on the clusters produced by the cluster variance program. In particular, we looked for local maxima for Ward's pseudo T-squared, a measure of the strength of the clusters, somewhere in the five-six cluster range.

Figures 2 through 4 show the structure of these clusters by state. A number of general observations are apparent from these results:

- The cluster analysis was in general, not particularly effective. The best cluster structure identified was able to account for only about 25 percent of the variance in the state fiscal data sets. In fact, at the particular cluster level shown (that is, four to six groups) generally less than 10 percent of total variance was accounted for by the cluster. In other words, while these clusters partition the states into separate groups, the within-group variance remains at 90 percent of the total. Clearly, assessments of similarity based on clustered data such as this should recognize that the variations in the data are not producing strongly similar groups.
- Partitioning the data by state size was found to be largely duplicative of the clustering effort itself, since the clusters based on all states tended to show groupings similar to those based on state size. Therefore, further analyses based on size alone are not reported, that variable being subsumed by the full cluster model.
- The number of variables used was a particularly important item in identifying the overall cluster picture. Clusters based on a small number of variables, particularly 7 and 13, showed less strength and greater variation in results. On the

TABLE 6 FACTOR ANALYSIS OF SHA DATA

	Variables	Loadings	Used in Cluster Scheme 19 Vars	12 Vars	7 Vars
Factor 1					
"Congestion, system size, agency size"	UPAVC1	.95	x	x	x
	UPAVC2	.95	x		
	UVC2	.90	x		
	UPATOT	.91			
	UVC1	.85	x	x	
	VMT	.87			
	COMPTR	.89			
	TOTENG	.88			
	RESPOP	.83			
	PLNRS	.87			
	UVCTOT	.80			
	UPSRMI	.80			
Factor 2					
"Expenditure/mile"	DISBPUSC	.94	x	x	x
	CORBPUSC	.93	x	x	
	ADMPUSC	.92			
	MTSPUSC	.90	x		
	PVC	.67			
Factor 3					
"Urban mileage, rural congestion, maintenance expenses, fuel tax revenues"	UUSC	.78			
	RVC2	.77	x		
	SERVMTS	.73	x	x	
	FUELTAX	.72	x	x	
	TOTUSC	.63	x	x	x
	RUSC	.62			
Factor 4					
"Bridges"	NOBRIDGE	.72	x	x	x
	NFANOBR	.72			
	FANOBR	.64			
	TRUM	.72			
	RITOT	.57			
	RPATOT	.59			
Factor 5					
"Deficient Bridges"	DEFBR	.86	x	x	x
	PDFNABR	.80			
	PDFABR	.76			
	FADEF	.63	x		
	NFADEF	.62	x		
Factor 6					
"Tax rates, disbursement rates"	DISBPRMV	.67	x	x	
	DIESLTAX	.67			
	DISBPOP	.65			
	RIPSRMI	.67	x	x	x
	GASTAX	.63	x	x	x
	PSR	.54	x		

TABLE 7 VARIANCE EXPLAINED BY FACTOR MODEL

Factor	Eigenvalues	Percent	Cumulative %
1	23.89	38	38
2	9.01	14	52
3	4.48	7	59
4	4.17	6.6	66
5	3.25	5	71
6	2.48	4	75

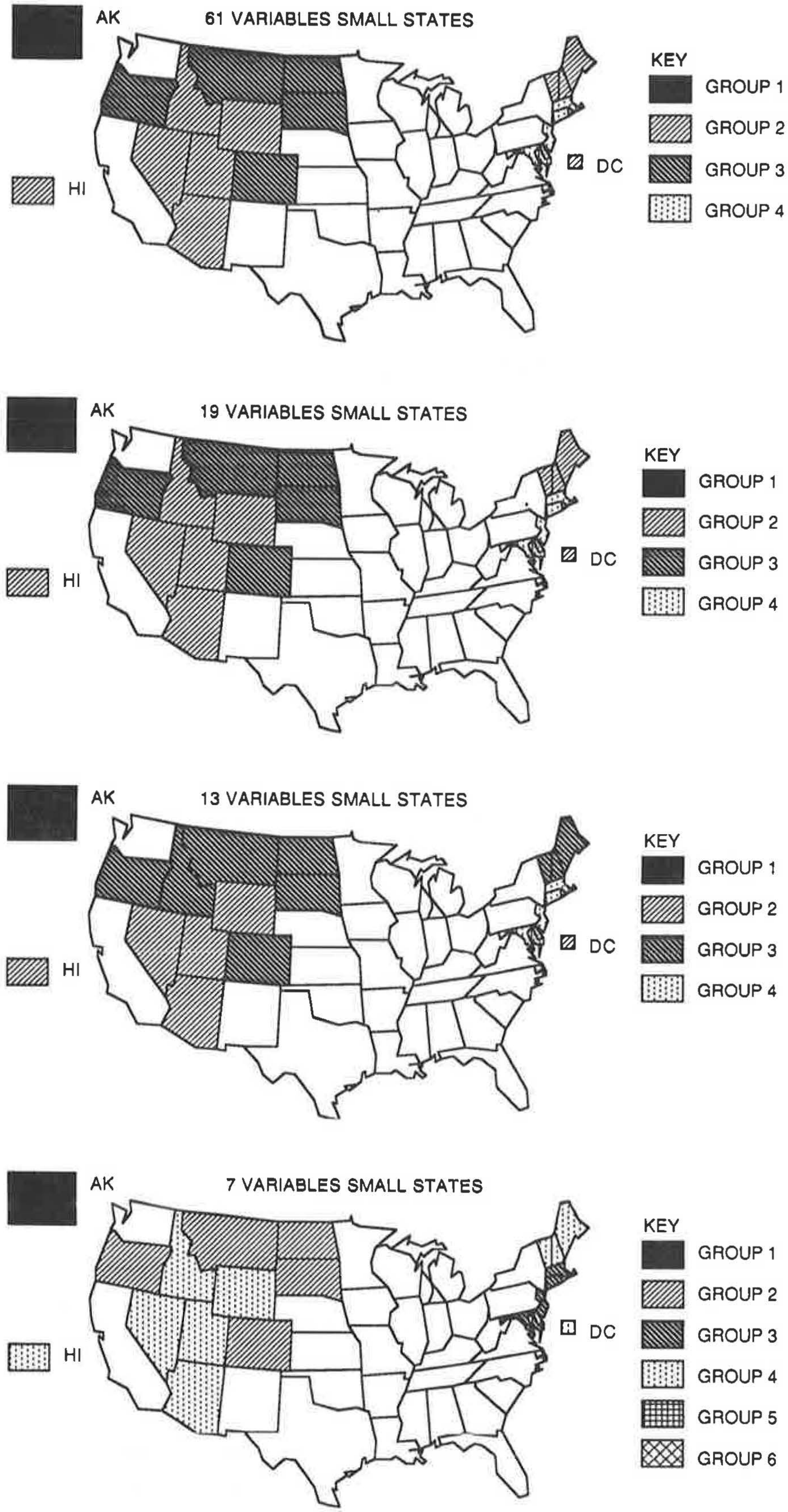


FIGURE 2 Clusters for small states.

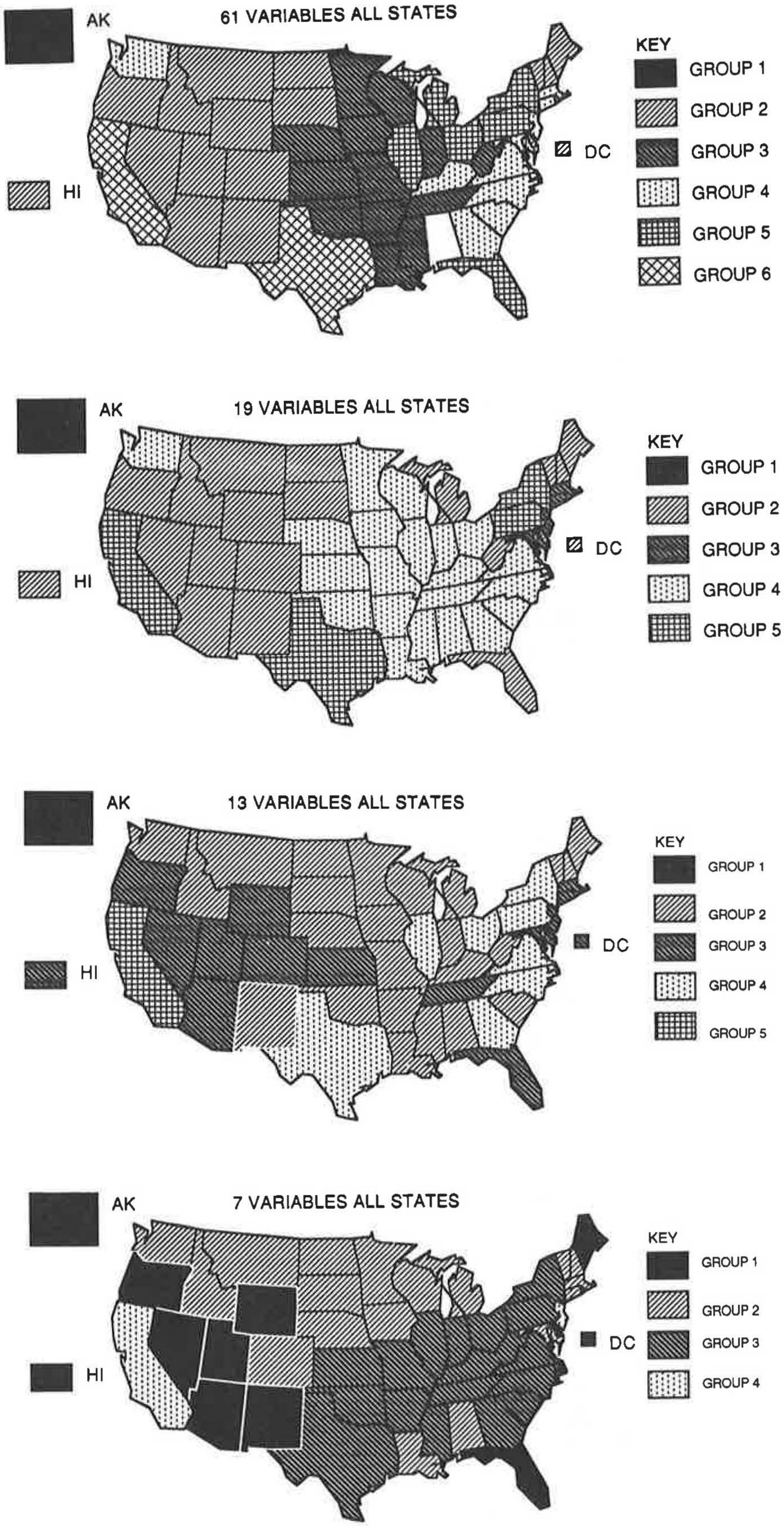


FIGURE 3 Clusters for all states.

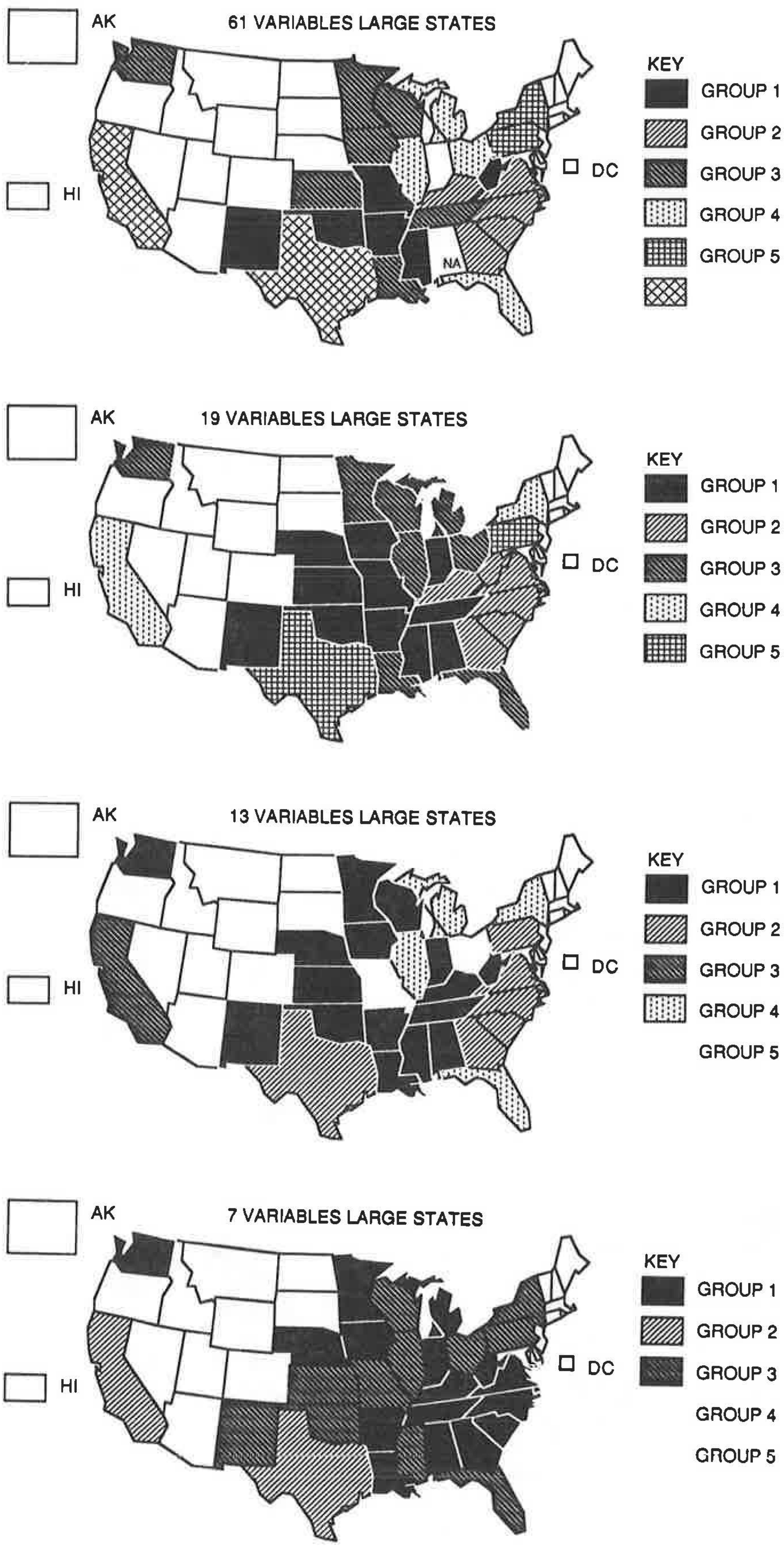


FIGURE 4 Clusters for large states.

other hand, clusters based on a large number of variables (61 variables) tended to produce more cluster groups and greater discrimination, particularly among large states. On balance, it was concluded that clusters based on 19 variables were approximately the right scale for further assessment. In particular, the results now to be discussed relate to those in Figure 3b (19 variables, all states). Figure 5 shows that the tree (dendogram) grouping for this clustering.

• The patterning shown in Figure 3b illustrates both the usefulness and the limitations of the technique. The Figure shows a fairly strong cluster structure in which states are grouped into five general categories:

1. Alaska, by itself.
2. A group consisting primarily of far western and northern New England states, supplemented by West Virginia, Michigan, the District of Columbia, Florida and Hawaii. Important subclassifications of this group are the western block (particularly Colorado, Idaho, and Montana), (North and South Dakota) and the northeast block (particularly Maine, New Hampshire, and Vermont).
3. A mid-Atlantic block of small states, particularly Connecticut, Massachusetts, Maryland, and New Jersey.
4. A large midwestern/southern block consisting of central state subgroups (Alabama, Tennessee, Indiana, Arkansas, Kentucky, and Iowa), southern states (Georgia, North Carolina, South Carolina, and Virginia) and midwest states (Kansas, Missouri, and Oklahoma).
5. Large states (California, New York, Texas, and Pennsylvania).

The dendogram (Figure 5) associated with this cluster shows particularly sharp breaks between Group 5 and other groups, with much less variation and much weaker breakings for the other groups.

While there is considerable overlap in these groups, Table 8 shows some interesting patterns. The mid-Atlantic and big states (Groups 3 and 5) spend two or three times as much per

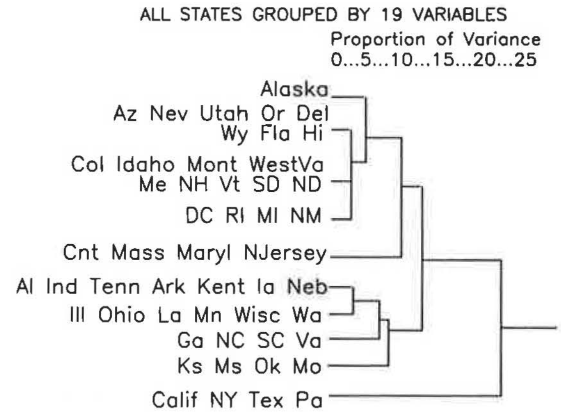


FIGURE 5 All states grouped by 19 variables.

mile as do the other groups, yet their overall condition is not much better. The mid-Atlantic block averages about as much traffic as the midwestern/southern block, and outspends it almost 4:1 per mile, but has about 3 times as many congested miles of road, and only slightly better road and bridge conditions.

This paper goes no further with this relative performance comparison of the states. That is a topic for later research.

DISCUSSION OF RESULTS

The findings above suggest a number of important policy implications for state decision making and political coalitions. First, the overall weakness of the results suggest that coalitions based on factors relating to current issues and other unaccounted for dimensions are certainly important and need to be nurtured. An analytical tool like this can provide some guidance as to what states may have similar problems, but it is not a substitute for good coalition building.

TABLE 8 COMPARISON OF CLUSTER MEANS

	Cluster				
Variables	1 Alaska	2 Far West/NE	3 Mid-Atlantic	4 Midwest/South	5 Big States
Miles under state control	11,426	7,627	3,978	24,752	37,513
Bridges	835	4,152	4,572	17,172	26,358
VMT (million)	3,589	15,830	35,906	35,298	123,959
Gasoline tax, ¢	8	12.3	11.9	11.9	9.75
Fuel tax revenues (000)	$24,197	$76,577	$150,729	$189,386	$471,626
Disbursements/mile	$28,697	$43,320	$222,503	$34,092	$63,767
Capital disb./mile	$14,544	$25,283	$77,065	$20,008	$31,856
Maintenance disb./mile	$8,816	$7,525	$32,346	$6,136	$14,328
Interstate PSR % <2.5	29.6	9.4	4.2	4.2	5.8
Mi. of congested urban Interstate	9.0	19.3	138.7	75.0	·341.0
Mi. of congested urban principal arterial	0.0	10.4	65.5	17.6	295.75
% Deficient bridges	19.9	33.5	41.0	48.0	42.1
Number of states	1	21	4	21	4

Secondly, the dendogram in the cluster analysis suggests that large states, particularly the four largest, are considerably different from the remainder in expenditures and power, as well as size. There is a considerable difference between these four and all the other states, even though the other states also contain numerous regional groupings. In looking at the differences between states, one should recognize that the other 47 are far more similar to each other than they are to these four.

Third, a number of interesting regional clusters appear that might have been expected. These include: western, southeastern, northern New England, and southern New England blocks. Within the dendogram one can also see strong regional affiliations in the Midwest as well. In a sense, therefore, the analytical tools here confirm the importance of geography in producing similar fiscal patterns. It should be noted that no geographic closeness measure was included in these statistics, and therefore the clustering of the state groups by geography is a testament to the importance of geographic similarity in producing fiscal similarity.

On the other hand, certain liaisons appear that might not have been anticipated. These include the northern-tier liaison between Maine, New Hampshire, and Vermont on the east and North and South Dakota on the west; the mountain liaison between Colorado, Idaho, and Montana on the west, and West Virginia in the east; the recreational liaison between Florida and Hawaii; and the western group consisting of Arizona, Nevada, Utah, Oregon, and Wyoming. Other important subregional groups are the southeastern seaboard core states of Georgia, North Carolina, South Carolina, and Virginia; a far-flung quadrangle consisting of Louisiana, Minnesota, Wisconsin, and Washington; a midwestern axis of Illinois and Ohio; and a broad southern midwestern axis of Alabama, Tennessee, Indiana, Arkansas, Kentucky, Iowa, and Nebraska.

The analysis raises more questions than it answers. Why are geographic liaisons so important in classifying agency expenditure patterns? What leads to far-flung similarities in apparently isolated environments? How can the similarity structures here be leveraged to identify and form working political coalitions? Would results be different if data were available over time, if data were more recent, or if additional information relating to modes of travel other than highways were included? Does the cluster methodology itself produce results which are different? What do the results say about the relative performance effectiveness or efficiency of the State highway agencies? To what extent can the effects of weather and climate be interpreted as influencing these results? Answers to these and numerous other questions must weigh the results of later research.

ACKNOWLEDGMENTS

The author acknowledges the assistance of Nathan S. Erlbaum of the New York State Department of Transportation for initial development of this data set. Ralph Campano and Steve Sonderman of UNC-Charlotte assisted in conversion and analysis. Norma Redmond painstakingly prepared the manuscript. The author of course retains the responsibility for errors of fact or omission.

REFERENCES

1. C. D. Phelps. Real and Monetary Disbursements of State and Local Highway Investment. *American Economic Review*, Vol. 59, No. 4, Sept. 1969, pp. 507–521.
2. K. Rao. *NCHRP Synthesis of Highway Practice 100: Managing State Highway Finance*. TRB, National Research Council, Washington, D.C., 1982.
3. H. R. Trumbower. "Collection and Disposition of Motor Vehicle Revenues," HRB, Proceedings, 1927.
4. *Highway Statistics, 1984*. FHWA, U.S. Department of Transportation, Washington, D.C., 1985.
5. *Highway Statistics: Summary to 1985*. FHWA, U.S. Department of Transportation, Washington, D.C., 1987.
6. C. W. Woodward. *Trends in Federal Domestic Transportation Programs, Revenues and Expenditures by State, Fiscal Years 1957-75*. U.S. Department of Transportation, Washington, D.C., 1979.
7. *Highway Taxes and Fees*. FHWA, U.S. Department of Transportation, Washington, D.C., 1984.
8. *Financing Federal-Aid Highways*. FHWA, U.S. Department of Transportation, Washington, D.C., 1987.
9. *Special Report 207: Transportation Professionals*. TRB, National Research Council, Washington, D.C., 1985.
10. *Sixth Annual Bridge Report to Congress*. FHWA, U.S. Department of Transportation, Washington, D.C., 1987.
11. K. Obeng. Application of Path Analysis Methodology to Transit System Maintenance Performance. *Transportation*, Vol. 15, No. 4, 1989, pp. 297–316.
12. *Statistics: User's Guide*. SAS Institute, Cary, N.C., 1985.

Publication of this paper sponsored by Committee on Statewide Multimodal Transportation Planning.

Consideration of Environmental Factors in Transportation Systems Planning: the North Carolina Experience

MARION R. POOLE

Incorporation of environmental factors and their consideration in transportation systems planning has become increasingly important in order to avoid conflicts and problems in future project development. Methods for incorporating environmental factors in thoroughfare planning studies in North Carolina are described and the most recent staff guidelines for incorporating these factors in studies are presented. Also discussed are the documentation of environmental inventories, the use of environmental data in development of the transportation plan, and the importance of system level environmental studies in right-of-way protection.

A goal at both the state and federal levels is to preserve and improve the natural and man-made environment. Improvements to the transportation system can help achieve this goal. Environmental factors have been a consideration in urban thoroughfare planning in North Carolina since its inception by the state in 1959. In the early years, however, environmental factors and considerations were not well documented and were somewhat different than environmental factors being considered today.

BACKGROUND

In the 1960s, major environmental factors considered in thoroughfare systems planning and in thoroughfare alignment decisions were neighborhoods, cemeteries, churches, historical areas, and parks. With enactment of the National Environmental Policy Act of 1969, the Clean Air Act of 1970, and other environmental laws and administrative orders, physical elements were introduced and consideration of environmental factors was formally incorporated into comprehensive thoroughfare planning studies. This was done even though federal regulations did not require an environmental impact statement (EIS) at the systems planning level.

The first thoroughfare planning study to document an environmental analysis of alternative plans was a study for Morehead City-Beaufort-Atlantic Beach, N.C. (*1*), which was completed in July 1971. The report's discussion of alternate thoroughfare plans relative to social, economic, and environmental effects included (1) a discussion of the physical characteristics of four alternative thoroughfare plans; (2) an evaluation of the social and economic effects of the alternative plans; (3) an evaluation of environmental effects; and (4) a discussion of the alternative plans relative to providing a feasible economic, safe, and efficient transportation system. The discussion of social and economic effects included effects on economy and employment; housing; parks, recreational facilities, and public utilities; public health and safety; and national defense. The number of homes and businesses damaged or taken by alternative plans were tabulated, but most of the evaluations were general in nature and not quantified. The discussion of environmental effects included air and noise pollution, aesthetics, conservation and water quality (including erosion, sedimentation, wildlife and general ecology of the area), and natural and historic landmarks. The majority of these evaluations was also general in nature.

The environmental analysis incorporated into thoroughfare planning studies at that time was careful to exclude terms such as "environmental impact" so as not to be confused with the EISs required for project studies. The highway engineer doing the thoroughfare planning study did not have access to environmental staff resources because of the environmental staff's heavy workload in project planning work. The thoroughfare planning staff used local staff resources, reference documents, and additional training to assist them in the environmental analysis work.

The identification of environmental factors, their application in thoroughfare planning, and their documentation in thoroughfare study reports has varied considerably since 1971. Some of the environmental factors tabulations included in a 1978 study report for New Bern (*2*) are shown in Tables 1 and 2 and are fairly typical of subsequent study reports. Identification and documentation has become more important as efforts to preserve and protect future rights-of-way well in advance of project planning and construction have become critical in order to reduce cost and disruption. The most recent staff guidelines for incorporating environmental considerations in thoroughfare planning were prepared in February 1989 (*3*). The following sections of this paper present the most recent guidelines used by North Carolina highway planning engineers in incorporating environmental considerations in transportation systems planning. Later sections discuss documentation of the environmental inventories, the use of environmental data in development of the transportation plan, and application of system level environmental studies in right-of-way protection.

It should be noted that, for the most part, the term "transportation systems planning" is synonymous with the term "thoroughfare planning" in North Carolina. Transit planning, other modes, and other transportation systems planning fac-

Division of Highways, North Carolina Department of Transportation, P.O. Box 25201, Raleigh, N.C. 27611-5201.

TABLE 1 AIR QUALITY ANALYSIS FOR THE NEW BERN STUDY AREA (1978 NEW BERN STUDY)

Year	System	Estimated VMT/a	Emissions in Tons per year/b CO	HC	NOx
1975	Base Year	426,721	7610	1241	1004
1982	Plan A	522,453	3741	730	803
1985	Plan A	563,481	2865	529	675
2000	Plan A	768,623	2738	474	766
2000	Recommended Plan B	753,198	2710	438	766
2000	Plan C	761,549	2573	438	766
2000	Plan D	741,950	2500	438	766

a/Based on average weekday travel
b/Emissions estimates were obtained from nationwide factors

Plan A--Existing plus programmed improvements
Plan C--Includes US17 Bypass to east of Bridgeton crossing Neuse River south of New Bern and connecting to existing US70 Bypass
Plan D--Includes US17 Bypass north of New Bern in Glenburnie Road corridor

TABLE 2 SELECTED ENVIRONMENTAL EFFECTS OF ALTERNATE THOROUGHFARE PLANS (1978 NEW BERN STUDY)

	Plan A	Plan B	Plan C	Plan D
Total Miles of Street Construction:				
Widening	1.06	6.66	6.66	7.79
New location	6.11	23.52	18.39	16.38
Estimated Number of:				
Businesses displaced	0	5	4	4
Homes displaced	0	69/c	68/c	68/c
Schools affected	0	1	1	1
Recreational areas affected	0	2	2	2

c/Includes families being relocated as part of New Bern's urban re-development program

tors are often included in the thoroughfare systems planning for the larger urban areas in the state. Transportation systems planning also typically involves an evaluation of transportation system management (TSM) alternatives in both large and medium size areas above 10,000 population. However, for small communities under 10,000 population and counties, transportation systems planning consists almost exclusively of planning for the road and highway system ("thoroughfare planning").

NORTH CAROLINA GUIDELINES FOR INCORPORATING ENVIRONMENTAL CONSIDERATIONS IN SYSTEMS PLANNING

Environmental considerations in transportation systems planning include (1) identifying critical environmental factors that need to be considered in systems planning and subsequent project planning; (2) developing alternatives based on the environmental inventories and environmental preservation objectives; (3) evaluating alternatives relative to travel service and environmental factors; and (4) developing a recommended plan that best meets environmental, travel service, and other public goals.

The North Carolina guidelines provide that inventories should be made and evaluations conducted on a number of environmental factors. These can be separated into three broad categories: (1) physical environment, (2) social and/or cultural environment, and (3) economic environment.

Physical Environmental Factors

These include air quality, water resources, soils and geology, wildlife, and vegetation.

Air Quality

Existing air quality problems both areawide and site specific are to be identified to the extent possible in the data gathering phase so that they can be considered in plan development. In the evaluation of alternative plans, air pollutant emissions are computed for each significantly different alternative plan and the "do nothing" alternative.

Impacts include carbon monoxide (CO), hydrocarbons (HC), and nitrogen oxides (NOx).

Measures include tons or kilograms emitted per day, computed by a computer program that utilizes pollutant emission factors and tabulations of vehicle miles of travel (VMT) by average speed groups.

Mitigation measures considered in systems planning include: (1) vehicle emission controls, (2) reductions in VMT, (3) increasing average speeds, (4) eliminating congestion and delay, and (5) reducing fuel consumption. Comparison of pollutant emissions of alternate plans is one basis used for evaluation of alternatives.

Sources of information are air pollutant emission factors from MOBILE II computer program and VMT by speed increment group from alternative traffic assignments.

Water Resources

These are considered both (1) as they relate to the support of population, agriculture, and industrial consumption; and (2) for fish and wildlife uses. Water resources and wetlands are to be defined in environmental inventories for subsequent consideration in systems planning. Impacts of transportation on water resources need to be quantified in the evaluation phase.

Impacts include watersheds, wetlands, pollutants from vehicles (oils, particulates, etc.), and hazardous waste spills—area specific in each case.

Measures include acres impacted (map measurements) and exposure (VMT, truck routes, and truck movements).

Mitigation measures can include construction procedures, facility design, regulations, avoidance, and replacement.

Sources of information on water resources include aerial photography, topographic maps, soils maps, and flood hazard studies.

Soils and Geology

Consideration of soils and geology in systems planning is primarily concerned with identification in the environmental inventories of areas and sites that need to be considered in the planning phase. Effects may or may not need to be quantified in the evaluation and documentation.

Impacts include soil resources (farmland), mineral resources, construction (streets and other), hazardous waste disposal areas, and stability (earthquake risks, etc.).

Measures include acres and other quantities.

Mitigation measures include avoidance and construction procedures.

Sources of information include soil maps, geological survey maps, old land use maps, and aerial photography both old and new.

Wildlife and Vegetation

Considerations include identification of endangered species and habitats in the environmental inventories, consideration of these elements in development of alternatives, and quantifying any impacts in the evaluation of alternatives.

Impacts include habitats and endangered species.

Measures include acres (habitats).

Mitigation measures include avoidance and replacement.

Sources of information include soils reports, the North Carolina Natural Heritage Program, and local project planning reports.

Social and/or Cultural Environmental Factors

These elements include housing, neighborhoods, noise, educational facilities, churches, park and recreational facilities, public health, safety, national defense and emergency evacuation, and aesthetics.

Housing

Impacts of alternative plans on housing are quantified in the evaluation of alternatives. Considerations include housing stock, displacement, and minority groups. The evaluation is to include the number of displacements of minority group housing.

Impacts include displacement and disruption.

Measures include the number of units displaced by condition, minority households displaced, miles of street widening, and right-of-way cost.

Mitigation measures can include replacement, acquisition, and availability of replacement units.

Sources of information include housing condition survey maps, field inventories, and right-of-way cost estimates.

Neighborhoods

Neighborhoods are identified through an inventory and are considered in plan development and analysis of alternatives. Impacts of alternative plans on neighborhoods are usually subjective and can be both positive and negative.

Impacts include division, disruption of cohesion, degradation or improvement of access, and traffic impacts.

Measures include traffic (average daily traffic or VMT), subjective changes, and formation of lobbying groups.

Mitigation tools include redevelopment, landscaping and beautification, traffic control, and avoidance.

Sources of information include neighborhood and land use studies, field surveys, and local knowledgeable sources.

Noise

Detailed analysis of transportation noise impacts on abutting land uses is normally done in the project planning and design stages, when more information is available on the design and operation of the facility. However, noise problems may in some instances be identified, or voiced by the public, in the systems planning study and will need to be evaluated and addressed in the evaluation stage. In systems planning the evaluation will in most cases be subjective and based on "rule-of-thumb" methods. However, there are more sophisticated tools and reference texts available if needed.

Impacts are on humans (more sensitive receptors).

Measures are truck volumes (an indicator of problems) from field measurements.

Mitigation measures include elevation or depression of roadways, spatial separation, acoustical barriers, and building construction.

Sources of information include project planning reports and housing and urban development project studies.

Educational Facilities

These are usually considered of significant community value and are typically major traffic generators. Included are various facilities ranging from the local neighborhood school (if any still exist) to the major college or university. Usual social environmental concerns include access, parking, safety, proximity to play areas, school bus routing and accessibility, pedestrian movements, noise, and aesthetics. Systems planning considerations may include school attendance districts, functional obsolescence of school buildings, pedestrian and vehicular movements, and functions of the institution.

Impacts include access, traffic operations, safety, noise, effect on attendance areas, and effect on school bus routes.

Measures include trip attractions, access time, number of students, and traffic counts.

Mitigation measures include project planning and construction with consideration to pedestrian and vehicular movement and the functions of the institution.

Sources of information include the education system, trip generation manuals, and travel models.

Churches

These institutions are usually considered community value factors. Individual churches are usually important to a wide range of people and often may have large service areas. Functional obsolescence may need to be considered in the planning process.

Impacts include disruption, noise, parking, pedestrian access, and effect on service area.

Measures include right-of-way cost and size of congregation.

Mitigation measures include avoidance, replacement in kind, and design enhancements of the facility.

Sources of information are the employment survey and local church associations.

Park and Recreational Facilities

If a transportation improvement project encroaches on public land devoted to park or recreational purposes, a Section 4(f) statement is required verifying that there is no prudent or feasible alternative. These facilities need to be identified in the inventory phase and considered in the development of the plan and improvement recommendations. Transportation system impacts can be both positive and negative.

Impacts include encroachment, traffic, access, and aesthetics.

Measures are travel time (access).

Mitigation measures include avoidance, replacement in kind, and design.

Sources of information are land use maps and property maps.

Historic Sites and Landmarks

These are treated similarly to park and recreational facilities. Properties listed on the National Register can be determined fairly easily. Properties that "may qualify" for listing are usually more difficult to identify. Complete avoidance is usually the only alternative. Local historical societies and planning staffs are usually the best source of information.

Archaeological Sites

These sites have significance similar to park and recreational facilities and historic sites and landmarks. However, in many instances these sites can be excavated, evaluated, information and data extracted, and then used for transportation purposes. Finding a site of significance that could be affected by a proposed transportation facility can delay the project and increase project costs associated with archeological work at the site. Archaeological sites of significance may be difficult to identify during the systems planning process due to lack of information. Historical publications may sometimes provide information on sites to the systems planning engineer.

Public Health and Safety

This factor includes access to health care facilities, ambulatory services, fire and police protection, and garbage collection. Often this element is not seriously considered in either systems planning or project development planning. Major considerations are access to health care facilities and response time.

Impacts are primarily in terms of access.

Measures include travel time, level of service, and average speed.

Mitigation is not applicable.

Sources of information include the travel models, speed and delay studies, and the providers of the services.

National Defense and Emergency Evacuation

This is a social environmental factor concerned with the ability of the transportation system to serve transportation demands during periods of national or local emergency such as war and natural calamity. In North Carolina, for example, hurricane evacuation is a significant consideration for coastal areas. This environmental factor is considered during the plan development phase.

Impacts include troop and military equipment movements and population.

Measures include travel time, level of service (capacity), and bridge weight limits.

Mitigation is not applicable.

Sources of information include the travel models, traffic capacity studies, bridge studies and inventories, and defense agencies.

Aesthetics

This is an important consideration in the plan development stage and is often involved in a number of other social environmental factors. Will the plan result in a future transportation system that will be pleasing to the public and contribute to an improved urban environment? Improving the aesthetics of the urban environment through transportation improvements is a mitigation tool in dealing with environmental considerations in systems planning.

Economic Environmental Factors

Economic environmental factors include businesses, employment and income, economic development, public utilities, transportation costs, capital costs, and operation and maintenance costs.

Businesses

The implementation of a thoroughfare plan has both positive and negative effects on business within the area. As new thoroughfares are constructed, or old ones widened, the improved mobility tends to improve the overall business climate and proves more attractive to the establishment of new business interests. The construction of new facilities often opens up new land areas for business expansion. On the negative side is the potential disruption or removal of existing businesses as a result of thoroughfare construction.

Impacts include accessibility, parking availability, traffic operations, access control, and truck and service vehicle access.

Measures include level of service, truck VMT, number of businesses displaced, number of employees displaced, right-of-way costs, and parking lost.

Mitigation measures for adverse effects include acquisition, relocation, and provision of off-street parking.

Sources of information include the employment survey, right-of-way cost estimates, and system level of service measures.

Employment and Income

Improvement in the level of service provided by the transportation system will reduce transportation costs for industry, facilitate industrial employment expansion, and contribute to area income through additional business activity and reduced transportation cost for workers.

Impacts include accessibility, goods movements, and disruption of operations.

Measures include level of service, truck VMT, and employment.

Mitigation tools include acquisition, relocation, and access improvements.

Sources of information include travel models, employment survey, goods movement surveys, and local business organizations.

Economic Development

This factor is very similar to the employment and income factor. New thoroughfares that open new areas for development will most influence this factor. In general, new industry prefers good access to at least two transportation modes (i.e., rail and highway). Economic development would most often be considered in the transportation plan development stage, but may be identified as a high priority goal for the urban area in the early study stages.

Impacts include access to developable land, goods movement, and transportation costs.

Measures include acres of developable land served, miles of new thoroughfares, employment potential, and total VMT and truck VMT.

Mitigation is not applicable.

Sources of information include local business organizations, industrial development commissions, public utilities, and land development and land use plans.

Public Utilities

This factor is a consideration in the transportation plan development and evaluation stage. The transportation plan can impact utility service areas and existing utilities.

Impacts include relocation and disruption.

Measures are miles of street widening.

Mitigation is not applicable.

Sources of information include right-of-way cost estimates, utilities, and local officials.

Transportation Costs

This factor is applicable in the transportation plan evaluation stage and is used in the comparison of alternative system plans and projects.

Impacts include personal disposable income, business costs, and area economy.

Measures include VMT, vehicle hours of travel (VHT), level of service, vehicle operating costs, user time costs, and accident costs.

Mitigation is not applicable.

Sources of information include travel model data and local cost data, which can be used to develop these measures.

Capital Costs

This factor is applicable in the plan evaluation stage and is used for comparison of plans. Capital costs usually include construction costs and right-of-way costs as separate elements.

Operation and Maintenance Costs

These are usually considered in the evaluation of alternatives but could be identified as problems earlier in the planning process. Examples of this would be ferry operational costs, and high maintenance costs associated with obsolete bridges, traffic control devices and systems, and roadways.

ENVIRONMENTAL INVENTORY DOCUMENTATION

Environmental inventories need to be recorded and documented on planning base maps both for use in the systems planning and for future reference. Environmental elements that are logical to record on maps include (1) watersheds, (2) wetlands, (3) prime farmland, (4) mineral resources deposits, (5) hazardous waste disposal sites, (6) wildlife habitats, (7) educational facilities and attendance areas, (8) churches, (9) public parks and recreational areas, (10) cemeteries, (11) historical sites and landmarks, (12) archaeological sites, (13) health care facilities, (14) ambulatory service facilities and service areas, (15) fire and police protection facilities and service areas, (16) national defense facilities, (17) industrial development sites, and (18) significant utilities. Some elements such as neighborhoods are not easy to delineate because of differing definitions.

It is also important that environmental factors be recorded in the study report, because it is the best long-term reference document. Recording information on maps to the extent possible is the best procedure. The old adage "a picture is worth a thousand words" is very applicable.

USING ENVIRONMENTAL FACTORS IN THE DEVELOPMENT OF THE TRANSPORTATION PLAN

Environmental factors identified and recorded in the inventories provide a data set for use in developing alternative plans along with many other data sets. Other environmental considerations that are directly related to the measures of the ability of the transportation system to provide a transportation service come into play when comparing alternative plans. Environmental factors and considerations may often be in conflict with each other and other transportation objectives. Different groups will place different weights on environmental factors based on their own priorities. Laws and regulations will place constraints on plan options and thoroughfare alignments. Urban area planning goals and objectives and community value factor studies can sometimes provide guidance on the weights that should be given to alternative environmental considerations.

A common problem that develops in the alternative plan development, plan evaluation, and plan adoption stages is that one environmental factor may become a focal point for a special interest group. Other equally important considerations may be pushed into the background. The transportation systems planner must be careful in these situations to use all available tools to ensure all important considerations are kept before the decision makers. Providing current maps and easily understood data tabulations on alternative plans, and responding quickly to requests for information and analysis, is the best method for ensuring adequate consideration of all important factors.

In the evaluation of alternative plans at the systems level, it is conceivable that the "best" systems plan from an environmental standpoint could have one key element which is "bad" from a project environmental impact standpoint. Good documentation of the environmental analysis at the systems level may be crucial to gaining approval for proceeding with the project when it reaches the EIS stage in project development.

APPLICATION OF SYSTEM LEVEL ENVIRONMENTAL STUDIES IN RIGHT-OF-WAY PROTECTION

One of the typical concerns in requiring reservation and/or dedication of rights-of-way for thoroughfares well in advance of programming and construction has been the concern with a potential conflict with the National Environmental Policy Act and other federal and state laws and regulations. This was one of the first issues addressed by a 1984 North Carolina Right-of-Way Task Force cosponsored by the North Carolina Division, Southern Section, Institute of Transportation Engineers, and the North Carolina Chapter of the American Planning Association.

Work on this issue was completed by the Task Force on February 11, 1986, through documentation of a set of questions and answers (*4*) that were cleared through the Federal Highway Administration's (FHWA's) state, regional, and Washington offices. One of the important findings that came from the question-and-answer approach was that an EIS need not be completed before right-of-way can be protected. However, it was noted by FHWA that a protected corridor may carry little weight in the selection of an alignment for federal funding once appropriate environmental studies were completed. If the protected corridor has serious environmental problems when compared to other alternatives, federal funds may not be available for construction in the protected corridor. It was recommended that a preliminary environmental screening be conducted prior to designation of an alignment for protection to minimize potential future problems. It was recommended by FHWA that the screening should consider the following factors:

- FHWA cannot approve a project which uses publicly owned parkland, historic properties, and certain other types of land (Section 4f land) unless it can be shown that there is no feasible and prudent alternative.
- Historic properties cannot be affected without first complying with the Advisory Council on Historic Preservation requirements.
- Wetlands cannot be taken unless there is no practicable alternative and the project includes all practicable measures to minimize harm to the wetland.
- The project cannot result in significant encroachments on floodplains unless there is no practicable alternative.
- Right-of-way actually acquired through donation or purchase must be acquired in accordance with Title VI of the Civil Rights Act of 1964.

In summary, if all "problem" land uses are avoided, there should be no problem with right-of-way protection through dedication, reservation, or advance acquisition prior to EIS studies.

The identification and documentation of environmental factors, using the environmental data in plan development, and documenting their consideration and effects on alternative plans in systems planning, should adequately meet the regulatory requirements and minimize the possibility that significant alignment changes will occur in project planning and implementation.

The current approach in North Carolina is to (1) identify and document the environmental factors affecting the thoroughfare system plan and thoroughfare alignments in the systems plan and report; (2) locate proposed thoroughfares as accurately as possible based on all available data and information; and (3) use all available planning tools such as the subdivision ordinance to protect the proposed thoroughfare alignments in anticipation that the thoroughfare improvement will be done as contemplated in the system plan. Adequate and thorough environmental studies done at the transportation systems planning stage minimize the possibility of problems at the project development stage.

It should be noted that it is always possible that some unanticipated problems or issues will occur that cause a change in alignment when a project reaches the construction stage. However, this will not happen often if a thorough job is done in the systems planning stage. Long-term protection of thoroughfare corridors is an absolute necessity. There is just no other feasible alternative.

SUMMARY

This paper has discussed the North Carolina approach and experience in incorporating environmental considerations into systems planning. Transportation systems planning should include inventories of environmental factors, consideration of these elements in plan development and evaluation, and documentation of their application. Documentation of inventories and their application is considered crucial to ensure that right-of-way protection actions are soundly based and no major pitfalls are encountered in subsequent project development.

REFERENCES

1. *Morehead City-Beaufort-Atlantic Beach Thoroughfare Plan.* Planning and Research Department, North Carolina State Highway Commission, Raleigh, N.C., 1971.
2. *New Bern-Bridgeton-Trent Woods Thoroughfare Plan.* Planning and Research Branch, Division of Highways, North Carolina Department of Transportation, Raleigh, N.C., 1978.
3. *Environmental Considerations in Transportation Systems Planning.* Planning and Research Branch, Division of Highways, North Carolina Department of Transportation, Raleigh, N.C., 1989.
4. *Right-of-Way Protection and Federal Funds.* North Carolina Division Office, Federal Highway Administration, Raleigh, N.C., 1986.

Publication of this paper sponsored by Committee on Transportation Planning Needs and Requirements of Small and Medium-Sized Communities.

Increasing the Capacity of Urban Highways: The Role of Freeways

PATRICK DECORLA-SOUZA AND CHRISTOPHER FLEET

An analysis of the supply and use of urban highways by functional class using the Highway Performance Monitoring System (HPMS) suggests that the potential for shifts in travel share from lower order facilities to freeways and expressways will increase as urban travel increases and urban population centers increase in size, and the greatest potential shifts will be in the suburbs of the largest urban areas. The results of an economic analysis suggest that, in general, a more optimum mix of highway facility types might be achieved by new capacity investments on freeways and expressways. On average, widening of freeways in large urban areas could be paid for by peak-period user charges ranging from 2.4 cents per vehicle mile in fringe areas to 8.9 cents per vehicle mile in the urban core. Analysis and inferences are presented that suggest that freeways/expressways have a powerful role to play in plans to expand urban highway system capacity, especially in heavily congested large urban areas, if public acceptance can be achieved. However, financing such improvements using the current tax structure, which relies heavily on motor fuel taxes, would not be equitable towards users who use the highway system during off-peak periods, when existing highway system capacity would suffice. Consequently, financing mechanisms which rely on road pricing, possibly using automated vehicle identification technology, should be investigated in plans to expand urban highway systems.

What will it cost to maintain today's quality of transportation service in urban areas for the next 20 years? What will be the best mix of new highway capacity by functional class? What financing mechanisms could be used to pay for investments in new capacity?

Questions such as these are being asked in efforts underway to examine the options for the Federal role in the post-Interstate era. It has been speculated that, to minimize traffic congestion, the optimum share of travel on limited access facilities is about 28 to 30 percent in rural areas; however, in urban areas, as density and size of an urban area increase, this optimum share increases, and the best service is provided with about 50 percent of travel on limited access facilities in the largest urban areas (*1*). This paper reviews current conditions of urban highway use, supply and level of service by functional class, and assesses the potential and economic efficiency of actions designed to shift travel between functional classes.

A 1987 FHWA study (*2*) indicated that significant amounts of new highway capacity will be needed in urban areas. Also, an economic analysis done by FHWA (*3*) to compare benefit-cost ratios of investments in new capacity revealed that, depending on functional class, each dollar invested in capacity improvements would return between \$5 and \$12 in benefits. The analysis indicated that investments in all functional classes would be cost-beneficial. However, the approach used in the study did not differentiate between higher-cost investments in the urban core and lower-cost investments in the fringe, and implicitly assumed that the shares of travel on each functional class would stay relatively the same in the future. Would investments designed to shift the share of travel between functional classes be even more cost-beneficial? We attempt to answer this question in this paper, by urban area size and by urban development density category (i.e., urban core vs. fringe).

Finally, we attempt to provide some scale of what it really costs to serve a vehicle mile of travel for the portion of a trip which uses new capacity. Such information could assist in choosing between alternative financing mechanisms, and in setting charges for facility users if tolls charged though mechanisms such as automated vehicle identification (AVI) technology were to be used to recover costs for new or widened facilities.

There are three parts to the analysis:

1. analysis of the current supply, use and level of service on urban highways;
2. analysis of the economic efficiency of new highway capacity on alternative facility classes at alternative levels of service; and
3. estimation of needed charges to users of new capacity, by type of investment and location within the urban area.

PART 1: SUPPLY, USE AND LEVEL OF SERVICE ANALYSIS

Analysis Procedures

This part of the study draws from and updates with more recent data the results of a previous study (*4*) of current conditions and recent trends in system supply and use by functional class. The study used sample data in FHWA's Highway Performance Monitoring System (HPMS) database (*5*), and sought to quantify the relationships between urban highway travel use, supply, and urban development characteristics (see Figure 1), using HPMS data. HPMS is a coordinated data base that requires annual reporting by state highway agencies to FHWA.

Use, supply and level of service information from HPMS data for a sample of 164 urbanized areas were cross-tabulated by urbanized area size, functional class, and urban development density characteristics. Urbanized areas were categorized into five groups: (1) 50,000–75,000 population, (2) 75,000–

Office of Planning, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C. 20590.

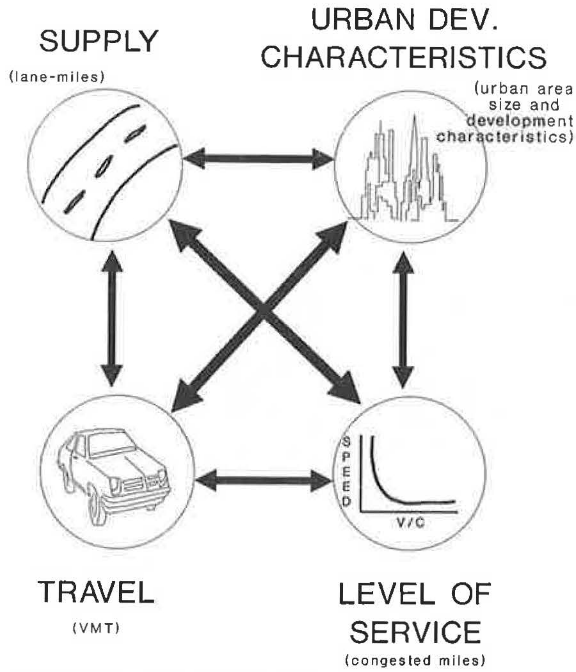


FIGURE 1 Overview of relationships investigated.

200,000 population, (3) 200,000–500,000 population, (4) 500,000 to 1 million population, and (5) more than 1 million population. Local street data are not included in HPMS. Non-local urban highways were categorized into four classes: (1) freeways and expressways, (2) other principal arterials, (3) minor arterials, and (4) collectors. Three urban development density categories were used: (1) urban core, (2) suburbs, and (3) urban fringe.

The representativeness of the sample urbanized areas used in the analysis is indicated by size category in Table 1. Although nearly half of the states were omitted from the analysis due to data limitations, only four of the 60 cells in the cross-tabulation (i.e., 5 urban area sizes × 4 functional classes × 3 density categories) contained less than 100 samples. It was therefore felt that the number of observations per cell was adequate for the analysis. Also, as indicated in Table 1, the distribution of urbanized areas within each population size group in the analysis closely matched the nationwide distribution for all urbanized areas.

Cross-tabulated data for the years 1982 through 1987 were obtained for non-local roadways with respect to the following characteristics:

- *Highway use:* Share of daily vehicle miles of travel carried by each functional class, and intensity of use (VMT per lane-mile) daily.
- *Supply:* Share of lane miles by functional class, and supply density in lane miles per 10,000 population (population data from the 1980 Census were used).
- *Level of service:* Share of route miles with peak hour volume-to-capacity ratio greater than 0.85.

Inferences

Share of Travel by Functional Class

About two-thirds of the total (non-local) travel in urbanized areas of all sizes is carried by the two highest functional classes of facilities (i.e., freeways/expressways and other principal arterials) (see Figure 2). As urban area size increases, however, the proportion carried by freeways/expressways alone increases from under 20 percent in the smallest areas to about 40 percent in the largest. From this pattern we may infer that as urbanized areas grow in the future there will be a tendency toward higher use of freeways and expressways.

Effect of Supply on Travel Shares

While the share of travel on freeways and expressways increases from medium size areas to the largest size areas, the share of supply shows a slight decline (Figure 3). This suggests that in large urbanized areas (greater than 1 million population) relatively lower levels of freeway and expressway supply are constraining further increase in use of such facilities. From this we might conclude that higher-order facilities have the potential to carry a larger share of urban travel than they currently do.

Figure 4 tends to confirm the pattern seen in Figures 2 and 3. This is a plot of supply and use of freeways/expressways on a per capita basis. Freeway/expressway use per capita (seen on the left) remains relatively constant in areas above 200,000

TABLE 1 REPRESENTATIVENESS OF SAMPLE

	Number of Areas		Percent of Areas	
	All	Sample	All	Sample
<75,000 population	111	41	29.8%	25.0%
75,000-200,000 population	146	71	39.1%	43.3%
200,000-500,000 population	62	28	16.6%	17.1%
1/2-1 million population	24	10	6.4%	6.1%
>1 million population	30	14	8.0%	8.5%
Total	373	164	100.0%	100.0%

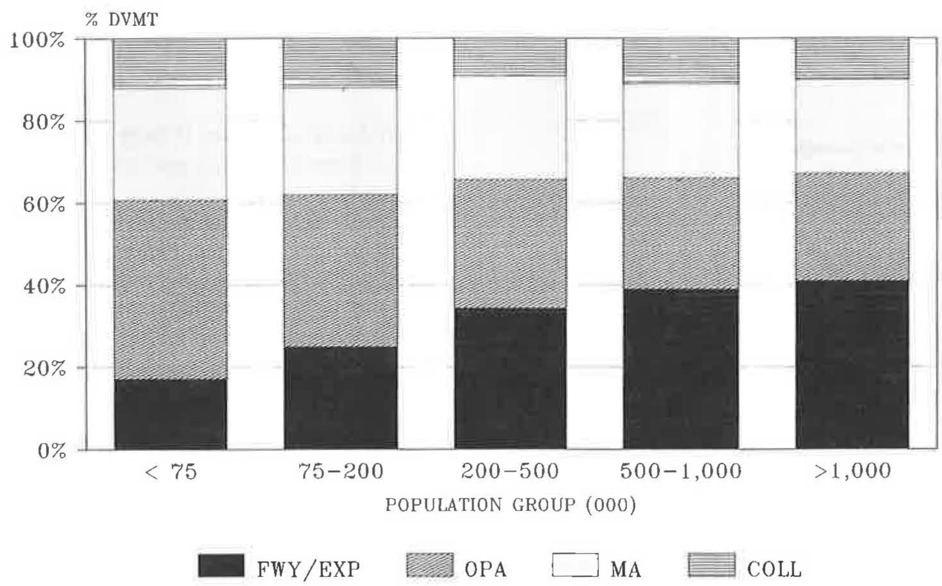


FIGURE 2 Share of daily travel by functional class and urbanized area size (1987).

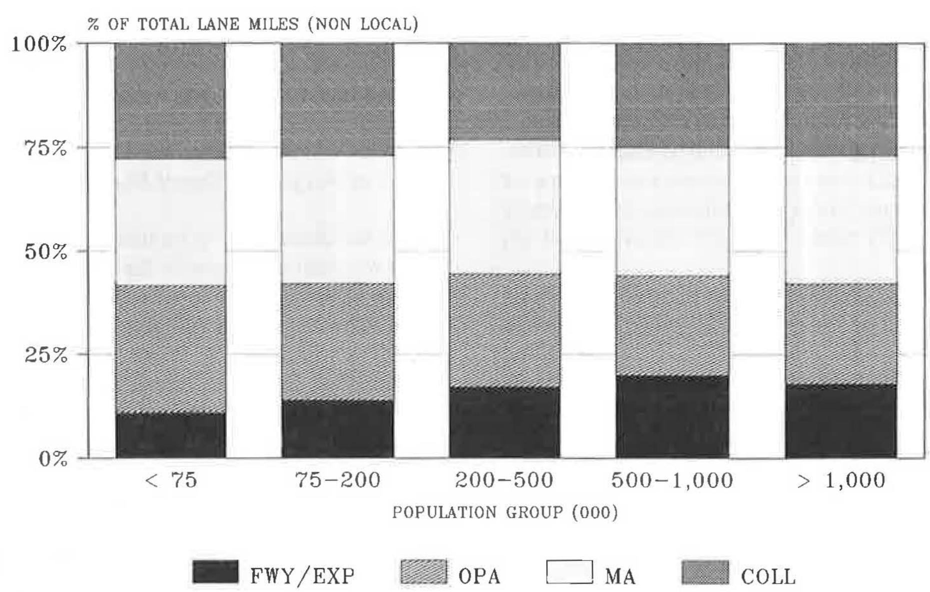


FIGURE 3 Share of supply by functional class and urbanized area size (1987).

population despite the sharp drop in supply per capita (seen on the right) in these areas as urban area size increases. This suggests that there is a preference to use higher-order facilities in the largest urbanized areas because, among other things, these facilities afford the user greater connectivity or accommodate longer trips.

Recent Shifts in Travel Shares

As total travel increases over time in the larger urbanized areas (with over 200,000 population), there is a tendency for it to shift toward freeways/expressways. Over the 6 years from 1982 through 1987, the percent of the daily VMT carried on freeways/expressways generally increased except in the smallest areas (Figure 5). The implication is that, as travel continues to increase in the future, there is a potential for freeways/expressways to carry a larger share of that travel, provided that supply constraints are not a factor.

Congestion and Travel Shares

Figure 6 shows freeway/expressway travel (intensity of use) in daily VMT per lane-mile and congestion levels expressed as a percentage of route-miles with peak-hour volume-to-capacity (V/C) ratios greater than .85. While the intensity of use triples from the smallest to the largest urbanized areas, the proportion of congested route-miles increases sixfold.

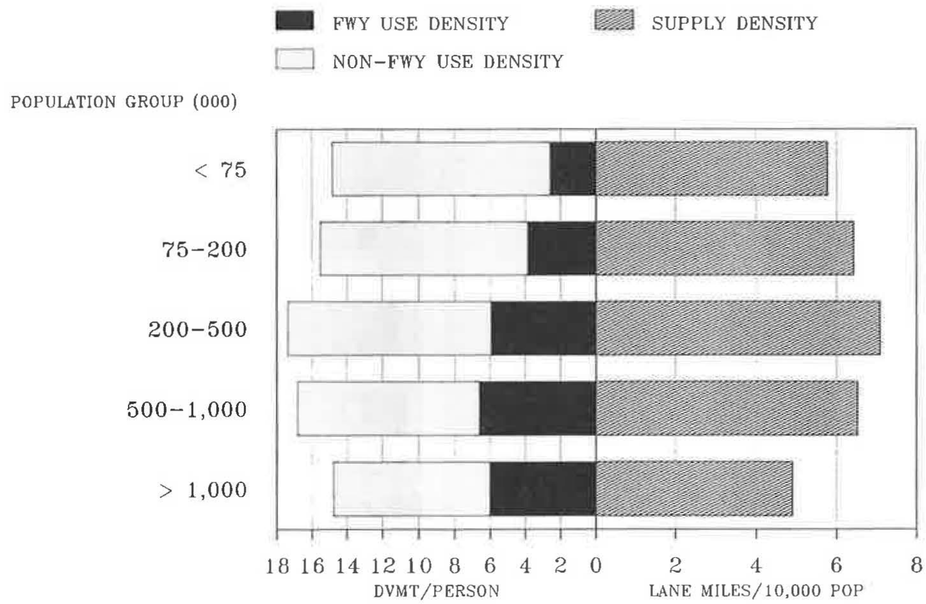


FIGURE 4 Freeway/expressway supply and use density by urban area size group (1987)

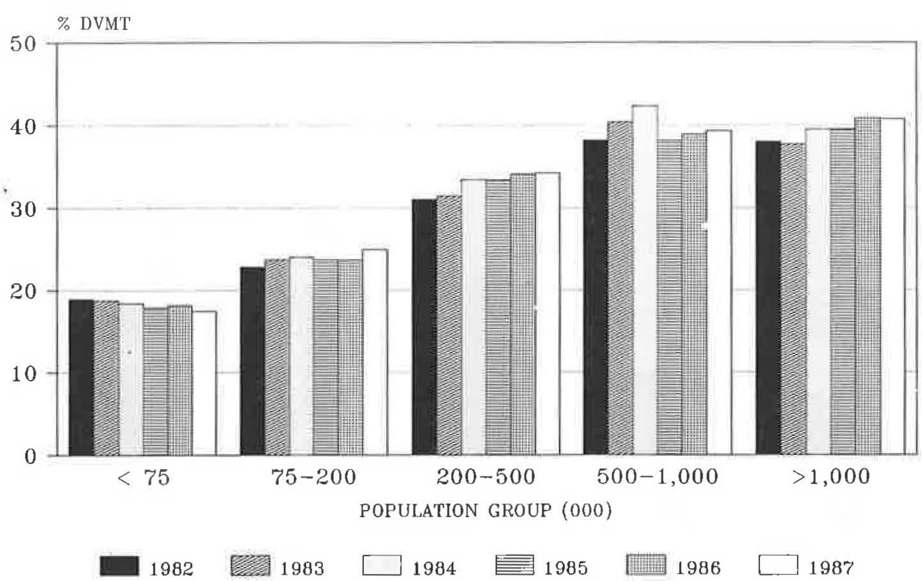


FIGURE 5 Share of daily travel carried by freeways and expressways, 1982–1987.

The increase in freeway/expressway travel shares with increasing urbanized area size, in spite of sharply increasing congestion, may be due in part to the relatively higher congestion levels on competing arterial facilities. Figure 7 shows congestion levels for four functional classes of facility by urban area size group. Except in the largest urban areas (over 1 million population), other principal arterials experience higher congestion levels than freeways/expressways. The exception in the largest areas may be explained by the preference to use freeways/expressways for longer trips or the greater connectivity between widely separated locations afforded by these facilities.

Share of Travel, Supply, and Congestion Levels by Urban Development Density

The share of travel and share of supply of freeways/expressways are considerably lower in the suburbs than in either the urban core or the fringe of urbanized areas of all sizes. Figures 8 and 9 show these patterns of supply and use respectively. These data suggest that the currently perceived congestion problems in the suburbs are at least partly the result of lower levels of freeway/expressway supply, and correspondingly lower travel shares for freeways/expressways in the suburbs, when compared to other development density categories in the

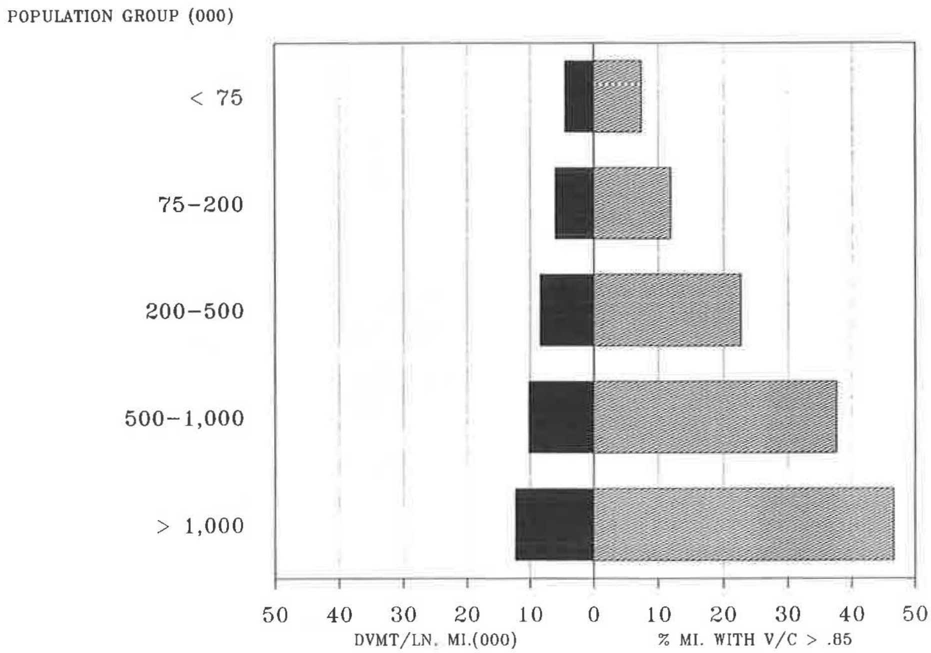


FIGURE 6 Freeway/expressway intensity of use and congestion levels by urbanized area size (1987).

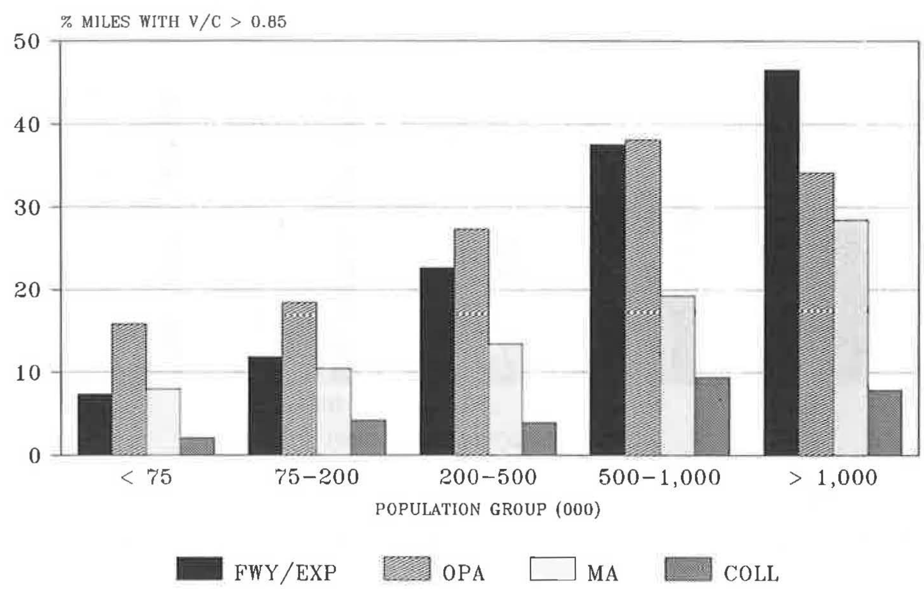


FIGURE 7 Level of service by functional class and urbanized area size (1987).

urbanized areas. This would imply that there is a potential to shift significant shares of travel in the suburbs to freeways/expressways if supply levels in the suburbs are increased to levels comparable to those in the core and fringe. Congestion levels on freeways/expressways (as shown in Figure 10) in the suburbs are comparable to those in the urban core, giving some credence to the perception of "suburban gridlock" in these areas.

The pattern of consistently increasing congestion on freeways/expressways as urbanized area size increases (Figure 10) is not matched by the pattern of congestion levels on other principal arterials (Figure 11). Arterials show a dropoff in congestion levels as urbanized area population increases beyond 1 million. An inference that can be drawn from these patterns is that there is a propensity for freeway/expressway travel in large areas despite the more severe congestion levels on these facilities relative to those on other principal arterials. One explanation may be that the more direct connection between widely separated locations afforded by the network of high-order facilities in the large urbanized areas is preferable to

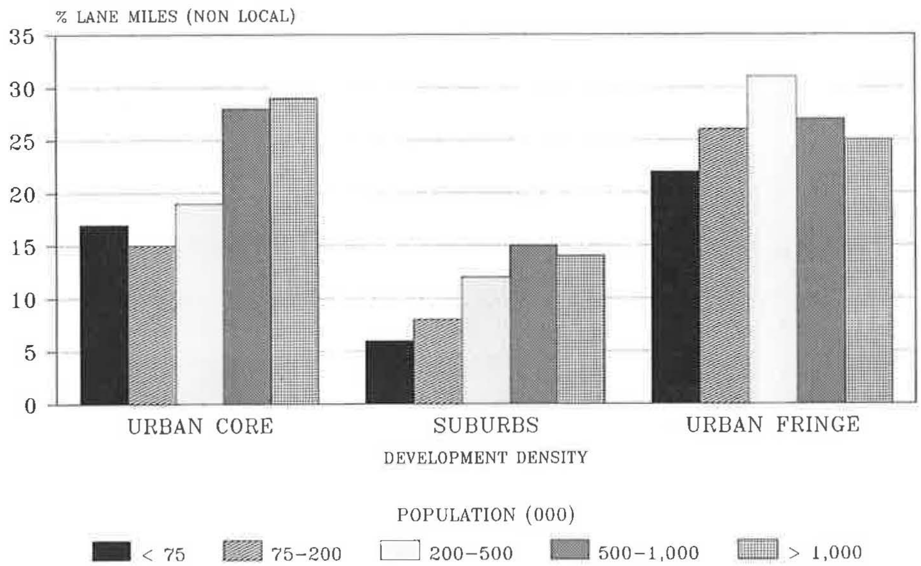


FIGURE 8 Freeway/expressway share of supply by development density (1987).

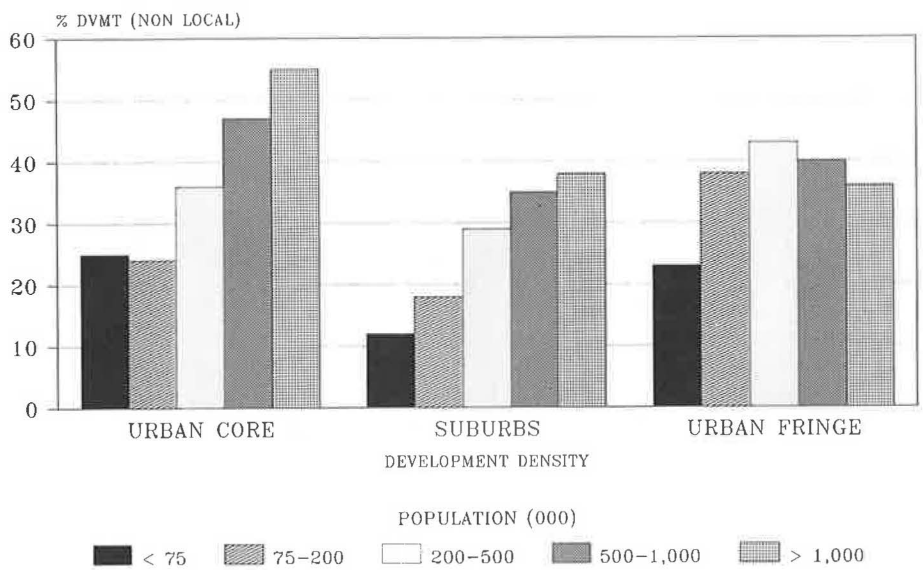


FIGURE 9 Freeway/expressway share of use by development density (1987).

the lower level of connectivity of other principal arterials, especially in light of the generally longer trip lengths as urbanized areas increase in size.

PART 2: ANALYSIS OF ECONOMIC EFFICIENCY OF NEW CAPACITY

Analysis Procedures

For this portion of the analysis, the four functional classes used earlier were reduced to the following three facility classes to conform with sources of data (*6*) used in the analysis: (1) freeways/expressways, (2) other principal arterials, and (3) collectors. The "collector" facility class combined the minor arterial and collector classes used earlier. The analysis was done for large (750,000 to 2 million population) urban areas for which nationwide average travel characteristics were available (*6*). Also, two alternative levels of service (LOS) were evaluated—LOS C and LOS D. New capacity additions for two urban development density categories—urban core and fringe—were compared. The effect of right-of-way availability was assessed by comparing two types of investment: new facilities on new rights-of-way, and widening of existing facilities.

The comparative economic efficiency of alternative functional classes was evaluated by estimating total cost per VMT

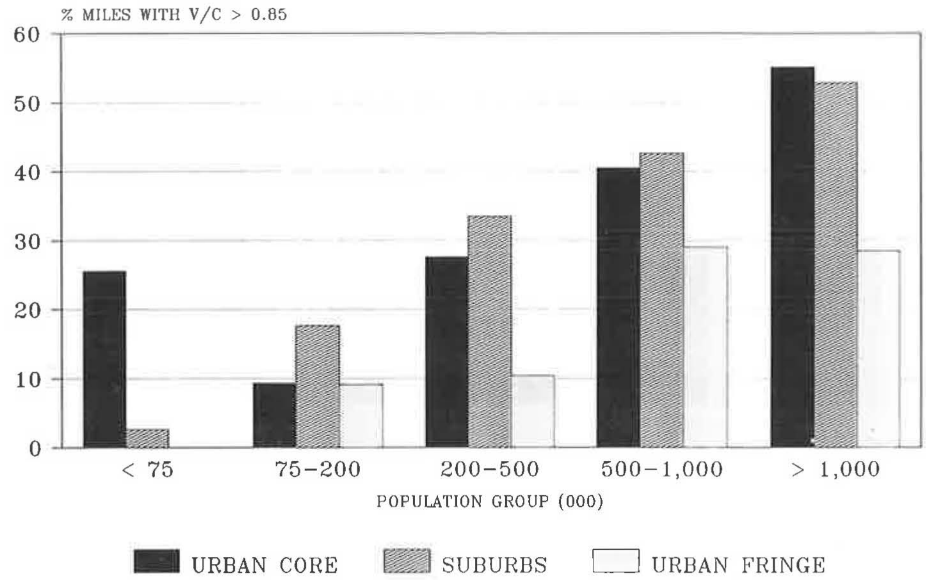


FIGURE 10 Level of service on freeways/expressways by development density (1987).

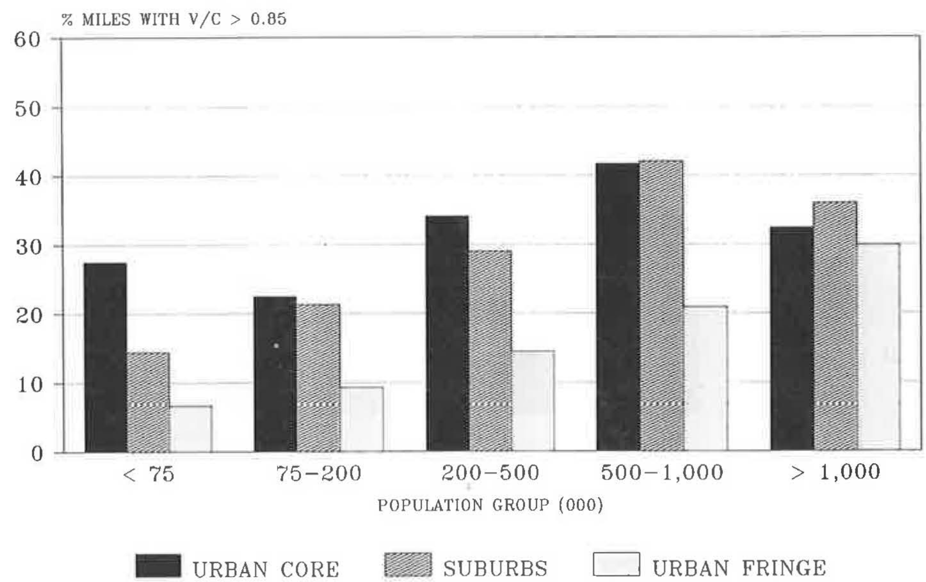


FIGURE 11 Level of service on other principal arterials by development density (1987).

served on each functional class. Total cost per VMT was obtained by aggregating highway facility costs per VMT and user costs per VMT. The lower the total cost per VMT, the greater is the economic efficiency of the facility class or level of service.

A basic assumption in the analysis is that the major part of a trip between any two zones of an urban area may be served by any of the three alternative facility types. An assumption is made that highway system volumes are currently at capacity conditions for acceptable LOS (either LOS C or LOS D, depending on the urban area's system performance standards). As trips between the zones increase, the decision maker may be faced with the question: Which facility type should be widened, or constructed on new alignment, assuming that the only consideration is economic efficiency? For example, if peak-hour vehicle trips from zone A to zone B increase by about 3,000, this increase may be served by 2 freeway lanes, or 4 arterial lanes, or 6 collector lanes.

Which investment would be the most efficient economically? The analysis assumes that the urban area's entire highway system currently operates at capacity and that the three facility type alternatives can be provided to serve areawide increases in travel between all pairs of zones. Lane requirements to serve projected increases in VMT can be developed on an areawide basis, if we assume that the new capacity to be provided will balance (i.e., be exactly equal to) the increase in volume of travel to be served. We may then compare the systemwide capacity improvement alternatives by simply using a single lane mile of each facility type to represent systemwide increases in capacity, and by using design service volume per lane of the facility to represent the volume of traffic served on the improvements. Costs per VMT served can then be estimated for each facility type to make comparisons between them.

The analysis procedure is presented in the flow chart in Figure 12. The procedure assumes, for purposes of illustration, that projected increases in traffic on the urban highway system will need to be served entirely by new capacity, either on new facilities or on widened existing facilities. Peak-hour traffic served by new or widened facilities is assumed to equal its hourly service volume at the selected LOS (C or D). Traffic served during the peak periods of the day and off-peak periods is then estimated based on time-of-day travel percentages and directional split percentages for the location within the urban area by urban area size, obtained from NCHRP 187 (*6*).

The total cost of travel over each lane-mile was obtained by adding estimates of user costs and facility costs. User costs considered were travel time, vehicle operating, and accident costs. Procedures used to estimate these user costs are documented elsewhere (*7–9*). Facility costs considered included life-cycle costs (i.e., annualized capital costs for a 20-year life at 10 percent interest) and facility operation and maintenance costs. The estimates of capital costs per lane-mile used in the analysis are presented in Table 2. Facility and user costs per VMT served were then calculated by dividing them by VMT served annually.

Inferences

Comparative Costs Per VMT—Large Urban Areas

Tables 3 and 4 present estimates of costs per VMT served on new capacity in large urban areas, at LOS D and at LOS C. The results indicate that:

- On average, freeways/expressways provide the most economically efficient mobility service for a given urban development density, level of service, and type of construction (i.e., new facility or widening) in large urban areas. The lowest total cost per VMT is 41.9 cents, on added freeway/expressway lanes for widening projects in the fringe, at LOS D.
- Generally, it is more economically efficient to design new capacity to operate at LOS C than at LOS D, with one excep-

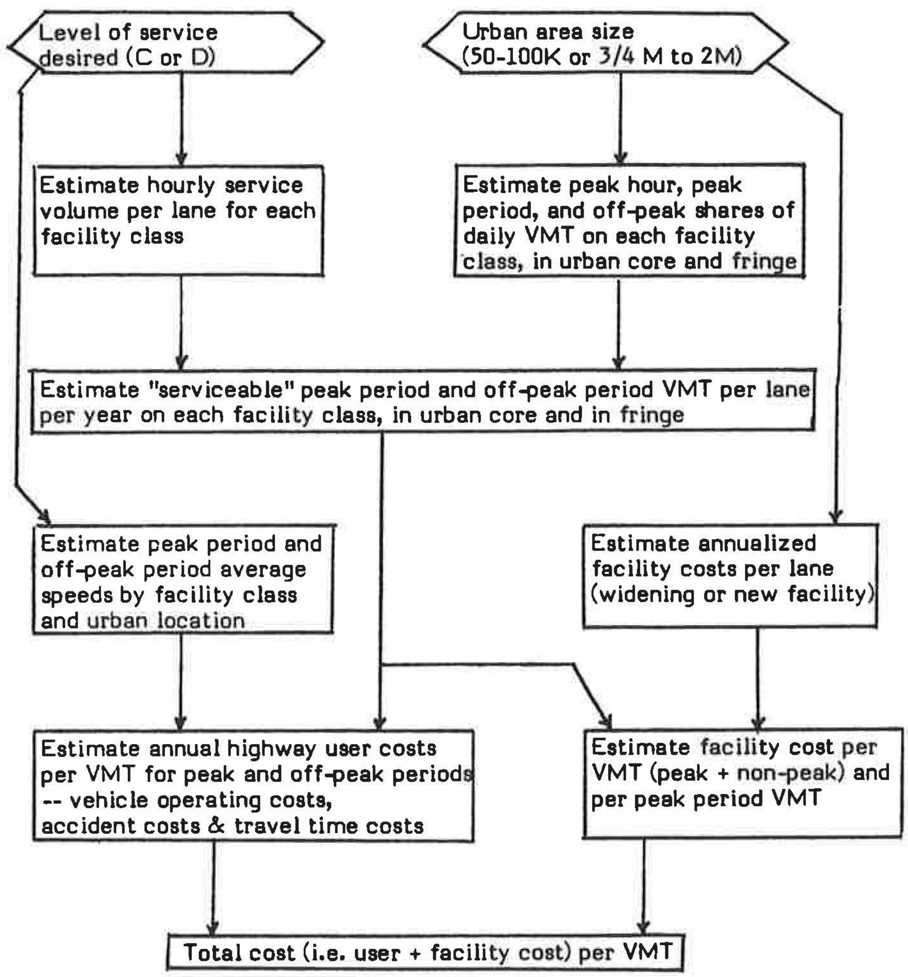


FIGURE 12 Economic analysis procedure.

TABLE 2 CAPITAL COSTS PER LANE-MILE FOR LARGE URBAN AREA (MILLIONS OF DOLLARS)

	New Facility 1/	Widening 2/
Core radial:		
Freeway:	10.10	5.0
Arterial	3.76	2.0
Collector	1.90	1.0
Fringe crosstown:		
Freeway	5.00	1.56
Arterial	2.00	0.74
Collector	1.00	0.29

1/ Estimated on the basis of nationwide new capacity cost estimates in Reference 2.

2/ Source: Reference 10

TABLE 3 COSTS IN THE CORES OF LARGE URBAN AREAS (CENTS PER VMT)

		Fwy/Exp	O.P.Art	Collector
New Facilities				
LOS D	User costs	44.6	94.2	128.1
	Facility costs	23.0	20.5	17.1
	Total	67.6	114.7	145.2
LOS C	User costs	40.8	80.2	100.5
	Facility costs	28.3	22.9	19.5
	Total	69.1	103.1	120.0
Widening				
LOS D	User costs	44.6	94.2	128.1
	Facility costs	11.5	11.0	9.1
	Total	56.1	105.2	137.2
LOS C	User costs	40.8	80.2	100.5
	Facility costs	14.1	12.3	10.4
	Total	54.9	92.5	110.9

TABLE 4 COSTS ON THE FRINGES OF LARGE URBAN AREAS (CENTS PER VMT)

		Fwy/Exp	O.P.Art	Collector
New Facilities				
LOS D	User costs	40.8	82.7	94.3
	Facility costs	9.6	9.0	8.0
	Total	50.4	91.7	102.3
LOS C	User costs	38.2	71.5	80.3
	Facility costs	11.4	10.1	9.1
	Total	49.6	81.6	89.4
Widening				
LOS D	User costs	40.8	82.7	94.3
	Facility costs	3.1	3.4	2.5
	Total	43.9	86.1	96.8
LOS C	User costs	38.2	71.5	80.3
	Facility costs	3.7	3.9	2.9
	Total	41.9	75.4	83.2

tion—new freeways and expressways provided in the urban core.

- Facility costs per VMT are a significant portion of total costs in the urban core (about 40 percent for new freeway/expressway facilities at LOS C), but are a smaller portion of total costs in the fringe (about 23 percent for new freeway/expressway facilities at LOS C).

Comparative Costs per VMT—Excluding Travel Time Costs

Since there is considerable disagreement regarding the translation of travel time savings into monetary benefits, the sensitivity of results to value of time was evaluated by excluding travel time costs from the previously computed costs per VMT. Table 5 indicates that, even if travel time costs are excluded, freeway/expressway costs per VMT are lower than those for lower-order facilities, irrespective of development density characteristics, right-of-way availability (i.e., new facilities vs. widening), and LOS at which the new capacity will operate.

PART 3: ESTIMATION OF NEEDED USER CHARGES

Analysis Procedures

If financing through user charges is desired, systemwide capacity improvements, such as the improvements considered in this economic analysis, may be financed by two basic types of user charges: (1) a motor fuel tax, or (2) road pricing. Road pricing may be implemented by using a variety of tools, such as area-wide licenses required for downtown travel in peak periods, conventional toll booths, or electronic pricing using automated vehicle identification (AVI) technology. Several trends are converging to make road pricing a serious consideration as a revenue source for expansion of transportation capacity (*12*). New technology will help enhance political acceptability of road pricing, as toll booths and the delays they usually involve are replaced by speedier automated collection.

In this part of the analysis, an attempt is made to provide some scale to the level of charges which would be needed to pay for new capacity under two pricing schemes—implementation of tolls either throughout the day or during peak periods only.

Where road pricing systems are feasible to recover the costs for providing expanded system capacity, the "fair" charges per VMT would be equal to the facility life-cycle costs per VMT. In cases where lanes are added to existing facilities, this amount may be divided by the total number of lanes after widening to get user charges, because users of *all* lanes (not just the added lane) can share in paying the cost. User charges for widened facilities were estimated assuming a typical urban highway widening from two lanes in each direction to three lanes in each direction, i.e. a 50 percent increase in capacity, which roughly corresponds to the approximately 50 percent increase in traffic projected by several urban areas over the next 20 years.

For comparison with user charges currently recovered from existing system users, the current average gasoline tax was converted to cents per VMT on each functional class as follows. Average state motor fuel tax receipts were estimated at 12 cents per gallon, based on 1986 adjusted net gallonage receipts, annual VMT, and average miles travelled per gallon of fuel consumed, from *Highway Statistics—1986* (*11*). Total tax receipts per gallon were then estimated at 21 cents, including the 9 cent federal tax. Actual user charges by facility type in cents per VMT were then estimated based on an estimated gas mileage of 30 miles per gallon (MPG) on urban freeways/expressways and 15 MPG on other principal arterials and collectors (*9*).

It is often argued that new capacity is usually needed only to serve peak-period users, and therefore the costs of adding new capacity should be allocated to peak users alone, even though the added lanes may be used both in peak as well as off-peak periods. Therefore, facility costs per peak-period VMT served were also computed, by dividing facility costs

TABLE 5 TOTAL COSTS EXCLUDING TRAVEL TIME COSTS (CENTS PER VMT)

		Fwy/Exp	O.P.Art	Collector
Large Urban Areas: Core				
New Facilities	LOS D	42.1	51.5	54.0
	LOS C	45.3	50.7	50.5
Widening	LOS D	30.6	42.0	46.0
	LOS C	31.2	40.1	41.5
Large Urban Areas: Fringe				
New Facilities	LOS D	28.5	39.4	39.1
	LOS C	29.1	37.6	36.9
Widening	LOS D	22.0	33.8	33.6
	LOS C	21.4	31.3	30.7

by peak VMT served on each facility type. The resulting costs per VMT are the needed user charges where road pricing is sought to be applied only to peak users.

Inferences

User Charges On Improved Facilities Only

Table 6 focuses on needed user charges per VMT for LOS D service in large urban areas when only improved facilities are tolled. The estimates indicate that user charges per VMT generally increase with facility class and are lower in the fringe than in the city center. Where road pricing is feasible and all users are charged, freeway/expressway user charges at design LOS D would range from 1.0 cent per VMT for a typical facility widening in the fringe to 23.0 cents per VMT for a typical new facility in the core. Where road pricing is feasible and is applied only to peak-period users, the new capacity user charges per VMT are more than double the charges when all users share the costs. These figures indicate the magnitude of the cross-subsidy between off-peak users and peak-period users of expanded highway facilities in cases when facility expansion is actually needed only to serve peak-period users.

Table 6 also presents user charges that are actually generated by the current gasoline tax in terms of cents per VMT on the facility. The relatively low levels of actual charges under the current tax structure indicate that users of new capacity (generally peak-period users) are heavily subsidized by users of existing capacity (generally off-peak users). Needed user charges for peak-period travel on widened freeways/expressways in the core of large urban areas, for LOS D design, could be as high as 8.9 cents per VMT when the charges are applied only to peak-period users. On the other hand, under the current tax structure, less than 1 cent per VMT would be recovered. In the fringe, the comparable costs are 2.4 cents per VMT with only peak-period users sharing costs, versus less than 1.0 cent per VMT under the current tax structure.

CONCLUSIONS

The analysis of current supply and use of urban highway facilities suggests that there is a tendency for freeways/expressways to carry a larger share of travel as travel demand and urban area size increase. However, relatively lower levels of freeway/expressway supply in the largest urban areas appear to be constraining further increases in share of travel on such facilities. The potential to shift significant shares of travel to freeways/expressways is greater in the largest urban areas and in the suburbs.

The economic analysis of alternative new capacity investments indicates that, given the extent and size of the freeway system currently in place, it is usually more cost-efficient to provide new highway capacity on freeways/expressways. Also it is usually more economically efficient to design new capacity to operate at LOS C rather than at LOS D.

The results from the analysis of needed user charges suggest that travellers for whom the existing system provides adequate service (mainly off-peak users) heavily subsidize travellers who need new capacity (mainly peak-period users). Where road pricing is feasible on improved facilities only, the needed user charge in the typical situation (e.g., widening a four-lane freeway to six lanes in the urban fringe to provide LOS D service) would be about 1.0 cent per vehicle mile if both peak and off-peak users share the costs, and 2.4 cents per vehicle

TABLE 6 USER CHARGES NEEDED ON IMPROVED FACILITIES IN LARGE URBAN AREAS AT LOS D (CENTS PER VMT)

		Fwy/Exp	O.P.Art	Collector
New Facilities				
Core:	All users charged	23.0	20.5	17.1
	Only peak users charged	53.2	50.7	37.8
Fringe:	All users charged	9.6	9.0	8.0
	Only peak users charged	22.1	23.8	19.3
Widening				
Core:	All users charged	3.8	3.7	3.0
	Only peak users charged	8.9	9.1	6.7
Fringe:	All users charged	1.0	1.1	0.8
	Only peak users charged	2.4	3.0	2.0
Charges with current tax structure				
Core and fringe		0.7	1.4	1.4

mile if only peak-period users bear the costs. On the other hand, the needed user charge to pay for a *new* freeway in the fringe of a large urban area could be as high as 9.6 cents per vehicle mile if both peak and off-peak users share costs, and 22.1 cents per vehicle mile if only peak-period users bear the costs.

Implications for Urban Highway Policy

The analysis and inferences presented in this paper suggest that, if public acceptance can be achieved, freeways/expressways have a powerful role to play in plans to expand urban highway system capacity, especially in heavily congested large urban areas, because of their superior economic efficiency relative to lower-order functional classes. However, financing such improvements using the current tax structure, which relies heavily on motor fuel taxes, would not be equitable towards users who use the highway system during off-peak periods, when existing highway system capacity would suffice. Consequently, financing mechanisms that rely on road pricing, possibly using AVI technology, should be investigated in plans to expand urban highway systems. AVI technology will help enhance political acceptability of road pricing as toll booths and the delays they involve are replaced by speedier automated collection. This technology will also facilitate the implementation of differential tolls by time of day. Availability of public transit options and preferential pricing for high occupancy vehicles (HOV) could make pricing more acceptable politically, while facilitating shifts in travel mode.

Recommendations for Further Research

The economic analysis focused on traditional types of highway facilities. It excluded consideration of HOV facilities, transitways, and non-highway type facilities (e.g., light rail, heavy rail, bikeways, etc.). Further research is needed to make comparisons with such alternative investment types. Also, the sensitivity of the results to some of the simplifying assumptions made for the analysis need to the investigated.

The analysis of supply, use, and level of service was based on data aggregated by urbanized area size group. Individual urbanized areas were not analyzed. The patterns, trends, and relationships developed in this study need to be verified based on more detailed data from individual case study urbanized areas.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Carol J. Roadifer, Highway Engineering Trainee, FHWA, for her fine work in constructing the spreadsheet for the economic efficiency analysis, and to Sharon Watkins of FHWA's Office of Planning for cross-tabulating and plotting the HPMS data.

REFERENCES

1. L. Mertz. *Thoughts on the Future of The Highway Program.* Prepared for the Federal Highway Administration, January 27, 1988.
2. *Overall Highway System Requirements and Investment Levels: The Future National Highway Program 1991 and Beyond.* Working Paper No. 14. FHWA, U.S. Department of Transportation, Washington, D.C., October 1987.
3. *Highway Performance and Investment Analysis: The Future National Highway Program 1991 and Beyond.* Working Paper No. 13. FHWA, U.S. Department of Transportation, Washington, D.C., October 1987.
4. C. Fleet and P. DeCorla-Souza. Supply and Use of the Nation's Urban Highways. Presented at the Second Conference on Transportation Planning Applications, Orlando, Florida, April 1989.
5. *Highway Performance Monitoring System Field Manual.* FHWA, U.S. Department of Transportation, Washington, D.C., January 1984.
6. A. B. Sosslau et al. *NCHRP Report 187: Quick-Response Urban Travel Estimation Techniques and Transferable Parameters, User's Guide.* TRB, National Research Council, Washington, D.C., 1978.
7. *A Manual on User Benefit Analysis of Highway and Bus Transit Improvements.* American Association of State Highway and Transportation Officials, Washington, D.C., 1977.
8. *Vehicle Operating Costs, Fuel Consumption, and Pavement Type and Congestion Factors.* FHWA, U.S. Department of Transportation, Washington, D.C., June 1982.
9. *An Evaluation of the Economic Efficiency of New Highway Capacity on Alternative Facility Classes, Technical Appendix.* FHWA, U.S. Department of Transportation, Washington, D.C., November 1988.
10. *Highway Performance Monitoring System Analytical Process, Vol. II—Version 2.1, Technical Manual.* FHWA, U.S. Department of Transportation, Washington, D.C., Dec. 1987.
11. *Highway Statistics—1986.* FHWA, U.S. Department of Transportation, Washington, D.C., 1986.
12. K. A. Small et al. *Road Work.* The Brookings Institution, Washington, D.C., 1989.

Publication of this paper sponsored by Task Force on Transportation Planning Applications.

Transportation Planning Methods for Improving Mobility in Developing Activity Centers in Orange County, California

CHRISTINE HUARD-SPENCER

The Orange County Transit District (OCTD) has a three-phase proactive planning process for developing transportation demand and systems management (TDM and TSM) actions at suburban activity centers, and implementing them with assistance from the private sector. The first phase involves coordinating with city planners and project developers to determine needed transit amenities (turnouts, shelters, pedestrian access, transit center) and preferential facilities for ridesharers, and then integrating these amenities in the project development plans through the local jurisdiction development conditions for the project. The second phase, formation of a transportation management association (TMA), is based on activity center employer and employee surveys. The key objective is to assist employers and the TMA (once formed) in planning and implementing various TDM and TSM strategies for the activity center. The third phase involves analysis of employee travel characteristics and a determination of transportation infrastructure and service deficiencies within the activity center, based on an analysis of existing and future travel demand. The demand methodology used involves locally developed procedures and models run on an in-house microcomputer system. The travel forecasting results are used to plan capital facilities including transitways, HOV/transit access ramps, and park-and-rides, and to develop transit services for the activity center.

Orange County is located in Southern California, between Los Angeles and San Diego, and covers approximately 750 square miles. In 1989, the County had approximately 2 million people and 1.4 million jobs. The County is a typical suburban environment, with a number of activity centers that approach the density and total employment of traditional central business districts, surrounded by areas of low density residential and other development. The County has ten major activity centers located within one mile of an existing freeway, with several of the activity centers located near two or more freeways. All the centers are heavily dependent on these freeways to provide access for employees, business patrons, and deliveries.

The Orange County Transit District (OCTD) is the countywide transit operator for Orange County and offers a wide range of services, including:

- local and express bus service;
- Dial-A-Ride (curb-to-curb) service; and
- commuter transportation services such as rideshare matching, vanpool organization and seed fleet vehicles, and development of employer-based strategies such as alternative work hours and telecommuting.

OCTD focuses on activity centers from three different levels in its efforts to improve and maintain mobility in Orange County:

1. coordination of private land development activities and proposals, and local jurisdiction public works projects, with services and facilities needs for transit and other high occupancy modes;
2. development and support of Transportation Management Associations (TMAs) in the major activity centers; and
3. long-range infrastructure and facilities planning focused on the freeways serving the major activity centers.

This paper focuses on OCTD's activities within the South Coast Metro Activity Center (the Metro) for two reasons:

1. The Metro is an established multiuse center that is still experiencing significant levels of new development in office, retail, and mixed uses. The Metro currently has in excess of 1,000 employers, with over 25,500 employees, and experiences significant peak-hour congestion on both local streets and the adjacent freeways.
2. All three levels of OCTD's planning process have been successfully implemented in the Metro area.

COORDINATION OF TRANSIT SERVICES AND FACILITIES NEEDS WITH LAND DEVELOPMENT ACTIVITIES AND PUBLIC WORKS PROJECTS

OCTD has several goals in coordinating its service and facility needs with private land development and public works projects proposed in Orange County:

- To ensure that needed physical facilities to support alternative travel modes such as bus, carpooling and vanpooling are incorporated in proposed development projects. As used in this paper, transit is defined to encompass all bus modes

P&D Technologies, P.O. Box 5367, Orange, Calif. 92613-5367. Former affiliation: Orange County Transit District, P.O. Box 3005, Garden Grove, Calif. 92648.

(public and private) and high-occupancy modes such as carpooling and vanpooling.

• To begin conceptual development of the need for a TMA or other activity center-oriented support systems, and to ensure participation in existing TMA type activities in an established activity center.

• To plan for future infrastructure improvements and to ensure the incorporation of design controls to protect opportunities to implement preferential facilities in the future.

OCTD has developed a straightforward process for coordination of its transit needs with proposed land use and public works projects. The process includes the following major steps:

• Identification of the legal authority for a transit agency to be involved in local land use planning. In California, the California Environmental Quality Act (CEQA) gives local agencies such as OCTD the authority and responsibility to comment on the effect of proposed land uses on each agency's services and/or facilities and their ability to continue to function effectively once the proposed project is implemented. The mitigation of any impacts on the agency generated by the proposed project is included as part of the final environmental finding for the project.

• Seeking an expanded role with the appropriate local jurisdiction, to cover types of land use plans not covered by CEQA, such as site plans and street improvement designs. The roles and responsibilities of the local jurisdiction and OCTD in this expanded review process can be detailed in a memorandum of understanding (MOU) or in a less formal manner.

• Involvement in the local land use proposal review process, including review of land use proposals for potential impacts on transit, submittal of formal comments on the proposal describing the impacts and the appropriate mitigation, and the incorporation of these comments in the conditions placed on the development proposal by the local jurisdiction.

• Development of professional working relationships with planning and public works staffs in the affected local jurisdictions to ensure that the process works smoothly with minimum delay to the proposed project.

OCTD considers a number of topical areas when reviewing proposed land use plans, including:

• The need for facilities to support a broad range of alternative travel modes, including:

—amenities for both public and private bus services, including shelters, benches, paved passenger waiting areas, pedestrian access to adjacent uses, lighting, turnouts, concrete bus pads, information signs and handicapped accessible ramps;

—amenities for other high-occupancy modes, including park-and-ride spaces (in residential areas), rider meeting and waiting areas, reserved parking spaces at work sites, and on-site services for employees;

—amenities for bicycle commuters, including racks and lockers, bicycle lanes and protected access to work sites, showers, and clothing lockers; and

—pedestrian amenities including lighted, paved, handicapped-accessible walkways between project buildings and adjacent uses (restaurants, banks etc.), and bus stops.

• The need to develop a TMA or for tenants of a proposed project to join an existing TMA.

• The need for design provisions to accommodate any long-term infrastructure improvements such as transitway access/egress ramps or local street preferential facilities.

The District conducts a number of activities to directly support this process:

• Assignment of staff to review and comment on proposed land use plans.

• Development and wide distribution of *Design Guidelines for Bus Facilities* to ensure that all facilities provided as part of private or public development projects are designed and built to standards for the types of public and private transit vehicles operated in Orange County. *Guidelines* is approximately 60 pages long and several thousand copies have been distributed at no cost to city staffs, developers, architects, planners, and engineers in Orange County.

• Development and wide distribution of *Consideration of Transit in Project Development*, a brochure that describes the benefits of coordinated planning in a straightforward and easy to understand manner for nonplanners, such as city councilpersons, other elected officials, and company executive officers. Several thousand copies of this document have been distributed and have been very positively received.

• Conduct of seminars on facilities planning, as part of a larger seminar series on alternative commute modes, largely focussing on major employers. The seminars provide a concise summary of facilities needs assessment, planning, and implementation in a nontechnical manner appropriate for audiences with nonengineering and nontechnical backgrounds.

The land use review and coordination process has been successful in the provision of facilities to support high-occupancy modes in the major activity centers for a number of reasons:

• OCTD staff have formed strong relationships with the local jurisdiction planning, engineering and public works staffs, developers, and other professionals in the land development process, resulting in open and trusting interactions among the participants in the process.

• OCTD has consistently provided clear, concise comments within the mandated project time frames, thereby minimizing any delay to proposed projects.

• The availability of the *Design Guidelines* ensures that designers have ready and convenient access to the appropriate design standards for needed amenities.

• The availability of the *Consideration of Transit* brochure has provided useful explanations to local decision makers to solicit their support and approval of needed HOV amenities as part of proposed land use projects.

DEVELOPMENT OF TRANSPORTATION MANAGEMENT ASSOCIATIONS: HELPING THE PRIVATE SECTOR TO HELP ITSELF

The next level in the OCTD planning process to improve mobility in activity centers is the development of TMAs, in order to focus the limited resources of the OCTD Commuter

Network Department in the high-density activity centers in Orange County. Commuter Network provides a wide range of employer-based services, including rideshare matching, technical expertise and support for specialized program development, and the development and early technical and staff support for TMAs. The goals of the OCTD TMA development program include:

- to organize activity center employers to participate in transportation demand management (TDM) planning through TMAs or other related groups;
- to assist employers and/or the TMA in developing goals and in implementing selected TDM strategies;
- to work with city and county planning agencies to include them in the TDM planning efforts and to assist them in developing and implementing municipal TDM measures; and
- to coordinate TDM plans in the activity centers with other transportation strategies planned by all the regional, county, and local agencies for major transportation corridors in Orange County.

The first step in the development of a TMA at the South Coast Metro was the development and conduct of a pre-TMA survey. The survey was intended to accomplish a number of purposes related to the desire to maximize the effectiveness of efforts in the South Coast Metro area, including:

- to provide reliable estimates of current employee commuting behavior at the activity center;
- to provide information on current employer initiatives and support concerning employee transportation and alternative trip modes; and
- to assess the employee and employer market potential for various TDM and transportation system management (TSM) techniques, including carpooling, vanpooling, alternative work hours, telecommuting, and parking management.

Three types of survey instruments, with different objectives for each survey, were used:

1. A written employee survey form, which was distributed to employees at their work sites. The survey objectives included collection of accurate data on current work trip characteristics (such as origin, destination, travel time, trip distance), work schedules, willingness to consider alternative trip modes, and employee need for an automobile before, during, and after work. Data were collected from a representative sample of employees (with a 56 percent response rate), which was approximately 10 percent of the workforce in the activity center.
2. A written employer survey form, which was distributed to employers at their work sites. The objectives of the survey included development of a profile of employers, including work schedule policies, parking availability and costs, availability of on-site services, and current ridesharing incentives.
3. Face-to-face interviews with company executives and senior management conducted by OCTD and consultant staff. The objectives of these interviews were to ascertain the perception by senior management of the area traffic conditions, the effect of traffic on the organization's ability to conduct business, and the willingness of the employer to participate in an areawide cooperative effort to help solve traffic problems. These interviews were conducted with 24 of the largest employers in the South Coast Metro Activity Center.

The survey results were used in a number of ways:

- A TMA, a public (OCTD) and private (South Coast Metro Alliance and the Executive Task Force) joint venture, was formed in late 1986.
- The preliminary programs the TMA would pursue in the areas of both transportation systems and demand management were developed.
- Express bus service was instituted, with marketing focussed on those candidate employees most likely to use transit.
- The new programs in the Metro, such as on-site services, parking management, a guaranteed return trip program, and intercompany vanpooling, were planned.

LONG-RANGE INFRASTRUCTURE IMPROVEMENTS: TRANSITWAY PLANNING ACTIVITIES

As part of its long-range planning activities, OCTD has developed a countrywide Transitway Development Program to implement and promote the use of a freeway-based system of commuter lanes and transitways, focussing on the major Orange County activity centers, including the South Coast Metro. The proposed system includes approximately 110 miles of preferential facilities for buses, carpools and vanpools. These facilities are being evaluated and implemented by OCTD and the California Department of Transportation (Caltrans).

A key part of the concept design activities conducted by OCTD for the transitway system was the development and implementation of a travel forecasting methodology sensitive to changes in corridor level and site-specific characteristics, such as the ability of transitways to provide higher levels of service to activity centers like the South Coast Metro. The forecasting methodology used the 1980 U.S. Census Urban Transportation Planning Package (UTPP) for the base data set. Forecast year trip totals were built up with an iterative distribution process constrained by the adopted Orange County growth forecasts for the origin and destination market areas.

For the model, the transit and HOV mode splits were primarily determined based on the degree of travel time savings that commute trips would achieve by using preferential facilities in the morning peak hour versus using the mixed flow freeway lanes. Origin and destination area characteristics, such as employment density and type, also affected the modal share.

The UTPP Base/Socio-Economic Growth Approach modeling effort included a number of very specific tasks:

- Task 1: Review larger scale system level analysis assumptions, process, and output including the zone system, speed assumptions, prior mode assumptions, and transit and HOV mode split factors.
- Task 2: Establish more specific zone systems and design an origin-destination matrix for microcomputer analysis.
- Task 3: Establish base data files and determine background assumptions, including aggregation of the UTPP files

on total person trips, 2-person carpools and 3-person carpools to the zone system, and development of background assumptions based on empirical travel behavior data.

• Task 4: Estimation of travel time savings on the transitway versus the freeway for all trip interchanges. Travel time savings were computed based on the higher speeds on the transitways, on preferential facilities through major interchanges and on exclusive ramps into the activity centers.

• Task 5: Deletion of origin-destination cells with short trip lengths, less than 7 miles, and production of 1980 trip tables.

• Task 6: Estimation of HOV person-trips on the transitways and commuter lanes. Total HOV person-trips were computed based on factors such as the amount of travel time savings, the degree of travel time savings, and nonwork HOV travel.

• Task 7: Production of year 2000 and 2010 person-trip tables, based on increasing the UTPP 1980 person-trip table by the adopted socio-economic growth factors for Orange County.

• Task 8: Estimation and assignment of years 2000 and 2010 HOV person-trips on the transitways. Trip totals were assigned using a microcomputer assignment application developed by OCTD staff using macroequations on a LOTUS 1-2-3 spreadsheet program.

• Task 9: Estimation of years 2000 and 2010 transit usage on the transitway. Origins and destinations that could potentially be served by transit were identified, analyzed based on various mode splits, and those cells with greater than 150 peak-period trips were assigned to express bus transit.

The modelling process resulted in the following estimates of transit and HOV demand for the system of freeway preferential facilities:

• Overall summary of HOV projections for the entire system of proposed preferential facilities in Orange County.

• Comparison of demand generated under different occupancy restrictions (2- versus 3-person HOVs) on the preferential facilities.

• The destination/activity center level analysis output was used to help determine locations for direct HOV access ramps and characteristics of facilities in the transitway system.

The development and evaluation of potential access locations along the transitway system was conducted as a separate planning element of the overall OCTD Transitway Concept Design Studies. For the access element, the goal was to develop and evaluate conceptual physical connections between the transitway (within the freeway corridors) and the major activity centers adjacent to those segments of the freeway/transitway. The study effort was conducted in three phases:

1. The first level of analysis identified possible access points between the transitway and the adjacent general purpose travel lanes. This type of access would require HOVs to travel across the general purpose lanes and merge with the transitway at locations identified for this type of merge movement.

2. The next level of analysis identified those locations where a transitway and commuter lane would meet, and the need for a connection between the transitway and commuter lane to allow HOVs to continue to experience the travel benefits of the transitway, without having to physically exit the transitway and then separately enter the commuter lane.

3. Finally, the analysis considered the need for separate and direct access between the transitways and the adjacent activity centers. This phase included identification and evaluation of alternative access/egress designs, coordination with the affected local jurisdictions, identification of land use coordination issues, impacts on arterials and local streets, and potential preferential treatments for arterials and local streets. Factors in identifying possible ramp locations included:

—understanding and knowledge of the existing activity centers and the potential changes in the activity center over time;

—previous OCTD studies on preferential access in activity centers, including consideration of local street impacts and possible preferential treatments;

—coordination with Caltrans and local jurisdiction staff;

—commitment to minimize the impact to existing interchanges and local street systems.

This ramp analysis was a fatal flaw analysis based on four factors:

1. the ability of the location to meet the estimated transitway demand for the identified activity center;

2. the ability of the arterials and interchanges to accommodate a ramp structure and traffic volumes;

3. the complexity and extensiveness of design problems; and

4. the proximity of the proposed access/egress ramp to existing freeway to freeway interchanges.

Based on this analysis, two of five possible locations in the South Coast Metro area were carried forward for further consideration, which included:

• development of alternative design concepts for each access location, based on ability to meet demand, construction costs, right-of-way requirements and environmental impacts; and

• review of design and ramp concepts with Caltrans and the affected local jurisdictions.

OCTD has begun an Alternatives Analysis/Environmental Assessment (AA/EA) for the two ramp locations identified in the Concept Design Study. The AA/EA is expected to be completed in late 1990.

SUMMARY

To develop a planning methodology responsive to the needs of suburban activity centers that approach the density and total employment of more traditional central business districts, the OCTD:

• considered the overall ability of transit to serve the activity centers, patrons, and employees effectively;

• evaluated and began implementation of a broad range of transit services, including local and express bus (both public and private), carpool, vanpool, bike, and pedestrian modes within the activity center;

- considered both the origin and destination ends of the employee's commute trip, to ensure that facilities and services are in place for the user markets; and
- continued to provide technical staff and expertise to employers and other public agencies, as well as educating the general public, to ensure that provided services and facilities meet the users' needs.

REFERENCES

1. *Design Guidelines For Bus Facilities*, 2d ed. Orange County Transit District, Garden Grove, Calif., November 1987.
2. *Consideration of Transit in Project Development*, 2d ed. Orange County Transit District, Garden Grove, Calif., December 1988.
3. *Forecasting Usage on Proposed High Occupancy Vehicle Facilities in Orange County, California*. Orange County Transit District, Garden Grove, Calif., July 1987.
4. *Comparison of Transportation Demand Management (TDM) Market Research Results and Transportation Management Association Development in Three Suburban Activity Centers*. Orange County Transit District, Garden Grove, Calif., 1988.

Publication of this paper sponsored by Task Force on Transportation Planning Applications.

Developing Defensible Transportation Impact Fees

WILLIAM B. MOORE AND THOMAS MULLER

Rapid growth in many parts of the country, combined with budgetary pressures at the federal and state levels, has increased the importance of pursuing alternative funding sources for the construction of transportation infrastructure, particularly roads. The belief that development should pay its own way is becoming prevalent in many communities. Transportation impact fees are being used to address both of these issues. An impact fee affects three groups: developers, new residents, and existing residents. Each group has an interest in ensuring that a proposed fee is affecting it equitably. Developers and new residents do not want to pay more than their share of infrastructure requirements and, conversely, existing residents do not want to subsidize growth. A methodology for developing a defensible transportation impact fee must address a number of issues. Among these are: quantifying the benefits that are derived from new transportation infrastructure; identifying the recipients of the benefits; and calculating the size of an equitable impact fee. A methodology is described that employs an equity-based approach utilizing net-present-value techniques and addresses the legal criteria of "rational nexus."

Impact fees (often referred to as exactions or development fees) had their origin in California during the economic boom that followed World War II. By the late 1970s, impact fees were being used by numerous jurisdictions, particularly in California and Florida. During the 1980s, additional states passed legislation authorizing local communities to impose impact fees. In 1989, for example, the conservative Virginia legislature gave certain urban counties permission to impose, after July 1, 1990, impact fees for roads. The legislation also allowed these counties to add a surtax to the state income tax for the explicit purpose of funding new road projects. Vermont, one of the most rural states in the nation, also passed impact fee legislation to become effective on July 1, 1989.

Although there is some concern that impact fees have been abused in certain communities, such fees, particularly for roads, are gaining acceptance. During the 1980s, numerous urban areas, particularly outer suburbs, have experienced rapid growth in traffic volumes. Local electorates are aroused by these higher volumes and the accompanying congestion. These localities are anxious to find new means of financing their infrastructure requirements. Increasing long-term municipal debt is unpopular with voters, who fear such debt will result in higher property taxes. Transportation impact fees are becoming an attractive alternative or supplement to traditional debt financing.

In the following sections, we discuss key issues, describe a methodology for calculating benefits, address the adequacy of existing revenue flows, show how the "lumpiness" of road projects can be treated, and, in the final section, present a model for implementing the methodology.

W. B. Moore, Logistics Management Institute, 6400 Goldsboro Rd., Bethesda, Md. 20817-5886. T. Muller, META Consulting Group, 3301 Mantua Dr., Fairfax, Va. 22031.

KEY ISSUES

Transportation impact fees are becoming the largest single source of local revenue from developers and, for developers, the largest infrastructure cost outside the boundaries of their own projects. In most communities, developers are required to construct water lines, sewer lines, and streets to serve new housing or commercial space they are constructing and to dedicate these facilities to the locality following construction. However, charging for transportation improvements outside the confines of their project is a relatively new phenomenon outside California and Florida.

The recent proliferation of transportation impact fees raises numerous issues such as legal concerns, issues relating to equity, and the establishment of impact fee area boundaries, which must be addressed if they are to be used successfully.

Legal

In all states, including California and Florida, there has been recognition that to apply impact fees, the community must be able to demonstrate that the fee will directly benefit those asked to pay the fee. This is the so-called "rational nexus" that distinguishes legally between a fee and a tax. Were a developer to demonstrate that the entire community benefits equally from an impact fee, or that the community cannot distinguish between benefits received by existing community residents and a proposed development, the fee would be considered a tax and would have to be levied on all individuals or businesses. Further, such a tax could not usually be levied without authorizing state legislation.

The legal tests vary significantly on the degree of linkage that has to be demonstrated between the payment of a fee and the benefits derived. A very strict test, that a fee is "specifically and uniquely attributable" to a project, is required in New York State for the imposition of transportation (and other) impact fees. On the other hand, California requires only that a "reasonable relationship to the public welfare" be demonstrated. The degree of linkage varies in other instances as well. For example, in Montgomery County, Md., payment of impact fees allows a developer to advance his project through the subdivision approval stage because the fees indirectly expand the road system capacity in the immediate vicinity of his project. This is accomplished by giving the area funding priority in the capital budget.

Impact fees for roads typically require, as in Vermont, that a level of service be established. Further, such a fee cannot be imposed when existing facilities are sufficient to accommodate new development. Thus, if a new project only absorbs excess capacity, an impact fee may not be legal, at least in some states. Nor can such fees be imposed on new development to expand existing roads to meet prior demand.

Impact fees can also be challenged constitutionally on the basis of equal protection. However, courts have usually not accepted the argument that, because developments prior to the imposition of the fee did not have to pay their fair share of road costs, its imposition is a violation of the equal protection clause. Regardless of the legal test, impact fees must be expressly earmarked and spent for the purpose for which they are charged. If funds are not expended within a reasonable period, such as 6 to 10 years, they frequently must be refunded along with accumulated interest.

Equity

Equity has both legal and economic implications. Legally, any impact fee methodology has to take this concept into account. A sound fee must allocate costs reasonably across all users based on benefits received. Nonetheless, in most instances, communities have the power to exempt certain types of development from transportation impact fees if it is in the public interest. In Vermont, the state allows localities to exempt affordable housing from impact fees, the retention of existing employment, or the generation of new employment. This broad exemption policy could be interpreted to include virtually all commercial and industrial development, as well as moderate income housing. In reality, it is unlikely that localities will find it politically acceptable to exempt commercial developers unless there are significant public benefits from the application of such a policy. In Virginia, however, legislation is silent on the issue of exemptions.

The development and application of a methodology that derives an equitable impact fee have several advantages. Perhaps the most important is to reduce the likelihood of costly litigation. If a substantial fee is levied and perceived to be inequitable, its application may be challenged in court.

Establishing equity requires a sophisticated process that incorporates methods to ensure the absence of "double payments," as well as "double counting." Another major concern is the derivation of costs directly linked to the traffic associated with a project. In an effort to address these concerns, some communities provide developers with the option of undertaking their own impact analysis. If this independent study derives a fee that is more equitable than the established standard, the community may accept the alternative impact fee. To avoid lengthy equity disputes, most communities apply impact fees that recover only a portion of the calculated cost. The process of discounting costs has the merit of discouraging litigation, but may not provide the revenue necessary to meet needed infrastructure expansion. Nonetheless, communities believe that the present state of the art of impact fee development is less than an accepted science. Communities believe that a conservative approach will reduce the risk of litigation.

The objective of our methodology is to create a process that will establish impact fees that meet legal tests and are also equitable to all parties.

Impact Fee Zones

The first test a community has to meet in establishing an impact fee is that there is a deficiency in traffic capacity, and that this deficiency cannot be met from existing revenue sources. In Massachusetts, a court ruled that a jurisdiction has to be divided into impact fee zones that localize benefits to meet this criteria.

The methods employed to establish geographic areas designated as "impact fee zones" vary. A few communities designate their jurisdiction political boundaries as the zone. This approach, however, may not localize benefits, because any development within the community is subject to transportation impact fees. Geographically small communities that are in their early stages of growth are the most likely to designate their boundaries as impact fee collection zones. Montgomery County, Md., selects only those areas that have substantial deficiencies in areawide traffic capacity. These are areas that will require significant levels of road construction in future years. Typically, such areas have considerable unimproved land currently zoned for development; and to accommodate this construction, the road system would have to be expanded. In these areas, current levels of service (LOS) standards are already below acceptable standards. Therefore, additional development, in the absence of new road projects, would further deteriorate traffic conditions. Freeway construction projects serving primarily through traffic are not subject to payments from local benefit assessment fees. These projects are generally financed exclusively by the public sector.

BENEFIT CALCULATIONS

The first step in the calculation of benefits is to establish LOS standards and project transportation demand by land use and category. If the analysis finds that one project, or a group of likely developments over an extended time period (e.g., 20 years) will cause the LOS to be exceeded, a set of benefit calculations have to be initiated to establish that the impact fee meets legal, equity, and other criteria. Thus, the second step in the process is to estimate the cost, over an appropriate period, of bringing the road network to the established LOS given the projected new development.

Establishing LOS and Development Projections by Land Use

An early task in determining a transportation impact fee is to establish a LOS standard for the areas designated as "impact fee zones." The LOS a community finds acceptable varies and often reflects their present experience. In low-density communities outside large metropolitan areas, even service level "C" may be considered unsatisfactory. In higher-density metropolitan areas, levels "D" or "E" may be a reasonable standard. Once an LOS is selected, the additional traffic a particular project generates has to be calculated to determine if its use of the road system will exceed the LOS.

Communities need to estimate future transportation demand over an extended time period (such as 20 years) in each of their transportation zones to facilitate the task of determining the applicability of impact fees. This task can be established by several methods. The most common estimating method is based on zoning "build out." That is, on the basis of existing zoning, the local planning department estimates the potential maximum number of dwelling units, square footage of retail space, office space, and commercial/industrial structures that zoning allows to be constructed. The problem with this approach is that the actual level of future building activity may be unrelated to what zoning permits. For example, just because a particular area is zoned for high-density offices or retail trade does not necessarily mean that sufficient demand will be present for such construction to take place.

An alternative approach to zoning build out is an economic analysis that estimates annually the level of demand by land use. Such projections incorporate anticipated migration to and from the area, natural increase, change in personal income, likely employment expansion, and related indicators of future economic activity. These data, in turn, can form the basis for projecting the level of future building activity. In some instances, the projected demand exceeds what existing zoning allows; in other cases, activity will be less than zoning can accommodate. If existing zoning is insufficient, the locality has to consider rezoning or, alternatively, use existing zoning as a means to limit growth.

Economic projections are subject to considerable uncertainty. One problem is that in most instances, communities imposing impact fees are part of a larger metropolitan area. As such, land use policies of nearby communities have an effect on future activity. The imposition of fees itself can affect the level of future development unless all jurisdictions in the region apply such fees. Nonetheless, using projected economic activity is, in most instances, preferable to zoning as a measure of future activity for purposes of transportation planning, because it takes expected demand for various land uses into account.

Once a community projects the most likely level of future activity, transportation planners must estimate the road network that will be necessary to accommodate new development and, concurrently, maintain a satisfactory level of service. Impact fees cannot be utilized to improve traffic to a level that is above the established standard or to improve existing conditions.

Estimating Added Traffic Load

Each existing road system carries a given level of traffic. In a designated transportation district or zone, total traffic is comprised of through traffic (neither origin or destination points are within district) and traffic that originates and/or terminates within the zone. This traffic, in turn, can be grouped by origin and destination by land use. For example, a shopping mall within the zone may attract persons from a radius of 30 to 40 miles, well beyond transportation district boundaries. Thus, the destination of trips is within zone, but most trips originate outside. In neighborhood shopping centers, most shoppers live in the immediate area. Thus, both origin and destination are within the zone.

Communities can estimate the number and length of trips from national data (such as ITE calculations) or by the use of local surveys. The two approaches can lead to significant differences in projected traffic volume. For example, a study in Montgomery County, Md., found that traffic mean trip generation rates reported by ITE were in some cases 40 percent more than rates based on an internal study.

Once a community selects the basis for estimating the number of trips and their length, several steps must be taken to determine the cost of improving the transportation system to meet the needs of a new development. This added cost, in turn, is related to the benefit level received by the development. The following example illustrates the process of estimating additional traffic generated by adding a single family residential unit within a typical county:

Average Daily Trips per Person	2.5
No. Persons per Unit	3.1
No. Trips per Unit (2.5 × 3.1)	7.75
Average trip length	8.2 miles
Total Miles per Day	63.6
Adjustment Factors:	
(a) Double Counting	50%
(b) Percent New Trips	100%
Net New Miles per Day	31.8
Vehicle Capacity per Day per Lane-Mile	7,000
No. Units Served per Lane-Mile (7000 ÷ 31.8)	220

In the above example, the average miles traveled by a resident of a typical single-family detached unit is 63.6 miles per day. Applying the same approach, we can estimate the number of miles driven by residents in other housing units, including apartments and mobile homes. Traffic associated with commercial facilities, such as shopping centers, offices, and restaurants, is typically expressed in terms of trips generated by 1,000 ft^2 of space.

Gross miles driven require two adjustments to minimize the likelihood of double counting. This problem can be illustrated as follows. The estimate of miles driven by occupants of the single-family home includes work trips, which account for nearly one-half the daily miles driven. But if, for example, the home resident takes a job in a newly constructed office building whose developer is also being charged for trips by office workers, payment for the same trip may be collected twice. In reality, it is not feasible to know where occupants of new housing may work, but an equitable system has to consider this issue. In our example, the problem is resolved by reducing gross daily miles by one-half to take the double counting phenomenon into account. This adjustment is appropriate when dealing with a large geographic impact area within which a vast majority of trips take place—those to work, shop, school, entertainment, and to obtain services. When the designated impact fee area is geographically smaller or land uses are limited by zoning, reducing the mileage by 50 percent may result in undercounting. For example, if the designated area includes mostly single-family housing, no payments would be received from nonresidential development that attracts residents. Therefore, the adjustment for overcounting has to take into account the likely place of employment for residents of new housing within the area.

Another adjustment is to take into account trips that are not new. A certain proportion of trips may be diverted-link or pass-by trips. For example, a person on his way home from work may stop to purchase a loaf of bread or a carton of milk. The stop may be on the way, or the detour may only add one mile to a much longer commute. If so, total miles driven need to be adjusted for this factor. Several investigators have addressed this issue, and adjustments can be made to reduce the risk of double counting trips.

Cost to Bring Existing Road System to LOS Standard

In some instances, the existing road system may not be adequate to meet LOS standards. That is, there may be an existing shortfall, necessitating a future stream of capital outlays to provide an adequate level of services to existing residents and commercial/industrial establishments. It is important at this point to distinguish between improvements required to meet the needs of existing residents and improvements triggered by new development. The former cannot be paid for through impact fees.

Past economic activity in a community is linked to a flow of annual investments. For example, a community may have approved a road bond issue, with proceeds used to fund road projects within a specific "impact fee" zone. The debt service forms a stream of annual payments to be paid for and by current residents. Concurrently, these capital projects, when undertaken, may create excess capacity as a result of scale economies and engineering factors. The cost of carrying excess capacity is normally absorbed by existing residents until the excess capacity is utilized by new development, which would "buy" their capacity in the form of impact fee payments.

Cost to Upgrade System to Accommodate New Development

If excess capacity for the LOS considered the standard exists and is used to meet the needs of new development, the benefits of this excess capacity should be assigned to new development projects. The benefits accruing to new development are the pro-rata costs to construct the excess capacity. In most instances, these costs are met by tax revenue gains accruing to the community as a result of new development. The fact that a community has excess capacity suggests that existing revenue sources are sufficient to meet the transportation needs of the jurisdiction. Had these sources been insufficient, LOS standards would not have been met. Therefore, it is unlikely that impact fees could be justified when excess capacity is present. (It may also be that as a result of engineering factors and economies of scale, a community has excess capacity, although the revenue flow is insufficient to meet debt service payments. In such a case, an impact fee may be defensible.)

In most metropolitan jurisdictions, large-scale new development will necessitate capital improvements to the road system to accommodate new growth. In this case, the benefits accruing to new development would be the cost of previously constructed excess capacity, combined with the cost of new capital improvements. Once benefits are determined, the revenues a jurisdiction can anticipate from new development must be calculated and the size of impact fees determined.

ADEQUACY OF EXISTING REVENUE FLOWS

As noted earlier, if communities, utilizing existing revenue sources at current tax rates, can provide sufficient facilities to accommodate new development, impact fees for transportation cannot usually be charged. The vast majority of roads continue to be funded by government. Thus, local and state governments in fiscal 1986 spent $23.2 billion on highway construction projects. But localities spend only 21 percent of this amount. On a per household basis, $273 was spent on construction, with $57 from local sources. Most federal funds for roads are channeled through states. Therefore, the $23 billion includes federal funding, with the exception of special projects, such as roads leading to military facilities.

In recent years, particularly in growing areas, localities have begun to assume a greater share of construction projects. For example, Fairfax County, Va., and Montgomery County, Md., prior to the early 1980s, depended primarily on state funds for highway construction. This pattern has changed in recent years. Montgomery County capital outlay for roads in 1985 totaled $19 million, but the county is allocating $71 million for such projects in fiscal 1990.

When existing funding is deemed insufficient to meet needs, impact fees can be considered as an additional revenue source. Impact fee derivations have to take into account revenues applied for road construction that a new development will generate over time. An example illustrates this process. Revenue data for motor-vehicle taxes have been converted on a revenue per gallon basis to simplify the example.

Motor-related taxes (Year 1):	
Motor vehicle, titling, licensing, registration	$.17 per gal
State and federal gasoline taxes (shared by locality)	$.13 per gal
Total revenue from motor vehicle-related taxes and fees	$.30 per gal
Proportion of funds allocation for capital (balance for maintenance and other noncapital outlays)	33%
Contribution per gallon for capital	$.10
Number of miles driven per private vehicle	11,500 miles
Number of vehicles per unit	1.6
Total miles driven per unit annually	18,400
Number of gallons purchased (19 miles per gallon)	968
Revenue derived per unit	$97
Local tax revenue allocated for roads (Year 1):	
Real property tax per new unit	$1,450
Local sales tax per new unit (based on income of residents)	$230
Total	$1,680
Proportion of local budget allocated for roads	12%
Total annual revenue for roads per new unit	$208
Percent of road funds allocated for construction	31%
Funds available for construction	$65
Total available revenue (Year 1):	
Road-related taxes (including intergovernmental revenue)	$97
Local general revenue	$65
Total Public Funds	$162

In this example, if 500 single-family dwelling units are constructed, the revenue available in Year 1 would be $810,000. In subsequent years, the revenue flow for the same units would rise. In particular, property tax revenue due to property appreciation may be rising by 10 percent each year. However, revenue from the state gasoline tax, which is based on gallons rather than price, may show no rise for several years.

If an additional 200 single-family units are constructed in Year 2, and public funds rise to $180 per unit, the cumulative revenue flow from new development would be $360,000 plus $810,000 or $1,170,000.

"LUMPINESS" OF ROAD PROJECTS

Most road projects are "lumpy." That is, when considering developments individually, a road project is likely to add more capacity than is required by a particular development. Thus, excess capacity may be created. Care must be taken to ensure that new developments are not made to pay the entire cost of this excess capacity, because they are clearly not the sole recipients of its benefits. The approach taken in our methodology, because it views capital projects over a long (e.g., 20-year) horizon, avoids this problem by estimating future impact fees for future development within the planning horizon. Nonetheless, a financing problem may occur in smaller communities with limited funding where few new developments take place.

Two financing approaches are applied. In the first instance and consistent with the framework of our model, the local community may "front" the total cost, and user impact fees are collected over time to offset the project cost. The first developer who triggers the need by exceeding the LOS pays the share attributed to his project, and those who follow in future years reimburse the community for the remainder. Some communities, however, may not have the resources or be willing to front the cost. In Vermont, state legislation specifies that a municipality may require a fee for the entire cost of a capital project that will be used initially by only the first beneficiary. The municipality has to require that beneficiaries of future development pay an impact fee to the owners of the development on which the impact fee was levied.

The Vermont approach means that the first developer who triggers the need for a new project has to carry its full financial burden and financial risk if further development does not take place. Therefore, some assurance may need to be given that, if no development takes place within a specified time period, the community has to reimburse the original developer. Future developments should also be assessed the carrying cost, in addition to the proportionate share of the project cost. Otherwise, an unreasonable financing burden would be imposed on the first developer.

REVENUE AND BENEFIT CALCULATION

The revenues and benefits for a development must be compared to determine if the development is paying for the benefits it receives. Both the timing of revenue and benefit flows and the time value of money should be considered in this comparison.

This calculation requires that costs and payments be adjusted to reflect inflation and that net-present-value techniques be used to account for the time value of money (i.e., future cash flows discounted to their value in today's dollars). If these two factors are not addressed, it is unlikely that an equitable impact can be determined, because the utility value of money over time is not considered. When the comparison is properly made, an equitable impact fee can be determined.

TRANSPORTATION IMPACT FEE MODEL

The following is a general description of a computer model that has been developed to calculate transportation impact fees based upon the previously described methodology. The model is made up of a main model and three support modules (see Figure 1). The modules forecast growth, analyze capital requirements, and estimate revenues. A summary of inputs and outputs is shown in Figure 1. The complexity of the mod-

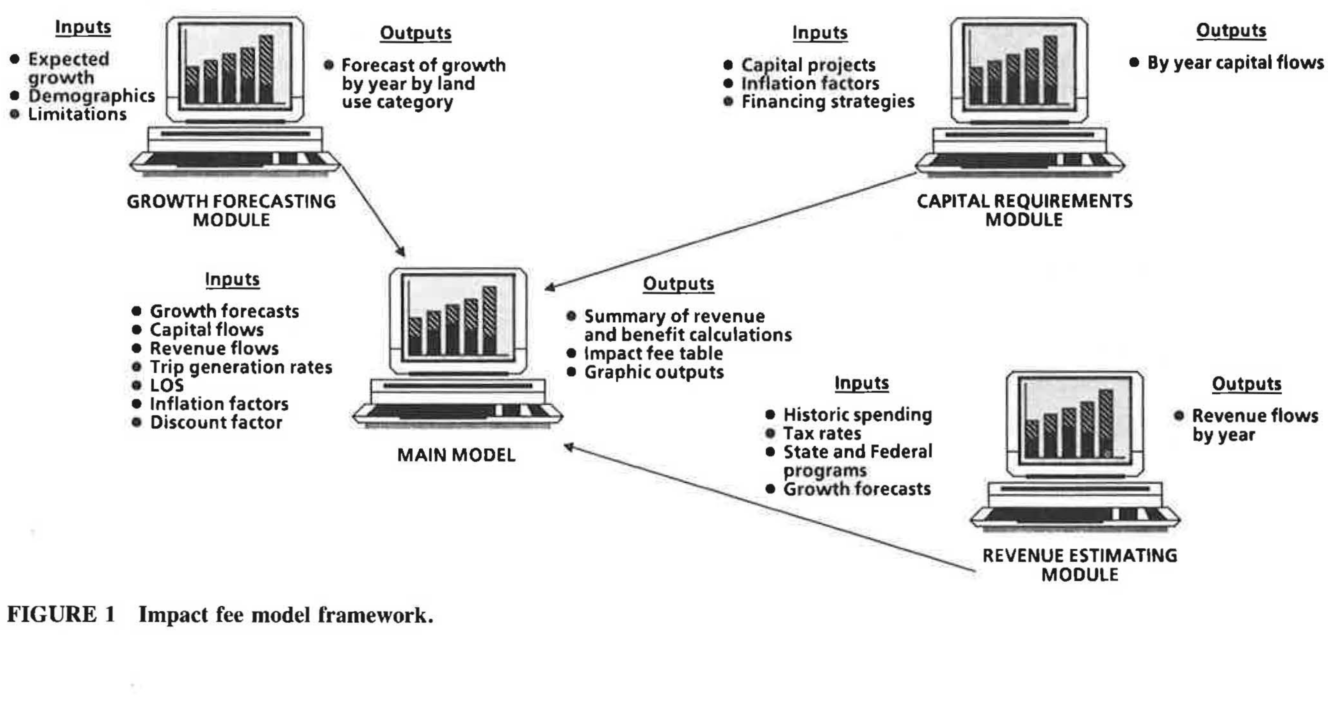


FIGURE 1 Impact fee model framework.

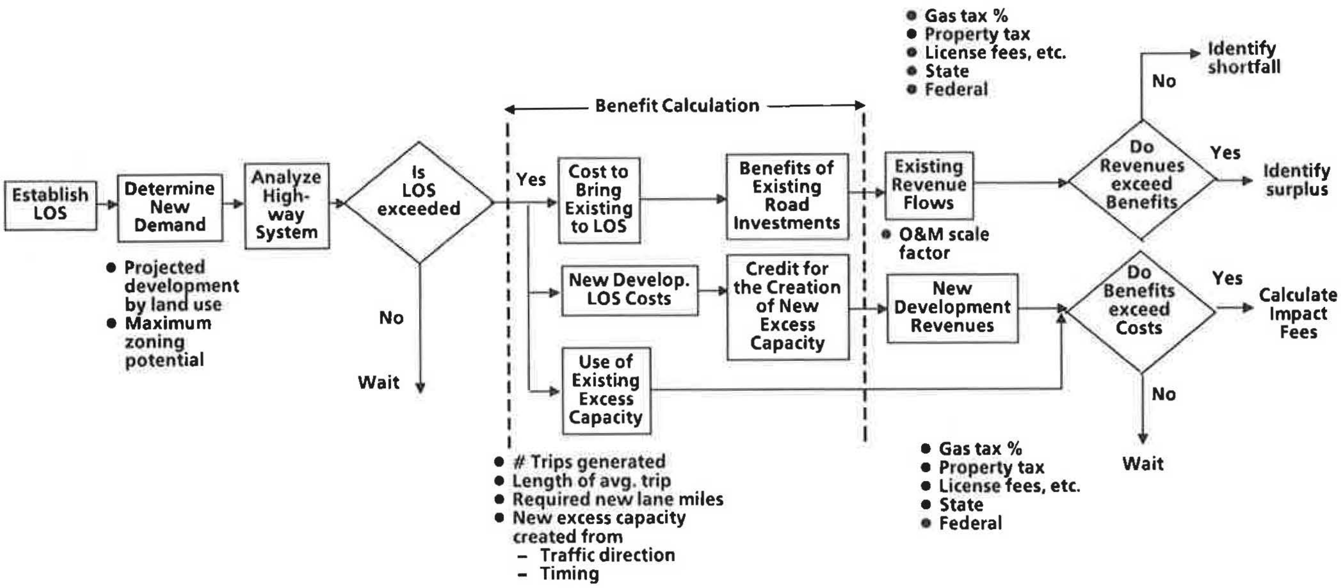


FIGURE 2 Impact fee methodology.

ules depends upon the size and complexity of the community to be analyzed. Small rural communities will generally have uncomplicated support modules, while the support modules of large urban communities will be nearly as complex as the main model itself.

The main model is where the majority of the calculations take place. The model has the capability of addressing the following issues:

- Variable trip generations rates by type of land use;
- Corrections for pass-by and diverted-link trips;
- Variable inflation factors;
- Variable (by year) growth rates;
- Variable mix of development (e.g., residential vs. commercial);
- Discount factors;
- Internal optimization routines to balance benefits and payments over a multiple-year period for both existing residents and new development; and
- Scenario analyses.

A summary of the methodology employed in the model is shown in Figure 2. The first step after determining impact fee zones is to establish a LOS standard. The new demand is then determined by either projecting development by land use or by the maximum zoning potential. The highway system is then analyzed to determine if the LOS standard is exceeded. This analysis is performed outside of the model. If the LOS standard is exceeded, the model is used to determine the transportation benefits expected to accrue to both the existing residents and the new development. This calculation captures the cost to bring existing residents to the new LOS, benefits of the existing road system, cost to meet LOS standards for the new development, charges for the use of existing capacity, and credits for the creation of new excess capacity. Once benefits are calculated, the transportation revenues are determined by considering contributions from gas taxes, property taxes, licenses and fees, and state and federal programs. The final step is to use net-present-value techniques to compare the yearly revenue and expenditure flows. Benefits and payments are equalized by establishing a system of equations with multiple unknowns and using the Newton-Rapson technique to find a solution. This solution yields impact fees that are equitable for both existing residents and new developments.

The strength of this approach is that it permits complex capital programs and growth scenarios to be analyzed in a detailed manner. The proposed methodology meets conceptual concerns that other techniques assume away by using average rates supplemented by large reductions in the calculated fee. We believe that this approach provides communities with the capability of determining equitable impact fees while simultaneously ensuring that the interests of existing residents are protected.

Publication of this paper sponsored by Task Force on Transportation Planning Applications.

Transportation Planning Data for the 1990s

David Glass, Ronald W. Tweedie, and Richard J. Zabinski

The New York State Department of Transportation (NYSDOT) is tasked with spending approximately $3 billion over the next three years for improvements to New York's transportation system. To carry out these responsibilities, NYSDOT must have good, reliable travel data on which to base its forecasts and, ultimately, its project and design decisions. Metropolitan area planners have expressed concern about the continued use of travel data gathered in the 1960s as the basis of travel simulation modeling in the 1990s. Because of the high costs, it is unlikely that comprehensive land use, transportation system, and travel surveys will be done in New York State in the 1990s. However, there are valid alternatives which meet regional data needs at lower costs. NYSDOT has begun to identify and collect the data necessary for travel forecasting, in addition to its regular monitoring and special data collection activities. New York's approach is to carefully coordinate any new travel survey activities to maximize their utility and to rely on the appropriate sharing and reuse of existing data when and where possible. Indications are that this approach will supply the required planning data to enable NYSDOT to formulate its capital program for the 1990s and shape the State's long-range transportation plans for the twenty-first century.

The New York State Department of Transportation (NYSDOT) is tasked with judiciously spending approximately $3 billion over the next three years for improvements to New York's transportation system. The majority of these funds will go to the implementation of highway system improvements. NYSDOT is also updating its long-range transportation plans for its urban areas in order to prepare strategies to meet the mobility needs of the 1990s and beyond. To carry out these responsibilities, the Department must have good, reliable travel and trip-making data on which to base its forecasts and, ultimately, its project and design decisions.

To support these two major activities, the Department, in cooperation with the state's 12 Metropolitan Planning Organizations (MPOs), and working within the Unified Planning Work Program process, has begun to define its transportation planning data needs beyond those to be met by the 1990 Census, while taking steps to maximize the use of existing data.

NYSDOT has also taken the initiative in acquiring necessary travel and trip-making data by supporting New York City MPO participation in the enrichment of the Nationwide Personal Transportation Study, offering to coordinate statewide acquisition of the 1990 Census Transportation Planning Package, working with universities to collect sample survey information and to explore data transferability issues, and providing technical and administrative help to local data collection initiatives. All these actions are part of an overall effort to effectively and efficiently acquire and supply the information needed to develop its capital programs and long-range plans.

Planning Division, New York State Department of Transportation, 1220 Washington Ave., Building 4, Room 15, Albany, N.Y. 12232.

HISTORICAL PERSPECTIVE

During the early and mid 1960s, NYSDOT designed and administered comprehensive surveys to inventory travel and land use in eight upstate New York urban areas (Buffalo, Rochester, Syracuse, Utica, Capital District, Elmira-Corning, Glens Falls, and Binghamton). The purpose of these surveys was to collect data to support the creation of a travel forecasting capability that would aid in the development of long-range plans and capital programs in the State's urbanized areas. NYSDOT collected information on typical weekday travel by persons and vehicles in the survey areas by conducting in-person home interviews of a sample of area households. The sample sizes of these surveys were large, ranging from 2,400 households in Binghamton to 13,000 households in Buffalo.

These surveys were designed to gather data on the auto occupancy, highway facility use, origin, destination, trip type, mode, land use, distance, and duration of all sampled household trips, along with detailed socioeconomic characteristics of the household, including household size, age of its members, income, and auto availability. Information on travel made within the survey areas by persons or vehicles located outside the areas was obtained by a roadside interview of a sample of the drivers of vehicles crossing cordon lines.

In addition, NYSDOT administered comprehensive telephone home interviews travel surveys in the Poughkeepsie and Jamestown urban areas during the 1960s and in 53 upstate counties (north of Westchester and Rockland) in 1970. A specialized telephone survey was administered in the Ithaca area during 1973 to determine the impacts of Cornell/Ithaca college student travel.

The Tri-State Regional Planning Commission (of which NYSDOT was a member) conducted a comprehensive origin and destination survey in the New York City area in 1963–1964. Nearly 50,000 households were interviewed at a cost of $2.5 million.

During the 1970s, NYSDOT updated the travel data collected in the 1960s through a resurvey of the Buffalo and Rochester urban areas. These resurveys were designed to detect changes in the trip generation rates and trip lengths over time in their respective areas. The type of information obtained was similar for both time frames and areas; in addition to home interviews and roadside (cordon) surveys, an opinion section was also included. These updates were done with smaller sample sizes than used in the 1960s with acceptable accuracy. The major difference between the 1960s and 1970s surveys was the absence of the land use inventory in the 1970 efforts.

Analysis of the resurvey results revealed that trip generation rates, disaggregated by household income, size, and auto availability (the parameters used for forecasting future travel) and by district, varied widely over the period in the respective survey areas; however, on an aggregate level, the areas show no major changes in trip rates (*1,2*). These travel surveys, along with inventories of existing land use and highway network configurations and their resultant data bases, are still used for estimating present travel.

Because of the high costs associated with comprehensive land use, transportation system, and travel surveys of the scope and size of those conducted in the 1960s, it is unlikely that major updates, of the same scale, will be done in New York State in the 1990s. However, there are valid alternatives that meet regional data needs at lower costs. For example, a 1983 survey of Capital District households effectively used a mail-out/mail-back technique resulting in 2,600 completed interviews. The cost of the survey administration was $14 for each completed interview, exclusive of pretest and instrument design costs. Data obtained from this survey have been applied in the calibration of a regional travel forecasting model for the Albany-Schenectady-Troy area (*3*).

Advances in survey design, technology, administration, and sampling procedures now allow for more cost-efficient updates. Smaller sample sizes, less detailed questionnaires, the availability of surrogate land use information (Census data, employment projections), and the use of multistaged samples (whereby households are stratified by a small amount of information collected from a large sample and a subset of that sample is targeted for in-depth interviews) may be cost-effective alternatives to administrating larger surveys.

NEED FOR DATA UPDATES

Forecasts of future travel are essential in the evaluation of alternative transportation system plans and in the selection of elements to the Department's capital program. Accurate travel simulation capabilities supported by current data are also necessary in order to identify the need for, and effects of, highway projects and to permit an assessment of the transportation impacts of proposed development.

Metropolitan area planners have expressed concern about the continued use of 1960s travel data as the basis of travel simulation modeling. The feeling among them is that changes in the highway and transit network, economy, land use, and demographics of the metropolitan areas in New York State may have rendered the 1960s data, and travel forecasts developed on the basis of those data, largely obsolete.

Demographic Changes

The metropolitan areas of New York State have experienced many changes in recent years. For example, the population of the New York City SMSA is now over 17 million, making it the largest SMSA in the country. Since the 1960s, it has seen major changes in the makeup of its population groups, and most likely their travel patterns, habits, and needs. The City's sheer size and dynamics have necessitated the involvement of numerous agencies in collecting travel data; their efforts are coordinated by the New York Metropolitan Transportation Committee (NYMTC). The effort in maintaining and updating this data base is enormous and requires constant attention. However, variations in the availability of resources have not allowed for a consistent maintenance of this data.

Societal Changes

Changes in the makeup and size of the work force and households have also caused a need for updating New York's travel data base. These changes include:

- increased labor force participation by women, which affects the number and types of trip demands on the transportation system;
- the aging of America, which has impacted travel patterns (the directions and magnitude of the impacts are yet to be determined: longer life span and better health care may translate into longer work force participation, maintaining the volume of work-related trips, but perhaps altering their spatial and temporal distribution; alternatively, early retirement, coupled with increasing discretionary time and financial resources and good health among the elderly, could result in more recreational travel); and
- smaller family size, which has led to an increase in the rate of formation of new households, the basic unit of trip-producers used in the development of trip generation forecasts (this increased rate of household formation has added trips to the network at a rate higher than attributable to population growth alone).

Land Use and Transportation System Changes

Major changes in the type and intensity of land use have affected the type, volume, direction, and temporal demand of travel on the state's transportation system. In turn, improvements to the transportation system have driven land development and use changes in both rural and existing developed areas. In combination, these dynamics have resulted in two general development trends that have affected our ability to forecast future travel with 1960s data: rapid urbanization in rural areas and changes in downtown land use.

Development/Expansion of Rural Areas

The high cost of land near the center of the urbanized areas, coupled with the availability of affordable land and modern transportation at the urban fringe, has aided in the residential, commercial, and industrial development of New York's rural areas. The provision of increased highway capacity, park-and-ride facilities, rapid rail, light rail, express bus transit services, and high occupancy vehicle lanes in New York State has shortened overall travel times and increased the accessibility to formerly rural areas. These areas have made available an affordable and convenient lifestyle for many New Yorkers who are now able to commute from residences far removed from downtown employment centers.

In certain instances, rapid urbanization of formerly rural areas has taxed the transportation system towards its designed capacity. Travel forecasts, fed by reliable data, are especially needed here to properly plan for this growth.

One example of the rapid urbanization of a largely rural area in upstate New York is the expansion of Fort Drum in Jefferson County to accommodate the U.S. Army's 10th Mountain Division. This military base and its associated growth will warrant designating the Watertown area as an MPO in 1990. This, in turn, will require the development of a data collection effort and forecasting capability where none previously existed.

Changes in the transportation system itself have also caused industrial location patterns to change. Traffic congestion and transportation system obsolescence in the inner cities have helped encourage industry to move to more accessible areas at the urban fringe. No longer tied to the central city for transport of labor, raw materials, and finished goods, modern industry is finding expressway interchanges and airport vicinities as prime locations.

Changing of Existing Developed Areas

The changing nature of the international economy has affected the ability of some of the State's heavy industry to compete. Nowhere is this more apparent than the Buffalo region's steel industry, which due to foreign competition now lies dormant. This has resulted in a shifting of employment and trips away from the industrial area towards other areas of the region.

There is also evidence of land use and travel pattern changes within New York's central business districts (CBDs). Once thriving retail stores headquartered downtown have scaled back or left for suburban malls. These have been replaced with smaller specialty shops, hotels, and restaurants. A resurgence of cultural and athletic facilities for entertainment and recreation has occurred in several downtown areas. Pedestrian and transit malls have also brought new economic activity geared to small package goods and shoppers without cars. Transportation terminals and warehousing have been transformed into restaurant, retail, and residential uses.

Recognizing that these economic, transportation system and associated land use changes have occurred, New York State is now taking steps to account for these changes in our forecasting methods.

Effect of Changes on Travel Demand

The location of business and population away from the central cities has blurred the pronounced directionality associated with the traditional suburban to downtown commute and may impact trip lengths and mode split relationships. The availability of flex-time, night-time work, and off-peak goods movement has caused a levelling of demand across time of day.

Such changes in the distribution of trips across time and space suggest the need for new origin-destination information.

NEW DATA TO SUPPORT CAPITAL PROGRAM AND LONG RANGE PLANNING

NYSDOT currently collects data for monitoring the performance and sufficiency of the state's highway system. This routinely includes continuous and coverage traffic counts, truck weights, speed, and vehicle classification data. In addition, special data collection activities result from legislatively mandated programs or cooperative efforts with other agencies.

In a major effort to support the Department's long-range planning initiatives, NYSDOT, in cooperation with the state's MPOs, plans to identify and collect the data necessary for developing and calibrating travel forecasting models in addition to its regular monitoring and special data collection activities. This will include collection of basic household, socio-economic (i.e., Census), and trip making data. Up-to-date land use and local highway network and transportation system configuration data is generally available from other sources and, where obsolete, would be updated.

Federal Data Sources

On April 1, 1990, the Bureau of the Census conducted the decennial census that collects general population and household information, including age, household size, auto availability, and income data. In addition, the Bureau will sample one out of six households for detailed "journey to work" data. The data set will provide the number of workers making work trips, number of work trips, work location (specific address), work trip time and departure time, mode used, carpool use, auto occupancy, and vehicle type. The Bureau expects to have summary data for commuters by place of residence, place of work, and journey to work for standard Census areas. The Census will be a key source of basic information used in the development of NYSDOT's travel forecasts.

The Census Transportation Planning Package (CTPP) will also be made available to agencies needing Census, journey to work, and socio-economic information summarized by Transportation Analysis Zones (TAZ) or Census tracts for use in travel forecasting applications. Working with and through the American Association of State Highway and Transportation Officials (AASHTO), NYSDOT plans to coordinate the needs of individual MPOs in New York State to acquire this package under one centralized purchase contract.

The U.S. Department of Transportation (USDOT) is also sponsoring the fourth Nationwide Personal Transportation Study (NPTS) in 1990. The NPTS is a household-based survey of the characteristics and personal travel patterns of the U.S. population. The 1990 version of NPTS is comparable in content to the 1983, 1977, and 1969 surveys: questions focus on weekday travel and intercity travel, household and personal socio-economic characteristics, driver information, vehicle information, and worker information; it thus offers a basis for comparison over time and by urban area size. Additional questions on accidents, seat belt usage, and highway type used for selected trips are new additions to NPTS. Unlike Census information, it will provide information on both work and nonwork trips, the latter an increasing share of overall travel. The 1990 version survey sample size will be 20,000 nationally. Based on a proportional sample of 80 samples per million population, New York State should be allotted approximately 1,400 samples. The NPTS will provide data which can be used to develop cross-classification tables for trip generation and trip length distributions, by trip purpose, for calibrating trip distribution models.

1990 NPTS Enrichment

The New York Metropolitan Transportation Council has contracted with the Research Triangle Institute (conducting the survey for USDOT) to increase the number of samples contacted in the New York City/Long Island region, thereby enriching the NPTS sample in that area. Questions will be the same as in the national survey. An additional 700 samples will be requested, which will bring the total size to nearly 1,600 samples in the NYMTC area.

New York State Data Sources

In addition to reliance on Federal data collection activities, NYSDOT is participating in several other initiatives to focus, develop and update its travel forecasting abilities.

University Research Transportation Center Project

NYSDOT has begun research designed to identify those underlying demographic, geographic, and transportation system features that characterize areas with similar travel and trip making attributes. The Department will have this research carried out as a collaborative effort by Rensselaer Polytechnic Institute and Cornell University, with funding coming from the USDOT's University Research Transportation Centers (URTC) program. The project will get underway in Winter of 1990.

The project will identify travel characteristics that can be transferred among "similar" urban areas, point out those that should not be considered transferable without more detailed area information, and establish "data transferability" guidelines. Ancillary suggestions pertaining to the maintenance and updating of fundamental travel data (including the frequency of resurveys, sample size requirements, and desired data stratifications) will also be developed.

In the near term, the project would provide NYSDOT with additional guidance in the overall design of future survey efforts and should result in a more focused, efficient, and effective resurvey. Over the longer term, basic guidance on the proper use of trip making data in nonsurveyed areas, and on the maintenance and updating of the fundamental trip making data base, should provide the Department, and other users, with increased confidence in the application of such data. It would also provide the Department with a well-developed data maintenance schedule with which it can plan for further data resurvey tasks.

Syracuse Area Resurvey

During the winter of 1989, the Syracuse Metropolitan Transportation Council (SMTC, the local MPO), in cooperation with NYSDOT and Syracuse University (SU), administered a telephone travel resurvey of the Syracuse area households. The purpose of the survey was to compare the rate of typical current weekday home-based trip making to 1966 rates (the year of the last Syracuse survey).

The resurvey was administered by SU students free of charge to the Department. SMTC provided administrative and telephone support, while NYSDOT provided the technical management. NYSDOT designed the resurvey to be consistent and comparable with the original 1966 Syracuse survey. Travel questions focused on "yesterday's" home-based travel (typically 75 to 80 percent of household trips are home-based) and included questions on home-based work travel and trip length. Socioeconomic questions dealt with household income, age of household members, household size, auto availability, and number of workers on the travel day. A question on household ZIP code helped tie responses to a geographic area.

Respondent confidentiality was assured during the interview; no information on respondent name or address was recorded. In order to keep respondent interest and cooperation, interviews were kept short (about 10 minutes) and limited to 8 questions. The survey was designed as a proxy report; one respondent was queried about the entire household's home-based trips on the day preceding the interview. It was felt that respondents were likely to know the number, type, and mode of home-based trips made by fellow household members. A sample size of 1,035 geographically dispersed households was established, considering the constraints of the University, method of administration, and past knowledge of the variability of the population.

In order to be sure to sample all parts of the study area, 45 telephone exchanges within the 1966 cordon line were identified. A LOTUS program function generated random four-digit extensions within these exchanges. An attempt was made to secure an equal number of successful interviews within each exchange. This approach approximated a sampling proportional to population, in that the larger communities have more exchanges assigned to them; it also assured that all parts within the study area were sampled.

A formal pretest was conducted to test the comprehension of survey respondents, timing of interviews, and the productivity of sample selection process. The results indicated that the survey questions were understood by the public, that the student interviewers understood and could properly carry out the survey, and that the target sample was probably attainable. However, unanticipated administrative problems, coupled with limitations imposed by the academic calendar, weather conditions, and the time students had available to work on the survey led to a smaller than anticipated sample size of 545 completed interviews.

Survey analysis and comparisons showed a modest 24 percent increase in the average home-based person-trip rate of Syracuse area households, despite a 40 percent increase in individual home-based person-trips when compared to 1966. The disparity is due to smaller average household size. Increases in the trip rates were greatest among the smaller and autoless households. Further comparisons showed that 1989 travel by households was relatively insensitive to income compared to 1966, and that travel by lower income households increased more than that of higher income households over the period.

The resurvey effort proved to be a valuable experience for NYSDOT, SMTC, and SU. It allowed NYSDOT to sharpen its skills in the planning, design, administration, and analysis of surveys in preparation for surveys in other urban areas during the 1990s. SMTC now has a limited updated trip generation capability in support of their travel forecasting efforts. Current areawide trip rates are now available for use with 1990 Census data. SU students gained valuable experience in all aspects of survey procedures and the University enhanced its image in the community through participation in meaningful projects that may otherwise, for lack of resources, go undone.

A cooperative working relationship was established during this pilot effort; based on the success of this effort, SMTC/NYSDOT has developed a complementary panel type survey, being administered by SU students during spring 1990. The survey will measure household total peak-hour travel, including origin and destination, for purpose of revising and calibrating the Syracuse area travel forecasting model.

New York Metropolitan Area Travel Surveys

The most comprehensive travel surveys in New York State are now underway in the New York Metropolitan area. Although planned and administered by various governmental agencies, these surveys together will provide the most complete picture of total travel in the New York City area since the surveys of the early 1960s.

Route 9A Project Since the early 1970s, NYSDOT has been involved in studying transportation alternatives to improve traffic flow on the west side of Manhattan. The initial effort to plan and construct an Interstate facility ended in 1985 with the "trade-in" or redesignation of Interstate funds for transit aid and the reconstruction of Route 9A as an arterial. A major data collection effort has been conducted to develop the basis for the engineering and environmental studies needed to advance the reconstruction project.

The primary effort, in addition to an extensive traffic volume count program, has been an origin and destination survey conducted through the use of a mail-back postcard survey distributed to drivers of passenger and commercial vehicles. The survey effort was divided into two phases: (1) a cordon survey conducted in November 1988 to survey vehicles entering/leaving the Manhattan CBD, and (2) a north/south and an east/west screenline survey conducted in February 1989 to survey internal travel (i.e., travel with both origin and destination in the CBD). The response rate for the cordon line survey was approximately 25 percent; the screenline response rate was somewhat lower.

Additional internal passenger vehicle travel was collected via an interview survey at on-street parking locations throughout the CBD. Also internal commercial travel was surveyed via mail-back questionnaires sent to vehicle owners registered in the CBD. Additional techniques were also used to collect data on taxis and transit services as they may affect the Route 9A corridor. Taxi data (origin and destination) were gathered from trip logs maintained by the taxi drivers. Transit route and usage data were obtained from the Metropolitan Transportation Authority.

The survey results are being coded to a zone system that is directly related to Census tracts, thus providing a linkage to the Census Transportation Planning Package and other Census products. Travel forecasts will be prepared using the TRANPLAN simulation model modified to accommodate Manhattan's unique travel characteristics.

MTA Total Travel Survey The Metropolitan Transportation Authority (MTA), operator of the New York City subway system, commuter railroads, a number of bus systems, and several highway toll facilities, is conducting a total travel survey to measure current person and vehicular movements throughout the New York City region. This will be the first comprehensive travel survey since the original travel surveys were conducted by the Tri-State Regional Planning Commission in the 1960s.

The survey objectives are threefold. First, transit service operators need information about current users of the system in order to make operational decisions at the local level on scheduling, routing, stopping patterns, and the like. Capital investment decisions will also be affected by this information. Second, the surveys will collect travel behavior information for use in near-term marketing activities designed to attract new riders. Third, MTA plans to develop forecasting tools to estimate future ridership and revenues resulting from alternative transit services. The components of the Total Travel Survey include:

- a regionwide random telephone survey of 24,000 households, conducted during April and May 1989, in which respondents were asked to provide an inventory of trips made on the day preceding the interview;
- intercept surveys on the Long Island Railroad, conducted during fall 1989 and winter 1990; and
- an intercept survey on NYCTA rapid transit system conducted in Spring 1990.

The survey results are being coded to a zone structure consistent with Census tracts to allow comparison with population, employment and other available information. The total cost of the project is $3 million.

Other Urban Area Surveys

Several New York State MPOs are planning to conduct, or participate in, travel surveys during the 1990s. The Capital District Transportation Committee (CDTC, the Albany area MPO) is reviewing the results of their 1983 mail-out/mail-back home interview survey for possible updating in 1990. Both the Rochester and Buffalo areas are exploring the possibility of conducting travel surveys in their respective areas during the early 1990s to update the base of data used to support travel forecasting capabilities. Maximum use of existing survey data (e.g., New York's Route 219 data) will be made, and coordination with other planning survey efforts (e.g., the Ontario-New York survey) will be undertaken as part of the design of these surveys.

Free Trade Agreement

The Free Trade Agreement (FTA) is an accord between the United States and Canada in which the two nations have agreed to a staged elimination (over ten years) of most trade tariffs. The agreement also provides for the elimination of nontariff restrictions, such as quotas, and the liberalization of cross-border service trade, travel, and investment.

The Province of Ontario and New York State both anticipate an increase in border crossing and regional traffic due to implementation of the agreement. Staffs from both governments are working on an approach to better gauge the nature of current and anticipated border area transportation problems and opportunities. Both groups are also concerned with the ability of other modes (rail, air, and water) to meet increased cross-border passenger and goods movement demands. Finally, the Department must devote attention to the support of the economic development opportunities offered by the FTA, wherever they may occur across the State.

A basic step in understanding potential transportation impacts is to determine the relevant characteristics of passengers and goods currently crossing the border, including their origin and destination patterns. New York and Ontario plan to share available data on travel and goods movement; New York's Route 219 survey data and information from Ontario's recent travel corridor studies should prove helpful.

SUMMARY

The New York State Department of Transportation, in cooperation with the State's local transportation organizations, is actively involved in defining and obtaining travel data needed to plan for the mobility needs of the 1990s. Because of a limited data collection budget, New York has had to examine very closely the data it needs and try to obtain it in the most efficient manner possible. New York's approach is to carefully coordinate any new travel survey activities to maximize their utility and to rely on the appropriate sharing and reuse of existing data when and where possible. Indications are that this approach will supply the planning data the Department needs to formulate its capital program for the 1990s and shape the state's long-range transportation plans for the twenty-first century.

REFERENCES

1. D. T. Hartgen. *1973 Buffalo Travel Survey: Design, Conduct and Processing*. New York State Department of Transportation, Albany, N.Y., 1975.
2. G. S. Cohen and M. A. Kocis. Component of Change in Urban Travel. In *Transportation Research Record 775*, TRB, National Research Council, Washington, D.C., 1980, pp. 42–47.
3. J. P. Poorman and D. J. Stacey. *Self-Administered Mailback Household Travel Survey*. In *Transportation Research Record 955*, TRB, National Research Council, Washington, D.C., 1984, pp. 10–17.

Publication of this paper sponsored by Task Force on Transportation Planning Applications.

Study of Transit Alternatives in the Dulles Airport Access Road Corridor

JEFFREY M. BRUGGEMAN

The median of the Dulles Airport Access Road has long been considered a potential transit corridor to connect Dulles Airport with the rest of the Washington metropolitan area. KPMG Peat Marwick was engaged by Fairfax County, Va., to examine the feasibility of transit development in the corridor. The study examined a variety of transit options including express bus, high occupancy vehicle facilities, and various rail technologies. Station locations, ridership estimates, operating costs, and capital costs were prepared for each alternative. Based on the results of the analysis, it was recommended that the county implement an enhanced express bus system utilizing the high level of service provided by the Dulles Access Road, with station sites identified for interim use for express bus operation and potential ultimate conversion to a rail or other higher technology when operational and financial conditions warrant such action.

The Dulles Airport Access Road corridor runs between Washington Dulles International Airport and the West Falls Church Metrorail station. The corridor serves the northern portions of Fairfax County, Va., including the "new town" of Reston, as well as a large area north and west of the airport in Loudoun County. The corridor also skirts the northern edge of the Tyson's Corner area, a major mixed-use suburban activity center and the largest "downtown" in Virginia.

The corridor was identified as a major transportation link in the early 1960s, when Dulles Airport was built in what was then an undeveloped part of the western suburbs of Washington, D.C. Development in the corridor occurred rather slowly until two rather recent events coincided in the late 1970s and early 1980s, namely, the deregulation of the airline industry and the explosive growth in suburban office development. Since then, the corridor has grown rapidly, and this growth is projected to continue into the next century (Figures 1, 2, and 3).

In addition to the rapid increase in population, employment, and airport-related travel, the corridor is projected to continue to be influenced by increases in car ownership and dispersal of commuting patterns, leading to even more rapid increases in work person-trips that must be accommodated on the transportation network (Figure 4). One of the major changes that profoundly influences travel and the transit opportunities is the shifting of travel emphasis away from a radial orientation (to downtown Washington and the inner suburbs of Arlington) and toward a much more diffused pattern. By far the largest increases in travel will occur in the intra-corridor market and in cross-county travel.

KPMG Peat Marwick, 8150 Leesburg Pike, Suite 800, Vienna, Va. 22182.

TRANSPORTATION SERVICES

The primary transportation feature of the corridor is the Dulles Airport Access Road and Dulles Toll Road, a unique "nested" pair of freeway facilities sharing a common right-of-way. The Access Road was built in the mid 1960s to serve Dulles Airport and consists of a four-lane freeway with access only to and from the Airport. This facility was built to accommodate an outer roadway system for local travel at a later date which was, indeed, constructed as the Dulles Toll Road in the mid-1980s. This latter facility has proved to be extremely popular, is running well ahead of traffic and revenue projections, and is currently under design for expansion from four to six lanes.

Traffic patterns in the corridor are also influenced by the connection between the corridor facilities and the rest of the region. The eastern end of the corridor connects directly with I-66, which is itself unique, being a four-lane freeway reserved for high occupancy vehicles (HOVs) and airport users in the peak direction during morning and evening rush hours. Other users of the corridor must exit at the Capital Beltway (I-495) or onto local streets serving the Tyson's Corner area.

Public transportation in the western part of the corridor features a system of buses that circulate in the Reston and Herndon areas and run express to the West Falls Church Metrorail system. Limited park-and-ride service is available from a small lot near the Access Road. Most of these express buses enter the Toll Road from general traffic ramps and then pass through a bus-only slip ramp to the largely uncongested Access Road for the express portion of the trip.

Service in Tyson's Corner area and the rest of the "inner" corridor consists of conventional local bus transit, with a major "hub" at a regional shopping center. Several routes connect from the Tyson's Corner area to the West Falls Church Metrorail station, although a well-developed shuttle system has yet to evolve. Some limited cross-county service is provided as well. Transit service to Dulles Airport is limited to express bus service from downtown Washington and van service from West Falls Church and various suburban activity centers in Northern Virginia and Maryland. There is almost no public transit service from Loudoun County.

The future of public transportation in the corridor is tied closely to the available right-of-way and planned highway improvements. As noted above, widening of the Toll Road is currently in the final planning stages, with a major local debate as to whether the additional capacity will be for general purpose traffic or restricted to HOVs. In addition, the Metropolitan Washington Airports Authority wishes to preserve enough right-of-way to add two additional lanes on the Access Road as air travel demands dictate. Thus, although the current

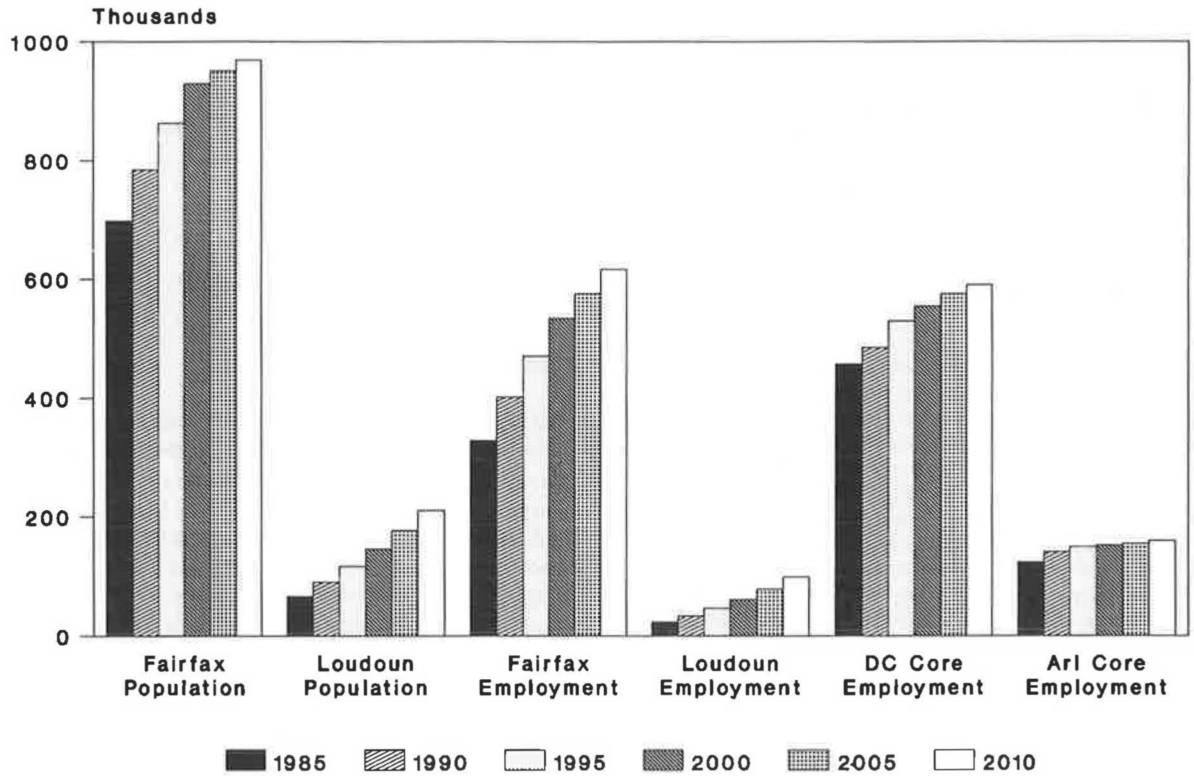


FIGURE 1 Population and employment.

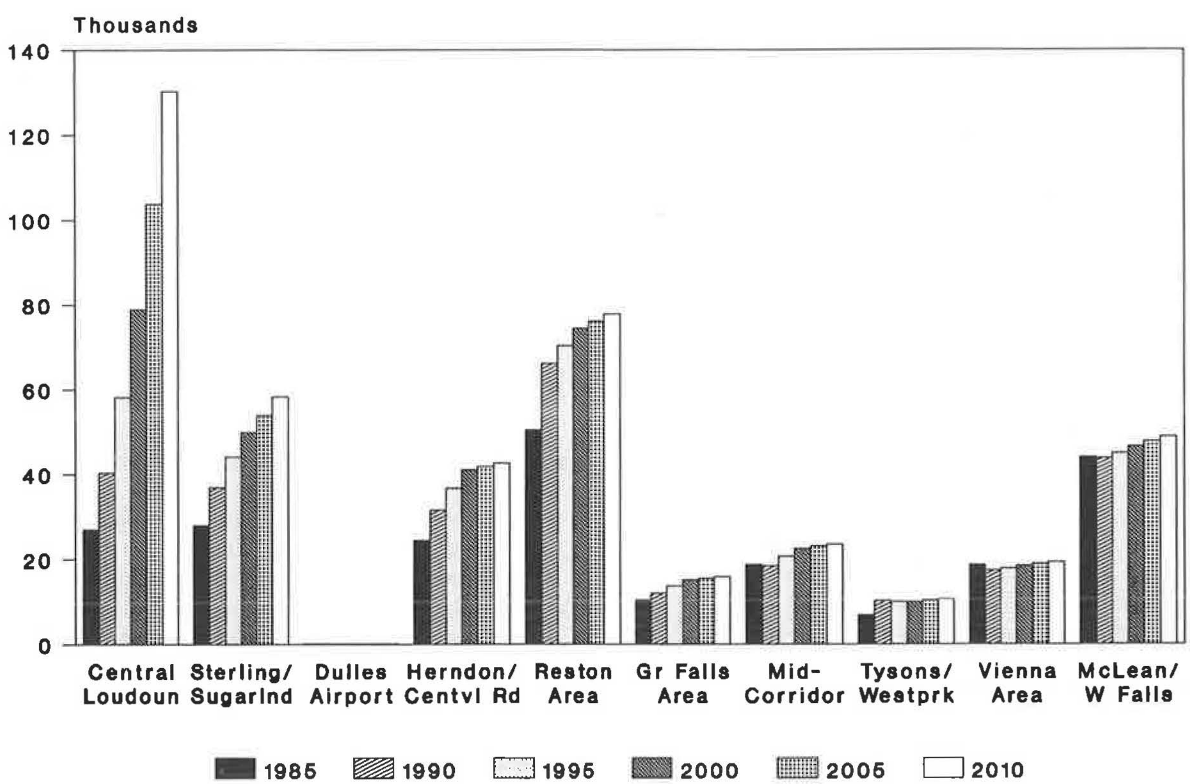


FIGURE 2 Corridor population.

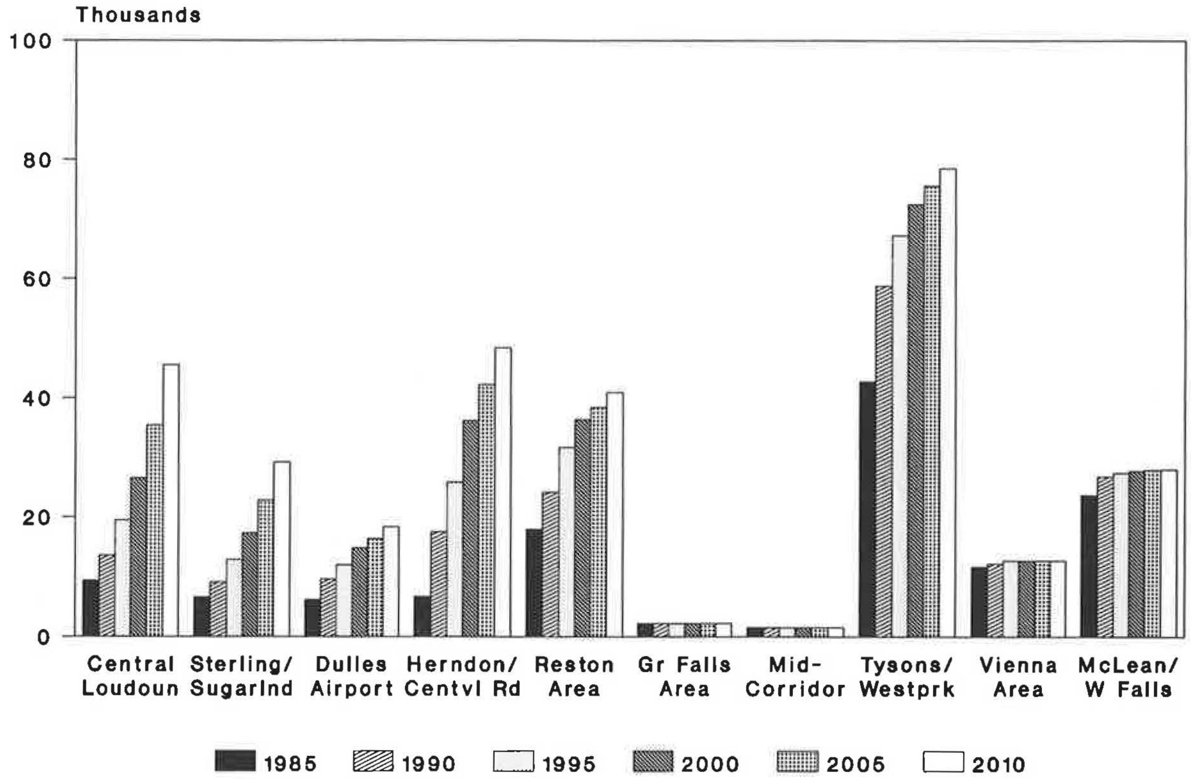


FIGURE 3 Corridor employment.

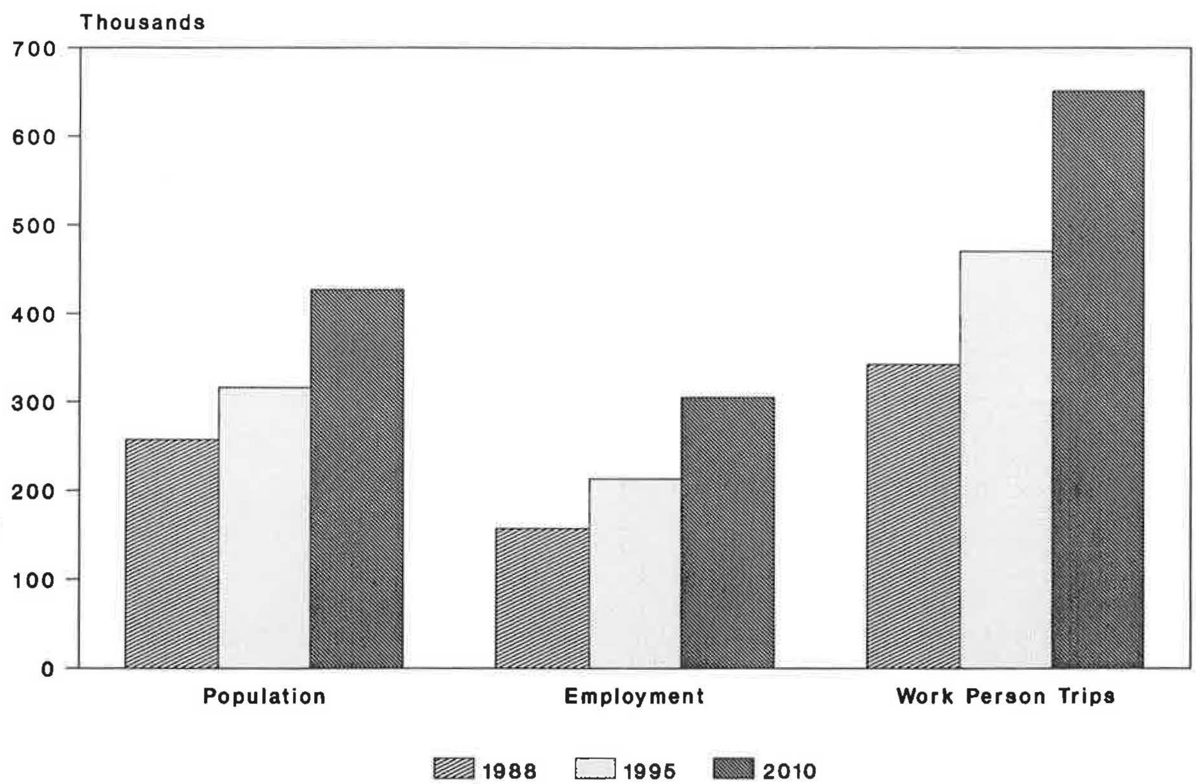


FIGURE 4 Work-related travel characteristics.

right-of-way would appear to be generous, once both the Toll Road and Access Road widenings are taken into account and accommodation is made for slip ramps, interchange improvements, and noise barriers, the space remaining for a transit facility is much more limited.

TRANSIT DEMAND ESTIMATION

The analysis of transit demand in the corridor required the development of procedures to deal with a variety of transit technologies as well as alternative station locations and operating plans. The demand estimation was performed using a microcomputer package, utilizing the logic contained in the regional planning models developed and maintained by the Council of Governments (COG). In a microcomputer environment, it was necessary to hold down the level of detail while still maintaining consistency with the regional process and providing appropriate sensitivity to the alternatives. The most important aspect of this consistency lay in the coding of access to transit, a common problem with other transit modeling studies.

The potential problem can be illustrated with a simple example. Area system geography is normally represented as a zone centroid connected to several adjacent intersections (Figure 5). Often, local bus service is provided through each connecting node (Figure 6). For some time, however, it has been standard modeling practice to recognize that not all of a zone may be able to walk to transit, and thus a transit walk area is defined around each transit route (Figure 7). The remaining part of the zone (and all zones not directly served by buses) is then assumed to be served using auto access to formal or informal park-and-ride facilities.

This type of coding is adequate for modeling conventional bus transit service, because the level of service at any given node is likely to be very similar and the all-or-nothing characteristic of conventional transit path building and assignment is not of major concern. An inconsistency arises, however, when a high-type transit service is introduced adjacent to part of a zone (Figure 8). In this case, the transit path is almost always via the high-type facility and, assuming that the percentage of the zone that can walk to transit still applies, will lead to overestimation of demand. In reality, a portion of the zone may have excellent service (Figure 9) while much of the zone is served only by local buses. In addition, the local bus service now may well require a transfer, further aggravating the problem.

The problem may be addressed in a number of ways, including developing a large number of small zones, particularly in the vicinity of transit stations, which can reliably be assumed to be either 100 percent walk to the high-type facility or require feeder access. This approach presents some problems in a transit alternatives study, such as that undertaken in the Dulles corridor, where station locations are unknown at the start and may be changed several times as the alternatives evolve. Also, increasing the number of zones results in some significant penalties with microcomputer processing.

The approach adopted for the Dulles study was to extend the two-path concept (walk and auto) used in the COG Washington model system to a system whereby the origin zone was divided into three potential areas: (1) walk to rail (or other high-type transit), (2) walk to feeder bus, and (3) drive to transit. At the destination end, two areas were identified: (1) walk directly from a transit facility, and (2) transfer to bus egress. Any destination area not served by transit is then considered unconnected, since outbound auto access is not a valid mode for most travelers.

This concept results in six potential transit paths (3 origin × 2 destination), as shown in Figures 10–16. Separate walk percentages for direct station access and bus access were esti-

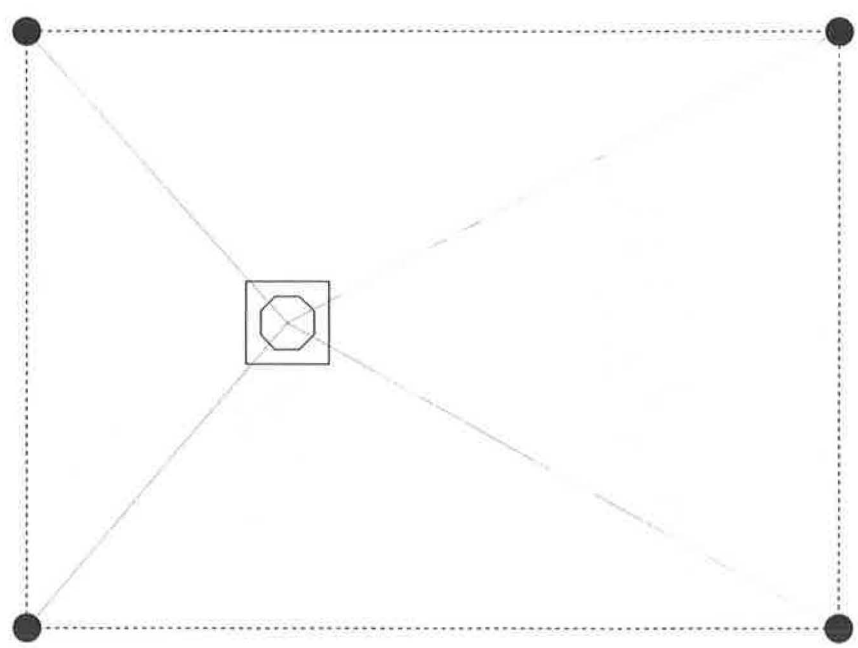


FIGURE 5 Area system example.

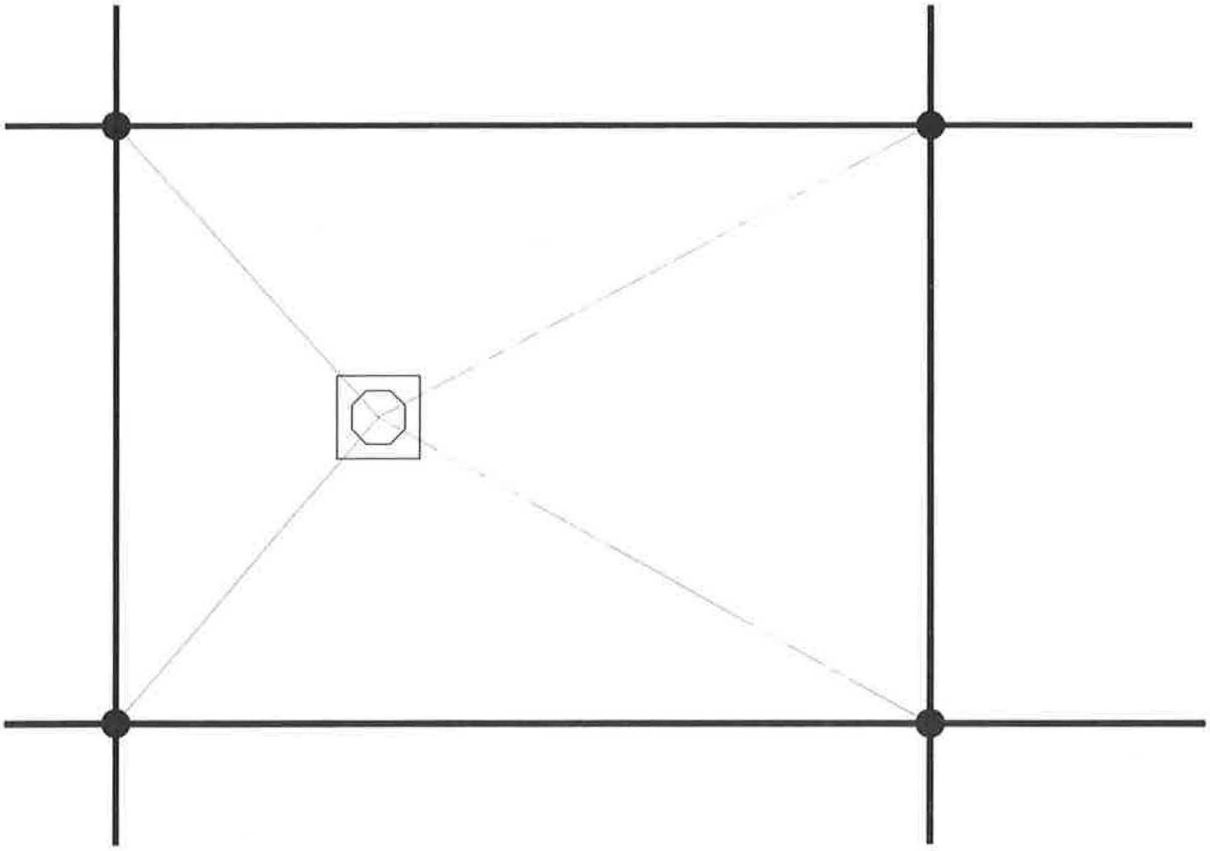


FIGURE 6 Local bus service through connecting nodes.

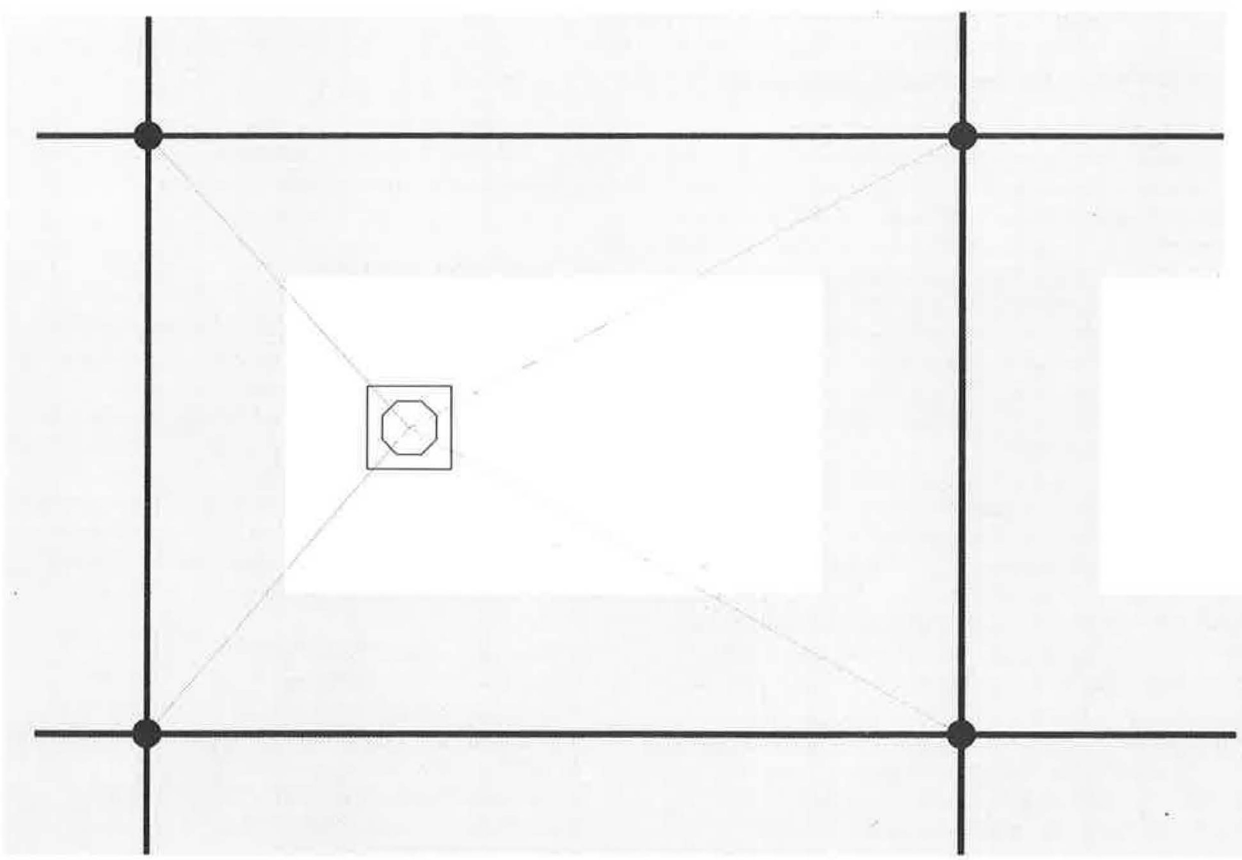


FIGURE 7 Transit walk areas around bus routes.

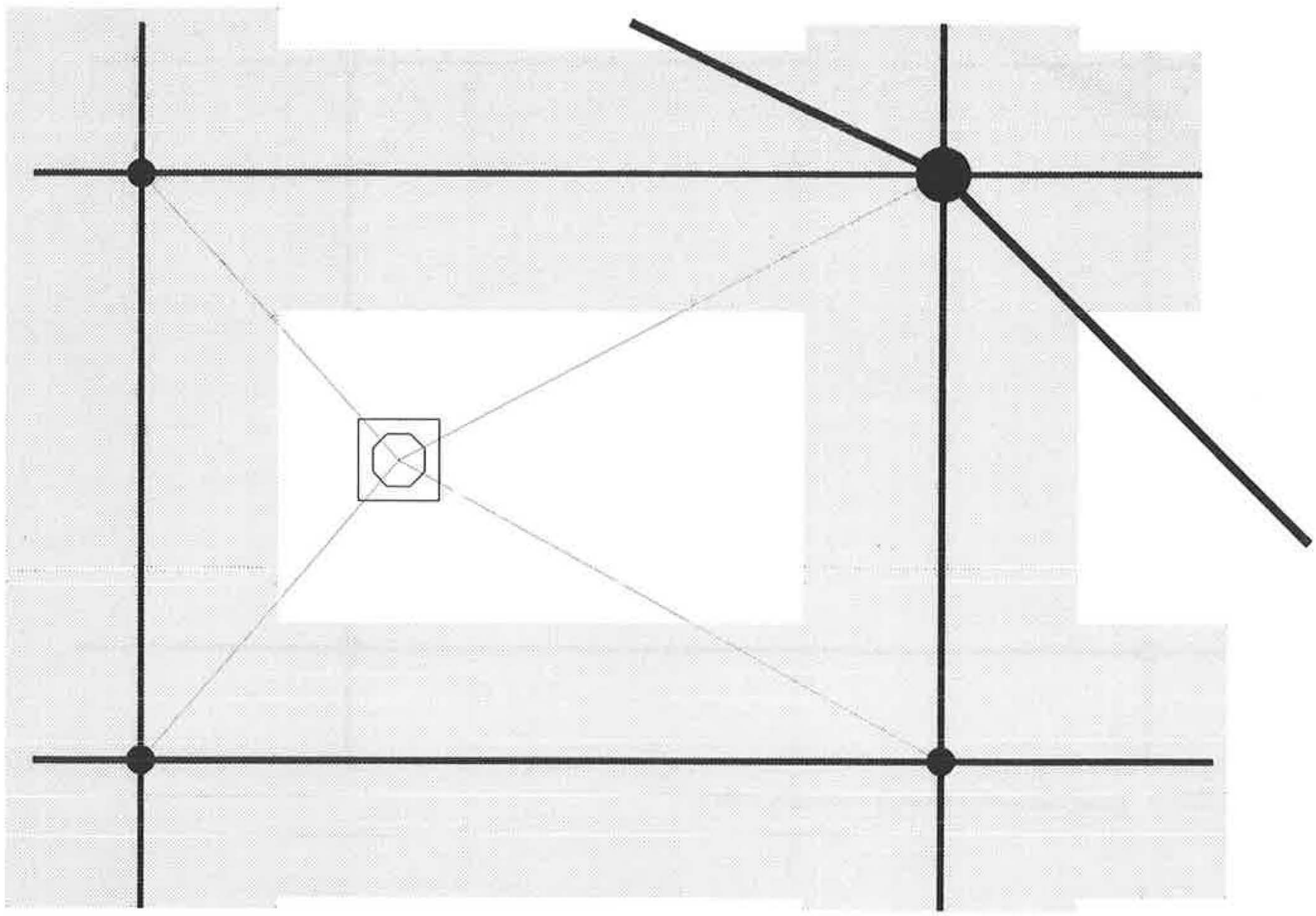


FIGURE 8 Addition of high-type transit service at one node.

mated for transit access and egress for each zone, and six potential transit paths were built and skimmed. The modal choice model was then applied for each set of path impedances, multiplied by the appropriate product of the origin and destination access percentages, and summed to produce a total interchange transit demand estimate. In practice, the transit trips on the six paths were maintained in the process for ultimate application of a simple mode of arrival model and assignment to the six individual paths. The assignment results were then combined to produce total loadings on the transit system components.

This approach offered great flexibility in the process, because a reliable estimate of the impact of moving or deleting transit stations could be determined simply by changing the access percentages, revising the transit line descriptions, and modifying the access links where necessary. The approach was also particularly relevant for this corridor, which did not include a direct downtown component, thus allowing the modeling to be undertaken using larger districts in the core area with negligible loss of accuracy.

The primary drawback to the approach was the need to generate six sets of transit paths, impedances, fares, and trip tables. Although each matrix remained modest in size, the number of matrices proliferated very rapidly and data storage, data management, and computer processing times increased significantly. The approach would thus seem to have great merit, but some offsetting consequences would need to be evaluated for any other potential application.

TRANSIT ALTERNATIVES

The development of the modeling procedures noted above and the identification of transit alternatives proceeded in parallel during the study, with decisions on the latter often influencing the design of the former. A total of nine transit alternatives were ultimately identified for analysis, combining the various technologies noted previously. These alternatives included an upgrading of conventional bus service as a transportation systems management (TSM) option, a grade-separated bus/HOV facility in the median of the Access Road/Toll Road, and a variety of rail options and technologies as follows:

- Base case (do nothing) [BASE];
- Transportation systems management [TSM];
- Busway/high occupancy vehicle lanes [HOV];
- Conventional light rail transit (LRT) [LRTA];
- LRT with Tyson's circulator [LRTB];
- Automated guideway transit (AGT) [AGTA];
- AGT with Tyson's loop [AGTB];
- Metrorail branch [METR]; and
- Rail hybrid [HYBR].

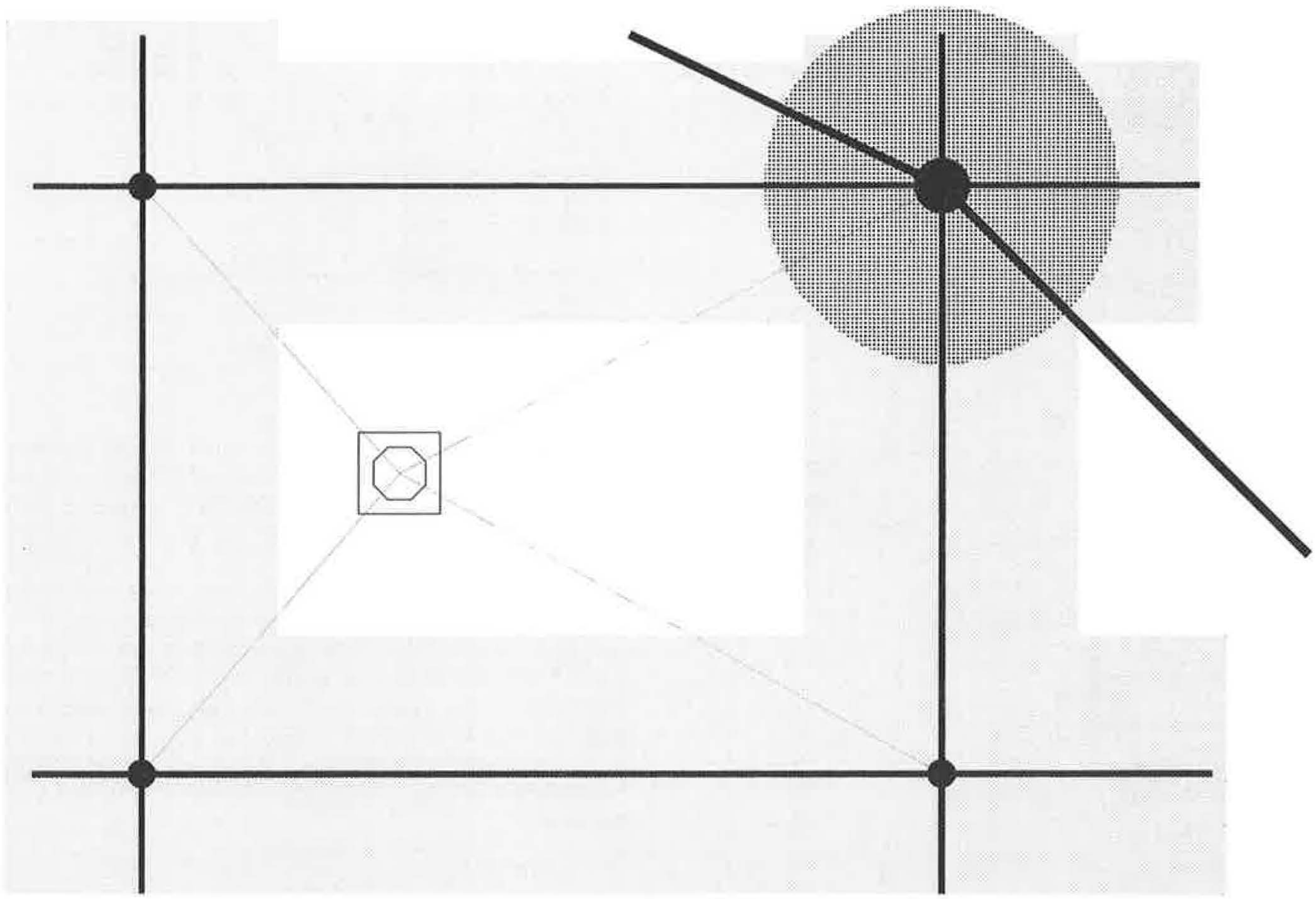


FIGURE 9 Improved service in part of zone may lead to overestimation of demand.

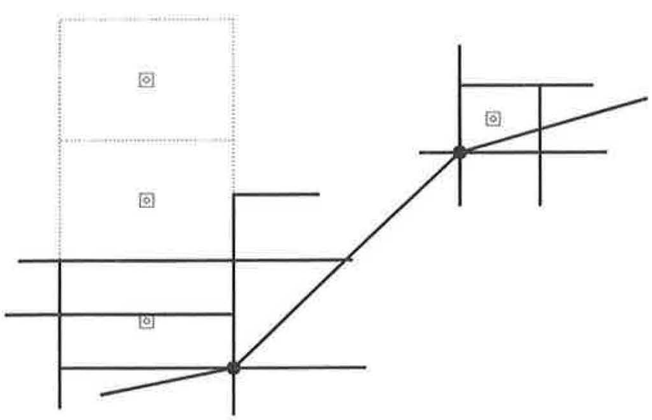


FIGURE 10 Path example.

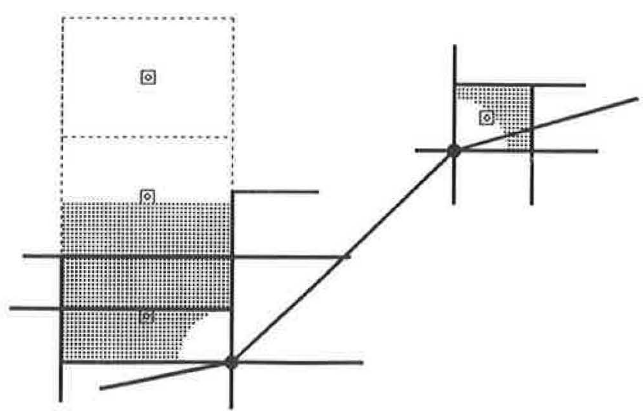


FIGURE 11 Path 1—bus-rail-bus.

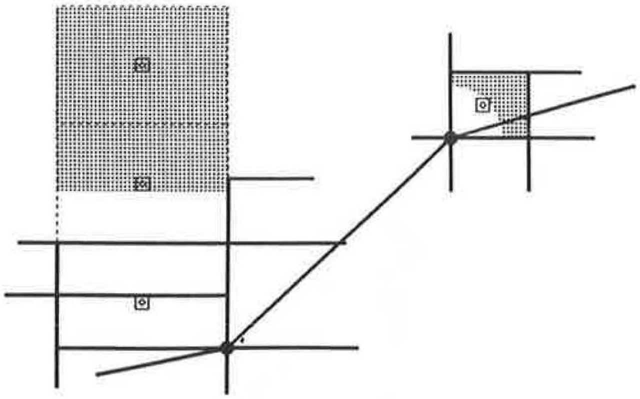


FIGURE 12 Path 2—auto-rail-bus.

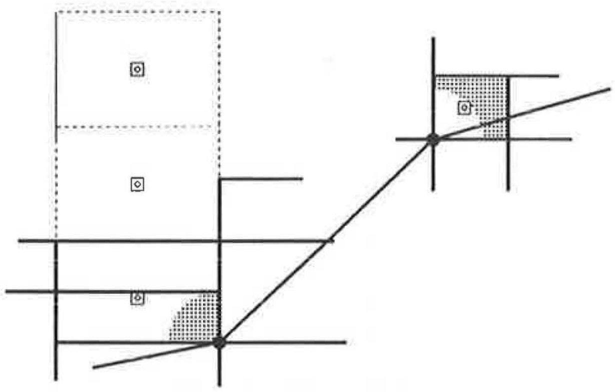


FIGURE 13 Path 3—walk-rail-bus.

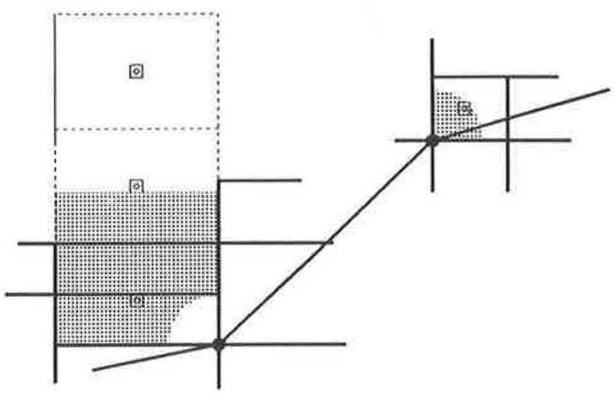


FIGURE 14 Path 4—bus-rail-walk.

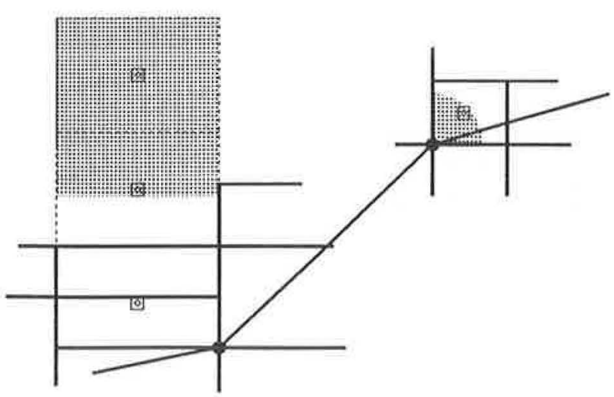


FIGURE 15 Path 5—auto-rail-walk.

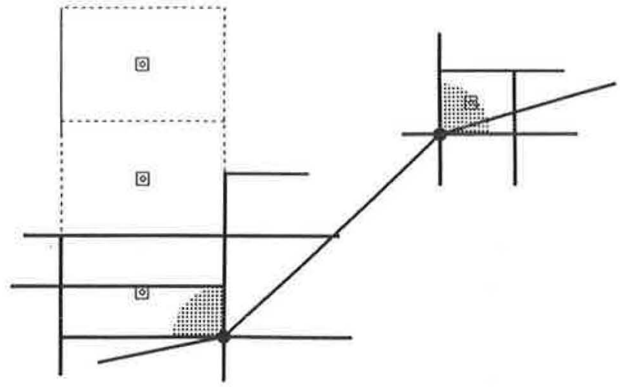


FIGURE 16 Path 6—walk-rail-walk.

All of the rail alternatives shared a common set of station locations, including the terminal stations at Dulles Airport and West Falls Church and six intermediate stations. A more limited set of stations was assumed for the TSM and HOV alternatives. Two alternatives also included an AGT system to serve the Tyson's Corner activity center. For demand estimation, an independent AGT loop was added to the baseline light rail alternative (LRTA) to produce a second alternative (LRTB) with a transfer at the Spring Hill Road station. Direct service through the Tyson's Corner area was explored with the AGTB alternative, which included two operating lines, one remaining in the Access Road right-of-way to West Falls Church and the other looping through the Tyson's Corner area.

TRAVEL DEMAND RESULTS

The various alternatives produced somewhat different travel times, although not nearly as different as those typically encountered in a corridor study due to the presence of the Access Road, which offers a largely uncongested roadway for express buses to West Falls Church in the TSM alternative.

Total corridor travel was the same in all alternatives (Figure 17), because a fixed person-trip table was used for all analyses. Overall transit demand levels were very similar, as might be expected from the similar travel times, and transit ridership only accounted for a rather small part of overall corridor demand. HOV usage was also very similar across the alternatives, since most HOV usage is "induced" by I-66 into downtown Washington, a common feature of all alternatives.

The analyses were undertaken for two horizon years, 1995 and 2010. The transit trip market is dominated by travel to the core (Figure 18), as is typical of a high-income suburban corridor, with the rail alternatives performing slightly better for internal travel than the bus alternatives. On a geographic market basis, transit is shown to capture a significant share of the travel to the core, but as noted previously this is a very low-growth market (Figure 19).

The total corridor travel is important in overall project evaluation but masks differences between the alternatives, because significant travel from the inner parts of the corridor is not affected by many of the alternatives. The differences in transit guideway ridership is greater among the alternatives,

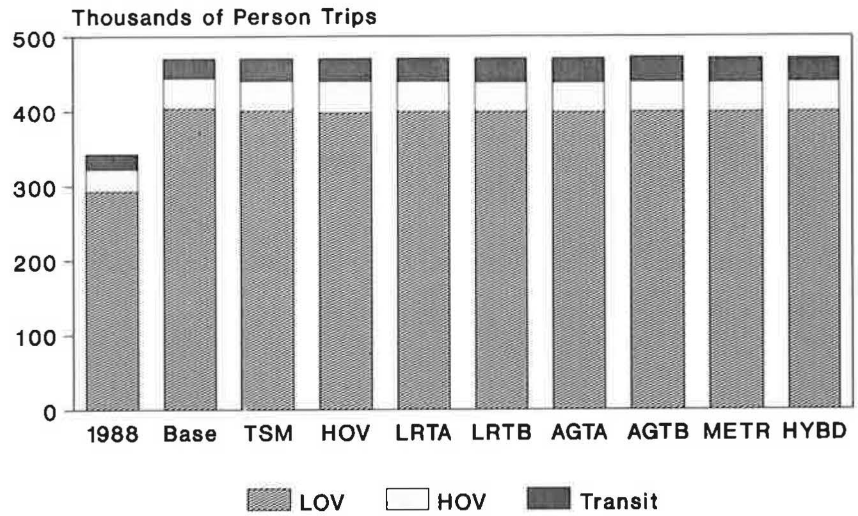


FIGURE 17 Total corridor work travel (1995 average weekday).

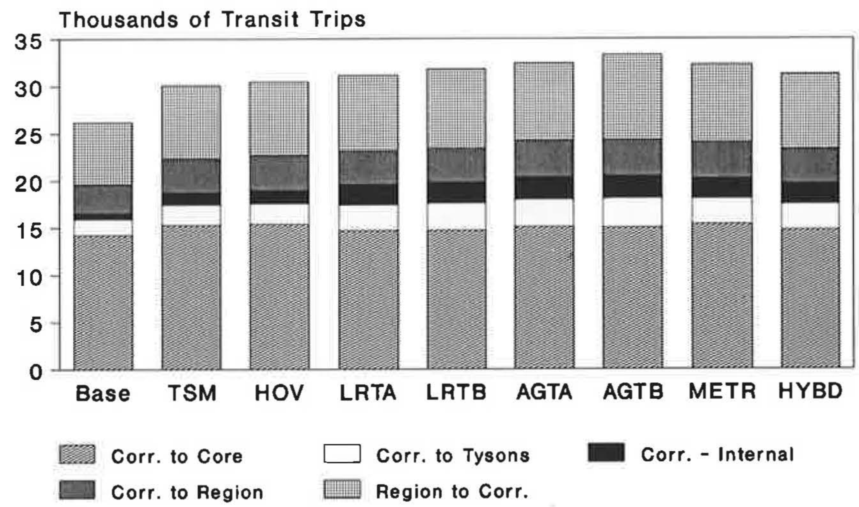
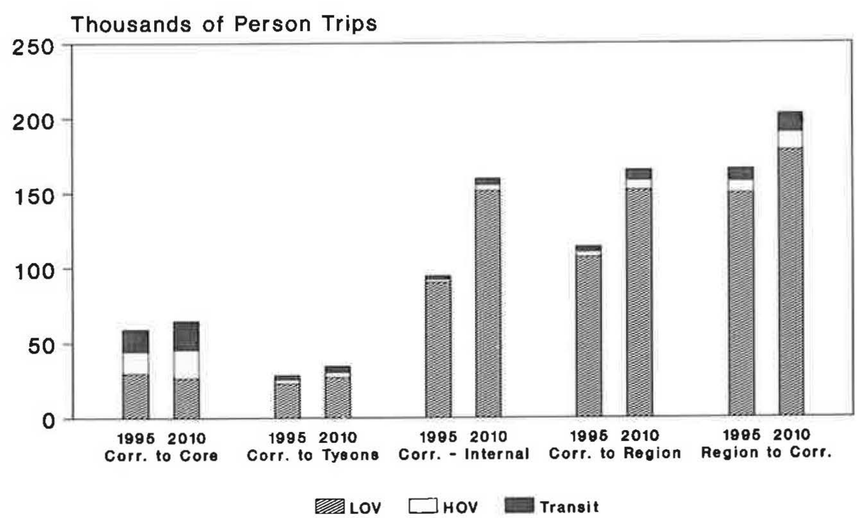


FIGURE 18 Corridor transit work trips by market (1995 average workday).



Assuming LRTA alternative

FIGURE 19 Summary of corridor work trips by mode by market.

with the AGTB alternative with direct service to Tyson's capturing the largest ridership (Figure 20). This result occurs since this alternative provides guideway service for virtually all outbound commuters from Metrorail to the entire Tyson's Corner area, while many commuters use local buses from the West Falls Church station in the other alternatives.

COSTS AND REVENUES

The operating costs for the various alternatives show that the additional rail operating costs are not fully offset by a reduction in bus operating costs (Figure 21), largely because the bus service that is replaced is high-speed, efficient service via the Access Road. Operating revenues are somewhat higher for the rail alternatives, however, due to higher ridership and a somewhat different fare structure, so the overall subsidy (Figure 22) for the simpler rail alternatives is about the same as for the bus alternatives.

The operating costs for Metrorail are much higher than for the other rail alternatives, not so much because of inherent differences in the technologies as the necessary differences in implementation. In the other rail alternatives, the service terminates at West Falls Church with a transfer. The Metrorail alternative, however, is a branch of the existing Orange Line, with alternating trains serving the current Vienna terminus and the proposed Dulles line. Because of downstream capacity problems, it is necessary to run longer trains on Metrorail than are required to serve the corridor demand alone, thus increasing car miles and operating costs.

The capital costs for the alternatives are vastly different, with the rail alternatives ranging from over $500 million to nearly $800 million in 1988 dollars (Figure 23). The HOV alternative requires approximately $200 million, much of it to rebuild the existing bridges over the Access Road and Toll Road to eliminate operationally unacceptable piers in the middle of the roadway. By contrast, the capital costs for the TSM alternative are less than $100 million.

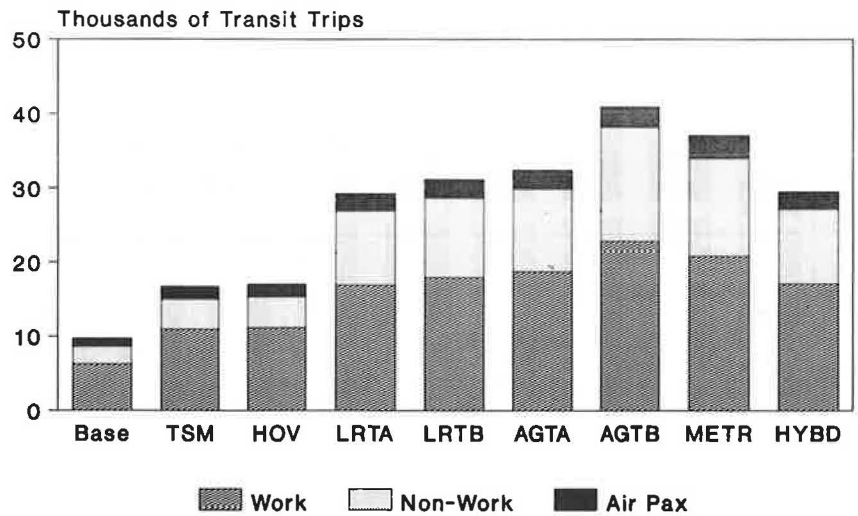


FIGURE 20 Transit guideway ridership by purpose (1995 average workday).

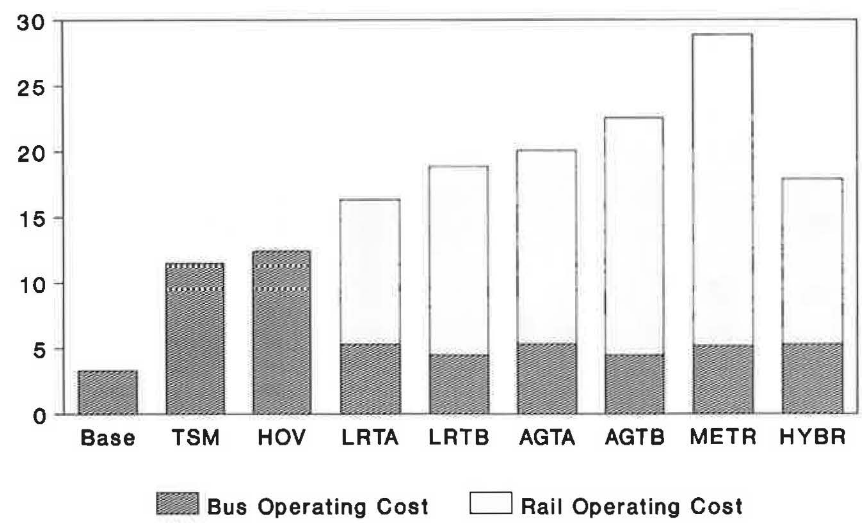


FIGURE 21 Annual corridor operating costs (1995, in 1988 dollars).

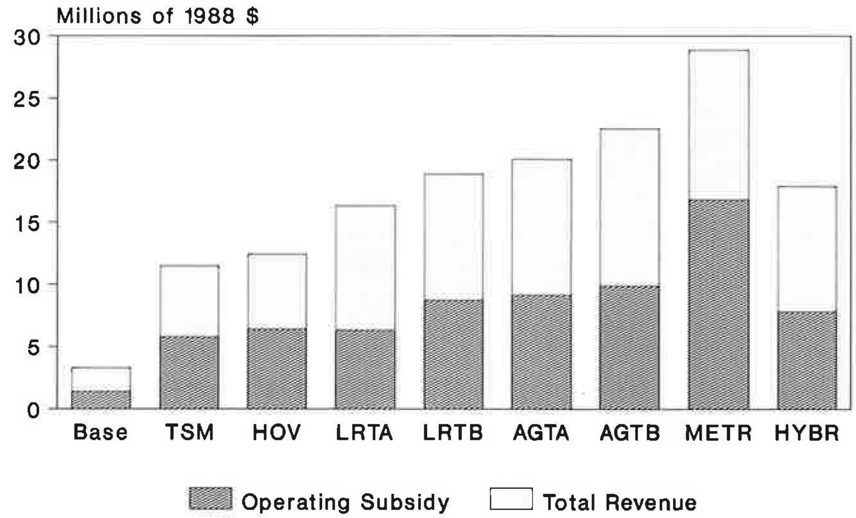


FIGURE 22 Annual operating cost, revenue and subsidy (1995, in 1988 dollars).

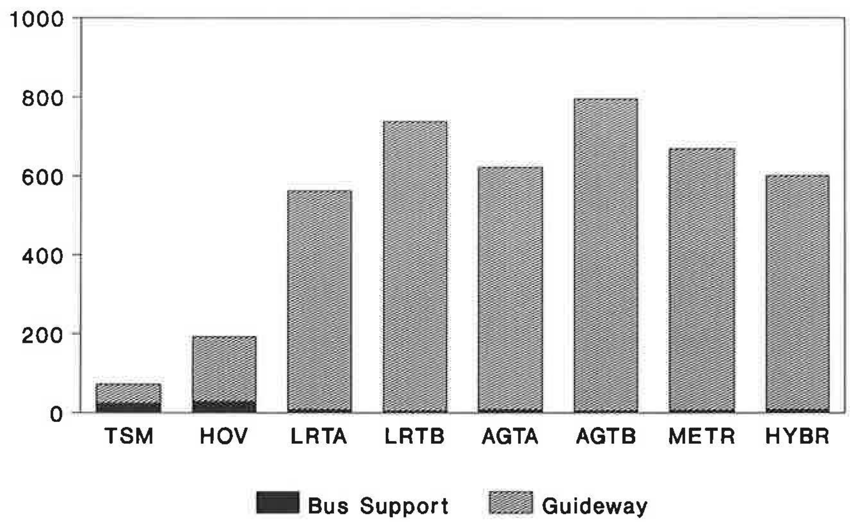


FIGURE 23 Capital cost summary (millions of 1988 dollars).

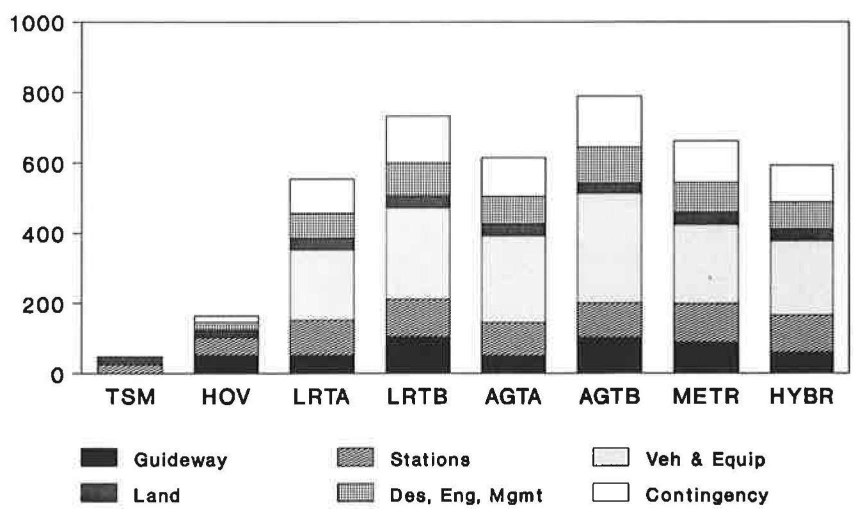


FIGURE 24 Guideway capital costs (millions of 1988 dollars).

The guideway costs for the various alternatives include the actual guideway facility itself, stations, vehicles and equipment, land, and allowances for design and contingencies (Figure 24). The vehicles and equipment costs are the largest part of the total and include an almost equal breakdown into trackwork, power, signalling, communications, vehicles, and maintenance facility components.

CONCLUSIONS

The unique features of the corridor profoundly affected the results of the analysis. A cost-effectiveness analysis following the guidelines of the Urban Mass Transportation Administration showed that rail investments did not reflect a cost-effective increment over the TSM condition. In addition, competing needs for transportation resources in the Washington area also appeared to argue against immediate pursuit of a rail option.

As a result, the study concluded that an aggressive bus-based transit system be developed in the corridor, while preserving the median alignment and identifying and preserving station sites for ultimate implementation of fixed guideway service at a future date. These recommendations were adopted by the Fairfax County Board of Supervisors (the client for the study), and steps are being taken to implement the recommended plan.

ACKNOWLEDGMENTS

The author wishes to acknowledge the valuable assistance of the staff of the Fairfax County Office of Transportation, particularly Andrew Szakos and Christopher Jenks; Samuel Zimmerman of the Urban Mass Transportation Administration, which funded a portion of the study; and the members of the technical advisory committee and citizens advisory committee.

Publication of this paper sponsored by Task Force on Transportation Planning Applications.

Simulating Traffic Control Devices in a Sub-Area Network

WILLIAM G. ALLEN, JR.

In a sub-area network, stop signs and signals at intersections can be a significant influence on travel paths. The presence of such devices often must be simulated in computerized highway networks so as to obtain more realistic assignments. One common way of doing this is to use "turn penalties." This is very resource-intensive, and has been known to confuse the path-building logic of some programs, resulting in poor assignments. An alternative method has been developed to avoid these problems. This method is based on theoretical models of delay, as a function of the V/C ratio of the controlled approach. Given the link's distance, the uncontrolled free speed, and the type of control device, a new "zero-volume" link speed and capacity are calculated to account for the delay induced by the control device. Further, the characteristics of the link are modified to account for delay caused by queuing: as traffic is assigned to the link, its constrained speed drops at a faster rate than if it had no control device.

A county-wide travel demand model using this technique was implemented using MINUTP, and the control device simulation method was applied with a FORTRAN program which modifies the speed and capacity of controlled links. A comparison of assigned volumes to ground counts by control device type showed reasonable results, and the constrained speeds on controlled links were also found acceptable.

Most highway network analyses do not attempt to simulate the presence of traffic control devices. This is mainly because highway networks have most often been used in regional studies, where intersection delay is usually not a significant part of total zone-to-zone travel time. Also, many planners are not satisfied with the usual methods of representing control devices: turn penalties or "microcoding." UTPS has had a history of difficulties in path building when turn penalties are used, and the documentation of at least one microcomputer package (MINUTP) suggests that this could still be of concern (*1*). Finally, the implementation of properly calibrated turn penalties on even a moderate scale would require resources well beyond the limit of most sub-area studies.

However, it is becoming much more common these days for computer-based highway network planning techniques to be applied to smaller analysis areas, such as sub-areas of a region or neighborhoods. One of the main difficulties in using planning software for such smaller areas is in defining the appropriate level of network detail. Generally, the smaller the area, the more significant intersection delay becomes in defining realistic travel paths. Since control devices are a powerful influence on delay at that scale, it thus becomes more important for a sub-area network to include them in some way.

This was the situation in the recently-completed Somerset Expressway Alternatives Study, a study of corridor highway alternatives in Somerset County, N.J. The analysis focused on a subsection of the county and was particularly concerned with future peak-hour traffic volumes. A very detailed network was used, covering almost all roads in the study area. The sub-area was small enough so that intersection delay was judged to be an important influence on path selection. However, the area was too large to permit a detailed investigation of intersection delay, so the use of turn penalties or microcoding would have been inappropriate. Thus, the project sought an alternative methodology that would permit the network to be sensitive to the presence of control devices.

Although there are numerous types of traffic control devices, this project dealt only with traffic signals and stop signs. Each signalized intersection was defined in terms of its major and minor approaches, based on known roadway and traffic characteristics. For intersections with stop signs, it was assumed that only the minor approaches were controlled. An inventory of the control device types for all intersection approaches in the study area was quickly prepared.

Thus, a method was needed that would permit the simplified simulation of intersection delay in a highway network with a minimum of new data required. The method would also have to be easily implemented within the MINUTP software package, which was being used for the Somerset Expressway project. The following sections of this paper describe the development of this method.

THEORY

The theory behind this project's simulation of control devices is that such devices affect link travel times in two ways. First, they reduce free-flow times by their mere presence. That is, the existence of a control device introduces a certain amount of delay, thus increasing link travel time and reducing effective speeds even if there is no traffic congestion on the link. A fixed delay value is associated with each control device type. This "zero-volume" delay would occur to a single vehicle as it approaches the intersection.

Second, it is theorized that as the volume grows, a link with a control device "behaves" differently than if that link had no such device, explained as follows. In the normal MINUTP capacity restraint calculation, the link running times are recalculated after each iteration as a function of volume using the standard default Bureau of Public Roads (BPR) curve, defined as follows:

$$T = T_0 \cdot \{1 + [S_c \cdot (V/C)^4]\} \quad (1)$$

125 Malcolm Rd., Mahwah, N.J. 07430.

where

T = new link time,
T_0 = free-flow link time (in MINUTP, this can optionally be taken as the time from the previous iteration),
S_c = sensitivity parameter (MINUTP default value = 0.15), and
V/C = ratio of volume to capacity (previous iteration).

Expressing this relationship in terms of speed on a link of a given distance, the equation becomes:

$$S = \frac{(D \cdot 3{,}600)/[(D \cdot 3{,}600)/S_0]}{1 + [S_c \cdot (V/C)^4]} \qquad (1.1)$$

where

S = new link speed (mph),
S_0 = free-flow link speed (mph), and
D = link distance (mi).

Thus, as volume increases on a given link, its speed is reduced. This function generates a curve of speed reduction factors that begin to decline sharply at V/C ratios above 0.75.

The second theory being explored is that the presence of a control device should reduce speeds even more sharply than equation (1.1) estimates. Not only does a control device reduce a link's capacity at each end (i.e., at the intersections), but it introduces additional delay based on queuing and increased friction from conflicting traffic. This additional delay should also be related to the V/C ratio, but the hypothesis is that the curve should be steeper than the default curve.

Of course, actual delay at nodes (intersections) is a function of many factors, not the least of which are the number of approach lanes, turning volume, opposing turns, signal progression, and cross-street volume. However, the challenge in this project was to develop a delay estimate for each link's B node based only on the characteristics of that link: length, free-flow speed, capacity, assigned volume, and control device type. No resources were available to modify the MINUTP software or write substantial new programs to access other network information. Thus, the main hypothesis of this method was that the delay at each intersection could be reasonably estimated from just the "known" data for each link.

In order to test these theories, detailed models of traffic delay were obtained from other sources. Specifically, the 1985 *Highway Capacity Manual* (2) provides an estimate for delay at a signalized intersection, assuming random arrivals. This function is as follows:

$$d = 0.38 \cdot C \cdot \left(\frac{(1 - G/C)^2}{1 - (G/C) \cdot X}\right) + 173 \cdot X^2 \cdot [(X - 1) + \sqrt{(X - 1)^2 + (16 \cdot X/c)}] \qquad (2)$$

where

d = average stopped delay per vehicle (seconds),
C = cycle length (sec),
G/C = ratio of effective green time to cycle length,
X = ratio of volume to capacity, and
c = capacity [vehicles per hour (vph)].

It is theorized that this function can be used to represent network delay at signalized nodes, as a function of volume and capacity. For simplicity, it was assumed in the Somerset project that all signals in the study area could be represented as having a cycle length of 90 sec. Further, it was assumed that the major approach legs all have a G/C value of 0.50 and the minor approach legs all have a G/C of 0.35 (the remaining 15 percent of the cycle is clearance time and was assumed to be unavailable for use by vehicles). This equation further assumes that there is no progression of signal timing. Although these conditions do not hold for every intersection in the study area, it was felt that they are, on average, sufficiently representative for the purpose at hand.

For intersection approaches controlled by a stop sign, it was theorized that delay could be modeled as if the intersection were uncontrolled. For these intersections, the ITE Handbook provides a method of estimating delay (*3*). Research by R. Ashworth indicates that average waiting delay can be estimated by:

$$E(d) = t1 - t2 - B2 + T \cdot \{e^{(q \cdot t2} - e^{[q \cdot (t1 - B2)]}\} \qquad (3)$$

provided that $t2 > t1 - B2$
where

$E(d)$ = delay per vehicle (sec),
$t1$ = critical gap of first driver in queue (sec),
$t2$ = critical gap of second driver in queue (sec),
$B2$ = time for second vehicle to move up to the head of the queue after the first vehicle departs (sec),
q = ratio of volume to capacity, and
T = $1/q$.

For simplicity, the following average values were assumed: $t1 = t2 = 8.5$ sec and $B2 = 3$ sec.

MODEL DEVELOPMENT

As noted above, MINUTP offers one method of determining link delay as a function of volume and capacity: the BPR equation [equation (1)]. After each traffic assignment iteration, the MINUTP ASSIGN program determines the V/C ratio for each link, and then calculates a new link time for the next iteration, using this equation. ASSIGN does, however, offer the user the opportunity to modify the value of the sensitivity parameter (S_c) in equations (1) and (1.1). This parameter affects the shape of the delay curve. A different value can be used for each of 63 possible speed categories.

The other parameter in equation (1.1) that can be adjusted is the link capacity (C). Before the assignment process even begins, the original link capacity in the unloaded network can be modified for links that have a control device. In practice, the original link capacity can be multiplied by a factor lower than 1, resulting in a lower effective capacity.

Since the BPR equation is built into ASSIGN and can be adjusted only as described above, it was decided to use this formula, modified as needed, to represent the delay that would be estimated by equations (2) and (3). The basic approach was to plot the values given by each equation as a function of V/C ratio, and then modify the S_c and capacity adjustment values until the BPR equation gave results that were sufficiently close to the theoretically derived values.

According to the theory postulated above, links with control devices should have S_c values greater than 0.15. Such values produce a steeper curve, indicating that speed drops more sharply with increasing V/C, compared to links without control devices. In addition, control devices generally reduce capacity compared to totally free-flow conditions. Thus, there is a sound theoretical basis for reducing the link's original capacity value in the presence of a control device. (Most traffic engineers would observe that signalizing an intersection's minor approach legs actually *increases* their capacity. However, in this case, as in most network models, nodes without traffic control devices have no delay, penalties, or constraints at all. Therefore, adding a control device to such a theoretical intersection would serve to restrict its capacity somewhat.)

Figures 1–3 show the speed reduction equations as a function of V/C ratio for major signalized approach, minor signalized approach, and stop sign control, respectively. In each figure, the upper curve represents the speed reduction that MINUTP would impose by default, based on the link volume and equation (1.1), without any control devices. This is referred to as the "Uncontrolled Speed." (Figures 1–4 and the following discussion demonstrate the methodology for a typical link with a 30 mph original free-flow speed, original capacity of 1,300 vphpl, and a length of 0.5 mi. The specific calculations depend on the link speed and length; 30 mph, 1,300 vphpl, and 0.5 mi are average values.)

Major Signalized Approach

As Figure 1 shows, the theoretical *Highway Capacity Manual* (HCM) curve calculated by equation (2) does have a steeper slope than the Uncontrolled Speed curve, particularly for V/C ratios of 0.8 to 1.5. The HCM curve also includes a factor for zero-volume delay. This is the delay that a signal imposes on a single vehicle, in the absence of any other traffic at the intersection. According to equation (2), when volume is zero, delay is 8.6 sec.

The development of the model, then, consists of fitting another BPR curve to match the HCM curve. This was done by modifying the S_c value in equation (1.1) and reducing the capacity by a fixed percentage until a reasonable match is obtained. This was done by entering the data into a spreadsheet and conducting several iterations manually until the two curves converged sufficiently. The resulting curves are labelled "Mod. BPR" in Figures 1–3.

Since equations (1.1) and (2) are fundamentally different, an exact match could not be obtained. However, in general, the Mod. BPR and HCM curves in Figure 1 are fairly close.

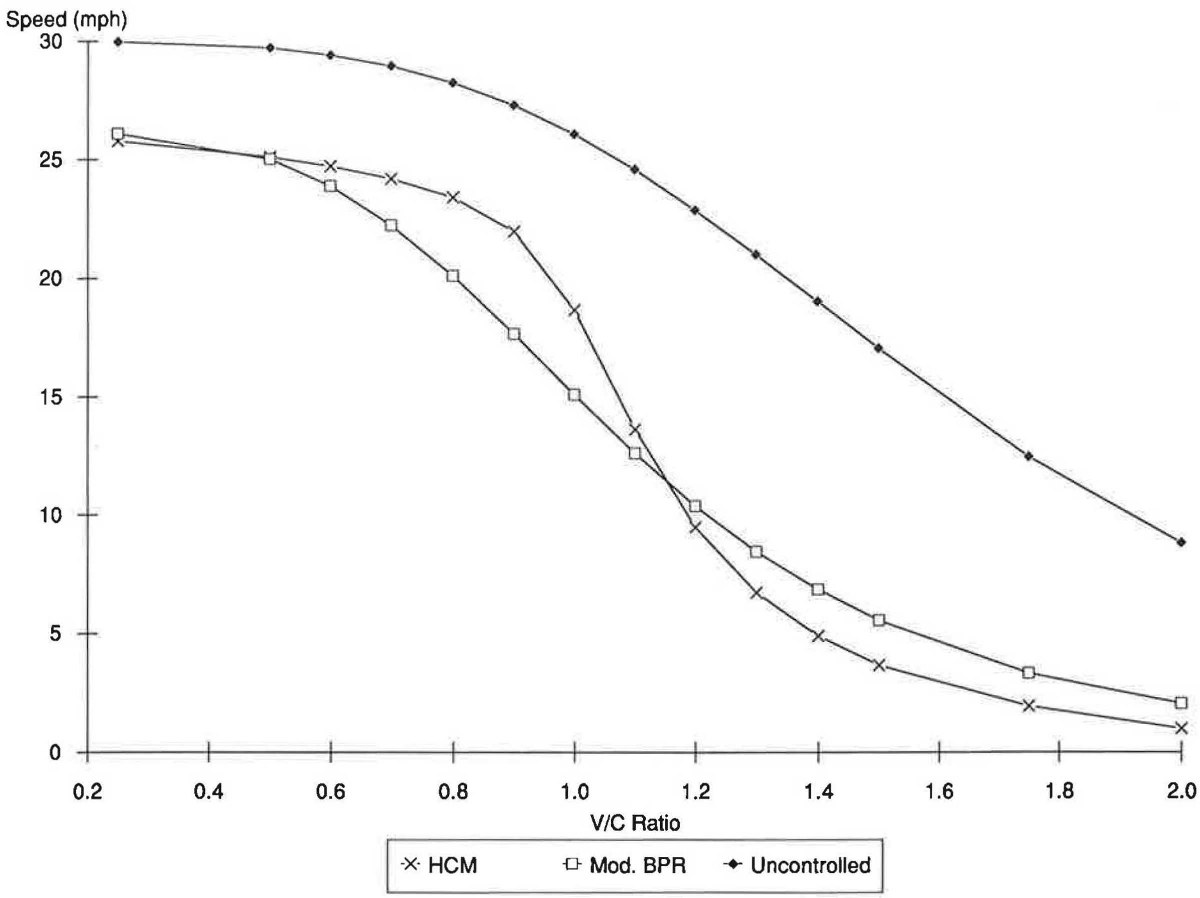


FIGURE 1 Signallized major approach speed reduction.

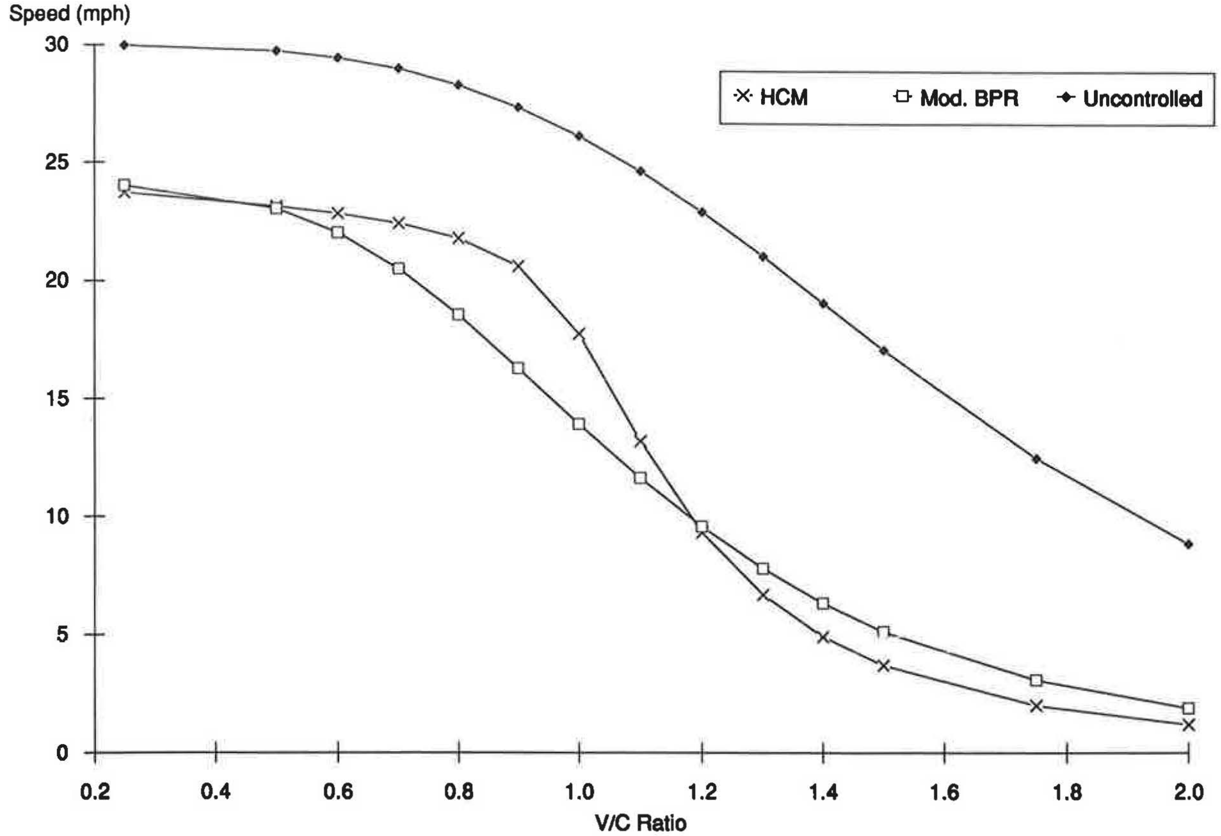


FIGURE 2 Signalized minor approach speed reduction.

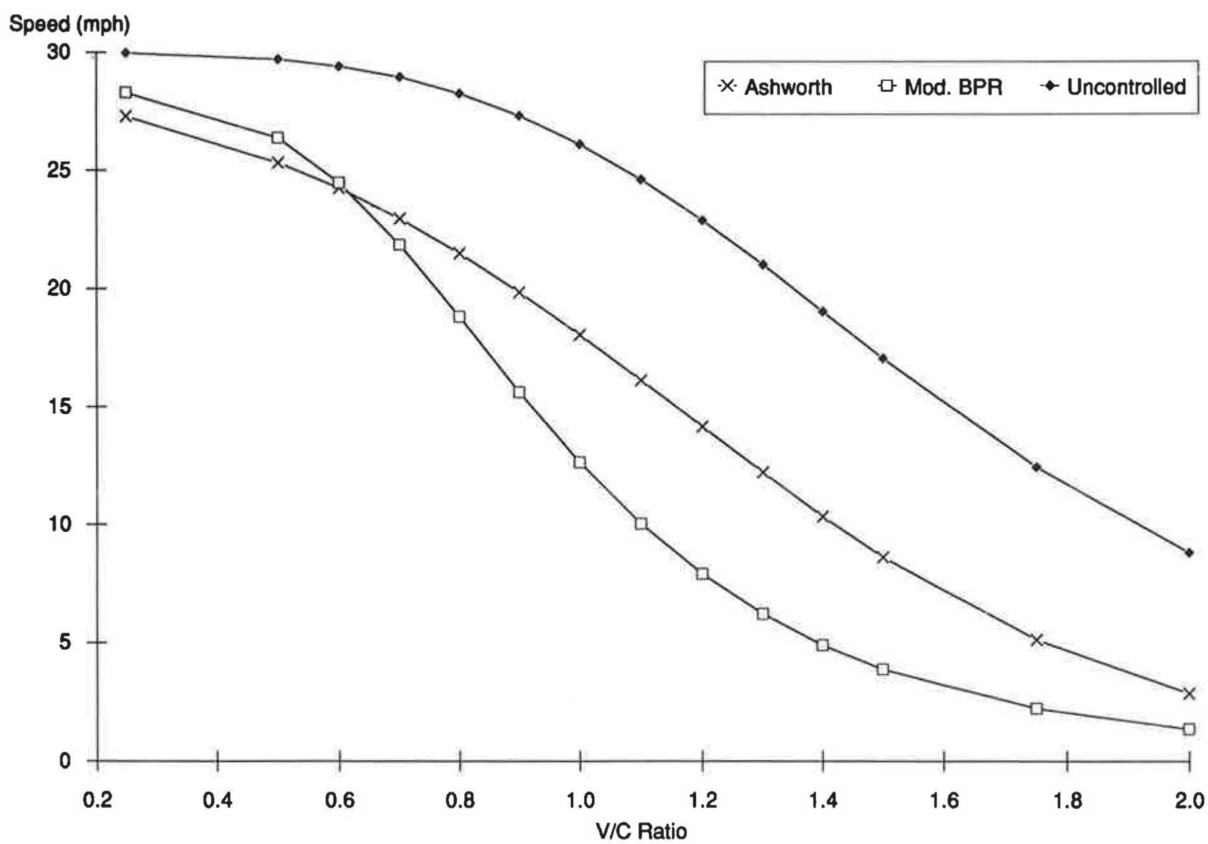


FIGURE 3 Stop sign speed reduction.

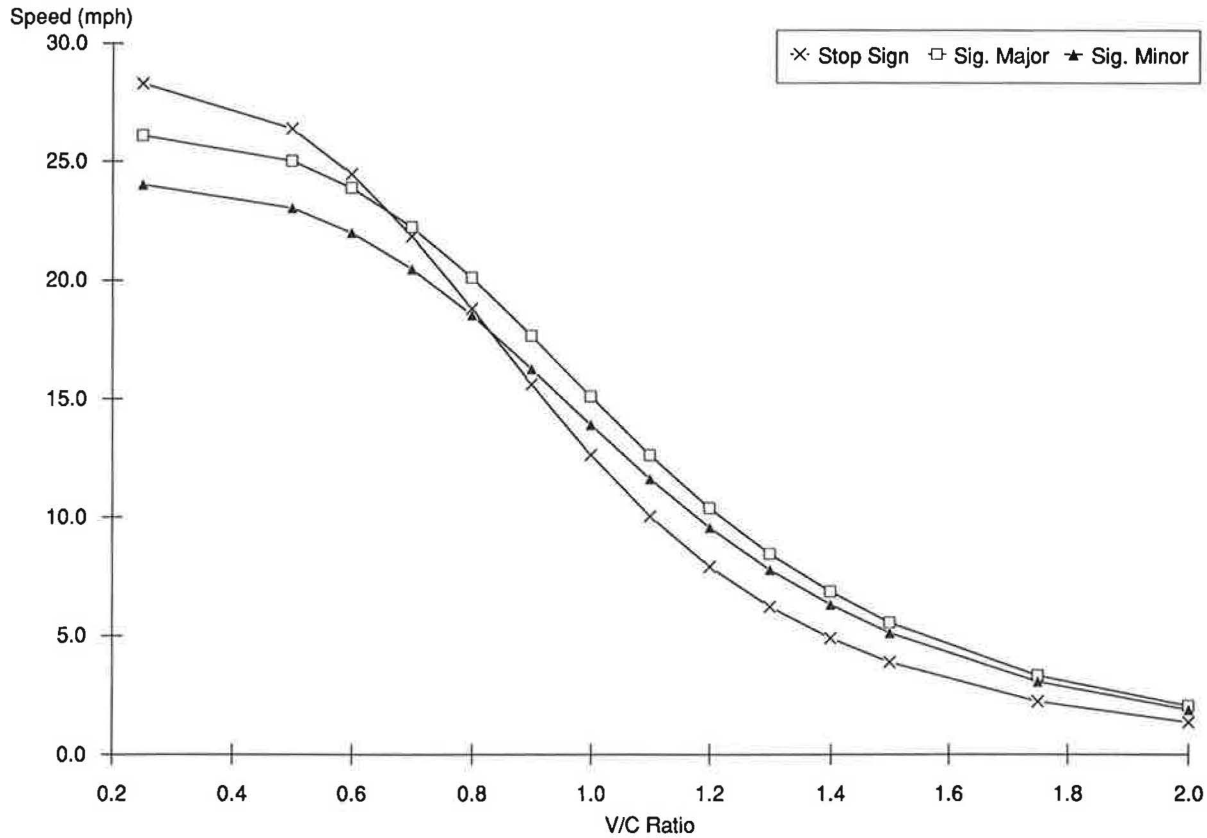


FIGURE 4 Speed reduction for all control devices.

The modified BPR formula underestimates speed for V/C of 0.5 to 1, and slightly overestimates speed for V/C above 1.2.

The final BPR Speed curve shown in Figure 1 was produced by modifying equation (1.1) as follows:

$$S = \frac{(D \cdot 3{,}600)/\{[(D \cdot 3{,}600)/S_0] + 8.6\}}{1 + 0.30 \cdot [V/(0.8 \cdot C)]^4} \quad (4)$$

where

- S = new link speed (mph),
- S_0 = free-flow link speed (30 mph in this example),
- D = link distance (0.5 mi in this example),
- V = link volume (from the previous iteration),
- C = link capacity (1,300 vph in this example).

As predicted, the S_c value of 0.30 indicates a steeper curve, compared with the default value of 0.15 for other links. The capacity is reduced by 20 percent, also confirming the above hypothesis. Thus, this equation was judged acceptable for estimating network speed reductions.

Minor Signalized Approach

A similar method was followed in developing the speed reduction function for the minor approach of a signalized intersection. The theoretical delay was calculated using equation (2), and the resulting reduction in speed was plotted as a tunction of V/C ratio (see the HCM curve in Figure 2). As noted above, the cycle length was assumed to be 90 sec and the minor approach G/C ratio was assumed to be 0.35. The only difference between the delay calculated for the major and minor approaches is that the zero-volume delay, based on equation (2), is approximately 14.5 sec.

As Figure 2 shows, the HCM curve (equation (2)) is somewhat parallel to the Uncontrolled Speed curve for V/C ratios of less than about 0.8. Above that point, the HCM curve shows a very sharp speed reduction with increasing volume, leveling off at a V/C of 1.5 and higher.

The fit of a BPR-type function to the curve calculated by equation (2) is also shown in Figure 2. The equation of this modified BPR Speed curve is

$$S = \frac{(D \cdot 3{,}600)/\{[(D \cdot 3{,}600)/S_0] + 14.5\}}{1 + 0.30 \cdot [V/(0.8 \cdot C)]^4} \quad (5)$$

As Figure 2 shows, the modified BPR curve matches the HCM curve well at V/C below 0.6. From 0.6 to 1.3, the modified BPR curve underestimates the HCM speed. Above 1.3, the two curves are similar. Overall, this fit was judged to be acceptable. The sensitivity factor (S_c) and the capacity reduction factors are the same as for the major signalized approach, with the only difference being the higher zero-volume delay.

Stop Sign Control

The same procedure was followed for the stop sign model, but using equation (3) as the theoretical basis. As Figure 3 shows, the "Ashworth" speed curve calculated by equation (3) also has a steeper downward slope than the Uncontrolled Speed, for V/C values below 1.0. Above 1.0, the two curves are just about parallel. However, this does not mirror actual field conditions, in which stop sign control starts to break down as the V/C approaches 1.0. In addition, comparing Figures 1, 2, and 3 suggests that the theoretical delay at stop signs is less than delay at signalized intersections for V/C above 1. This also seems illogical.

It was felt that at a V/C of about 0.75, the speed should drop even faster than equation (3) suggests. Thus the parameters of equation (1.1) were adjusted so as to produce the bottom curve of Figure 3. Note that the Mod. BPR and Ashworth curves begin to converge again at V/C ratios above 1.2.

The Mod. BPR curve also include a factor representing zero-volume delay. This is the delay that a stop sign imposes on a single vehicle, in the absence of any other traffic at the intersection. This delay is fixed at 3.3 sec, which is calculated assuming an initial speed of 30 mph and a deceleration rate of 4.6 mph/sec. The Ashworth equation does not include a factor for zero-volume delay.

The modified BPR speed curve in Figure 3 is represented by this equation:

$$S = \frac{(D \cdot 3{,}600)/\{[(D \cdot 3{,}600)/S_0] + 3.3\}}{1 + 0.30 \cdot [V/(0.7 \cdot C)]^4} \qquad (6)$$

Although the fit between the modified BPR and Ashworth curves is not as good as for the other models, the overall result is still consistent with the other models. The sensitivity factor is the same (0.30) as for the signalized approaches, but the capacity adjustment factor is slightly lower (0.7 vs. 0.8), properly suggesting that replacing a stop sign with a signal increases capacity for a minor approach.

OVERALL COMPARISON

Figure 4 presents the curves for equations (4), (5), and (6) in one graph. Below a V/C of about 0.7, the signalized approaches show a higher delay (lower speed) than the stop sign approach. Above 0.7, the stop sign-controlled link starts to break down faster, resulting in lower speed that either signalized approach. At very high V/C ratios, all three equations provide similar results.

Another test of the validity of the control device model is the comparison of assigned volumes to counts for the sub-area. The Somerset Expressway project focused on hourly volumes. Table 1 provides a comparison of the total of estimated vs. observed a.m. and p.m. peak hour traffic volumes, by type of control device, for all sub-area links which had counts posted in the network.

As this table shows, the control device model did not, by itself, lead to perfect assignments. However, neither did it unduly distort the network. These results suggest that more time could have been spent to fine tune the methodology for this project. For example, the parameters for the signallized approaches (zero-volume delay, capacity factor, BPR sensitivity) could have been adjusted to decrease the delay associated with the minor approaches to signallized intersections and slightly increase the delay for the major approaches. The stop sign delay function would appear to need no further changes.

TABLE 1 ASSIGNED VERSUS COUNTED VOLUME FOR DIFFERENT CONTROL DEVICES

	AM			PM		
Control Device	Counted	Assigned	Difference	Counted	Assigned	Difference
Signallized Major Approach	44,000	51,400	+17%	51,800	52,030	0%
Signallized Minor Approach	6,740	5,200	-23%	9,540	7,130	-25%
Stop Sign	5,680	5,830	+4%	7,940	7,160	-10%

IMPLEMENTATION

This model was implemented in MINUTP as follows:

- The highway network was coded as usual in MINUTP, with speed and capacity look-up table values coded in the SPDC and CAPC fields.
- A new data field was added to the network, called "CDEV."
- For each link, by direction, if there is a control device at the link's B node, a number was entered into the CDEV field: 1 for stop sign, 3 for major signallized approach, and 4 for minor signallized approach.
- A short FORTRAN program was written to read the binary network file and see if a control device was coded. If so, then the link's capacity was reduced and a new speed calculated based on the link's distance and original free-flow speed. (This could also have been done using the MINUTP NETMRG program, but the flexibility of FORTRAN was needed for other reasons.)
- In the batch file of set-up commands to run the ASSIGN program, the value of S_c for the links with control devices was changed from the default of 0.15 to 0.30.

From this point forward, the assignment process proceeded as usual in MINUTP. It should also be possible to implement this procedure, or a variation thereof, using other planning software packages.

CONCLUSIONS

The presence of traffic control devices is rarely simulated in regional network modelling. This is mainly because (1) such networks are not detailed enough that control devices would make a difference, and (2) the available means of simulating such devices (e.g., turn penalties or microcoding) are usually cumbersome or otherwise unsuitable. In sub-area networks, however, it is often desirable to represent the effects of intersection delay, and specifically delay associated with traffic control devices, in order to obtain more realistic paths and assignments.

This analysis demonstrates that an alternative approach is available. The normal BPR-based link capacity restraint function built into MINUTP can be modified to provide an approximation of the delay associated with traffic control devices. This estimate of delay is based on detailed, theoretical models of delay and is internally consistent among control device types. The method is relatively simple and can be implemented within existing planning software systems. This method provides an adequate compromise between data and resource limitations, and theoretically proper delay models, and should be useful in other sub-area network applications.

ACKNOWLEDGMENT

The research described in this paper was conducted by the author under contract to Garmen Associates, which performed the Somerset Expressway Alternatives Study for the New Jersey Department of Transportation. The author wishes to acknowledge the efforts of the Garmen staff in assembling the data for this work, and particularly their project manager, Gary Davies, whose determination provided the basis for developing this control device delay model.

REFERENCES

1. *MINUTP Technical Users Manual.* COMSIS Corporation, Mountain View, Calif., February 1988.
2. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
3. W. S. Homburger, ed., *Transportation and Traffic Engineering Handbook*, 2d ed. Institute of Transportation Engineers; Prentice-Hall, Englewood Cliffs, N.J., 1982.

Publication of this paper sponsored by Task Force on Transportation Planning Applications.

Evaluation of Three Inexpensive Travel Demand Models for Small Urban Areas

C. J. KHISTY AND M. Y. RAHI

Conventional urban travel demand models, which are data-hungry, costly, and mainly meant for use in large cities and metropolitan areas, are not suitable for small urban areas with a population of 500,000 or less. These small urban areas generally lack the staff, expertise, and budget to operate the conventional models. Three simplified travel demand models are evaluated that are suitable for small urban areas and make use of routinely collected ground counts. These three models are applied in a common setting to the City of Pullman (1980 population 23,579) in the State of Washington. Socioeconomic data and routinely collected ground counts for 1970 were used as inputs to run the models, and the outputs (travel forecasts) were compared with 1980 ground counts, to determine their forecasting capability. All three models tested performed very well. The RMS error ranges between 9 and 15 percent, and the link volume forecasting capability for most of the links ranges between 10 to 15 percent of the observed volumes. Contacts with selected planning organizations in the State of Washington reveal that such methods will be useful in small urban areas, considering their staff, expertise, time and budget limitations. Currently, these small urban areas use unproven heuristic methods. The models described in this paper will considerably help small urban areas to forecast travel demands, using routinely collected traffic ground counts and socioeconomic data, with confidence.

A primary purpose of the transportation planning process is to generate information useful to decision makers on the consequences of alternative transportation-related actions (*1*). The objective of this process is to provide information necessary for making decisions on when and where improvements should be made in the transportation system, thus satisfying travel demands and promoting land development patterns that are in keeping with community goals and objectives (*2*). Much of the attention of transportation planners and policy makers in the past has concentrated on the problems of metropolitan areas. In recent years, however there has been a greater awareness of transportation problems in small urban areas. Small urban areas, generally those with populations under 50,000, have somewhat different transportation planning needs as compared to large metropolitan areas. These needs require planning techniques that are less data hungry, less costly, and less time consuming.

This paper addresses the transportation planning needs of small urban areas with populations under 50,000. Nationwide, these small urban areas with populations between 2,500 and 50,000 have been gaining in their population share, as indicated in Table 1 (*3*). These population changes suggest a need for a renewed focus on the transportation planning requirements of small urban areas.

C. J. Khisty, Department of Civil and Environmental Engineering, Washington State University, Pullman, Wash. 99164. Current affiliation: Department of Civil Engineering, Illinois Institute of Technology, Chicago, Ill. 60616. M. Y. Rahi, Associated Traffic Consultants, 99 South Chester Ave., Suite 200, Pasadena, Calif. 91106.

One of the most important pieces of information for making decisions regarding transportation improvements is the horizon-year traffic volumes on the major links of a city's transportation network. It is customary for most cities and counties to collect traffic counts on their street system on a routine basis. This data base, consisting of base-year ground counts, can be put to good use in forecasting horizon-year traffic flows. Some inexpensive techniques for forecasting travel demands have been developed in recent years, using routinely collected ground counts and socioeconomic data. However, they have, to date, not been evaluated and tested in a common setting.

This paper describes, discusses, and evaluates inexpensive travel demand models using routinely collected ground counts. In a report (*4*) prepared for the Washington State Department of Transportation, several methods were examined of which four were found to be promising. Three of these four models are applied in a common setting to the City of Pullman (1980 population 23,579) in the State of Washington. Socioeconomic data and routinely collected ground counts for 1970 were used as inputs to run the models, and the outputs (link forecasts) were compared with 1980 ground counts to determine their forecasting capability. The reason for comparing only three of the four promising models is that the three evaluated in this paper are "calibrated" using the base-year traffic counts, while the fourth procedure (FHWA's Synthetic Traffic Simulation Method) is not "calibrated" or adjusted using the base-year traffic counts.

OVERVIEW

Travel demand models using routinely collected ground counts are a comparatively recent endeavor. The development of a travel demand forecasting model based on base-year traffic ground counts was first initiated as recently as 1972 (*5*). In these models, the link traffic volumes in the base year are used for calibration; the horizon-year socioeconomic variables and base-year calibrated models are then used to predict the traffic volumes in selected links of the network for the horizon year (Figure 1).

The transportation characteristics of small urban areas as well as their prevailing transportation problems have been described in a U.S. Department of Transportation (USDOT) document (*6*). As seen in Table 1, the proportion of total population living in small urban areas increased from 28.7 percent in 1950 to 39.6 percent in 1980. These population

TABLE 1 POPULATION SHARE OF COMMUNITIES OF DIFFERENT SIZES IN THE UNITED STATES (3)

Size of Areas	% of Total Population			
	1950	1960	1970	1980
Large Urban Areas 50,000+	35.3	36.2	35.9	34.1
Small Urban Areas 2,500 – 50,000	28.7	33.7	37.7	39.6
Rural Areas Less than 2,500	36.0	30.1	26.4	26.3

changes have generated an awareness of the transportation problems of these areas, which were all too frequently overlooked in the past. For instance, the existing transportation systems must accommodate new land use developments and handle the local traffic impacts resulting from the development of new activity centers such as shopping centers, industrial parks, etc., which were not too common in the past. This trend demands the development of planning tools and methodologies suitable for use in such communities.

Unfortunately, the growth of population and the subsequent demand for transportation facilities and services in small urban areas have not induced a proportional growth in the planning staff, budget, and other resources in these areas. As a result, most of these communities lack the resources necessary to run sophisticated conventional transportation models to make appropriate decisions. Most recently, in a survey conducted in the State of Washington, a selected number of county and regional planning councils, conferences, and commissions representing small- and medium-size urban areas were requested to express their experiences in forecasting travel demands in their study areas. In their responses, most of them indicated that the conventional models are inappropriate for use and that simplified, easy-to-understand, and inexpensive travel demand models would be useful, if available. They also mentioned that the lack of qualified staff, expertise, time constraints, and a limited budget often added to their problems (*4*).

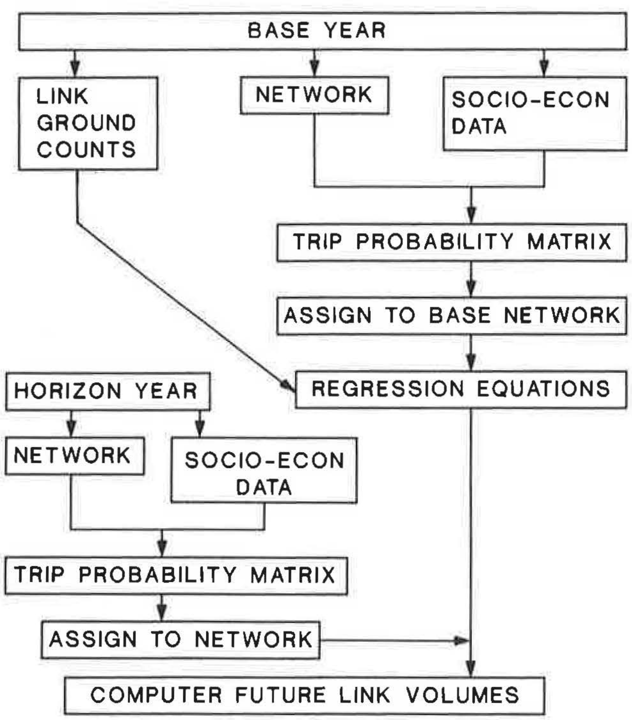


FIGURE 1 Flow chart of a travel demand model based on ground counts.

DESCRIPTION OF THE MODELS

It is customary for most cities and counties, including small urban areas, to collect traffic counts on their street system on a routine basis. For most small urban areas these counts serve the purpose of comparing and validating the estimates of the link volumes forecasted by some heuristic methods. Since these data are readily available in any small urban area, it was felt that demand models that make use of these data as basic inputs in travel forecasting would be the best choice for the small urban area planner. Several such models have been identified, of which four have been found applicable for small urban areas (*4*). These four models can be easily applied to small urban areas, because the input requirement of these models, besides link ground counts, is minimal. Moreover, the computational resources and expertise needed to run these models are also within the capability of the small urban area planner.

The four models identified as applicable for small urban areas are:

1. Low's model;
2. FHWA's Synthetic Traffic Simulation Method;
3. Neumann et al.'s technique; and
4. Khisty-AlZahrani's model.

A brief description of each of these models follows.

Low's Model

In this model, traffic volumes are determined one link at a time, primarily as a function of the relative probability that one link will be used in preference to another (*5*). Interzonal trip probabilities are assigned to the network, using traffic assignment procedures, to produce estimated trip probabilities on a link-by-link basis. Regression equations are developed to relate the counted link volumes to assigned trip probabilities and other link characteristics. These equations are then used to estimate link volumes for the horizon year, after determining and assigning new interzonal trip probabilities to reflect those conditions. The procedure was applied as a volume forecasting model for a metropolitan area in West Virginia. Hogberg (*7*) as well as Smith and McFarlane (*8*) evaluated the model by applying it to various small urban areas.

Some theoretical limitations have been pointed out in Low's model. In contrast to the conventional urban transportation demand (UTD) models, it does not exhibit model stability over time because of incorrect specifications. One misspecification is in representing trip productions and attractions by only production and attraction characteristics. The result is that the changes in the trip-making propensity of the study area population over time are not included in the model. For example, the population of an urban area could remain stable over time, but the number of trips could increase dramatically because of increases in, say, auto ownership.

A step-by-step procedure of the model follows:

1. Estimate trip productions and attractions for the zones. For work trips, zonal productions P_i equals the number of workers living in zone i, and zonal attractions A_j equals the number of employments located at zone j. This has to be done for both base year and horizon year.
2. Calculate f_{ij} (friction factors between zones i and j) using the equation:

$$f_{ij} = P_i A_j t_{ij}^{-2}$$

where t_{ij} is the travel time between zones i and j.

3. Assign f_{ij} (friction factors) to the network by all-or-nothing technique.
4. Develop regression equation, using the available link counts, as follows:

$$V^{kl} = a + b \sum P_{ij}^{kl} f_{ij}$$

where V^{kl} is the base year traffic volume on link $k - l$; P_{ij}^{kl} is equal to 1 if trips between $i - j$ is found on link $k - l$, 0 otherwise; and a and b are calibrated constants.

5. Estimate the horizon year link volumes using the regression equation and the assigned horizon year f_{ij} values.

FHWA's Synthetic Traffic Simulation Method

This simulation procedure has essentially the same components as that of the conventional transportation planning process, namely, trip generation, distribution, and assignment (*9*). Modal choice is not considered since the role of mass transit is relatively minor in most small urban areas. The basis of the approach is the borrowing of information and experiences from other studies to develop trip generation and trip distribution models. By borrowing travel relationships, the home interview survey and much of the data editing and analyses are eliminated. This elimination results both in reduced costs and in great time savings. Small cities are defined as those having less than 100,000 population.

The process begins by collecting standard socioeconomic and traffic count data. An internal origin and destination survey is eliminated. The highway network and appropriate zones are coded. Trip generation relationships (cross-classification trip rates or equations) are selected from other cities of similar size and characteristics. Productions and attractions for each trip purpose by zone are computed. These results are checked for reasonableness by the use of selected control zones, comparison of estimated vehicle-miles of travel (VMT) with actual VMT, and comparison of trip rates per dwelling unit (DU) and per capita with other similar studies. The productions and attractions are distributed by purpose using the gravity model and friction factors transferred from other transportation studies. The resulting trip length frequencies for each trip purpose are checked for reasonableness by comparison with the frequency curves from similar urban areas. The trip assignment is then made to the existing network. Gross checks and fine tuning are done to adjust the model.

Several small urban areas have successfully used the approach, achieving satisfactory reproduction of travel patterns. With reasonable care, the procedure should produce results which are good enough to be used in making decisions regarding future transportation plans and for the evaluation of the current system adequacy. Experienced transportation planning personnel and data analysts are needed to apply the method. Standard socioeconomic data, traffic counts (standard cordon, central business district cordon, screenline crossings, etc.), and highway network details are needed, as well as a comprehensive knowledge of models, procedures, and results of similar transportation studies.

The merits of the procedure lie in the large savings in cost, time, and effort, due to the elimination of the internal survey and minimal data editing and analysis. The limitations, however, include dependence on other transportation studies of similar dimensions, which may be difficult to come by. Inaccuracies from borrowed relationships may also be carried over. On the face of what has been stated here, this procedure is not as simple and efficient as it sounds. The saving in time and money lies in transferring rates from other studies.

A step-by-step procedure of the model follows:

1. Borrow trip rates for zonal production and attraction from other studies.
2. Apply the gravity model to distribute the zonal productions and attractions. The friction factors can be borrowed from NCHRP 187 (*10*).
3. Assign distributed trips to the network.

This model is not evaluated in this paper along with the others for reasons mentioned before.

Neumann et al.'s Technique

The model directly estimates areawide, all-purpose trip production rates (*11*). The method distributes and assigns zonal socioeconomic variables (autos, dwelling units, and population) directly to the study area network. External trips are deducted from the total ground counts to obtain the internal trips of the area. These ground counts are entered into a linear regression model as the dependent variables, and the assigned socioeconomic variables are entered as the independent variables. The resulting regression coefficients are the estimates of the areawide, all-purpose trip production rates.

The method is sensitive to the friction factor curves used in the trip distribution step. The methodology was tested in Lynchburg, Va. (1970 SMSA population 146,000), and Lexington, Ky. (1970 SMSA population 295,000). Estimated production rates obtained were within 96 percent of true rates. It produces accurate results and can be used in verifying borrowed production rates in the synthetic procedures. Since the method is sensitive to the accuracy of the friction factors utilized, a large proportion of the count stations should be located outside the central business district (CBD). In addition, data must be available on external-external and external-internal trip volumes for the city. This suggests that the methodology might find application in cities where the resources are sufficient to conduct external surveys. If a coded network is already available, the effort required for analysis could involve about three man-days, and the cost of data collection would be negligible.

A step-by-step procedure of the technique is shown below:

1. Assign base year socioeconomic variables directly to the network. The socioeconomic variable available for application can be the zonal population. The gravity model is used to distribute the variables. The zonal population and employment (as attraction) for base year are used for this purpose. The friction factors for the model are borrowed from NCHRP 187 (*10*). Using the travel time matrix, the gravity model is applied to distribute the socioeconomic variable. Assignment of this matrix gives the link totals for the socioeconomic variable (population).

2. Develop regression equation. The available link counts are used as the dependent variable while the corresponding link totals for population are used as independent variable for regression analysis. The following regression equation results:

$$V^{kl} = a + b \sum P^{kl} X_p^{kl}$$

where V^{kl} is the base year traffic volume on link $k - l$, a and b are the calibrated constants, and X_p^{kl} is the contribution of socioeconomic variable p to link $(k - l)$, and P^{kl} is either 1 (if link is contributed by p) or 0 (otherwise). Constant b is the areawide trip rate per person.

3. Assign socioeconomic variables for the horizon year. The horizon-year zonal population and employment are used to distribute the socioeconomic variable in the horizon year. The same set of friction factors are used in the gravity model. When the socioeconomic variable is assigned to the network it gives the link totals for the horizon year. These totals are used in the regression equation developed from the base-year data to estimate the horizon-year link volumes.

Khisty-AlZahrani's Model

This is an internal volume forecasting (IVF) model based on Low's model (*12*). The model incorporates improvements suggested by Smith and McFarlane (*8*). To eliminate the errors in Low's model, the model recommends replacing the zonal production and attraction characteristics with direct estimates of trip productions and attractions, and including origin zone accessibility in the denominator of the probability factor. The production and attraction variables and the friction factors are modified in this model as follows:

1. The production zone variable (P_i) is replaced by ER_i, the number of employees residing in zone i.

2. The attraction zone variable (A_j) is replaced by EW_j, the number of employees working in zone j.

3. The friction factor values $[F(C_{ij})]$ are calculated in a manner similar to that used in trip distribution models. $F(C_{ij})$ is an indirect indicator of the cost of travel (t_{ij}) between zones i and j

$$F(C_{ij}) = \exp(-0.10t_{ij})$$

The second theoretical limitation is avoided simply by dividing the attraction term EW_j at the destination by the total attractions of the study area, $\Sigma_j EW_j$.

The model was applied to the City of Spokane, Wash. (1980 population 171,300). The 1970 data were used to calibrate the model. The calibrated model was used to forecast the traffic volume for the horizon year 1980. It was assumed that the major street network would not change significantly between the base year and the horizon year. The only data that were necessary for the application of the model were the two variables ER_i and EW_j for 1980, which were obtained exogeneously. These values as well as the friction factor matrix were used to obtain the horizon-year trip interchange indices matrix. External-external and external-internal volumes on the links were obtained from suggestions provided in NCHRP 187 (*10*).

A comparison of the model with Low's original model indicated that the model output gave somewhat better results. Also, a comparison of the observed and the estimated horizon-year link volumes was made. Although the actual to estimated volumes for the horizon year ranged from 0.93 to 1.27, most volume groups were within 10 percent of the actual volumes.

Because network configuration and census data are generally available, the effort required to work the model for a small- or medium-sized city might involve 10 to 15 man-days. The model combines several conventional submodels into one process, and the output in terms of traffic volumes can be statistically described and tested. The model is quick, reliable, and transparent for forecasting travel in small urban areas.

A step-by-step procedure of the model follows:

1. Determine the number of workers residing in (labor force) and number of jobs available at each zone of the area under study.

2. Calculate values of the trip interchange index I_{ij} according to the formula:

$$I_{ij} = ER_i \frac{EW_j}{\Sigma EW_j} F(C_{ij})$$

where

ER_i = number of workers residing in zone i,
EW_j = number of jobs available at zone j,

$F(C_{ij}) = e^{-0.1t_{ij}}$ = friction factor between zones i and j, and
t_{ij} = travel time between zones i and j.

I_{ij} values are calculated for all the interchanges for both the base year and the horizon year.

3. Assign the I_{ij} values to the network. The all-or-nothing assignment is used to calculate the sum of the values of $P_{ij}^{kl}I_{ij}$ for each link. P_{ij}^{kl} values are equal to either 1 (if the trips between $i - j$ is found on link $k - l$), or 0 (if it is not).

4. Develop regression equation. Using link counts as the dependent variable $\Sigma P_{ij}^{kl}I_{ij}$ as the independent variable in the regression, the following equation is developed for use in the horizon year:

$$V^{kl} = a + b \sum P_{ij}^{kl} I_{ij}$$

where V^{kl} is the base year volume on link $k - l$, and a and b are calibrated constants.

5. Calculate horizon-year $\Sigma P_{ij}^{kl}I_{ij}$ and forecast for link volumes. The horizon-year values of I_{ij} are assigned to the network, and the values of $\Sigma P_{ij}^{kl}I_{ij}$ are calculated for all the links in the horizon year. These values are used in the regression equation developed from the base-year data to obtain the forecast-year link volumes.

APPLICATION OF THE MODELS

Three of the four models described in the previous section were selected for application in a common setting, because they emerged as the most promising techniques out of a total of 13 techniques considered for use in small urban areas (*4*). The criteria used to screen these techniques out of the currently available techniques were the following:

1. amount and type of data required;
2. transparency of the methodology used (i.e., absence of "black-box" effect);
3. simplicity in application;
4. ease of updating;
5. use of ground counts;
6. avoidance of large-memory computers;
7. use of hand-held calculators and/or personal computers;
8. cost of running the models;
9. time required in application; and
10. expertise required to run the models.

These ten criteria provide a rough measure of the suitability and applicability of the models for small cities. Table 2 shows applicability scores of these models. The application of the three models in a common setting was considered important, because the result of this experiment would provide a comparative evaluation of the models in terms of their performance, i.e., ease of their application, resource and time requirements, and accuracy. The City of Pullman in the State of Washington was chosen as a suitable free-standing small urban area for this evaluation. The city had a 1970 population of 20,509, which grew to 23,579 in 1980. Ground counts were available for this city for both the 1970 and the 1980 transportation network. The socioeconomic data required by the models were also available from the census records of 1970 and of 1980. The 1970 data were used to obtain the forecasted trips and these were compared to the actual 1980 ground counts. In the following sections, the data collection methodology and the performance of the models are described.

TABLE 2 APPLICABILITY SCORES OF THE METHODS [RANGE: 0 (BAD) TO 5 (GOOD)]

	Criteria										
	Data Required	Transparency	Simplicity	Ease of Updating	Use of Counts	Computer Required	Calculators/PC	Cost of Running	Time Required	Expertise Needed	
Model	1	2	3	4	5	6	7	8	9	10	Total
Low	4	1	3	0	5	4	4	5	5	4	33
FHWA	4	4	4	4	2	3	2	3	4	3	33
Neumann	4	4	2	3	4	3	3	4	4	3	34
Khisty-AlZahrani	4	4	3	4	5	4	4	5	4	3	40

DATA COLLECTION FOR THE MODELS

The minimum data needed by each of the three models are basically same. Data for both the base year and the horizon year were needed to apply the models and test their performances. The common set of data collected for the City of Pullman included:

- transportation networks for both base year (1970) and the horizon year (1980), showing the zones of origin and destination of trips along with travel times on the links,
- traffic counts on the links of the network, for both base year and the horizon year,
- population of each of the planning zones, for both base year and the horizon year, and
- employment data at each of the planning zones, for both base year and the horizon year.

The transportation network did not change significantly between the base year (1970) and the horizon year (1980) in the City of Pullman. The population and employment data for the city, obtained from the Bureau of Census by census tracts (CT), enumeration districts (ED), and small census districts (CD), were used in demarcating five planning zones. Zonal population and employment data were derived for both the base and the horizon years. These zones, along with the transportation network, are shown in Figure 2. For this analysis, the transportation network consisted of five planning zones, eleven major links and ten major nodes. There were also four external zones that contributed traffic to the link volumes on the major links.

The one census record that was easily accessible and useful was the Summary Tape File (STF) 3A. In particular, the data coded at Summary Level 15 in the STF 3A were sufficiently good for deriving zonal data. Although the zonal population data were readily available for both the base year and the horizon year, this was not the case for zonal employment opportunities data.

As mentioned earlier, the ground counts on the links of the transportation network of the city were readily available from the city engineering department. These data were recorded link-by-link in two city traffic study reports, one conducted around 1970 and the other around 1980. These reports provided road-link inventories as well.

Because travel demand forecasting models described here depend on exogeneous information on external trips, and their appropriate assignment to the internal network of the city, it was essential to seek an appropriate method to accomplish the external trips assignment on the network of the small urban area. The City of Pullman has four external neighboring communities, and traffic cordon counts in these neighborhoods were available in the traffic studies. Therefore, the methodology described in NCHRP 187 (*10*) to account for external trips was adopted. However, extreme care had to be

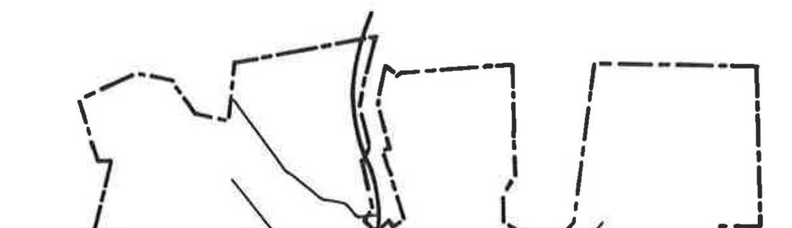


FIGURE 2 Transportation network of Pullman, Wash.

taken in assigning the external trips on the internal networks, although for most small urban areas, this task should not be a tedious exercise.

PERFORMANCE OF THE MODELS

All three models performed sufficiently well, although individual performance varied. All the models were first calibrated using the 1970 socioeconomic and transportation ground count data. This calibration yielded regression equations that could be used to forecast the horizon year (1980) link volumes. For the calibrated models, the goodness-of-fit of the regression equations was justified by the coefficient of determination (R^2) value and the Student's t-test. These values are shown in Table 3. The values appear to be quite reasonable, since the R^2 values are all above 0.80 and t_0 exceeds the threshold t of 3.25 at 99 percent confidence level with nine degrees of freedom.

After the horizon-year link volumes were forecasted by each of the models, they were compared with the actual 1980 traffic volumes of the corresponding links. Another set of regression equations was developed to relate the 1970 actual traffic volumes to the 1980 estimated traffic volumes of the corresponding links, so that the link volumes of minor street links (which were not used in the model calibration step) could also be forecasted. Once again, the goodness-of-fit of these equations was tested for validity. These values are shown in Table 4. The R^2 values range from 0.65 to 0.76, while t_0 once again exceed the threshold t of 3.25 at 99 percent confidence level with nine degrees of freedom.

Finally, for the comparison of the link volumes estimated by each of the models with those observed on the corresponding links, the root mean square (RMS) error and the percent RMS error values of each of the models were calculated and compared. The Percent Root Mean Square Error (% RMS) is defined as the ratio of the RMS error to the mean of the observed output variables. The RMS error, known as standard error, is the standard deviation of the sampling distribution. These are given as:

$$\text{RMS error} = [\Sigma(V_0^{kl} - V_e^{kl})^2/(N - 2)]^{1/2}$$

TABLE 3 GOODNESS-OF-FIT OF MODEL CALIBRATION

Model	R^2	t-test
Low	0.82	$t_0 = 6.29 > t_{9,.005} = 3.25$
Neumann	0.80	$t_0 = 5.94 > t_{9,.005} = 3.25$
Khisty	0.86	$t_0 = 7.50 > t_{9,.005} = 3.25$

TABLE 4 GOODNESS-OF-FIT OF MODEL FORECASTING

Model	R^2	t-test
Low	0.76	$t_0 = 5.38 > t_{9,.005} = 3.25$
Neumann	0.65	$t_0 = 4.05 > t_{9,.005} = 3.25$
Khisty	0.76	$t_0 = 5.31 > t_{9,.005} = 3.25$

and

$$\% \text{ RMS Error} = \frac{\text{RMS Error}}{V_{\bar{o}}} \times 100$$

where V_o^{kl} is the observed volume on link $k - l$, V_e^{kl} is the estimated volume on link $k - l$, N is the number of links, and $V_{\bar{o}}$ is the arithmetic mean of the observed link volumes. These values gave an individual performance of the models. The results of these evaluations are shown in Table 5. The range between 9 and 15 percent appears to be most satisfactory.

For a critical examination of the performance of the models, the ratios of the 1980 observed to estimated link volumes (V_o/V_e) are tabulated for each of the models in Table 6. The range of V_o/V_e is 0.76 to 1.11 for the Low model, 0.82 to 1.17 for the Neumann technique, and 0.81 to 1.09 for the Khisty-AlZahrani model. Low's model generally estimated link volumes lower than observed, with half the links within 10 percent of actual volumes. Neumann's model fared about the same, while the Khisty-AlZahrani model produced most of the link volumes within 10 percent of the actual volumes. In Figures 3 through 5, the observed 1980 link volumes are plotted against the corresponding link volumes estimated by each of the models. While all three models perform well, the performance of the Khisty-AlZahrani is good from all aspects.

TABLE 5 EVALUATION OF THE MODEL ESTIMATES

Model	RMS Error	% RMS Error
Low	1693.85303	15.04043
Neumann	1542.41626	13.69575
Khisty	1063.20752	9.44066

TABLE 6 RATIOS OF OBSERVED TO ESTIMATED VOLUMES (V_o/V_e), BY LINK, FOR THE THREE MODELS

Link No.	Low	Neumann	Khisty
1	0.929	1.098	1.039
2	0.784	0.938	0.888
3	0.980	1.171	1.086
4	0.837	0.968	0.919
5	1.114	1.073	1.075
6	0.905	0.821	0.907
7	0.804	0.837	0.883
8	0.812	0.868	0.905
9	0.764	0.769	0.805
10	1.046	1.049	1.094
11	0.924	0.847	0.909

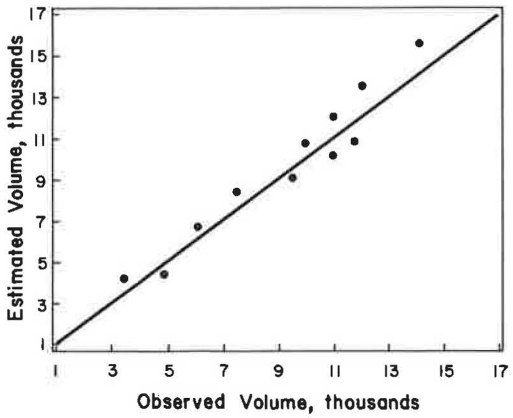


FIGURE 3 Performance of Khisty-AlZahrani model.

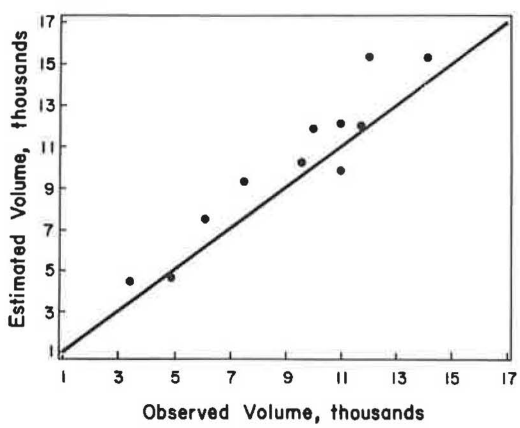


FIGURE 4 Performance of Low model.

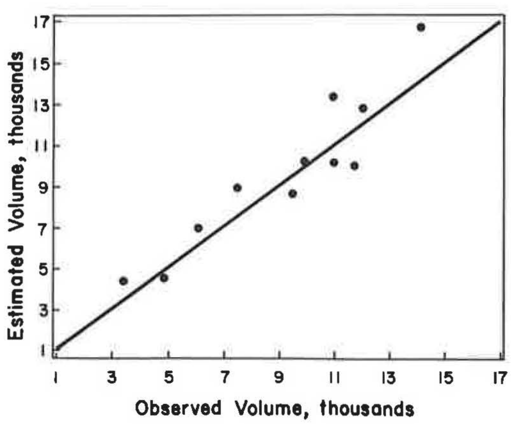


FIGURE 5 Performance of Neumann technique.

DISCUSSION OF RESULTS

The intention of this evaluation was not to proclaim a winner. Any of the three models can be easily applied in any small urban area setting with a population less than 50,000. The techniques can also be applied to larger urban areas.

These models would be particularly useful to transportation planners at the regional and local levels who are concerned with the planning and operation of transportation facilities and services in small cities or urban areas, especially those that are free-standing in rural regions but also those that are extensions of large metropolitan areas. The use of these models will result in considerable savings in time, money, and manpower, besides enabling decision makers to examine a variety of alternative plans reflecting broad policy. The use of these models can be highly recommended for modeling small urban areas, where transportation system management analysis is needed but qualified full-time transportation planners are not available on the planning staff. Caution, however, needs to be exercised regarding:

- the choice of socioeconomic variables (more experience needs to be gained from further application of the models in cities of varying sizes);
- the application of the models to cities having high percentage of mass-transit patronage;
- the fact that the outputs from these models are in terms of trips for all purposes (home-based, nonhome-based, and other trips could possibly be worked out with additional data);
- the exclusion of external-external, external-internal, and internal-external trips (count stations located on the cordon line would be most helpful in dealing with the forecast of such trips); and
- the use of this type of model for crucial policy options (the model is sensitive only to network changes).

CONCLUSION

Since the transportation characteristics, problems, and needs of small urban areas are quite different from large urban areas, special attention must be given in selecting techniques or models to forecast the demand for transportation facilities and services. The fact that these small urban areas generally lack the staff, expertise, and budget to operate the conventional models must be taken into account. Three inexpensive travel demand models discussed in this paper show sufficient promise in their performance to warrant adoption. All these models advantageously use traffic ground counts, an important input which is routinely collected by small urban areas and readily available to the traffic planner. The choice of adopting any one of these models should depend on the individual traffic planner, and on manpower and time availability.

REFERENCES

1. M. D. Meyer and E. J. Miller. *Urban Transportation Planning: A Decision-Oriented Approach.* McGraw-Hill, New York, 1984.
2. *Computer Programs for Urban Transportation Planning—PLANPAC/BACKPAC, General Information Manual.* FHWA, U.S. Department of Transportation, Washington, D.C., 1977.
3. *1980 Census of Population, Number of Inhabitants, United States Summary.* Report PC 80-1-A1, PP. 1–37. Bureau of the Census, U.S. Department of Commerce, Washington, D.C., 1983.
4. C. J. Khisty and M. Y. Rahi. *Inexpensive Travel Demand Techniques.* Report WA-RD-125.1. Washington State Department of Transportation, Olympia, Wash., 1987.
5. D. A. Low. A New Approach to Transportation Systems Modeling. *Traffic Quarterly,* Vol. 26, No. 3, 1972.
6. *Transportation Planning and Problem Solving for Rural Areas*

and Small Towns: Student Manual. FHWA, U.S. Department of Transportation, Washington, D.C., 1983.
7. P. Hogberg. Estimation of Parameters in Models for Traffic Prediction: A Nonlinear Regression Approach. *Transportation Research*, Vol. 10, No. 4, 1976.
8. R. Smith and W. McFarlane. Examination of Simplified Travel Demand Models. *Journal of Transportation Engineering*, ASCE, Vol. 104, No. 1, 1978.
9. A. Sosslau, A. Hassam, N. Carter, and G. Wickstrom. Travel Estimation Procedures for Quick Response to Urban Policy Issues. NCHRP Report 186, Transportation Research Board, 1978.
10. A. Sosslau, A. Hassam, N. Carter and G. Wickstrom. *NCHRP Report 187: Quick Response Urban Travel Estimation Techniques for Transferrable Parameters.* TRB, National Research Council, Washington, D.C., 1978.
11. E. Neumann, J. Halkias, and M. Elrazaz. Estimating Trip Rates from Traffic Counts. *Journal of Transportation Engineering*, ASCE, Vol. 109, No. 4, 1983.
12. C. J. Khisty and A. AlZahrani. Inexpensive Travel Demand Model for Small- and Medium-Sized Cities. In *Transportation Research Record 980*, TRB, National Research Council, Washington, D.C., 1984.

Publication of this paper sponsored by Committee on Transportation Planning Needs and Requirements of Small and Medium-Sized Communities.

Structuring Expert Systems and Other Computerized Models in Transportation

ARDESHIR FAGHRI

A thorough analysis is presented of how expert systems and existing simulation, evaluation, and optimization models in traffic and transportation engineering can be combined. The problems and shortcomings of the currently available computerized models are discussed, and potential applications of expert systems in curtailing these shortcomings are analyzed. Different taxonomies and configurations of how expert systems and other computerized models may be combined are shown, and the advantages and disadvantages of each configuration are stated. Finally, an application problem in transportation issues dealing with disaster evacuation and response planning is presented. It is concluded that expert systems and other computerized models work best when combined together to create intelligent computer programs.

For the past 20 years, the U.S. Department of Transportation through the Federal Highway Administration (FHWA) has sponsored a number of computerized models in traffic and transportation engineering (*1*). These powerful models, though useful, have had significant limitations:

- User cost—most models were originally designed for mainframe computers whose costs impaired the effectiveness of these programs.
- User access—turnaround times were often lengthy, sharply reducing the productivity of the professional. Furthermore, since traffic and transportation models are data intensive, their use often mandated onerous input-preparation efforts.
- Interpretation of results—model output takes the form of extensive statistical tabulations. Skill, insight and considerable knowledge on the part of the user are needed in order to extract those results which can provide a basis for decision-making.
- Models are evaluators—these models do not directly offer guidance for the decision maker. It is necessary for the skilled professional to execute these models repetitively, within the context of a well-designed experiment, in order to assess the relative benefits of candidate solutions and to select the best approach.
- Design models—models such as TRANSYT (*2*), MAXBAND (*3*), and PASSER (*4*) do provide results that can be used for improving traffic performance directly. They must be used by the traffic professional, are subject to similar user access difficulties, and are limited to components of a corridor system.

For these reasons, these evaluation and design tools have not fulfilled their potential. There have been some recent developments, however, which are ameliorating these shortcomings:

- These models have been "ported" to personal computers (PCs), thus providing improved access to the planning community. Moreover, PCs are becoming faster, less costly, and are now equipped with adequate memory.
- User-interactive input interfaces have been developed and combined with data base management systems (DBMSs). These greatly ease the input preparation activity, allowing technicians to replace engineers for this effort. Furthermore, the DBMS greatly reduces the extent of this effort.
- A user output interface is currently under development for the NETSIM (*5*) model, using interactive computer graphics (ICG) for PCs and workstations. This "Fifth Generation" technology, which requires no familiarity with computers on the part of the user, will greatly reduce the effort and cost associated with analyzing the model results.

While these recent advances represent significant improvements that enhance the appeal (and, hopefully, usage) of these models, they lack the ability to provide the decision maker with a direct response to his needs. Furthermore, there remains the need for highly skilled professional engineers to interpret the results and translate them into design or policy decisions. Finally, there is presently no organized and readily accessible inventory of current knowledge and experience which can be used to guide the decision-maker.

POTENTIAL ARTIFICIAL INTELLIGENCE APPLICATIONS

Advances in artificial intelligence (AI) offer the potential for solving or ameliorating these shortcomings and producing a new generation of traffic and transportation management tools. One of the major strengths of AI is that it can process knowledge, judgment, and opinion for application to such areas as game playing, theorem proving, general problem solving, robotics, natural language comprehension, and expert problem solving.

Some potential applications that employ artificial intelligence techniques might improve the limitations of current traffic and transportation management methodologies and evaluation models. These techniques can provide:

- Knowledge bases;
- Intelligence interfaces;

KLD Associates, Inc., 300 Broadway, Huntington Station, N.Y. 11746.

- Knowledge-based traffic and transportation management decision systems; and
- Real-time traffic management tools.

A "knowledge base" contains *facts*, or *rules* that use these facts as a basis for decision making. The core of all expert systems (ESs) is a knowledge base that contains information describing a specified "domain."

An "intelligence interface" can further relieve the user from tedious chores of data preparation, input-output, and model interfacing, and possibly help the user select appropriate analytical models for a given problem. For any particular traffic management strategy, the "intelligence interface" can (1) check the data requirement and data availability, (2) assist with data preparation if necessary, (3) do retrieval and preparation if data from the knowledge base is available, and (4) suggest analytical models appropriate to the problem.

The lack of determinacy (i.e., closed-form solutions) in many traffic management problems is beyond the capabilities of present algorithmic traffic management tools. Knowledge-based ESs offer a new approach for analyzing and solving nondeterministic problems for the highway corridor transportation manager. Such an ES would include an explanation module, knowledge acquisition module, context (also called workspace), knowledge base, and inference machine.

Traffic management experts can store their knowledge about any particular implementation of traffic management strategy to the *knowledge base* using the *knowledge acquisition module*. This information can be divided into two classes: (1) the factual or causal knowledge of the traffic management, and (2) the empirical associations, rules, or experiences. The knowledge base can also contain long-term historical traffic data, as well as network information. All information in the knowledge base is organized so that it may be effectively utilized by the other components of the system.

The *context* contains all the information that describes the problem currently being solved, including both problem data and solution status. The user can get access to the context through the explanation module. The traffic problem data may be divided into facts provided by the user and those derived or implied by the problem. The use of knowledge-based transportation management decision ES begins with the user entering some known facts about the problem (e.g., traffic corridor network configuration, traffic data, and other related information).

The *inference engine* is the knowledge processor. It operates on facts contained in the knowledge base and in the context, utilizing rules in the knowledge base to deduce new facts, which then can be used for subsequent inferences. The objective of the inference engine is to arrive at a global conclusion (goal), and the process continues until either the problem is solved and the context is transformed into the desired goal state, or when there are no more rules remaining to be invoked.

Many ES inference engines can deal with imprecise or incomplete knowledge. Associated with the data may be "certainty measures" indicating a level of confidence in the data. Rules are conditionally invoked, based on the certainty of the premise. The inference mechanism can then propagate certainty about the inferences along with results of the inferences.

The *explanation module* provides the traffic management decision ES with the capability to explain its reasoning and problem-solving strategy to the user. At any point the user may interrupt the system and inquire what it is doing and why it is pursuing the current line of reasoning. In addition, the system can explain how any fact was deduced and how knowledge was applied.

The *knowledge acquisition module* is provided to facilitate the knowledge input process. The information in the knowledge base is in rigid format, and the translation of knowledge obtained from experts to the required internal format may be tedious. Although it is desired that eventually the human expert be able to enter knowledge directly into the system, this goal is currently not achieved.

In addition, a *user interface* can be added. The user accesses the system through a friendly interface, often using a problem-oriented subset of English or computer graphics. The interface provides capabilities for the user to monitor the performance of the system, volunteer information, request explanations, and redirect the problem-solving approach used by the expert system.

With a knowledge-based traffic management decision system, even an inexperienced traffic manager can sit down in front of a terminal and state his goal for a certain traffic highway corridor. The manager will then enter some facts about the highway corridor under consideration, such as peak-hour traffic volume, highway arterial configurations, and other relevant information. Then the ES would recommend feasible transportation management strategies and the necessary procedures and additional resources required to implement the suggested management strategies. The solutions suggested by the traffic management decision system are deduced from knowledge obtained from a variety of experts, based on experiences from both successful and ill-fated previous similar projects.

AI techniques may also be useful in real-time traffic control. For example, current computerized urban signal control systems are based on the ideas of either minimizing vehicular delays or queues at signalized intersections or maximizing the progression band width of arterial signals. These ideas work well for an undersaturated network but not for congested networks. Currently, some researchers think that two simple principles—(1) keep intersections clear of spillback vehicles, and (2) give right-of-way to the direction in which traffic can move—might work better for congested networks. The ES type of program is ideal for either this simple system or a decidedly more complex rule-based system. In addition, traffic diversion and simulation capabilities can be added. The traffic manager can then have a real-time traffic management decision tool which will suggest traffic congestion or incident diversion schemes. This type of system can aid the traffic manager in making human decisions that require the exercise of certain models, only much faster. Such an ES was developed recently in France (*6*).

Finally, there are advantages to developing application programs in ES style. For a good conventional application program, it is a major undertaking to establish all the causal relationships, algorithms, or rules. The rules within a good conventional application program must be complete, unique, and correct. The program developer has the responsibility to insure that these three criteria are met. Due to the complexity of the transportation congestion problem, it is almost impossible to attain these three criteria. This makes conventional program development expensive. Additional costs are incurred when programs are updated, because major code modifications are generally required in order to accommodate new rules and locate the affected rules. For an ES program, it would be much easier to add more or change existing capabilities. New traffic control measures, such as incident diversion, demand/time-of-day tolls, on-board vehicle guidance system, and roadside two-way communication system, might be more easily incorporated into an existing ES program. The new traffic control measure would be first translated into appropriate rules, and then the new rules would be added to the existing rules. No algorithmic logic would need to be restructured and no existing code would need to be modified.

EXPERT SYSTEMS AND SIMULATION

Researchers and practitioners in the field of simulation and those in AI have had to face similar problems in creating models of complex and sometimes partially understood systems. To a large extent, solutions have been developed independently in each area, leading to techniques and software tools that differ markedly in terminology but often overlap in terms of concepts. The recent stress on knowledge representation in AI has emphasized a common ground, modeling of reality, but each group maintains a slightly different emphasis: dynamic behavior for simulationists, and logical inference for AI workers.

The purposes of simulation models and of ESs are similar: to provide a computer model that aids decision making. Their methods are similar in that they are both based on modular representations of physical systems and on "inference mechanisms" that drive these representations.

An inference mechanism, as opposed to user-programmed control structure, works with any number of modules, and these modules can (in theory at least) be presented in any order and are entirely independent of other modules. For instance, the inference mechanism of an event-based simulation is independent of the number of events and their order within the simulation. The inference mechanism, for simulation models, includes next-event-scheduling algorithms and interval-scanning procedures. For an ES, the inference mechanisms include backward and/or forward chaining with some form of uncertainty updating.

There are differences between these two disciplines:

- Simulation applications involve an iterative process wherein a model is designed, inputs are specified, an experiment is executed, the results are analyzed, a new run is designed, executed and analyzed, etc., until sufficient insight is gained to render a decision. In ES, on the other hand, the modeler constructs a knowledge base; the user defines the goal and lets the computer work to identify the decision rules.
- In simulation models, the data base is integrated with the program logic. For ES, the knowledge base is distinct from the inference engine that controls the logical flow.
- The data base for simulation models is generally numeric and formally structured. In an ES, the knowledge base is symbolic and represents facts, rules, judgment, and experience (i.e., heuristic knowledge) about a narrow problem area.
- Simulation employs algorithms; ES employs symbolic inference.
- Simulation languages are procedural or imperative (FORTRAN, GPSS, SIMSCRIPT). ES may use shells (OPS5, ROSIE, Expert-Ease) that, in turn, are usually written in functional or descriptive languages (LISP, PROLOG).

It has been well documented that knowledge acquisition is slow and costly due to the need for large amounts of knowledge. Simulation models, properly applied and integrated with information provided by experts, offer the potential for expanding the traffic engineering and transportation planning knowledge base *and* affecting important savings in cost and in time. The next section explores approaches that can both exploit the similarities between simulation and ES and accommodate their differences.

Structuring an Expert System Simulation Environment

Figure 1 shows a taxonomy for combining simulation and expert systems. Perhaps the most obvious way in which the two can be combined is by embedding an expert system within a simulation model as in (a), or vice versa as in (b). It is arguable that many simulation models already use knowledge, as opposed to data. For instance, a queue priority rule is knowledge. It may be pertinent to keep such rules in a knowledge base, rather than embedded in code. It may be necessary to embed a simulation within an ES for two reasons. First, the ES may need to run a simulation to obtain some results for the user. Second, and more important, the ES may use one or more time-dependent variables, and thus needs a simulation to update their values. This situation has occurred in some real-time military applications, where the system needs to know the position of ships, aircraft, etc.

Simulations and ESs that are designed, developed, and implemented as separate software, in parallel, may interact. A simulation model could interrogate an ES, as in (c). This may be useful where a simulation is developed for a complex system, and an ES already exists for part of the decision making within that system. The simulation can then access this, rather than mimic or encode the decision rules. ESs that execute and use the results from simulations, as in (d), are of increasing interest to knowledge engineers. Rather than test an ES on a user or a real environment, the ES can be tested on a simulation. Not only is development time reduced, but also testing can be more comprehensive. Further, if the simulation is a valid model, then an ES that adequately controls or responds to the simulation is, perhaps, valid itself. Application domains where this approach may be useful include real-time control, where an ES developed to ultimately control the process can be tested on a continuous simulation of that process.

Whereas in (a) through (d) the user of the main tool does not have direct access to the other, in many instances both an expert system and a simulation will be used together to do some task, as in (e). Each may well share some data; in effect, the simulation and ES will cooperate in the task. Given the

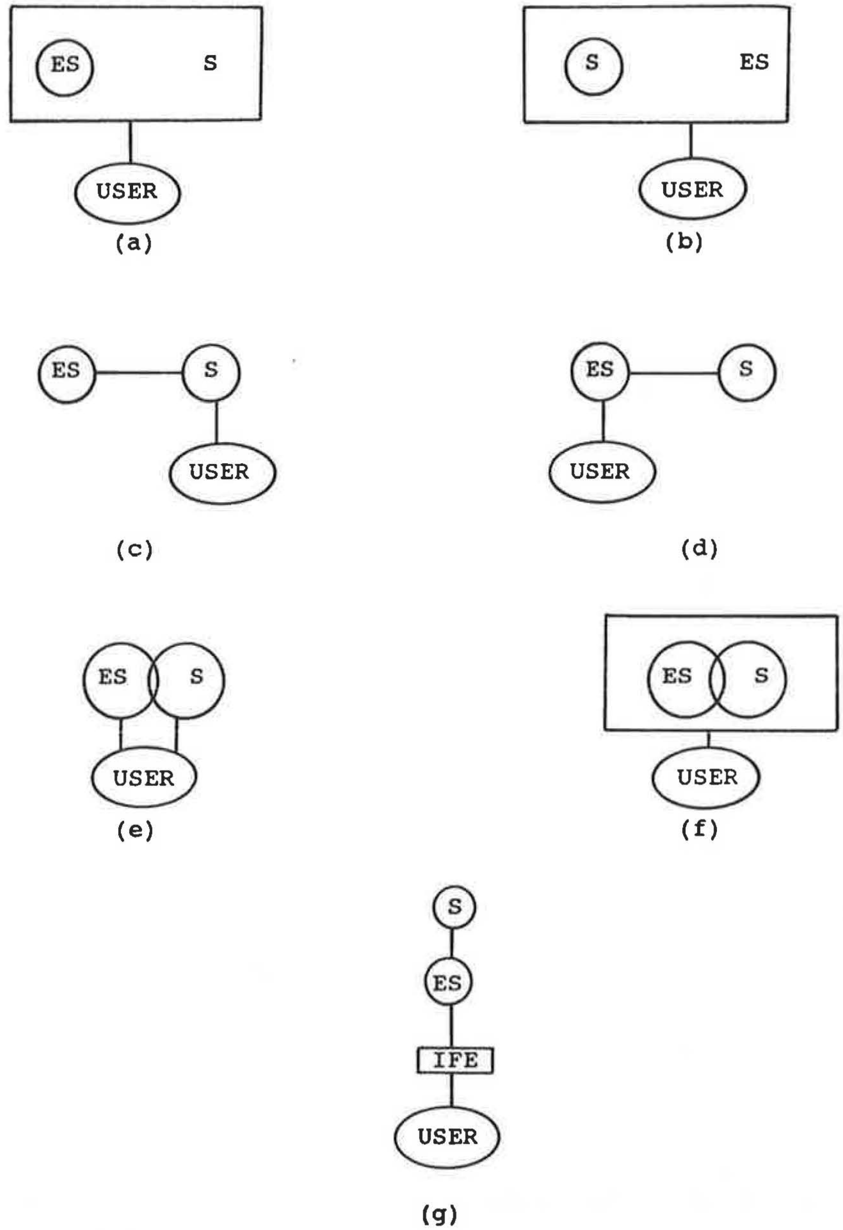


FIGURE 1 Taxonomy for combining expert systems (ES) and simulation (S).

increasing trend to hand over simulation models to users who may be inexperienced simulation users, there is a growing need to support the use, and guard against the misuse, of such models.

The cooperative simulation and ES may be surrounded by a larger piece of software, as in (f). Each may be part of a simulation environment or part of a large decision support system used directly by decision makers. New tools based on both simulation and knowledge-based methods fall into this category, for example, as does the Rule Oriented Simulation System (ROSS) developed by the Rand Corporation (*7*).

One of the most important application areas for knowledge-based methods is intelligent front ends (IFEs) or user interfaces (g). This generates the necessary instructions or code to use the package following a dialogue with the user and interprets and explains results from the package. Simulation programs, including TRAF (*8*), perform some of the functions of IFEs, although they do not actually execute the simulation and interpret the results. Useful intelligence for an ESSE would include:

- Dialogue handling (a natural language interface or at least user-directed free format input);
- Some model of the user, so that the system adjusts its requirements of the user, given evidence that the user is inexperienced, experienced, or whatever; and
- A model of the target package, so that some decisions can be taken by the IFE rather than referred to the user.

In considering how simulation and ESs can be combined, the above taxonomy is useful. For evaluating the pros and cons of different simulation/expert system configurations, several philosophies must be considered:

• The ES employs the simulation model to generate results that are added to the knowledge base, as required to respond to a user's needs, in the event this additional data is required. Here, the simulation model is executed as a precursor activity, prior to the interactive session between user and ES.

• The ES and simulation model are "cooperative," in that both may be executed during the session. Depending on the taxonomy selected, the user may communicate directly with the simulation model, *or* the ES will determine the need for the simulation model, prepare its input stream and execute it to generate the required "knowledge." In the latter case, the simulation model is transparent to the user who communicates only with the ES or IFE.

Note that the software system can be designed to accommodate both approaches, depending on the skills and objectives of the user. For example, macroscopic simulation models may be employed "on-line," while microscopic models used "off-line" to augment the knowledge base.

Integration of Expert Systems and Other Computerized Models

Expert systems may also be combined with other computerized evaluation or optimization models. An example of this type of integration in traffic engineering is the design of an expert system and the required interface to complement the existing intersection design models such as EVIPAS (*9*). Seven potential engineering design cases for the integration of expert systems and other computerized models are depicted in Figure 2 and discussed below.

In Case I, the engineering model controls the user interface. The graphics package (if separate from the model) and the ES shell are both run in a background mode and the user would not directly interact with them. There is one consistent interface. If the chosen model is very robust, easy to maintain, and has a built-in graphics package that is user friendly, then the expert system should run in background mode behind the model. However, the model must be able to handle graphics. The 80386 machines have a batch mode capability. The model must be multitasking, able to run itself (for hours as required), and still be available to the user for other tasks.

In Case II, the graphics package (if separate from the engineering model) controls the user interface. The model and the ES shell run in background mode, and the user would not directly interact with them. This case would be most appropriate if the desired model were very unfriendly to a casual user and had no acceptable built-in graphics capability. Again, there is one consistent interface. If the graphics package integrates well with model and ES, this may be the most appropriate user interface. However, the graphics package must be able to handle text and draw tables of output from the model.

In Case III, the ES shell controls the user interface. The graphics package (if separate from the model) and the engineering model run in background mode, and the user would not directly interact with them. There is one consistent interface, but experience has shown that it is often necessary to disable the ES tool graphics capability, and coding must be done in a graphics tool. Typically, expert systems have problems in the graphics (other than simple business charts) presentation area.

In Case IV, the graphics package and ES jointly participate as the user interface. Only the engineering model runs in background mode, and the user would not directly interact with it. This arrangement offers the user potentially better control over the graphics and ES. However, the user also has two separate interfaces to handle. Only if the ES can control the model and all text from the graphics package will an acceptable solution of a text and a graphics format be presented to the user.

In Case V, the ES shell and engineering model jointly control the user interface. Only the graphics package runs in background mode, and the user would not directly interact

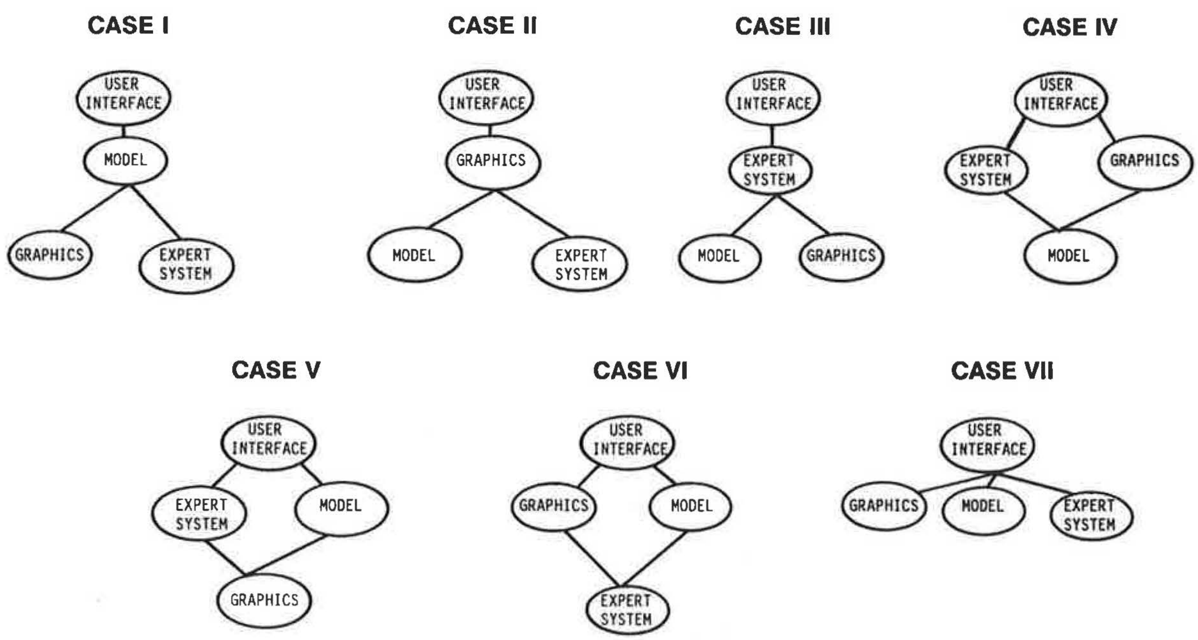


FIGURE 2 User interface cases.

with it. The user has potentially better control over the ES and model, but still has two separate text interfaces to handle. This option assumes that the model or the ES can completely control the graphics package. That assumption is valid only if the graphics package is built into the model.

In Case VI, the graphics package and engineering model jointly control the user interface. Only the ES shell runs in background mode, and the user would not directly interact with it. This offers the user potentially better control over the graphics and engineering model. The user still has two separate interfaces to handle. However, if the graphics package and model are built together then this reduces to Case I.

In Case VII, the graphics package, engineering model, and ES shell all directly interact as the user interface. None of the system components run in background mode. A very knowledgeable user will be able to most freely interact with the system, but the user must contend with three separate interfaces. For example, "Help" may be ^H, Fl, or H, depending on the environment. The user presentation would be very choppy, because the screens would all be different. Based upon the ability of the three separate packages to run together and the potential confusion to the practicing traffic engineer user, this may not be a viable option.

In the next section, an application problem—disaster evacuation and response planning—is described, the shortcomings of the existing simulation models in this area are discussed, and ways in which expert systems can be combined with existing models are analyzed.

APPLICATION

The state-of-the-art method and technique for dealing with highway network evacuation under disaster is to apply the available computer simulation and evaluation models (*10*). Some of the most popular models that are in use are: Network Emergency Evacuation (NETVAC) simulation model (*11*), EVAC PLAN PACK (*12*), and the Dynamic Network Evacuation (DYNEV) and its derivative IDYNEV computer models (*13*). These models, in general, have been designed to simulate the traffic flow and estimate the vehicular traffic evacuation time around a nuclear power plant in case of radioactive leaks similar to the Three Mile Island accident.

The state-of-the-art in evacuation modeling is IDYNEV (*14*). Table 1 summarizes the measures of effectiveness output by IDYNEV and also presents the input data required for simulation. The measures of effectiveness output data are provided for each network link and are also aggregated over the entire network.

TABLE 1 MEASURES OF EFFECTIVENESS OUTPUT BY IDYNEV (*14*)

Measure	Units
Travel	Vehicles miles and vehicle trips
Moving Time	Vehicle minutes
Delay Time	Vehicle minutes
Efficiency: moving time/ total travel time	Percent
Mean travel time per vehicle	Seconds
Mean delay per vehicle	Seconds
Mean delay per vehicle mile	Seconds/mile
Mean speed	Miles/hour
Mean occupancy	Vehicles
Mean saturation	Percent
Vehicle stops	Percent

The input data required for IDYNEV are summarized below:

- Topology of the roadway system.
- Geometries of each roadway component.
- Channelization of traffic on each roadway component.
- Motorist behavior that, in aggregate, determines the operational performance of vehicles in the system.
- Specification of the traffic control devices and their operational characteristics.
- Traffic volumes entering and leaving the roadway system.
- Traffic composition

Limitations of Current Procedures

Radwan et al. (*10*) have reviewed the aforementioned models and note several important features that the existing models lack. Among the deficiencies are:

- Since most of the models have been designed and tested for nuclear power plant accidents, they lack the capability of handling different disaster types, particularly natural ones.
- The shelter areas are defined as anything beyond the hazard area boundary lines, with no capacity constraint.
- The evacuation routes are treated solely on the traffic flow conditions, with no risk factors attached to them that can be influenced by the available evacuation time, the direction of disaster propagation, and the route topography. (IDYNEV is an exception; it takes into account time and route topography.)
- The models are not capable of testing different population density strategies.
- The models cannot be used for future land use management planning.

Another potential problem of applying the current models is the lack of communication between emergency management professionals and transportation professionals, which has resulted in an incomplete integration of resources. Transportation professionals usually lack the knowledge and experience of emergency management, and emergency management professionals often lack the expertise of running the simulation models and interpreting the output results.

Expert Systems Applications

The concept of a knowledge-based expert system (KBES) can be applied in conjunction with the current simulation models (in this case IDYNEV) to curtail some of the deficiencies described in the last section. A complete technical description of how a suitable ES can be built to interface with the IDYNEV model has been presented elsewhere (*15*). In brief, the different disastrous events are presented in a frame-based representation format in the ES, as shown in Figure 3, and the expertise of how to control traffic during each event is presented as a rule-based format. The decision-making process for dealing with the specific emergency event starts at the "appropriate actions to be taken" slot (shown in Figure 3). The slot contains appropriate heuristic knowledge in the form of rules in dealing with specific disasters.

This process is accomplished by (1) going through the frame; (2) having the frame automatically revise the local network (the If-Added procedures accomplish this); (3) feeding the revised network to the simulation model; and (4) combining the results of the model, the features of the network, and the features of the disaster, so that the expert system can make the final recommendation of what to do. The If-Needed procedure within the "appropriate action to be taken" slot trig-

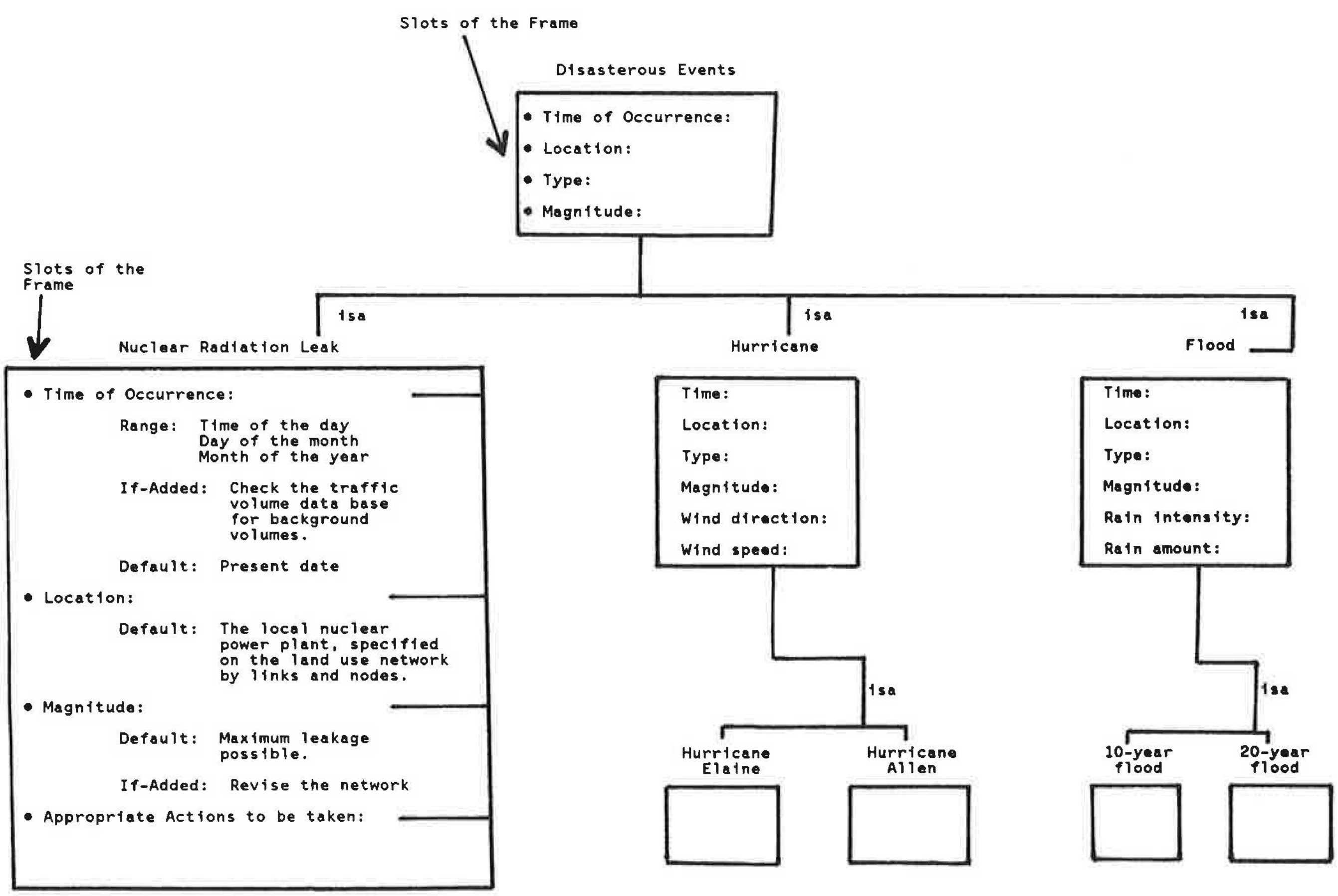


FIGURE 3 Frame-based representation of disastrous events for expert systems.

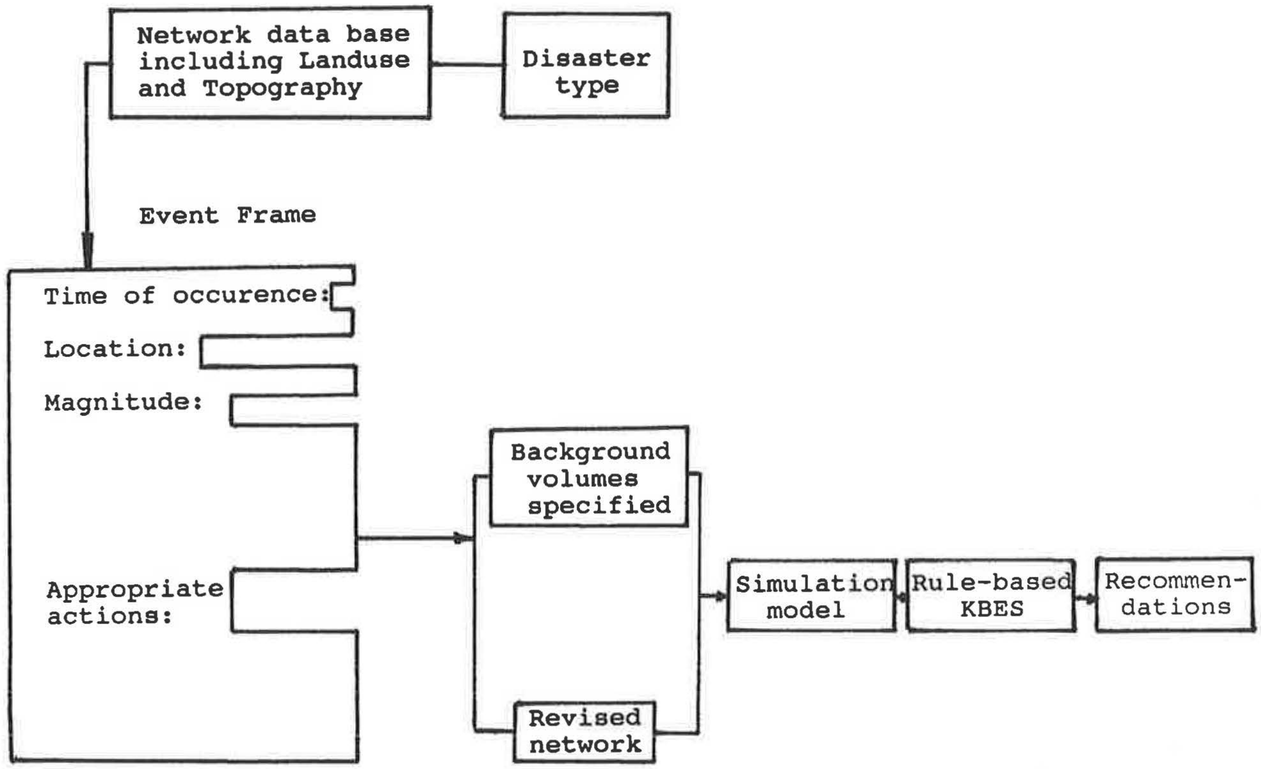


FIGURE 4 Knowledge-based expert system for disaster response planning.

gers this chain of events. Figure 4 shows the diagram for this process.

SUMMARY AND CONCLUSION

Because of the availability of many powerful traffic and transportation engineering evaluation, optimization, and simulation models, the applications of expert systems should be investigated in terms of how best they can be integrated with the available models. Different configurations for combining ES and other computerized models were presented, and the advantages and disadvantages of each configuration were discussed in this paper. The existing computer models consist of mostly algorithmic, sequential procedures with little or no symbolic or heuristic knowledge. Expert systems, on the other hand, are capable of handling symbolic knowledge excellently. Integration of available computer models and expert systems leads to the development of intelligent computer models for solving the problem.

REFERENCES

1. *Handbook of Computer Models for Traffic Operations Analysis.* FHWA, U.S. Department of Transportation, Washington, D.C., December 1982.
2. *TRANSYT-7F: Traffic Network Study Tool (Version 7F), User's Manual.* FHWA, U.S. Department of Transportation, Washington, D.C., June 1984.
3. *Maximal Bandwidth Traffic Signal Setting Optimization Program for Microcomputers (MAXBAND-PC) User's Manual.* University of Kansas Transportation Center, Lawrence, Kan., October 1985.
4. *PASSER II and PASSER III User's Manuals.* Texas Transportation Institute, Texas A & M University, College Station, Tex., September 1988.
5. *TRAF-NETSIM User's Manual.* Report FHWA-RD-89-150. FHWA, U.S. Department of Transportation, Washington, D.C., March 1989.
6. B. Forasté and G. Scenama. *An Expert System Approach to Handle Traffic Congestion.* Institute de Reserche des Transports, Paris, France, September 1985.
7. D. McArthur and P. Klahr. *The ROSS Language Manual.* Report N-1854-AF. Rand Corporation, Santa Monica, Calif., September 1985.
8. *TRAF User Guide, Implementation Package.* Report FHWA-RD-85. FHWA, U.S. Department of Transportation, Washington, D.C., July 1985.
9. A. G. R. Bullen and N. P. Hummon. EVIPAS: A Computer Model for the Optimal Design of a Vehicle Actuated Traffic Signal. In *Transportation Research Record 1114*, TRB, National Research Council, Washington, D.C., 1987.
10. A. E. Radwan, A. G. Hobeika, and D. Sivasailam. A Computer Simulation Model for Rural Network Evacuation Under Natural Disasters. *ITE Journal*, Vol. 55, No. 9, September 1985.
11. Y. Sheffi, H. Mahmassani, and W. Powell. *NETVAC: A Transportation Network Evacuation Model.* Center for Transportation Studies, Massachusetts Institute of Technology, Cambridge, Mass., 1980.
12. P. R. C. Vorhees. Evacuation Planning Package. Presented at the 61st Annual Meeting of the Transportation Research Board, Washington, D.C., 1982.
13. KLD Associates, Inc. The DYNEV Model: An Effective Tool for Developing and Evaluating Evacuation Plans. Presented at the 61st Annual Meeting of the Transportation Research Board, Washington, D.C., 1982.
14. T. Urbanik. Transportation Analysis for Evacuation: State of the Art. *ITE Journal*, Vol. 56, No. 3, March 1986.
15. A. Faghri. *Computerization of Heuristic-Based Decision Making Problems in Transportation Engineering.* Ph.D. dissertation. University of Virginia, Charlottesville, Va., May 1989.

Publication of this paper sponsored by Task Force on Expert Systems.

Expert System for Selection of Network-Based Transportation Planning Software Packages

K. NABIL A. SAFWAT AND KHALED EL-ARABY

The rapid proliferation of microcomputer software packages for network-based transportation planning, with different capabilities and limitations, makes it difficult to evaluate and select a package to satisfy the needs and constraints of a particular agency or transportation planner. The software selection process is complex because it involves a multi-objective decision-making process with ill-defined tradeoffs between the objectives as well as the capabilities and limitations of alternative packages. An expert system is described to assist practicing transportation planners and engineers in selecting microcomputer packages for network-based transportation planning to satisfy their agencies' needs and constraints. The Network-Based Transportation Planning Software Selection Advisor (NETSSA) is implemented in LISP on a VAX computer. NETSSA is highly interactive and user friendly. Its current knowledge base includes nine software packages; however, it can easily be expanded by its developers to include additional software packages and/or heuristics. The advice provided by NETSSA is supported with full explanation and reasoning. The user can accept it in whole or in part. NETSSA allows the user to change the relative weights placed on the different aspects and options, assign new weights, and declare specific requirements as being absolutely critical to the user. These flexibilities enable the user to use NETSSA interactively until he or she arrives at a recommendation that would optimally satisfy his or her agency's "realistic" needs and constraints.

Almost all transportation studies involve systematic analysis processes that include the forecasting of traffic flow patterns on several elements of transportation systems (*1*). Microcomputer technology is developing very rapidly and is becoming increasingly available to more transportation planners and engineers. Consequently, several microcomputer software packages have been developed recently to facilitate network-based traffic forecasting on transportation networks (*2*). Representative packages include TRANPLAN/NEDS, MicroTRIPS, EMME/2, TMODEL2, MINUTP, TRANSPRO, MOTORS, JHK SYSTEM II, CARS, etc.

Although these packages predict traffic flow patterns on transportation networks, their capabilities and limitations differ in several aspects such as maximum allowable network size (i.e., number of zones, nodes, and links), hardware requirements, trip generation capabilities, trip distribution options, modal split approaches, traffic assignment techniques, output reporting options and formats, interactive graphic facilities, prices, maintenance costs, technical support, ease of use, compatability with existing computational facilities, etc. Furthermore, each package is constantly being enhanced over time to increase its capabilities and reduce its limitations.

The selection of the software package that best suits the needs and constraints of a particular transportation agency is certainly a challenge that all transportation planning agencies face. The evaluation and selection process is a complex, time-consuming, and costly one. It involves multi-objective decision making with ill-defined tradeoffs between the objectives of that agency, as well as between the capabilities and limitations of alternative packages. Furthermore, existing packages are constantly being enhanced and therefore, by the time the decision is made, it may very well be already out of date. The Texas State Department of Highways and Public Transportation (SDHPT) is using the "Texas Travel Demand Forecasting Package," installed on the mainframe computer at the Texas SDHPT, and has recently completed a two-year study for the evaluation and selection of a microcomputer software package to be used in different urban areas within the state and to be linked with the Texas Package (*3*). The study involved a detailed comparison of several microcomputer network-based transportation planning packages. The Tri-County Regional Planning Commission, Lansing, Mich., conducted a similar thorough investigation for Michigan Department of Transportation (*4*). The study involved evaluation of 11 network-based microcomputer packages and even included actual site demonstration by several venders. The State of Florida addresses the problem of traffic forecasting in its urban areas through a standardized model structure that can perform a variety of forecasting functions—the Florida Standard Urban Transportation Model Structure (FSUTMS), developed over the last ten years by the Bureau of Multimodal Systems Planning, Division of Planning and Programming, Florida Department of Transportation (*5*). The microcomputer version of FSUTMS is based on one of the microcomputer packages indicated above. The first and second National Conferences on Transportation Planning Applications (1987, 1989), which were attended by over 300 participants, most of whom were transportation planners and practitioners from various agencies (federal, state, local, public, private, MPOs, universities, etc.), and the 1st, 2nd, and 3rd International Conferences on Microcomputers in Transportation (1985, 1987, 1989) were indeed rare occasions for professionals to share their experiences with applications of alternative techniques and software packages to their respective problems.

Artificial Intelligence (AI) is a branch of computer science concerned with simulating human intelligence in a computing

Department of Urban and Regional Planning, Texas A&M University, College Station, Tex. 77843-3137.

machine environment, including natural language understanding, vision, learning, robotics, and expert systems. Expert Systems (ES) is another branch that attempts to simulate or reproduce intelligent problem-solving behavior in a computer program. An ES may be defined as an interactive computer program that incorporates judgment, experience, rules of thumb, intuition, and other expertise to provide knowledgeable advice about a variety of tasks (*6*). An ES differs from a conventional computer in several aspects, including representation and use of knowledge, separation between knowledge base (knowledge that defines the specific problem to be solved) and the inference mechanism that uses the knowledge base to solve the problem at hand, the heuristic nature of the problem-solving process, and the orientation toward symbolic rather than numerical processing of information (*7*).

The appropriate problems for using US are those that require expert knowledge for their solution; the reverse is not true. That is, many problems solved by human experts may not be easily solved by ES techniques. Criteria for selecting an appropriate problem for an ES application have been developed (*8,9*) and include:

1. the problem should focus on a narrow specialty area and should not involve a lot of common-sense knowledge;
2. the problem should not be too easy or too difficult to solve by human experts;
3. algorithmic solutions are impractical because of complex physical, social, political, environmental, and/or judgmental aspects of the problem, which generally resist precise description and deterministic analysis;
4. the problem-solving process should be adapted to handle different types of problems and dynamic situations;
5. knowledge transfer from scarce human experts to other humans is too difficult or may take too long;
6. the problem-solving process involves extensive basic and background knowledge that only a few experts can possess; and
7. high performance results are required in a short time while a human expert is not available.

It should be emphasized, however, that ES should be used as a tool to advise human experts, but not to replace them.

Based on the above review of literature and problem definition, ES technology is believed to be most suited to address the problem of transportation planning software selection.

This paper describes an effort to develop an expert system to assist transportation engineers and planners in selecting the microcomputer network-based transportation planning software packages that would most effectively meet their specific needs and constraints. The bulk of this paper will describe the structure and operation of the Network-Based Transportation Planning Software Selection Advisor (NETSSA). The next section gives an overview of NETSSA, including a listing of the software packages to be considered for analysis, the set of evaluation criteria utilized by the expert system, the assignment of default and user-defined relative weights for different options and evaluation criteria, the operation of NETSSA, and its architecture. The following section demonstrates the flexibilities and capabilities of NETSSA through an example application. The final section includes summary and conclusions.

AN OVERVIEW OF NETSSA

Network-Based Transportation Planning Software Packages in NETSSA

The initial phase of the study included a detailed examination of recent studies where network-based transportation planning software packages were reviewed and evaluated. Two studies that tested and compared the software capabilities and features to a great extent were identified:

1. *A Comparison of Microcomputer Packages for Network-Based Highway Planning*, conducted by the Texas Transportation Institute (*3*);
2. *Michigan Microcomputer Traffic Forecasting Model Evaluation*, conducted by the Tri-County Regional Planning Commission (*10*).

The first study performed a detailed evaluation of four packages and the second included 11 packages. In this paper, NETSSA included the nine packages identified below by their names and developers. It should be clear, however, that other existing or newly developed packages can be added very easily, provided of course that their characteristics are known. In NETSSA, new information can only be added by the system developers in order to maintain accuracy of the knowledge base.

Package	*Developer*
TRANPLAN/NEDS	Urban Analysis Group
MicroTRIPS	PRC Voorhees/RVA/MVA Systematica
EMME/2	Centre de Recherche sur les Transports
MINUTP	COMSIS Corporation
MOTORS	M. M. Dillon Ltd.
TRANSPRO	Transware Systems
JHK SYSTEM II	JHK and Associates
CARS	Roger Creighton Associates
TMODEL2	Metro Transportation Group, Inc.

It should be clear that the authors do not endorse any particular software (or company) included in (or to be added to) NETSSA or recommended by the example application in this paper.

Software Evaluation Criteria in NETSSA

Each of the two studies mentioned above used its own set of evaluation criteria. Although there was a great deal of overlap between the criteria used in both studies, it was clear that the Michigan study involved more detailed classification. In this paper, the most significant evaluation criteria, particularly those which were clearly identified in both studies, were classified into ten basic categories; each category included several related characteristics and/or options. The ten categories were also classified into two major groups. The first four categories are basic modeling analysis options:

1. Trip generation (trip generation equations, category rates, and trip generation rate estimation);

2. Trip distribution (fratar, exponential gravity, friction factor gravity, K-factors, and gravity calibration);
3. Modal split (binary logit, diversion curves, and full network transit modeling); and
4. Traffic assignment (all-or-nothing, dial multipath, incremental multipath capacity-restrained, iterative capacity-restrained, iterative multipath capacity-restrained, equilibrium assignment, and turn prohibitors).

The other six categories are supporting analysis options:

5. Network size (maximum allowable number of zones, nodes, and links);
6. Package limits (maximum allowable number of trip purposes and trip generation variables);
7. Package prices;
8. Hardware requirements (IBM-PC or compatible, floppy or hard disk drives, screen displays, dot matrix printer, plotters, and high-resolution color monitor);
9. Report options (unloaded network, selected paths, link loads, turning movements, assignment convergence, VHT/VMT summaries, and ground count comparisons); and
10. Graphics capabilities (highway networks and loads, interactive network editing, highway paths, transit networks and loads, zonal data and matrices, and node volumes and delays).

Assignment of Weights to Software Evaluation Criteria in NETSSA

In order to reflect the relative importance of alternative options within different categories of the evaluation criteria, NETSSA includes default weights assigned to each of the four basic modeling analysis options identified above. This is particularly useful in reflecting the tradeoffs involved in the selection process. For example, in the traffic assignment category, the iterative capacity-restrained assignment option was given a weight of 0.6, while the equilibrium assignment option was given a weight of 0.8. A package that has equilibrium assignment capability but does not have capacity-restrained is therefore evaluated higher compared to another one that does not have equilibrium assignment but has capacity-restrained. In situations where none of the packages satisfies all user-defined modeling needs, NETSSA will recommend as its first selection the package that has the highest overall weight in satisfying as many user-defined modeling options as possible, and the one with the next higher weight as the second selection.

NETSSA will also identify the options that the recommended packages do not satisfy and the requirements for their operation. This is also true in case there are two or more packages which satisfy all user-defined modeling needs. If the user does not agree with NETSSA's default weights or recommendations, he or she can specify his or her own weights for alternative modeling options. In fact, the user can also add his or her own weights to any other supporting analysis options. Furthermore, if there is an absolutely critical criterion that must be satisfied, the user can assign a very high weight to that particular option or constraint. The user-assigned weight will override NETSSA's default weight, and the recommendations will be based on the user's own weight. Of course, multiple critical criteria may be specified simultaneously.

Operation of NETSSA

NETSSA is basically an advisor that mimics a search by a knowledgeable transportation engineer or planner. It incorporates judgment, experience, and other expertise to provide advice about the selection of the software that meets the user needs and constraints. An overview of the five main tasks performed by NETSSA is shown in Figure 1. These steps are:

1. Identification of satisfied and unsatisfied evaluation criteria. NETSSA queries the user to input the software packages to consider in the selection process, his chosen modeling analysis needs (i.e., trip generation, trip distribution, modal split and traffic assignment options), and other supporting analysis needs and constraints (i.e., network size, trip purposes and trip generation variable limits, hardware requirements, budget and graphics options). Then through the incorporation of both the heuristic knowledge present in the system and the user input, the evaluation options that are satisfied, not satisfied and/or may not be satisfied due to insufficient information are identified for each chosen package and shown accordingly to the user.
2. Evaluation of each package. Given the current standing of each software package with respect to the different user-

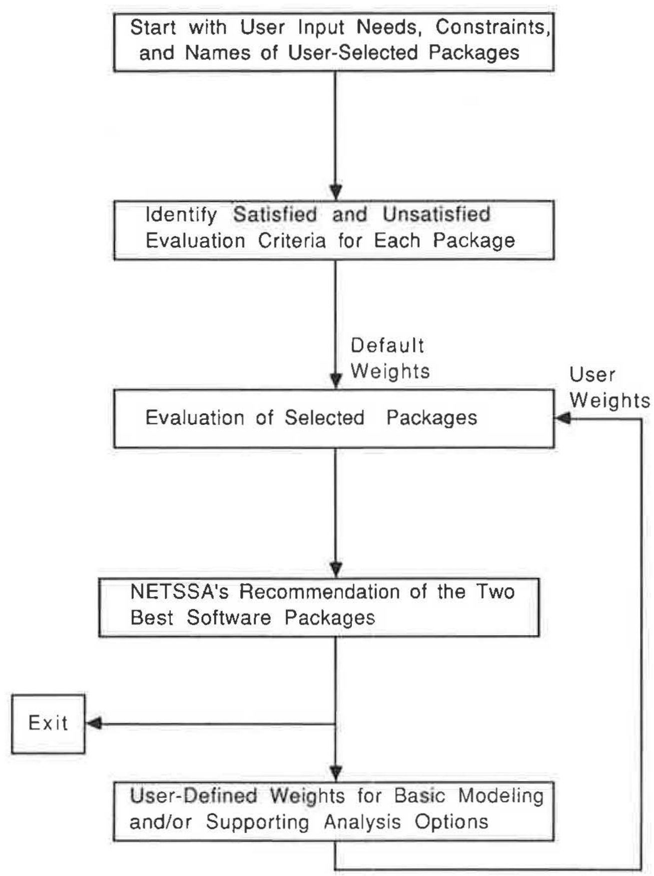


FIGURE 1 Operation of NETSSA.

defined needs and constraints, and given the default weights assigned by the system developers to each modeling analysis option, the overall weight of each package is evaluated.

3. Recommendation of software packages. The software package with the highest "relative merit" or highest overall weight is then recommended to the user. NETSSA identifies the options that the recommended package does not satisfy and the requirements (package limits, minimum memory, etc.) for operating the package. NETSSA also recommends a second-best software package together with its merits and requirements.

4. Modification of weights. NETSSA then allows the user to modify its default weights of the modeling analysis options provided by the system developers. At this point, the user can also assign new weights to other supporting analysis options (e.g., hardware requirements, budget limit, and maximum network size). These flexibilities enable the user to reflect his or her own judgment about the relative importance of all options. The user can even identify some options as absolutely critical requirements that *must* be satisfied in the selected software. The user, however, has to be cautious when specifying his own weights. Assignments of weights to supporting analysis options can jeopardize the effect of the basic modeling analysis options on the selection process.

5. Recommendation based on user-defined weights. After NETSSA receives the user-specified weights, the system searches for the software package that meets all the user critical requirements, if any. If no package satisfies the critical requirements, the user is notified and has the opportunity of either making changes to the critical options and weights or to exit the system. If the user selects to continue, NETSSA repeats Steps 2 and 3 with new user weights from step 4 and recommends the software package with the highest relative merit together with the second-best choice. The system also gives the merits and operating requirements of each recommended package.

Interactive analysis of options and their respective weights may be repeated as many times as the user wishes until he or she reaches an "optimum" solution for his or her "realistic" needs and constraints. An example application described later in this paper will provide more insight into the operational flexibilities and interactive capabilities of NETSSA.

Architecture of NETSSA

One fundamental characteristic of expert systems is that the knowledge used to solve the problem is expressed primarily in symbolic terms. NETSSA is written in LISP language and is implemented on a VAX mainframe computer. LISP was chosen because of its power for representing, storing, and retrieving symbol structures and its flexibility with respect to the problem-solving strategies. Figure 2 shows the architecture of NETSSA. Below is a description of each of its components. A more detailed and general description of components of expert systems may be found elsewhere (*6–9,11*).

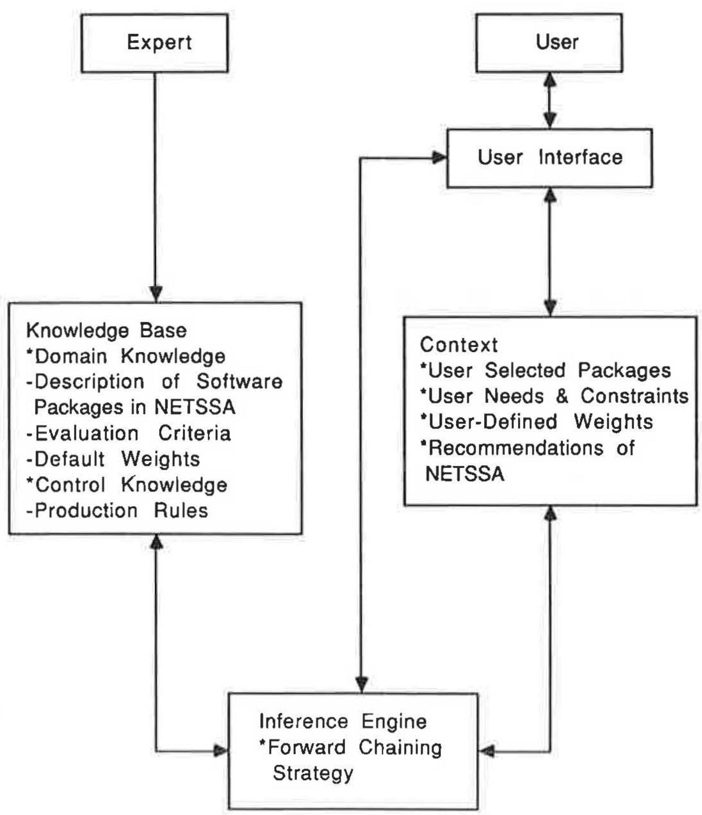


FIGURE 2 Architecture of NETSSA.

Knowledge Base

The knowledge base is the powerhouse of the expert system. It contains all relevant information, facts, causal knowledge of the domain, and heuristics used for problem-solving activities. In NETSSA's knowledge base, the domain and the control knowledge of the problem are included.

The first type of knowledge (i.e., domain knowledge) consists of the features, capabilities, package limits and analysis features for each of the nine software packages. Table 1 shows an example of the domain knowledge on traffic assignment features offered by several packages. Other types of data, such as default weights assigned to basic modeling analysis options, are also present in the domain knowledge. The domain knowledge is represented as object-attribute-value triplets or OAV. In this scheme, object may be a particular software package. Attributes are general characteristics or features of objects (e.g., network size limits, trip distribution features). The final member of the triplet is the value or nature of that attribute (e.g., trip distribution options may include friction factor gravity distribution or fratar distribution). Figure 3 shows an example of an OAV representation.

The second type of knowledge (i.e., the control knowledge) includes all the rules and procedures that determine how the domain data are related and the basis for assigning weights and selecting the softwares. In NETSSA, this knowledge is represented in the form of production IF-THEN rules, in which the satisfaction of one or more premises leads to one or more actions or consequences.

Context

The context is the component of the expert system that contains all data, symbols, true facts, or rules that reflect the current status of the problem. The context in NETSSA will initially contain information about the packages specified by the user to be evaluated and the user's set of needs and constraints. The context would expand as the problem-solving process expands to include, for example, information about the options that are satisfied or not satisfied by each package. It can also include information about the overall weight of each package and the user defined weights.

TABLE 1 TRAFFIC ASSIGNMENT FEATURES OF SOME PACKAGES (*3,10*)

Packages	All-or-Nothing	Dial Multipath	Incremental Multipath	Iterative	Iterative Multipath	Equilibrium
TRANPLAN/NEDS	x	x	No	x	No	x
MicroTRIPS	x	x	x	x	x	x
EMME/2	x	No	No	x	No	x
MINUTP	x	x	x	x	x	x
JHK-System/II	x	x	x	x	x	x
CARS	x	No	No	x	x	No

* x=Feature is available
No=Feature is not available

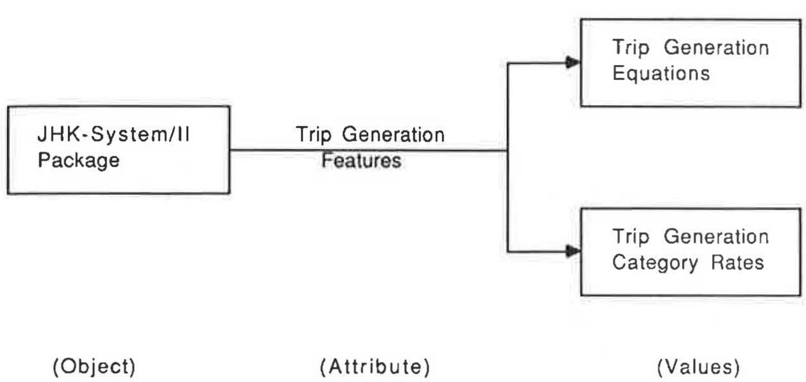


FIGURE 3 Example of object-attribute-value (OAV) representation.

Inference Engine

The inference engine is responsible for the execution of the expert system through manipulation of the knowledge base and context. The inference engine locates and executes an "active" rule (one in which the premise is satisfied) and executes the required action. This is done by detecting changes in the context, comparing the context with the knowledge base, and deciding which action would be most appropriate. The process of selecting active rules is repeated until no rule can be satisfied or until a prediction indicates the end of the session.

In NETSSA, the forward chaining processing strategy is utilized. According to this strategy, the expert system works from an initial state of known facts (user input and domain knowledge) and searches for the best conclusion (i.e., most appropriate software that fits these facts). In Steps 3, 4, and 5 described above, the expert system begins with the recommended package based on the system default weights. It determines whether it is appropriate for the user-defined weights and critical requirements. If it does not support the user requirements, the expert system pursues the validity of other packages in the knowledge base.

User Interface

The user interface allows the user to interact directly and efficiently with the system. It prompts the user to input his facts and data, and provides him with explanation and reasoning for the decision reached. In NETSSA, user inputs include the available packages, evaluation constraints, user weights, etc. NETSSA's outputs include (1) identification of packages that satisfy all user input needs and constraints, if any; (2) the 1st and 2nd recommended software packages; (3) the features satisfied, unsatisfied, and possibly satisfied by each package; (4) the default or user-defined weights based on which the decision was made; and (5) the requirements for operating the recommended packages.

Example Application

In this section, an example application of NETSSA is demonstrated. NETSSA begins by welcoming the user and listing the available software packages in the system. It then prompts the user to select the packages that he would like to include in the selection process (see Figure 4). In this example, the user chooses five software packages to include in the selection process. The chosen packages are: TRANPLAN/NEDS, MicroTRIPS, EMME/2, TRANSPRO, and MINUTP.

The user then specifies his basic modeling and supporting analysis options that fit his specific needs and constraints. In this example, the user identifies the following options:

1. Trip generation equations, trip generation rate estimation;
2. Friction factor gravity distribution, K-factors, gravity self calibration;
3. Binary logit modal split, full network transit modeling;
4. Equilibrium assignment, and turn prohibitors.
5. Max. no. of zones < 1500, max. no. of nodes $< 16{,}000$, max. no. of links $< 16{,}000$;
6. Max. no. of trip purposes < 15, max. no. of trip generation variables < 26;
7. Budget limit = $3000;
8. IBM-PC or compatible, hard disk drive, screen displays, dot matrix printers;
9. Report VHT/VMT summaries, report turning movements, report links loads and selected paths, report unloaded network; and
10. Interactive network editing.

Once the user inputs his options as defined above, NETSSA evaluates each of the five packages and displays for each package the options which are satisfied, not satisfied and/or may or may not be satisfied for insufficient information. In this example, the information displayed for TRANPLAN/NEDS is shown in Figure 5. Again, similar information is displayed for each of the five packages.

Using the default weights in NETSSA, it recommends two software packages. For each of the recommended packages, NETSSA displays the user-defined options not satisfied by the package, if any, together with the requirements for operating the package. In this example, the NETSSA made the recommendation shown in Figure 6.

At this point the user may elect to exit NETSSA or to continue by specifying his own set of weights for basic modeling options, as well as new weights for other supporting

```
WELCOME TO THE NETWORK-BASED TRANSPORTATION PLANNING SOFTWARE SELECTION ADVISOR

                              * NETSSA *

THIS PROGRAM WILL ASSIST YOU IN CHOOSING THE SOFTWARE PACKAGE THAT
WILL BEST SUIT YOUR NEEDS FOR YOUR TRANSPORTATION NETWORK ANALYSIS PROBLEM
==========================================================================

THE SOFTWARE PACKAGES AVAILABLE ARE :

TRANPLAN/NEDS        MICROTRIPS           EMME/2
MINUTP               MOTORS               TRANSPRO
CARS                 JHK-SYSTEM/II        TMODEL2
```

FIGURE 4 NETSSA's welcoming message.

TRANPLAN/NEDS SATISFIES THE FOLLOWING OPTIONS

(MAX. NO. OF ZONES 1500) (MAX. NO. OF TRIP PURPOSES 15)
(BUDGET LIMIT $ 3000) (DOT MATRIX PRINTERS) (SCREEN DISPLAYS)
(HARD DISK) (IBM-PC OR COMPATIBLE)
(TRIP GENERATION EQUATIONS)
(GRAVITY SELF CALIBRATION) (K-FACTORS)
(FRICTION FACTOR GRAVITY DISTRIBUTION)
(FULL NETWORK TRANSIT MODELING) (TURN PROHIBITORS)
(EQUILIBRIUM ASSIGNMENT) (REPORT VHT/VMT SUMMARIES)
(REPORT TURNING MOVEMENTS) (REPORT LINK LOADS)
(REPORT SELECTED PATHS) (REPORT UNLOADED NETWORK)
(INTERACTIVE NETWORK EDITING)

TRANPLAN/NEDS DOES NOT SATISFY THE FOLLOWING OPTIONS

(MAX. NO. OF NODES 16000) (MAX. NO. OF LINKS 16000)
(MAX. NO. OF TRIP GENERATION VARIABLES 26)
(TRIP GENERATION RATE ESTIMATION) (BINARY LOGIT MODAL SPLIT)

FIGURE 5 Information display for TRANPLAN/NEDS software package.

==

AS A FINAL OUTCOME AND BASED ON OUR OWN DEFAULT WEIGHTS FOR THE BASIC MODELING ANALYSIS OPTIONS YOU HAVE SELECTED , IT WOULD BE RECOMMENDED TO USE :

* 1st. SELECTION :
MINUTP SOFTWARE PACKAGE

* 2nd. SELECTION:
EMME/2 SOFTWARE PACKAGE

MINUTP DOES NOT SATISFY THE FOLLOWING USER SELECTED OPTIONS :

(MAX. NO. OF TRIP PURPOSES 15) (GRAVITY SELF CALIBRATION)

MINUTP REQUIRES THE FOLLOWING OPTIONS :

(BUDGET NOT LESS THAN $ 3000)
(MINIMUM MEMORY 256 K RAM)
(FLOPPY DISK OR HARD DISK DRIVES) (MAX. NO. OF ZONES 2000)
(MAX. NO. OF NODES 16000) (MAX. NO. OF LINKS 32300)
(MAX. NO. OF TRIP PURPOSES 9)
(MAX. NO. OF TRIP GENERATION VARIABLES 26)

YOUR SECOND CHOICE WOULD BE TO USE :

EMME/2 SOFTWARE PACKAGE

EMME/2 DOES NOT SATISFY THE FOLLOWING OPTIONS :

(MAX. NO. OF ZONES 1500) (MAX. NO. OF NODES 16000) (BUDGET LIMIT $ 3000)
(TRIP GENERATION RATE ESTIMATION)

EMME/2 REQUIRES THE FOLLOWING OPTIONS :

(BUDGET NOT LESS THAN $ 7500)
(MINIMUM MEMORY 640 K RAM) (HARD DISK DRIVE)
(MAX. NO. OF ZONES 800) (MAX. NO. OF NODES 5000) (MAX. NO. OF LINKS 16000)
(MAX. NO. OF TRIP PURPOSES 40)
(MAX. NO. OF TRIP GENERATION VARIABLES 100)

==

FIGURE 6 NETSSA's recommendation based on default weights.

options. In this example, the user elected to continue. The weights specified by the user are:

Option	*Weight*
Gravity Self Calibration	0.4
Exponential Gravity Distribution	0.6
Trip Generation Rate Estimation	0.3

Based on these user-defined weights and the default weights not altered by the user, NETSSA's recommendation changed as shown in Figure 7.

As indicated earlier, the user can assign new weights to other supporting analysis options. In this example, the user elected to exercise this option. The user-specified weights are:

Option	*Weight*
Budget Limit	1.0
Max. No. of Trip Purposes 15	0.8
Max. No. of Links 16,000	0.7
Max. No. of Zones 1500	0.9

Based on these user-defined weights and system default weights for the basic modeling analysis options, NETSSA's recommendation again changed as shown in Figure 8.

Finally, the user can specify some critical requirements. In this example the user elected to specify the following critical requirements:

- Budget Limit $3000,
- Max. No. of Zones 1500, and
- Max. No. of Trip Purposes 15.

Again NETSSA offers recommendations and information on the package(s) that satisfies the user critical requirements, as shown in Figure 9. The process could continue as many times as the user wishes until he or she arrives at a recommendation that would "optimally" satisfy his or her needs and constraints.

```
==========================================================================
AS A FINAL OUTCOME AND BASED ON YOUR OWN SET OF WEIGHTS, IT
WOULD BE RECOMMENDED TO USE :

* 1st. SELECTION :
EMME/2  SOFTWARE PACKAGE

* 2nd. SELECTION :
MICROTRIPS  SOFTWARE PACKAGE


 * FIRST SELECTION *

EMME/2  DOES NOT SATISFY THE FOLLOWING OPTIONS :

(MAX. NO. OF ZONES 1500)  (MAX. NO. OF NODES 16000) (BUDGET LIMIT $ 3000)
(TRIP GENERATION RATE ESTIMATION)


EMME/2  REQUIRES THE FOLLOWING OPTIONS :

(BUDGET NOT LESS THAN $ 7500)
(MINIMUM MEMORY 640 K RAM)                          (HARD DISK DRIVE)
(MAX. NO. OF ZONES 800)   (MAX. NO. OF NODES 5000)  (MAX. NO. OF LINKS 16000)
(MAX. NO. OF TRIP PURPOSES 40)
(MAX. NO. OF TRIP GENERATION VARIABLES 100)


 * SECOND SELECTION *

MICROTRIPS  SOFTWARE PACKAGE


MICROTRIPS  DOES NOT SATISFY THE FOLLOWING OPTIONS :

(MAX. NO. OF LINKS 16000) (MAX. NO. OF TRIP PURPOSES 15)
(MAX. NO. OF TRIP GENERATION VARIABLES 26)          (BUDGET LIMIT $ 3000)
(TRIP GENERATION RATE ESTIMATION)


MICROTRIPS  REQUIRES THE FOLLOWING OPTIONS :

(BUDGET NOT LESS THAN $ 4250)
(MINIMUM MEMORY 256 K RAM)
(FLOPPY DISK OR HARD DISK DRIVES)                   (MAX. NO. OF ZONES 2000)
(MAX. NO. OF NODES 40000) (MAX. NO. OF LINKS 14000)
(MAX. NO. OF TRIP PURPOSES 3)
(MAX. NO. OF TRIP GENERATION VARIABLES 18)
==========================================================================
```

FIGURE 7 NETSSA's recommendation based on user-defined weights for basic modeling options.

```
===============================================================================

AS A FINAL OUTCOME AND BASED ON YOUR OWN SET OF WEIGHTS, IT
WOULD BE RECOMMENDED TO USE :

* 1st. SELECTION :
MINUTP  SOFTWARE PACKAGE

* 2nd. SELECTION :
TRANPLAN/NEDS  SOFTWARE PACKAGE

 * FIRST SELECTION *

MINUTP  DOES NOT SATISFY THE FOLLOWING OPTIONS :

(MAX. NO. OF TRIP PURPOSES 15)                     (GRAVITY SELF CALIBRATION)


MINUTP  REQUIRES THE FOLLOWING OPTIONS :

(BUDGET NOT LESS THAN $ 3000)
(MINIMUM MEMORY 256 K RAM)
(FLOPPY DISK OR HARD DISK DRIVES)                  (MAX. NO. OF ZONES 2000)
(MAX. NO. OF NODES 16000) (MAX. NO. OF LINKS 32300)
(MAX. NO. OF TRIP PURPOSES 9)
(MAX. NO. OF TRIP GENERATION VARIABLES 26)


 * SECOND SELECTION *

TRANPLAN/NEDS  SOFTWARE PACKAGE


TRANPLAN/NEDS  DOES NOT SATISFY THE FOLLOWING OPTIONS :

(MAX. NO. OF NODES 16000) (MAX. NO. OF LINKS 16000)
(MAX. NO. OF TRIP GENERATION VARIABLES 26)
(TRIP GENERATION RATE ESTIMATION)                  (BINARY LOGIT MODAL SPLIT)


TRANPLAN/NEDS  REQUIRES THE FOLLOWING OPTIONS :

(BUDGET NOT LESS THAN $ 2500)
(RECOMMENDED MEMORY 640 K RAM)                     (HARD DISK DRIVE)
(MAX. NO. OF ZONES 1500)  (MAX. NO. OF NODES 10000) (MAX. NO. OF LINKS 12000)
(MAX. NO. OF TRIP PURPOSES 15)
(MAX. NO. OF TRIP GENERATION VARIABLES 25)

===============================================================================
```

FIGURE 8 NETSSA's recommendation based on user-defined weights for supporting analysis options.

```
===============================================================================

AS A FINAL OUTCOME AND BASED ON YOUR OWN SET OF WEIGHTS: TRANPLAN/NEDS SOFTWARE
PACKAGE IS THE PACKAGE THAT SATISFIES ALL YOUR -CRITICAL-
REQUIREMENTS

TRANPLAN/NEDS  REQUIRES THE FOLLOWING OPTIONS :

(BUDGET NOT LESS THAN $ 2500)
(RECOMMENDED MEMORY 640 K RAM)                     (HARD DISK DRIVE)
(MAX. NO. OF ZONES 1500)  (MAX. NO. OF NODES 10000) (MAX. NO. OF LINKS 12000)
(MAX. NO. OF TRIP PURPOSES 15)
(MAX. NO. OF TRIP GENERATION VARIABLES 25)


TRANPLAN/NEDS  DOES NOT SATISFY THE FOLLOWING OPTIONS :

(MAX. NO. OF NODES 16000) (MAX. NO. OF LINKS 16000)
(MAX. NO. OF TRIP GENERATION VARIABLES 26)
(TRIP GENERATION RATE ESTIMATION)                  (BINARY LOGIT MODAL SPLIT)

===============================================================================
```

FIGURE 9 NETSSA's recommendation based on user-defined critical requirements.

CONCLUSIONS

Several microcomputer software packages have recently been developed to facilitate network-based traffic forecasting on transportation networks. The selection of the software package that best suits the needs and constraints of a particular transportation agency is a complex, time-consuming and costly challenge. Expert system technology was thought to be most appropriate to address this issue of transportation planning software selection.

The Network-based Transportation Planning Software Selection Advisor (NETSSA) developed in this paper was implemented in LISP language on a VAX mainframe computer. NETSSA's current knowledge base includes the nine software packages. Evaluation criteria were identified, and NETSSA's developers further classified these criteria into basic modeling options (for which default weights were assigned) and other supporting analysis options. The default weights in NETSSA were based on the expert opinion of a few transportation systems analysts including the developers. The flexibility of NETSSA was demonstrated by allowing the users to alter these default weights and to assign new weights to any of the supporting analysis options. The user can even declare some options to be absolutely critical requirements that must be satisfied by the selected software.

An example application of NETSSA was described to illustrate the basic features and capabilities of NETSSA. The example application showed that there is no one package that would satisfy all the "desired" user needs and constraints, and the NETSSA's recommendations are quite sensitive to the user-defined needs, constraints, and weights. Therefore, the user would have to be very careful in evaluating the tradeoffs between the capabilities, limitations, and requirements of alternative packages, on the one hand, and his agency's objectives, needs, and constraints, on the other hand. NETSSA can be very easily expanded to include additional software packages or updates of existing ones in its domain knowledge base.

Though NETSSA includes several important evaluation criteria, some may need additional refinements. For example, the modal split option of "full transit modeling" may require further detailed definition and classification. In addition, a particular user may want to consider additional criteria that may not currently be included in NETSSA. Further research and user comments should help identify additional criteria to be added to NETSSA. Furthermore, some evaluation criteria would be difficult to include and measure, such as "ease of use." That is, two packages may be capable of performing the same function (i.e., both satisfy a particular criterion), but one may be a lot easier compared to the other. NETSSA includes some "proxy" measures for "ease of use," but more enhancement is needed to explicitly consider this important criterion. It should also be very useful to test NETSSA against human experts involved in an actual process for software selection.

Future research for the development of expert systems for similar and related problems may prove to be fruitful. For instance, a similar software selection advisor is needed for site impact and corridor analyses. A more general expert system for the selection of transportation planning techniques to address a broader range of transportation problems would be a more challenging system to develop. In fact such an expert system may include NETSSA and similar systems as components.

ACKNOWLEDGMENTS

The authors would like to express their deep appreciation to Paul Hamilton, Tri-County Regional Planning Commission, Lansing, Michigan, and to George Dresser, Vergil Stover, and Duk Chang, Texas Transportation Institute, for their cooperation in providing the results of their invaluable studies. The comments of Walter Daugherity, Knowledge Systems Research Center, Texas A&M University, and TRB referees on the first version of the paper are greatly appreciated.

REFERENCES

1. M. Manheim. *Fundamentals of Transportation Systems Analysis: Volume I: Basic Concepts.* MIT Press, Cambridge, Mass., 1979.
2. *Microcomputers in Transportation: Software and Source Book.* UMTA, U.S. Department of Transportation, Washington, D.C., 1983, 1984, 1985, 1986, 1987, 1988.
3. D. Chang, V. Stover, and G. Dresser. *A Comparison of Microcomputer Packages for Network-Based Highway Planning.* Report 1110-3. Texas Transportation Institute, Texas A&M University, College Station, Tex., October 1988.
5. H. E. Pelt. Florida Standard Urban Transportation Model Structure. Presented at 1st Transportation Planning Applications Conference, Orlando, Fla., April 1987.
6. J. Gashing, R. Reboh, and J. Reiter. *Development of a Knowledge-Based Expert System for Water Resource Problems.* SRI International, Menlo Park, Calif., August 1981.
7. M. Maher, ed. *Expert Systems for Civil Engineers: Technology and Applications.* American Society of Civil Engineers, New York, N.Y., 1987.
8. F. Hayes-Roth, D. A. Waterman, and D. B. Lenat, eds. *Building Expert Systems.* Addison-Wesley, Reading, Mass., 1983.
9. C. Yeh, S. Ritchie, and J. Schneider. Potential Applications of Knowledge-Based Expert Systems in Transportation Planning and Engineering. In *Transportation Research Record 1076*, TRB, National Research Council, Washington, D.C., 1986.
10. P. Hamilton. *Michigan Microcomputer Traffic Forecasting Model Evaluation: Final Draft Comparison Matrix.* Tri-County Regional Planning Commission, Lansing, Mich., August 31, 1988.
11. A. Faghri and M. Demetsky. *A Demonstration of Expert Systems Applications in Transportation Engineering: Volume I: Transportation Engineers and Expert Systems.* Virginia Transportation Research Council, Charlottesville, Va., June 1987.

Publication of this paper sponsored by Task Force on Expert Systems.

Strategic Planning of Freight Transportation: STAN, An Interactive-Graphic System

TEODOR GABRIEL CRAINIC, MICHAEL FLORIAN, JACQUES GUÉLAT, AND HEINZ SPIESS

STAN is an interactive-graphic system for national or regional strategic planning of multimode multiproduct transportation systems. Briefly described are the general concepts underlying the system, the data base structure, the assignment procedures, and the results. Data and results from an actual application are used to illustrate these concepts.

STAN is an interactive-graphic, multiproduct, multimode method for national or regional strategic analysis and planning of freight transportation. The strategic level of planning implies a medium- to long-term horizon and a rather aggregate level of detail for the representation of the transport services provided. The contemplated alternatives usually represent anticipated (or predicted) changes in the origin to destination demands for the products considered or major changes to the transportation infrastructure. At the strategic level of planning, the aim is to obtain an overall estimation of the movements of freight of most (or all) products, by most (or all) of the modes available; thus tactical decisions (such as when to schedule the trains that are operated) are not considered at this level.

A typical use of STAN is the comparison and evaluation of alternatives. Usually, an existing situation is described by one of the scenarios of the data bank (base scenario). A simulation of the freight flows is carried out. Future situations are described by other scenarios, which are constructed and contained in the data bank. The simulation of freight flows is carried out on these scenarios as well. Then flows, link costs and delays, intermodal shipments, and other performance indicators may be compared to those of the base scenario in order to carry out various analyses and comparison evaluations.

STAN is an interactive-graphic system specially designed for the strategic planning of freight flows at the national or regional level. The system offers a comprehensive and flexible modeling framework coupled with up-to-date algorithmic techniques and powerful computing capabilities. It allows the planner to engage in the planning process by using terminology that he is familiar with and obviates data processing type of tasks. Once the data bank is set up, the planner need not know or concern himself with the technical details of computer programming or computer systems. It also permits him the instantaneous visualization of input data, results of computations, and information retrieved from the data bank, all in graphic or list form.

The goal of this paper is to describe the main features of the system. The following sections describe the general concepts that underly the design of STAN, present the most relevant characteristics of its implementation, and show an example of actual application that illustrates the features and capabilities of STAN. We conclude this paper with a discussion of the relative advantages of STAN as compared to other planning methods and a brief enumeration of planned developments and functional enhancements of STAN.

T. G. Crainic, Centre de recherche sur les transports, Université de Montréal and D.S.A., Université du Québec à Montréal, Montréal, Canada. M. Florian and H. Spiess, Centre de recherche sur les transports, Université de Montréal, Montréal, Canada. J. Guélat, Université de Lausanne, Lausanne, Switzerland.

GENERAL CONCEPTS THAT UNDERLY THE DESIGN OF STAN

The STAN system is composed of a series of modules used to input, modify, display, and output information pertinent to strategic analyses of multimode multicommodity transportation networks. All data may be entered and modified either in batch or interactively; while using the second alternative, graphic worksheets are provided for entering network data. Figure 1 gives a schematic representation of a STAN data bank, while Figure 2 presents the list of modules currently making up the system. Both are helpful in clarifying the descriptions that follow.

In STAN, the data that may be entered, modified and used for calculations fall into three main categories: networks, matrices and functions. The network data may be entered, modified, and displayed by using the *network editor*. The *matrix editor* performs the same operations for matrices and also provides facilities for matrix calculations and matrix balancing. All the relevant function sets and functions are entered and evaluated by the *function editor*. The STAN data bank may also contain demarcation lines, utility data, and a log book, which we refer to as the auxiliary data. Demarcation lines may be superimposed on graphic displays to make them more readable. Utility data contain titles, module parameters, predefined windows, etc. A log book keeps track of the use made of the various STAN modules. Brief descriptions of the networks, matrices, and functions that may be entered and manipulated by STAN are given in the following sections.

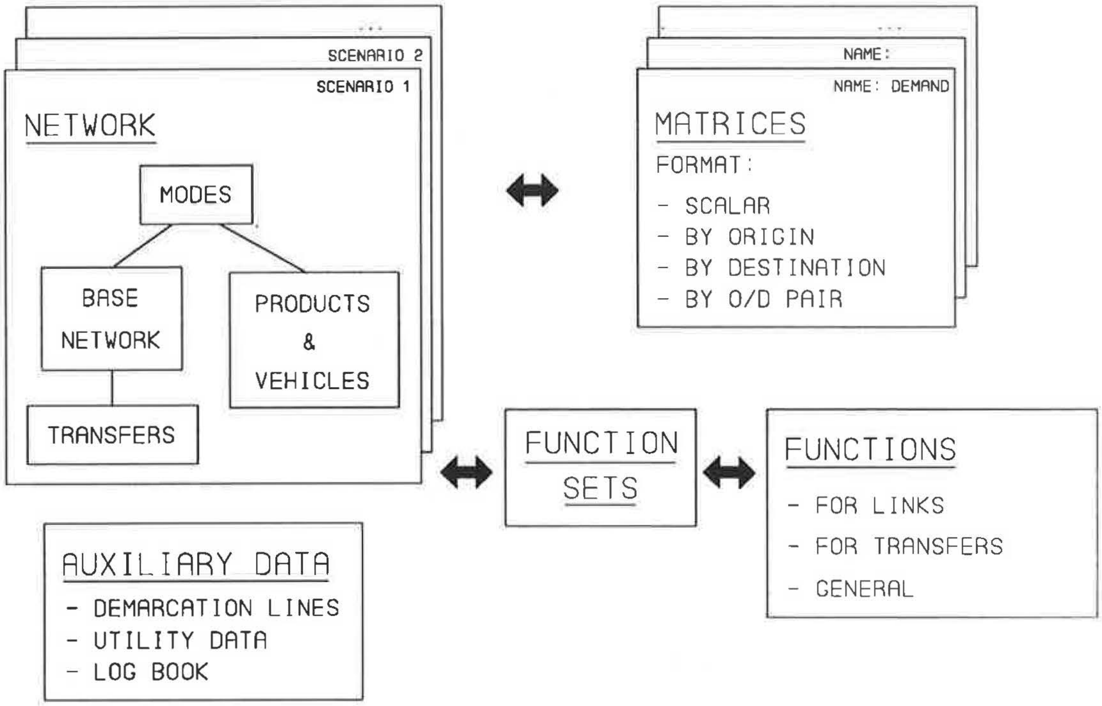


FIGURE 1 STAN data bank.

The Networks

The transportation infrastructure that spans the region studied is represented by a multimodal network. The flows that can occur over this transportation network are described by specifying the products and the modes used for their transport. A network scenario is described by a set of modes, products, vehicles, base network, and transfers. Any of this data may be modified, at any time, provided that the hierarchy of the data structure is respected, as shown in Figure 3.

A product is any commodity (collection of similar products), good, or passenger that generates a link flow specifically associated with it. A mode is a means of transportation that has its own characteristics, such as vehicle type and capacity, as well as a specific cost function. Depending on the scope and level of detail of the contemplated strategic study, a mode may represent a carrier or a part of its network representing a particular transportation service, an aggregation of several carrier networks, or specific transportation infrastructures, such as highway networks or ports.

The base network is defined by the list of all the nodes and all the links between those nodes. In STAN parallel links are allowed between two nodes with the convention that a single mode is allowed on a given link. A path may contain successive links of different modes only if transfers are specified between them. A transfer is defined at a node, as a sequence from the mode of the incoming link, via the node, to the mode of the outgoing link. Thus, transfers represent intermodal transhipments at nodes.

For each product, a vehicle type is specified for each mode. A vehicle is described by its weight, by the weight it can transport, and by an optional convoy weight (used, for instance, to specify train length). The number of vehicle types equals the number of products multiplied by the number of permitted modes.

The user may specify up to three additional data elements for each mode, product, vehicle, node and link. Such user data may be observed flows or any associated data that the planner intends to use in subsequent analyses.

Each complete network data set (modes, products, vehicles, base network, transfers) makes up a scenario. The base period is one of these scenarios. The user may define a new scenario, by duplicating the data present in an existing scenario and by making the appropriate changes in any one of the network data components. All these manipulations may be performed interactively, graphically when appropriate. In STAN, the network data that correspond to a base period is not conceptually different from data of contemplated scenarios. Various examples of modelizations of actual situations, performed by using this framework, may be found elsewhere (*1–3*).

The Matrices

The matrices that are handled in STAN may be full matrices, origin or destination vectors, or scalars. These may contain various data related to the zone subdivision of the area studied, such as origin/destination demand matrices, production by origin, and attraction by destination. A matrix may be both an input to or an output from a computational procedure.

Matrices may be entered, modified and graphically displayed by using the matrix editor. The graphical display of the content of a matrix is performed by means of bars of width proportional to the value associated with an origin/destination (O/D) pair or with a node, considered as an origin or a des-

```
1.   U T I L I T I E S
1.11 Direct data bank manipulations
1.21 Log book consultation
1.22 Scenario manipulations
1.23 Input / modify / output utility data
1.31 Input          demarcation lines using batch entry
1.32 Input / modify demarcation lines interactively
1.34 Output         demarcation lines

2.   N E T W O R K   E D I T O R
2.01 Input / modify / output modes
2.02 Input / modify / output products
2.11 Input          base network using batch entry
2.12 Input / modify base network interactively
2.13 Plot           base network
2.14 Output         base network
2.21 Input          transfers using batch entry
2.22 Input / modify transfers interactively
2.23 Plot           transfers
2.24 Output         transfers
2.33 Plot / output  shortest paths

3.   M A T R I X   E D I T O R
3.01 Input / modify / display zone groups
3.11 Input          matrices using batch entry
3.12 Input / modify matrices interactively
3.13 Plot           matrices
3.14 Output         matrices
3.15 Plot           matrix histograms
3.25 Matrix calculations
3.26 Matrix balancing

4.   F U N C T I O N   E D I T O R
4.01 Input / modify / display function sets
4.11 Input          functions using batch entry
4.12 Input / modify functions interactively
4.13 Plot           functions
4.14 Output         functions

5.   A S S I G N M E N T   P R O C E D U R E S
5.11 Prepare scenario for   assignment
5.21 Multiproduct multimode assignment

6.   R E S U L T S
6.01 Result summary
6.13 Plot           product costs and volumes
6.14 List           product results
6.15 Plot           product vehicles and convoys
6.23 Compare        product costs and volumes
6.33 Plot / output  shortest paths

9.   E N D   O F   S E S S I O N
```

FIGURE 2 Modules of STAN.

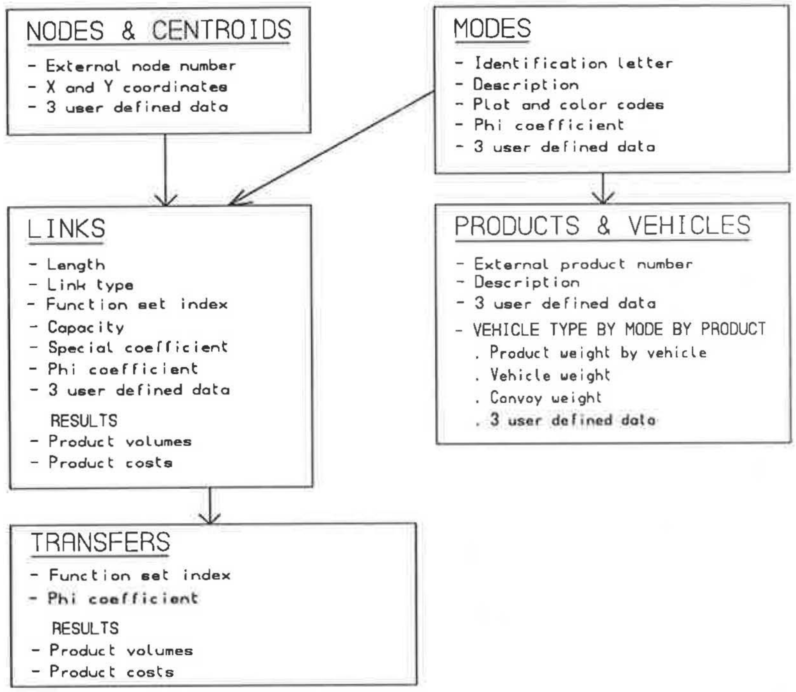


FIGURE 3 Network scenario hierarchy.

tination. Histograms, based on the contents of a matrix, may also be computed, compared and graphically displayed.

The matrix editor, which is the same as that of EMME/2 (*4,5*), has been designed to permit the user to choose a matrix format that avoids duplication of data in the data bank. For instance, if a full matrix has the same value in all its cells, then the user may specify only a scalar matrix as its value. The matrix editor will then use only the space required to keep the scalar, until a full matrix is requested in a STAN module. Then, it will automatically expand the scalar value to a full matrix with that value in all of its cells for the execution of that module. The same concept applies to given values for origins or destinations.

The matrix editor does not define any matrices by default. The definition of a matrix and its contents is left to the user. Thus, if an assignment routine expects the availability of O/D matrices, the user defines these matrices and ensures that they contain the correct data. This provides great flexibility in evaluating a given scenario with different O/D matrices.

Any network scenario created by the network editor may be used with any of the matrices present in the data bank; the user identifies the relevant matrices that he wishes to use with a given scenario.

The matrix editor includes a module that permits the user to perform matrix calculations. The user specifies the operations that he wants to perform as an algebraic expression. Such algebraic expressions may be stated by choosing from 16 logical and algebraic operands and 9 intrinsic functions. The variables that may be used in these expressions are the matrices of the STAN data bank, centroid numbers, and zone groups. The matrix editor of STAN also contains matrix balancing procedures in two and three dimensions. Their judicious use, with the appropriate matrices, permits the user to scale matrices or to implement some versions of classical spatial interaction (trip distribution) models.

For strategic planning purposes, the demand for transportation cannot be individually considered for each possible node of the network. This would generate unmanageably large O/D matrices and would explode both the model size and the computational time required by the algorithms, without improving the forecasts made. Zones, which contain several nodes in a contiguous geographical area, are defined and the transportation demand is aggregated accordingly. Each zone is then represented in the network by a special node, called a centroid.

Whenever the matrix editor expects a zone number, a zone group can be specified, indicating that the operation applies to all zones within the group. A zone group is a set of zones. A set of zone groups that includes all the zones is called an ensemble. Within an ensemble, a zone may be in only one group. The user of STAN may specify up to 26 ensembles, each containing up to 100 zone groups. The advantage of using zone groups is that selection of subsets of zones may be accomplished efficiently, which is useful in many applications.

Most of the operations which involve matrices are usually performed on the entire matrix or on a submatrix defined by a subset of origins or a subset of destinations. In addition, the matrix elements to be considered may be selected by using a constraint matrix and a constraint interval. In this case, only those elements that correspond to elements of the constraint matrix that satisfy the constraint interval will be considered.

The Functions

STAN allows the use of a wide variety of functions, which may be specified for links as well as for transfers. Up to three functions may be specified on a link (transfer) for each product defined in the current scenario. The functions associated with links and transfers are unit cost functions; they will be multiplied by the product volumes to form the total generalized cost function used in the assignment algorithm.

These functions are not part of the code and have to be specified by the user as algebraic expressions by using the usual arithmetic and logical operators and intrinsic functions. Several keywords may be used as operands, each representing either a particular network attribute (vehicle characteristics, link properties, user defined data, etc.) or a specific combination of variables (total link or transfer flows, number of vehicles or convoys on links, etc.). When required, a function is evaluated with data that correspond to the variables specified by the user in the algebraic expression that defines it. The authors have previously described particular functional forms used in actual applications (*1*,*2*).

A function set associates to a link (transfer) up to three functions for each product in the current scenario. The order of the functions in a function set corresponds to that of the components of the generalized cost used in the assignment. Function sets are an efficient and compact way to associate functions to links and transfers. Several links (transfers) may share the same function set. In addition to the types of functions described above, the user is free to define and display functions that are not employed in any of the standard calculations of STAN.

Assignment Procedures

The most general assignment procedure provided by STAN is a multimode multiproduct assignment method, which minimizes the total cost of shipping the products considered, from origins to destinations, via the permitted modes (*1*,*2*). For each product, a subset of modes, which are permitted for a given assignment, may be specified. Moreover, the O/D demand matrix of a product may be split into several matrices for which subsets of permitted modes may be defined, respectively. This mechanism allows the introduction of predefined mode choices into the STAN system. It is always possible to assign all the products to all the modes, if desired.

The assignment does not permit explicit capacities on the network. Thus, when the services of certain modes are subject to capacity restrictions (e.g., rail, ports), these may be modeled as congestion and penalty functions, which are usually integrated into the delay cost functions (*1*). The assignment procedure is being carried out by using a total generalized cost function, which is a weighted combination of up to three components of the function sets. In order to reflect the vocabulary most used in practice, these components are called operating cost, delay, and other, respectively. The interaction between the flows of different products on different modes and its effect on costs is considered explicitly in the assignment procedure of STAN.

The Results

STAN permits the user to obtain a wide variety of results in interactive graphic form or as a printed output. The main feature of the results is the possibility of interactive comparison of scenarios with graphical displays. While the main results pertaining to comparisons of scenarios are related to flows and costs, a wide variety of other results may be obtained by using the user defined data. Many of these features are illustrated below.

Worthy of emphasis is that, unlike a "batch" code, where each successful execution terminates with a particular, predefined set of results, an interactive-graphic system allows the user to obtain results of different types (graphic displays, reports, plots, etc.) during a particular session. The notion of "result" is thus different from that of a batch code and may be considered to consist of the entire gamut of displays, results of computational procedures, data bank queries, and scenario comparisons.

Auxiliary Data

The auxiliary data consist of demarcation lines, log book, and utility data. The user may define demarcation lines that may be superimposed on a graphical output. A demarcation line may identify geographical characteristics of the area such as rivers or mountains, or certain regions of the country such as provinces and states. A log book keeps record of the identity of the user and of the modules and elements of the data bank used during all STAN sessions. The log book is not created automatically; this has to be done by the user. Utility data refers to certain data which are related to the use of STAN. These data may be transferred from one STAN data base to another to facilitate the construction of new data banks.

CHARACTERISTICS OF IMPLEMENTATION

STAN has been coded in standard ANSI FORTRAN 77 and has been designed for easy transferability to various computer makes. It is currently available on IBM PC/ATs (and compatibles) equipped with the DSI-780+ and PM-030 coprocessor cards on Intel 80386/7-based PC's in protected mode and SUN Sparc workstations and servers. It may also be easily adapted for other computing environments. At present, STAN is implemented for use on a wide variety of graphic terminals that support the Tektronix protocol. It may also be used with several high-resolution graphic adapters and suitable monitors on personal computers operating under MS-DOS. Plots created with STAN may be drawn on Calcomp, Hewlett Packard, Houston Instruments, or Tektronix digital plotters. Plots may also be produced on dot matrix printers.

The code of STAN has a modular structure. Each one of the modules (see Figure 4) is an independent program, but all the modules share the STAN subroutine library. All data transfers between modules occur only via the data bank. Figure 4 gives a schematic representation of the program structure.

Worthy of mention is the fact that STAN provides the user with a language aimed at building macro procedures (*7*). This language, which is the same as that used in EMME/2, is quite versatile; it allows, for example, to specify parameters that are replaced by actual data at execution time. This is a powerful feature, one that may significantly facilitate the user

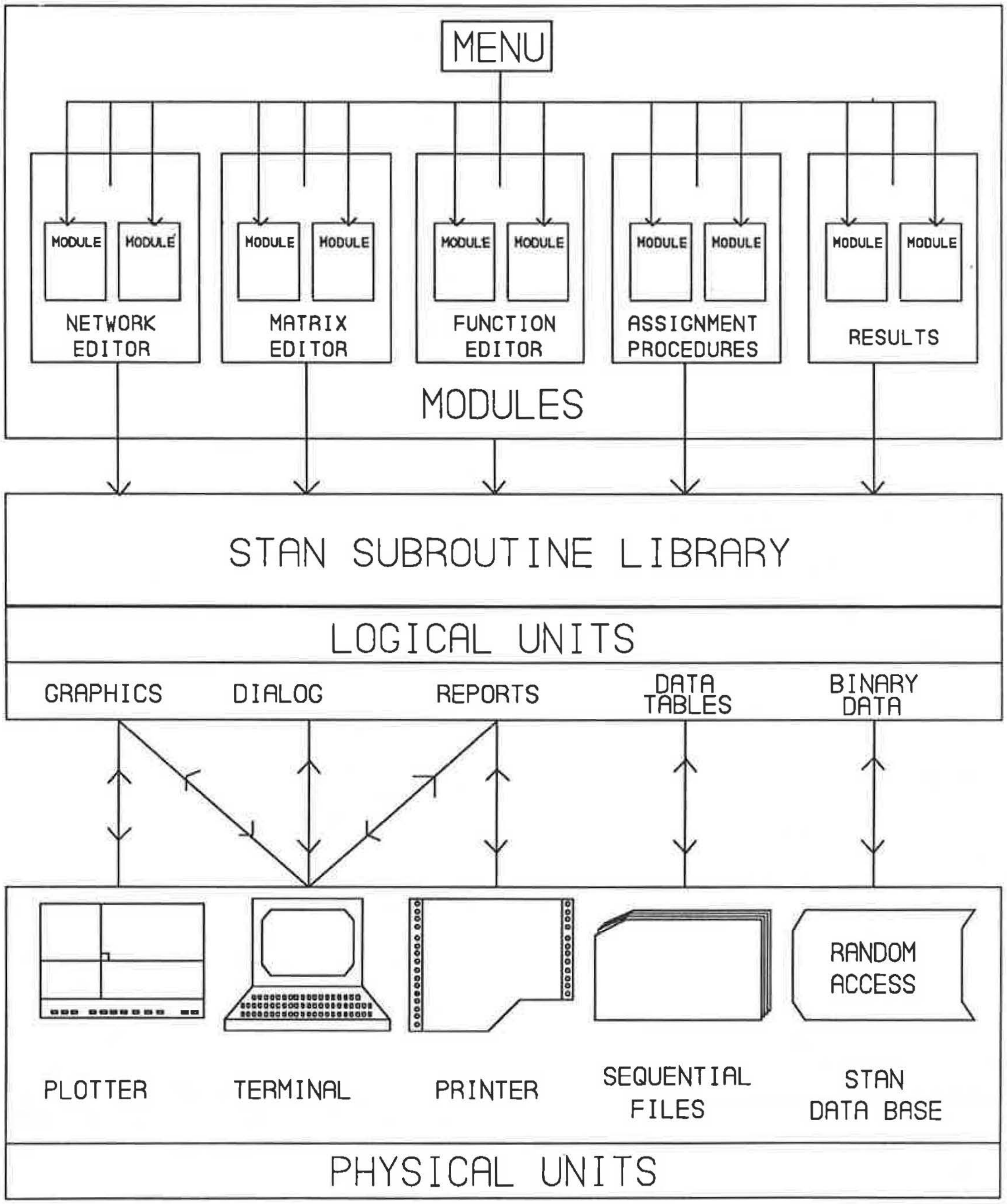


FIGURE 4 Program structure.

applications by substituting macro calls for sets of frequently used commands.

APPLICATIONS

This section is dedicated to the presentation of an actual application, in order to illustrate the capabilities and potential of STAN for the modeling and analysis of national or regional multimodal transportation networks. The object of this application was the study of the transportation network of the South-East region of Brazil. This region corresponds to the States of São Paulo, Rio de Janeiro, Minas Gerais, and Espirito Santo (regions IV, V, and VI in Figure 5), which represent the heart of the country with respect to all criteria: economics, demographic, social, political, etc. Figure 5 also illustrates the utilization in STAN of demarcation lines to identify specific subregions and geographic features of a study area.

Ideally, the data necessary for constructing planning scenarios are determined in order to permit a policy-sensitive representation and analysis of the system under study. In practice, one has to take into account the data available, the requirements of the modeling and algorithmic approach, and also

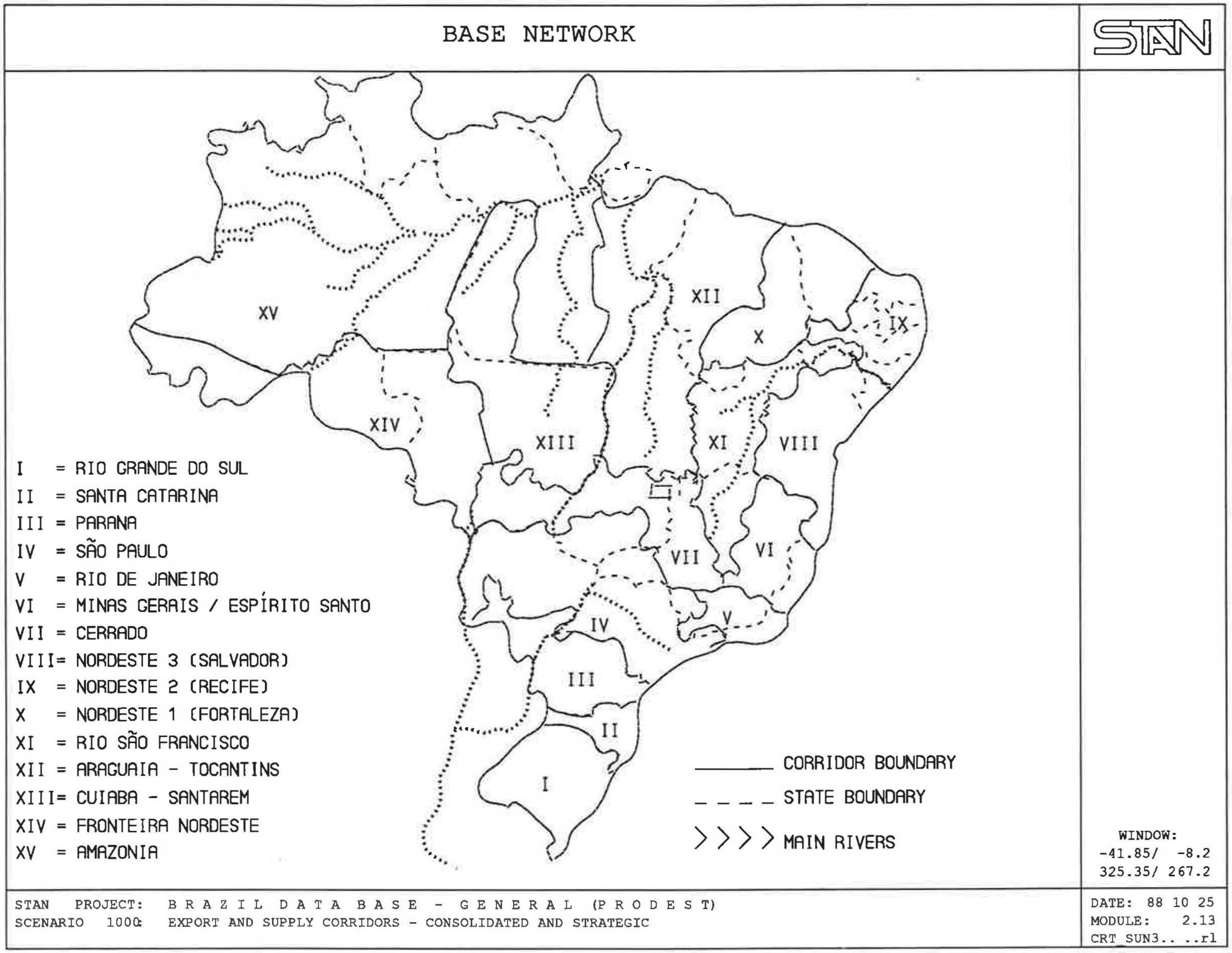


FIGURE 5 Brazilian Regions.

the anticipated efficiency of the solution and evaluation methods of the available computer-based system. The collection of data is a multistep process that requires that choices be made at each step of developing the data base for a particular application.

STAN requires data describing the components of the multimodal transportation network (modes, nodes, links, transfers) and their attributes (vehicles, lengths, capacities, unit costs, associated functions, etc.) and data quantifying the transportation demand that is to be shipped, on the permitted modes, from each origin to each destination.

The first step in the definition of the data bank for a new application is to divide the region under study into traffic zones and to associate a centroid to each zone. Zones are generally divided according to several criteria, such as homogeneity, concordance with natural boundaries, and compatibility with the main available data sources. For the Brazilian applications, we have adopted the zone definitions used by GEIPOT, which is the planning agency of the Brazilian Ministry of Transportation.

The zones defined by GEIPOT served to subdivide the South-East region, while outside the region several groups of zones have been aggregated into so-called external zones. To specify the external zones, the zones in the other regions of the country that have significant trade flows with the South-East region were retained. The other zones were subsequently aggregated into appropriate groupings. Figure 6 illustrates the main connections between the study area and these external zones.

The network for the South-East region application was defined by starting with a rather detailed network representation developed by GEIPOT. This network was subsequently aggregated, simplified, and updated by using data from various governmental organizations that operate or supervise railways, roads, ports, and navigation operations. A list of the data sources, as well as numerous details on how data for this application was collected, validated, and represented in STAN, has been published (*1*).

Ten modes were defined to represent the main transportation infrastructure of the region: five for rail (large or metric gauge with diesel or electric traction, plus a narrow gauge diesel traction mode), one for road, another for ports, and three for navigation (inland, coastal, and ocean). Figure 7 illustrates the final multimodal network; modes are identified by line pattern and by color when a color monitor is used. Node numbers may or may not be displayed. When modes are not identified, the system displays the bare network (see Figure 9) that shows all possible connections in the network.

An important part of the network definition process has been its interactive-graphic validation. By using the graphic capabilities of STAN, Brazilian technicians, who knew the network well, could focus on relatively small areas one at a time and, by using the graphic worksheet, modify any error they found. Figure 8 illustrates the use of the graphic worksheet to interactively modify the base network: node 9999 and the links between it and nodes 1441 and 1990 have been added, by using the "Reg. Nodes," "Add," "Links," and "Use Same Data" commands.

Transfers have been similarly displayed, analyzed, and eventually modified by using a graphic worksheet similar to the one previously presented for the base network. Figure 9 illustrates how transfers may be graphically selected (this may be performed on either the bare or the base networks), while Figure 10 displays the plot of a transfer. Several transfers may be simultaneously displayed on the same plot.

In order to complete the specification of the transport services provided on this network, a variety of attributes had to be defined. These include the characteristics of the typical vehicles that transport each product on each mode, the physical characteristics and the functional forms for the transportation delay cost, and energy consumption for each link and transfer of the network. In particular, functions were used to model the impact of the road's condition (unpaved, paved), horizontal alignment, profile, and maintenance state on the operating costs, energy consumption, and travel time of various vehicle types. Similar studies were conducted for the navigation modes. Appropriate functions were also used to represent the impact of limited infrastructure capacity and congestion conditions on the traffic flows and productivity measures (costs, travel times, etc.) for the rail and port modes. Figure 11 shows various examples of link and transfer delay functions for the rail modes, as well as the keywords of STAN that may be used to define them. Figure 12 displays the plot of a typical rail link delay function (such as fl22 in Figure 11) for various capacities and congestion levels. This exercise required long, arduous, and often frustrating efforts, often due to the lack of homogeneity among the data sources and sometimes due to lack of reliable data.

Data that permit a meaningful visual representation of the network were also included, such as demarcation lines for the national and regional boundaries, as well as natural (rivers and mountains) or artificial (hydroelectric works) geographical features. Demarcation lines constitute a salient feature of most of the figures of this chapter, mainly Figure 5.

Building the origin-destination (O/D) matrices, containing the demand (in tons) of each product that was included in the study, could not have been successfully accomplished within the time and budget resources of the project without the help of existing data. In this respect, we were fortunate to have access to the O/D matrices that had already been estimated by GEIPOT.

For the purpose of the pilot phase of the project, eight products were chosen for validation purposes. These products were chosen for their importance and impact on the transportation system in the South-East region. The eight products are: iron ore, steel products, coal, cement, fertilizers, soya oil, soya grain, and soya meal. These products generate the major flows on the links of all the modes present in the region.

These matrices were validated, updated, and aggregated, according to the network definition of the application, by using the graphic and analytic tools provided by STAN. Several graphical representations of matrix data were used: desire lines representing O/D transportation demands (Figure 13), chimneys proportional to the total supplies or demands or both (Figures 14 and 15), and histograms or histogram comparisons of matrix and cumulative values (Figures 16 and 17).

The analytic tools provided by STAN were used to validate the sectoral interdependencies observed for the selected products by answering the question: Are the quantities of products arriving at a center sufficient to generate the final output of product, according to the transformation coefficients? For example, the matrix calculator was used to transform the

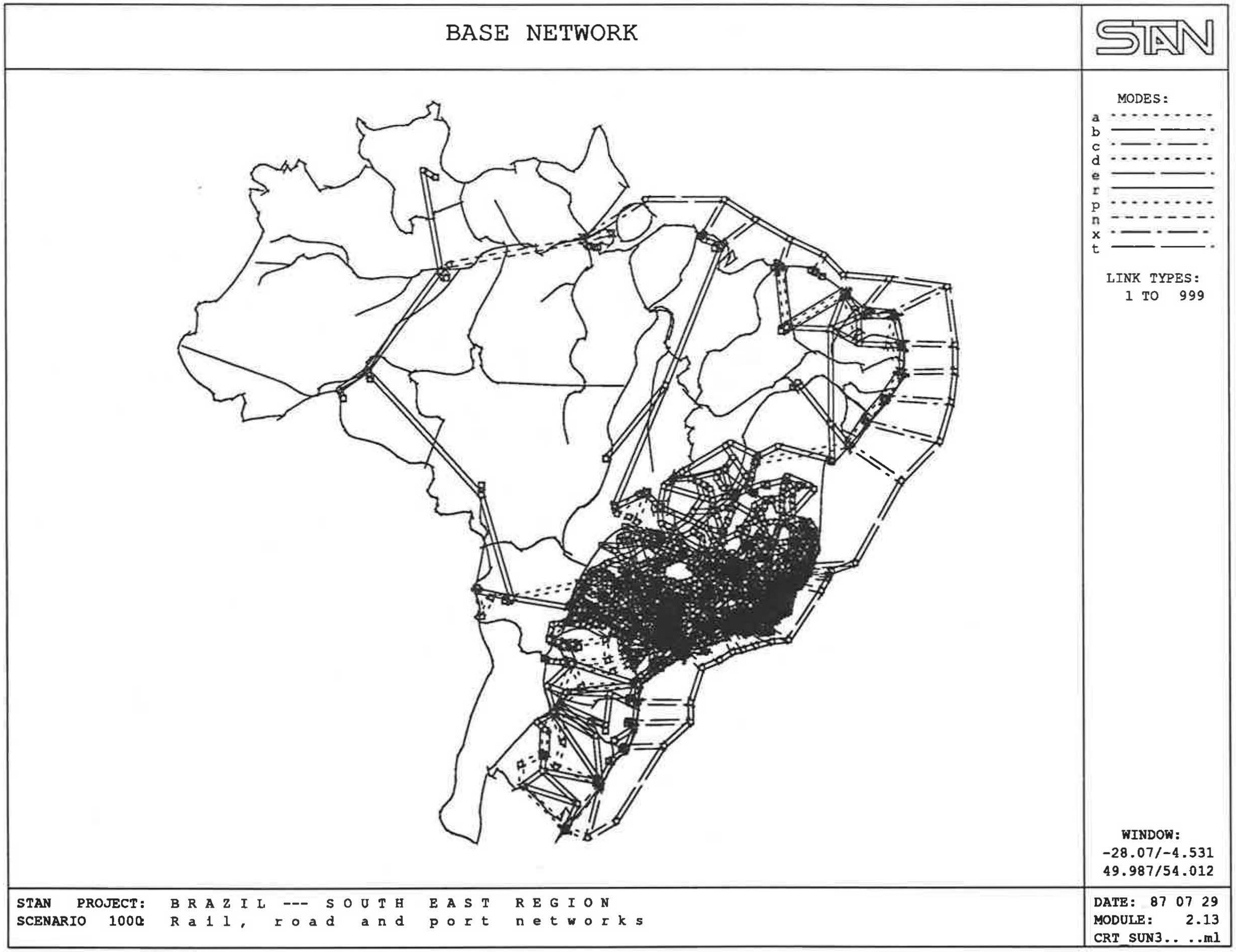


FIGURE 6 External zone connections.

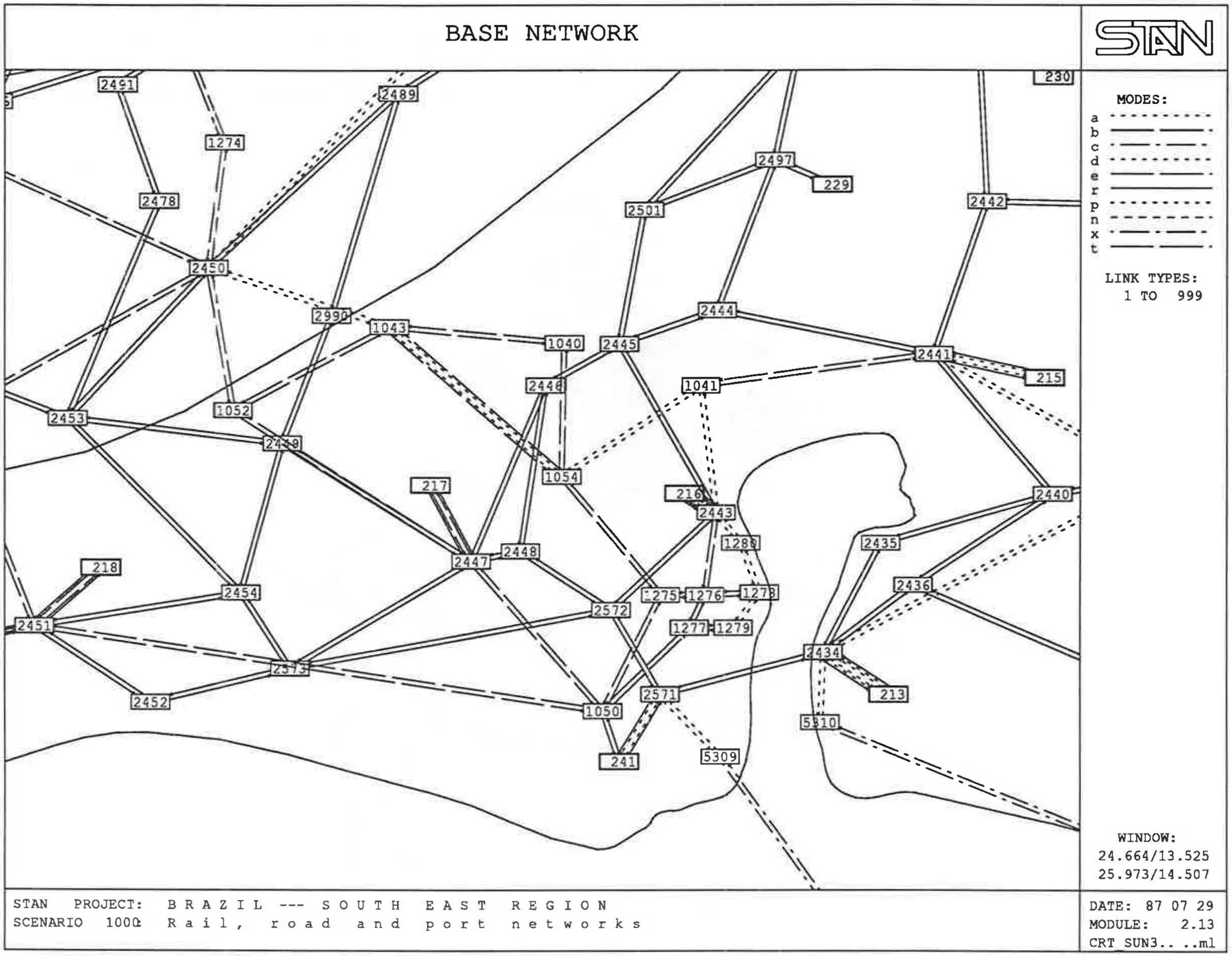


FIGURE 7 Multimodal network representation.

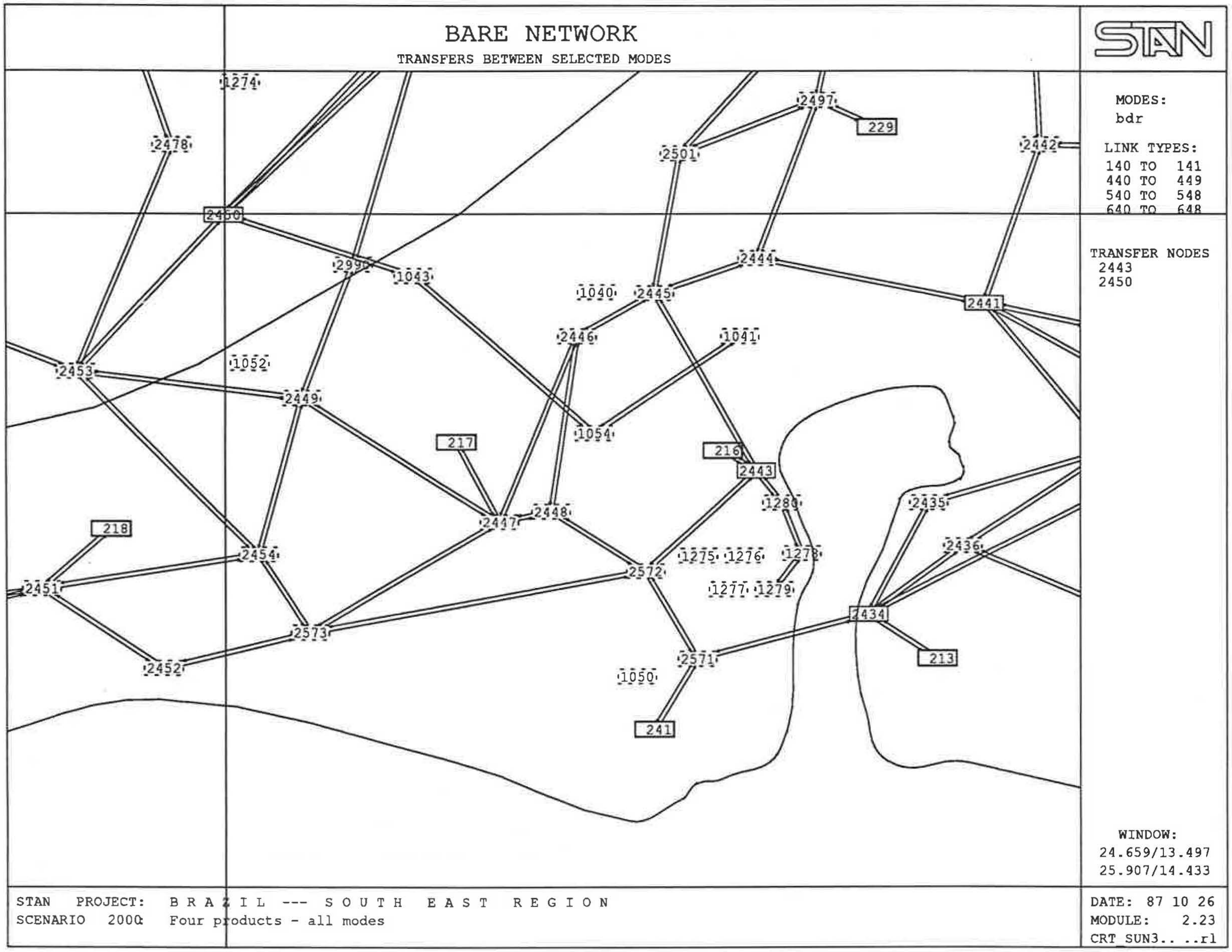


FIGURE 9 Interactive-graphic selection of transfers.

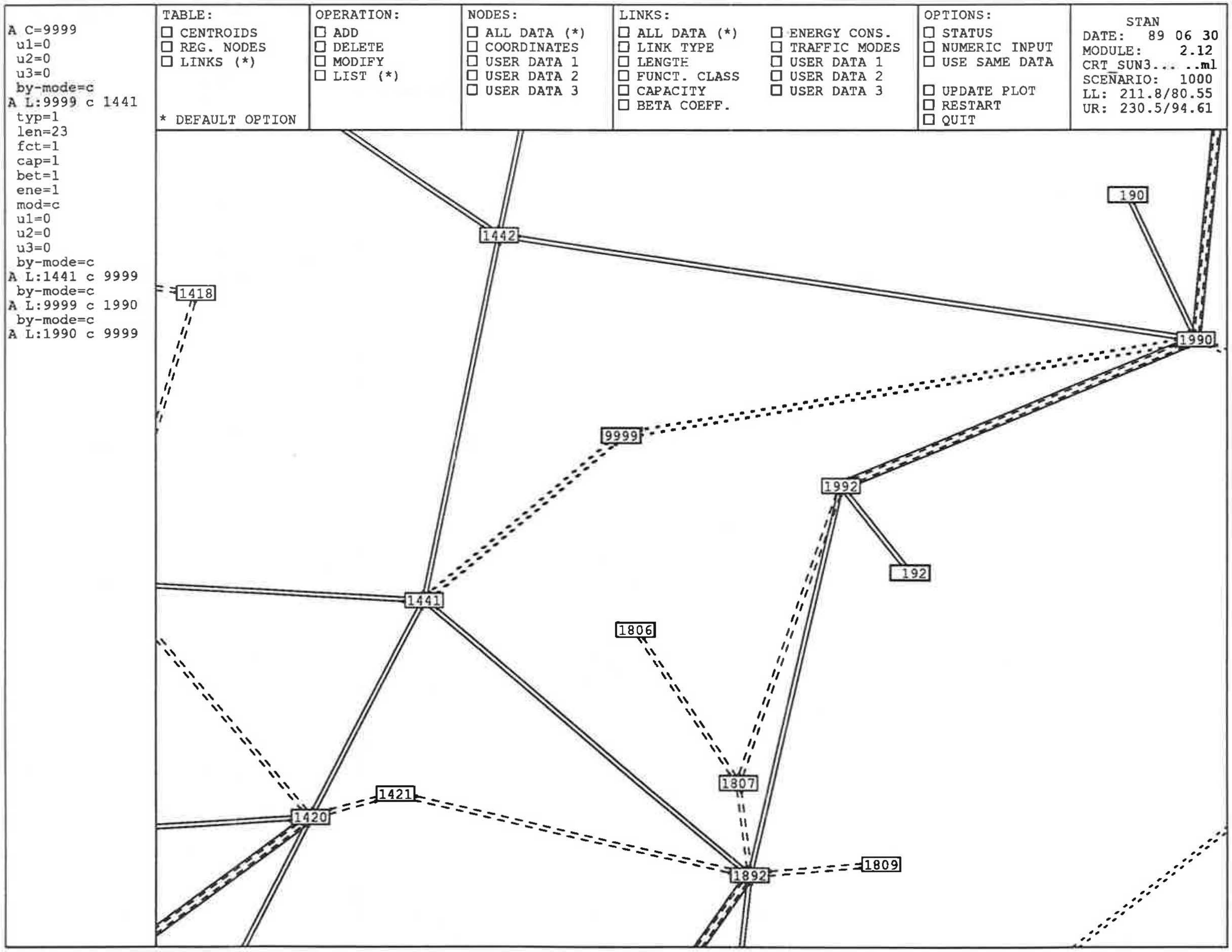


FIGURE 8 **Graphic worksheet.**

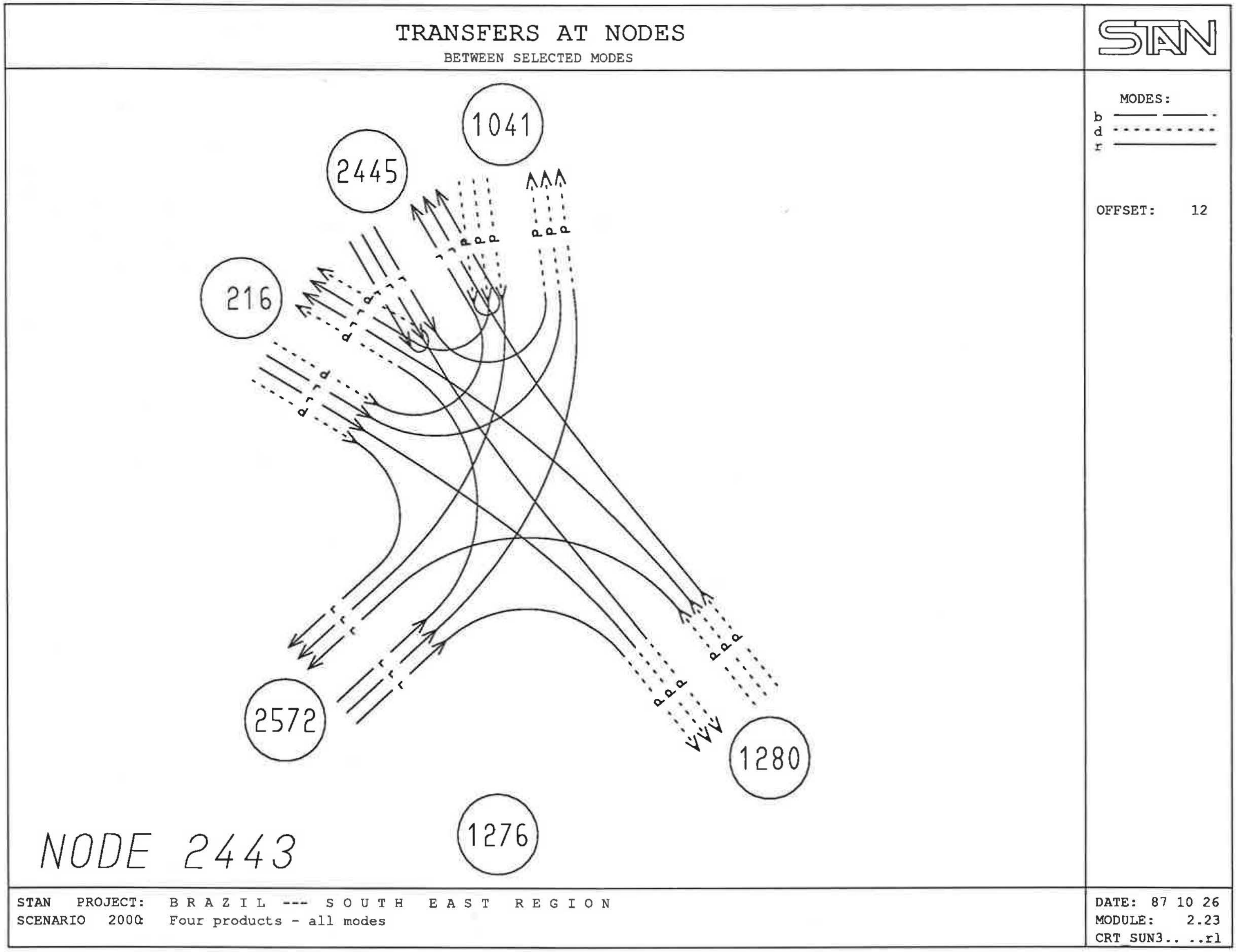


FIGURE 10 Plot of an intermodal transfer.

```
STAN Module:   4.14          Date: 91 02 28          User: SOOO/STANDEMO...SG
Project:       B R A Z I L --- S O U T H E A S T R E G I O N
------------------------------------------------------------------------------------------------

Function definitions
--------------------


link cost                 ( length   cap      beta     ul1      ul2      ul3      volprod
                            conwgt   wbyveh   vehwgt   up1      up2      up3      voltot
                            vehprod  vehtot   train    strain   um1      um2      um3
                            ump1     ump2     ump3     phil     phim )

 fl21  = 0

 fl22  = .992 * ((1 + vehwgt / wbyveh) / (conwgt * beta)) * .0182 *
         length * (1 + .055 * strain + 6.18 * ((strain / cap) ^ 6))

 fl23  = .992 * ((1 + vehwgt / wbyveh) / (conwgt * beta)) * .0333 *
         length * (1 + .055 * strain + 6.18 * ((strain / cap) ^ 6))

 fl24  = .0017 * length / wbyveh


transfer cost             ( volprod  voltot   up1      up2      up3      un1      un2
                            un3      phit )

 ft31  = 0


user general              ( x0       x1       x2       x3       x4       x5       x6
                            x7       x8       x9 )

 fu1   = .5 * x1 * (sin(.5 * 3.1416 * x1))

 fu2   = sqrt(0 .max. (x1 - 1))
```

FIGURE 11 Link and transfer delay functions.

matrices of the final products (e.g., soya meal and soya oil) into the matrices for the raw product (e.g., soya grain). These matrices were then added up, and the result was aggregated by origin, giving the demand vector. The O/D matrix for the raw product was aggregated by destination, giving the supply vector. The supply-demand differences were then computed and recorded. All these operations were performed by using the matrix calculator. The differences are displayed in Figure 15 and the corresponding histogram in Figure 17. The excess in the supply of soya grain indicated, in most cases, export flows by maritime ports. Deficits pointed to the need to verify the data by consulting the industries that were involved. In the particular case of soya products, the differences were small and relatively easy to reconcile.

The matrix calculator and the matrix balancing procedures provided by the STAN matrix editor were also used to generate O/D demand matrices for intermediate years, starting from a base-year O/D matrix and forecast supplies and demands. Similar procedures were also used to generate the O/D matrix for the empty car traffic that has to be considered on the rail modes (*4*).

Several freight transportation scenarios were analysed by using this data base. The scenarios were designed to concord with the needs and priorities of the Ministry of Transportation and were concerned with increasing the productivity and efficiency of the transportation facilities, the adaptation of the transportation system to the requirements of the domestic and foreign market, and the selection of the investment projects and improvements in the operation of the system, as well as the control and reduction of fuel consumption in the transportation sector. Assignments were performed either by using several O/D matrices by product, reflecting an exogenous mode choice, or on the multimodal network, according to a chosen measure of generalized cost.

Quantitative results, tabulated or listed in conventional reports, were useful for detailed cost, benefits, and infrastructure utilization analyses. Of an equally great importance, for the instantaneous grasp of the situation and of the impact of the policies that were proposed, were the graphical representation of results. Plots of product flows in tons on the bare network (Figure 18; this figure also indicates the selection of the window that appears in Figure 19) or base network proved

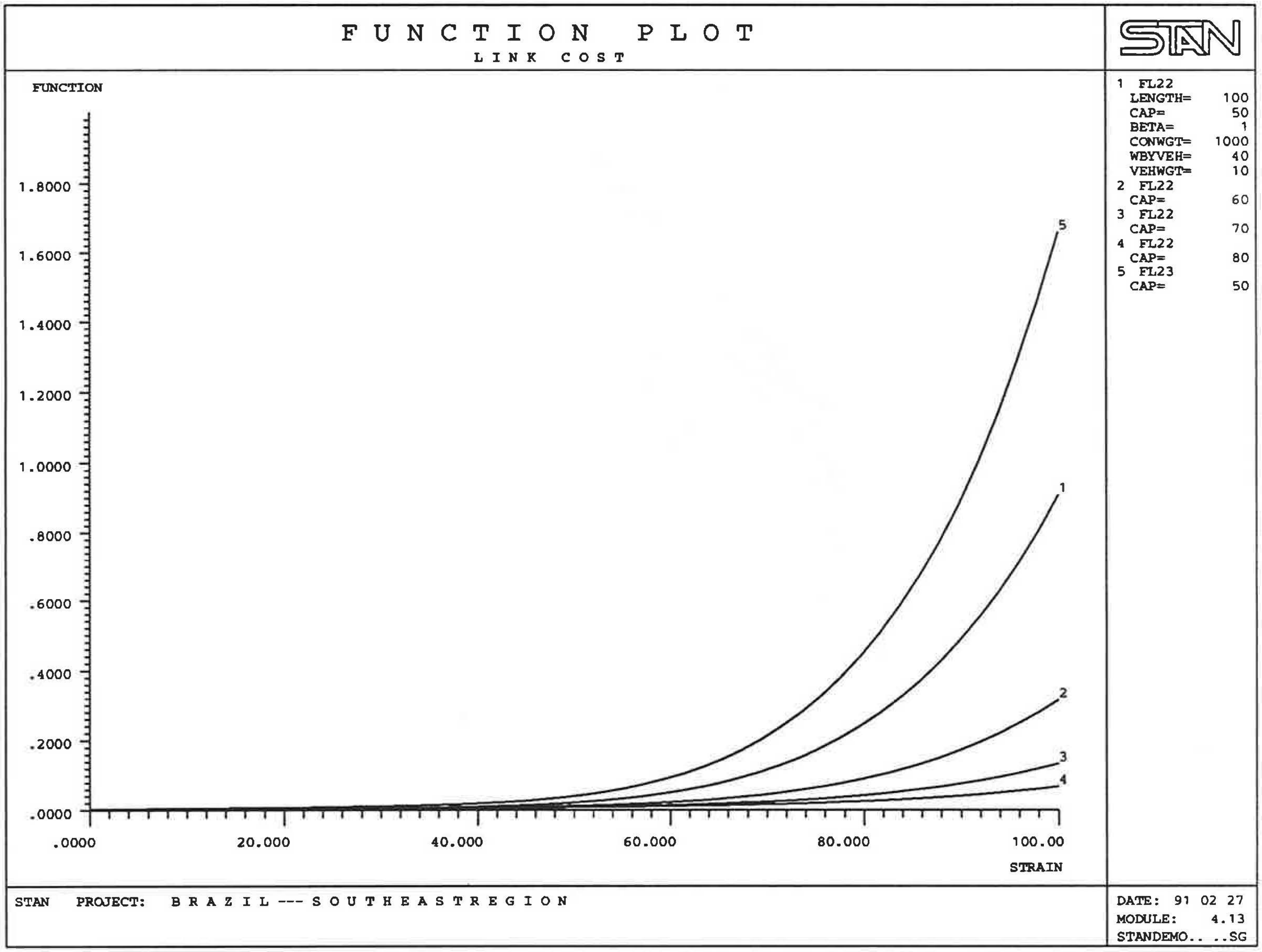


FIGURE 12 **Rail link cost function.**

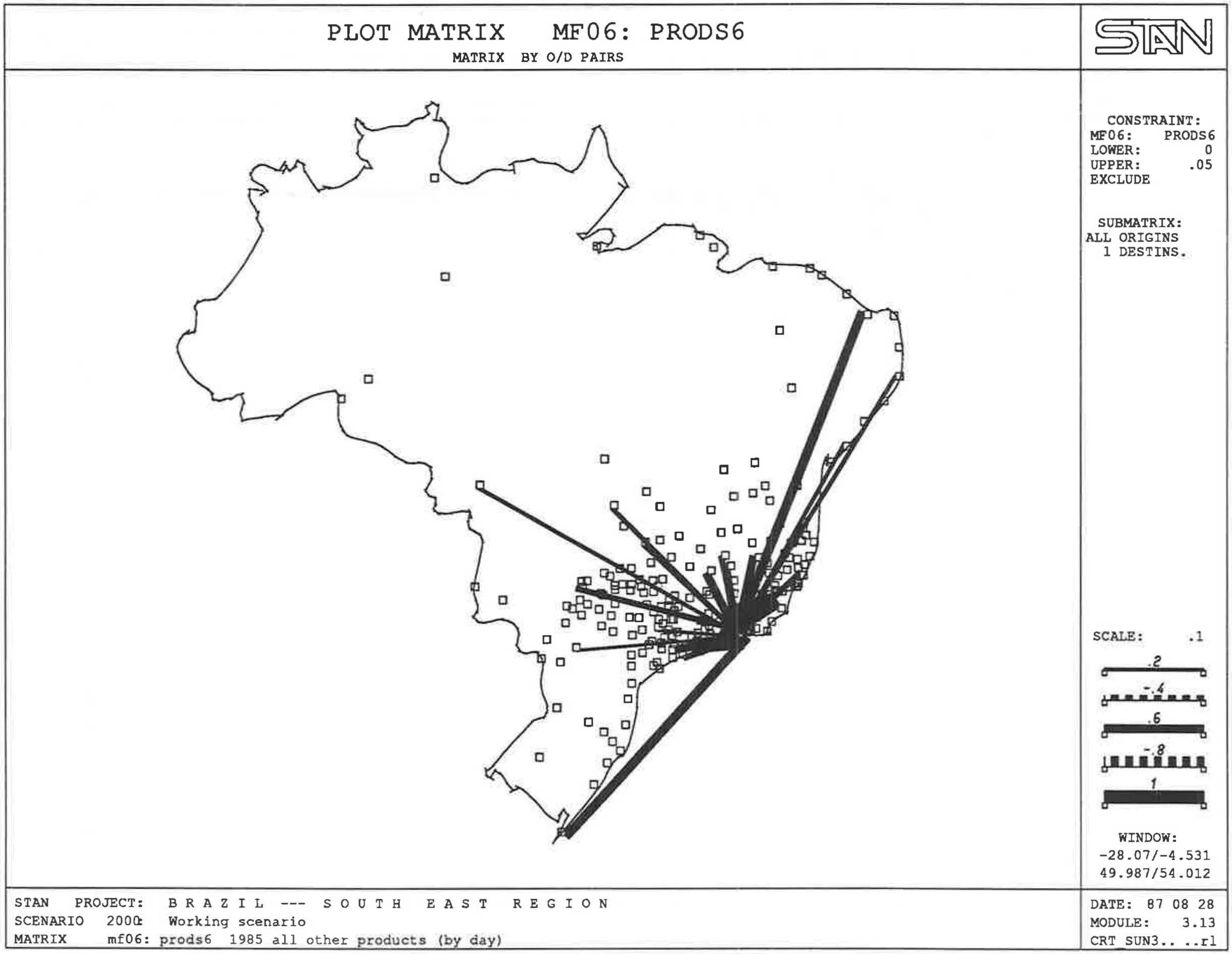


FIGURE 13 Matrix plot of O/D desire links.

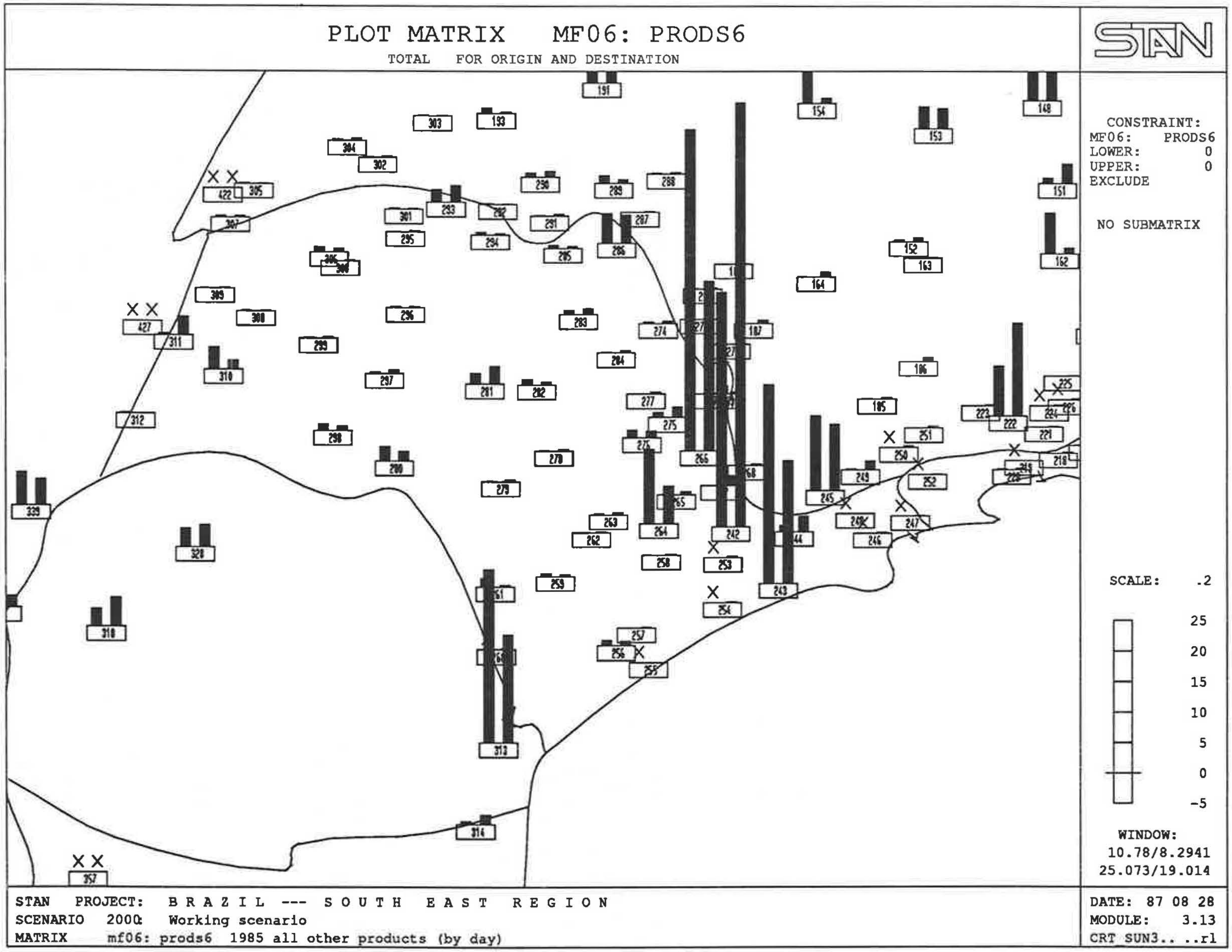


FIGURE 14 **Zone supplies and demands.**

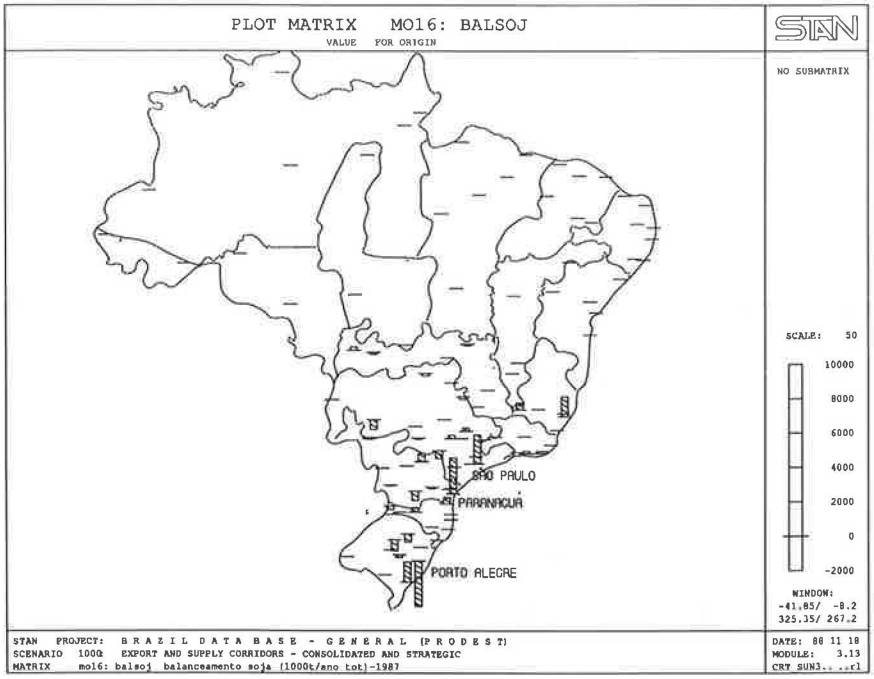


FIGURE 15 Soya grain supply-demand differences.

useful. Besides product volumes, the corresponding unit costs or number of vehicles or convoys (Figure 20) were also plotted on the links and transfers of the network.

The most useful graphical displays, however, were the plots of comparisons of scenarios, which vividly bring to life the impact of a specific policy or investment on the flow distribution and on the utilization of the infrastructure, when compared to the base year, for example. Comparisons of the product volumes or costs may be displayed on the links (Figures 21 and 22) or the transfers (Figure 23) of the network. Other displays present shortest unit (marginal) cost paths (Figure 24).

The STAN system proved to be an analytic tool of great use for the strategic planning of the transportation sector in Brazil. The application had a practical use, as it was integrated in studies that evaluated the feasibility of investment projects for upgrading the infrastructure and improving the operations and the energy consumption.

THE STAN SYSTEM IN PERSPECTIVE

We conclude this presentation with a brief analysis of the main differences between STAN and other methods aimed at long-range planning of transportation networks, and with an enumeration of planned developments, both methodological and functional, of the system. First, it is worth mentioning that although STAN is specifically designed for freight transportation, it is based to a large extent on the design of EMME/2 (*4,5*), which is an interactive-graphic multimode method for urban and regional planning for the transportation of persons. In particular, the user interface and the matrix editor are similar to that used in EMME/2.

From a methodological point of view, our approach offers a modeling framework that is richer than most models previously proposed (*2,3,6*). It is clear, however, that for strategic planning purposes, we do not seek the detail of a model that identifies shippers and carriers explicitly. Rather, we consider a model that is adequate for scenario comparisons when major investments are considered. Compared to the model of Friesz, Gottfried, and Morlok (*8*), which is a formulation of breadth and scope similar to ours, we also assume an exogenous total demand by product but, unlike their model, our formulation does not include a carrier choice, or equivalently an endogenous mode choice component. We assume therefore that the shippers' behavior is reflected in the O/D matrices used and the specification of the mode choice, as remarked above. Note, however, that when specific information is known concerning some shippers and their mode choice, this information may be introduced into our model by the proper definition of the permitted modal subset for the corresponding O/D matrix. The resulting flows are obtained by mode, and no explicit information is available by carrier, unless a carrier and a mode coincide.

In particular, we obtain the second component of the Friesz, Gottfried, and Morlok model, which deals with the assignment of freight flows on carrier networks, by defining a mode for each carrier and by using the same functions and demand specification. Thus, the network model is much more refined, because it allows detailed representation of the transportation infrastructure, facilities, and services, as well as the simultaneous assignment of multiple products on multiple modes, thus capturing the competition of products for the service capacity available, a feature that is of particular relevance when alternative scenarios of network capacity expansion are considered. While our aim is to provide a valid method for

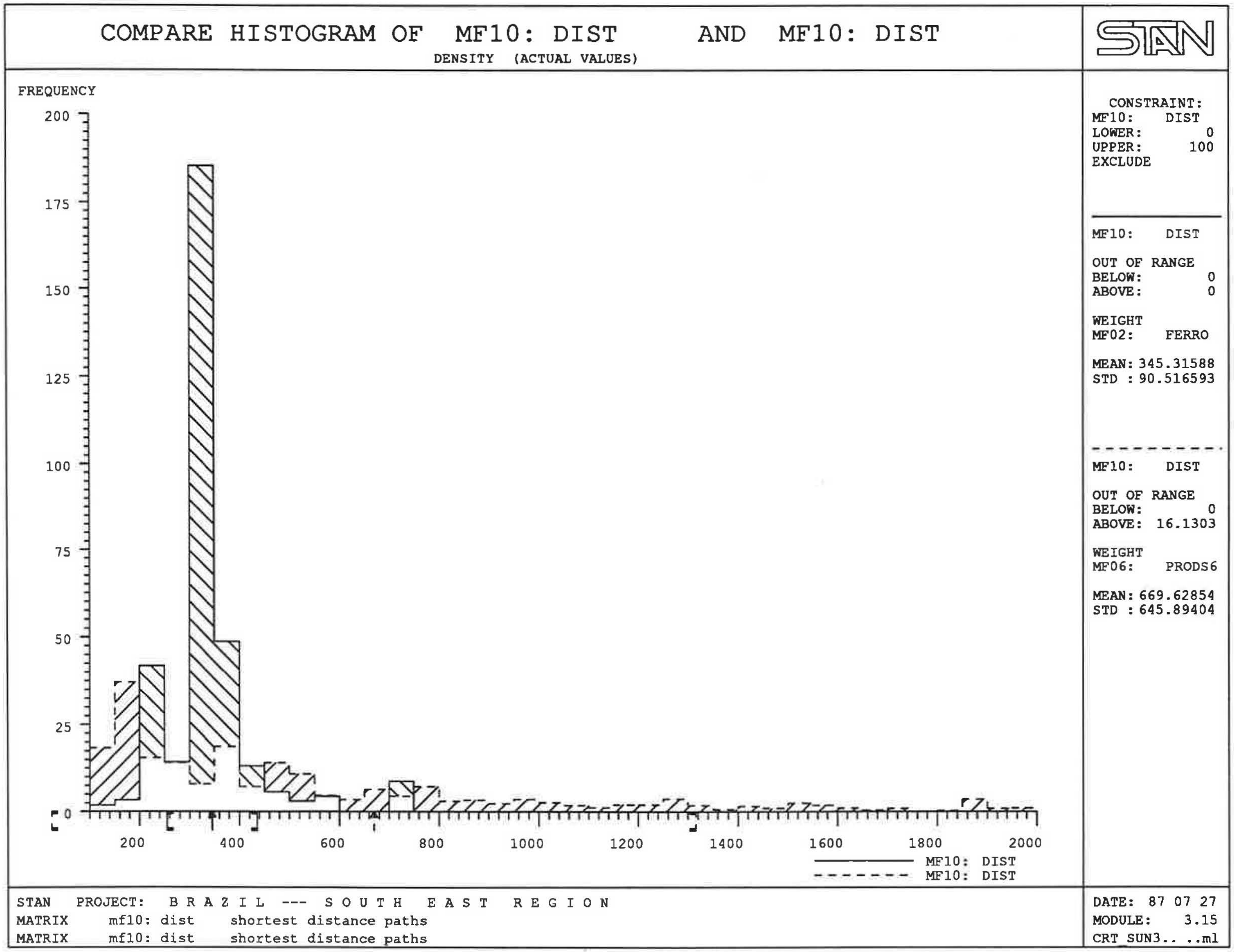


FIGURE 16 Comparison of matrix histograms.

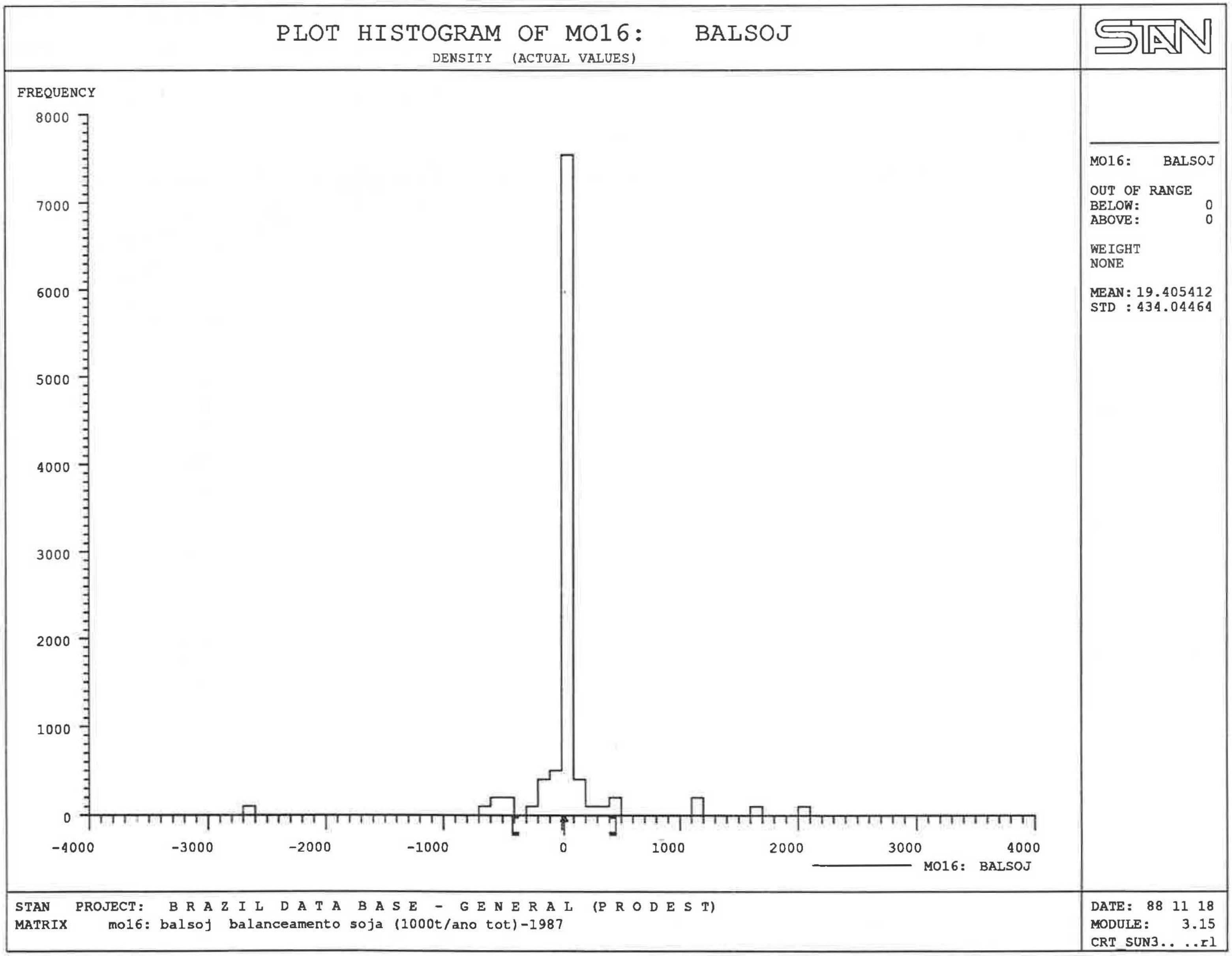


FIGURE 17 Histogram of soya grain supply-demand differences

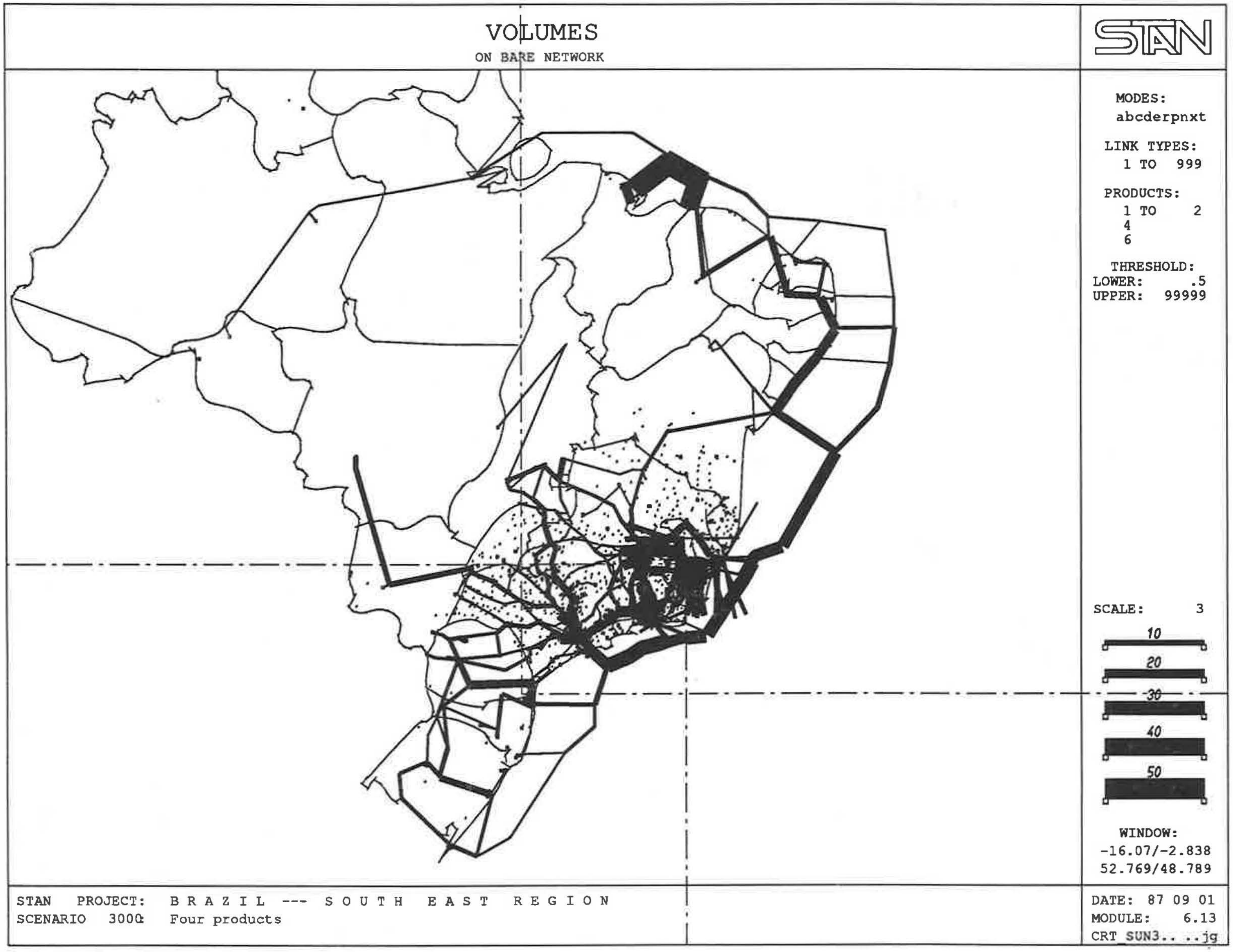


FIGURE 18 Plot of total link flows.

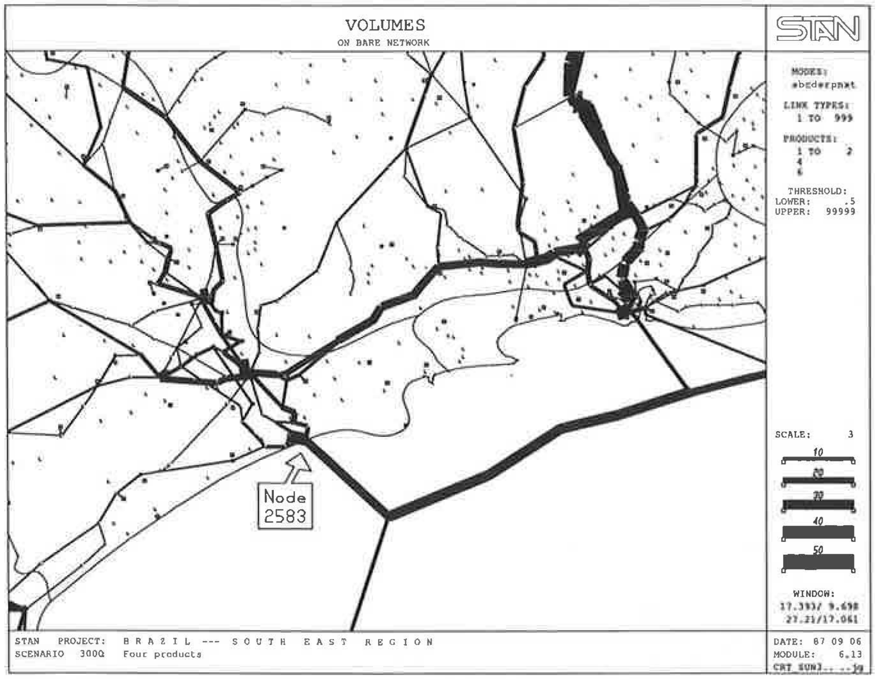


FIGURE 19 Subregional window defined in Figure 18.

strategic planning at a national level, the model is sufficiently flexible to represent the transport infrastructure of one carrier only. In summary, the model that we propose provides a refined representation of a large multimode, multiproduct transportation system for strategic planning purposes, but it may also be used to analyze the freight transportation carried out by a single carrier.

With respect to graphic displays and plots, one must also consider the *Princeton Transportation Network Model and Graphic Transportation Model* (*9,10*), which provides relatively simple optimization procedures. However, this system does not offer a general modeling framework, since it consists of a very detailed representation of the North American transportation network. As such, it is not adaptable to different contents and its scope is thus limited.

An important characteristic of STAN is the fact that it is an "open" system, in the sense that new developments and enhancements may be added both to its methodological core and to its functionality.

The STAN system is now used by several organizations in Brazil, Sweden, U.K., Finland, Spain, and the United States. The feedback that we receive from the users and the progress in our own research leads us to define new research and development objectives that result in enhancements of the system. We briefly mention some of them.

The basic model that is implemented in STAN assumes that the demand for transportation is specified for each product by a number of O/D matrices. However, many strategic analyses for freight networks take as given only the vectors of total supply and demand of products, at origins and destinations, respectively. This problem is particularly important for the analysis of transportation of raw materials such as coal, where a customer is indifferent as to the origin of the product. Thus, an algorithm for the simultaneous determination of flows and demand matrices is being developed and will form the basis of a new module of STAN.

Another computational algorithm that will be provided by the STAN system is a multiproduct multimode maximum flow method (*1*). This facility allows the evaluation of the maximal amounts of certain commodities that can be transported with the existing transportation infrastructure, and thus may be very useful when considering major changes in demand.

The actual implementation of the model does not save the path information, due to memory considerations. However, this information is often required in transport analyses of particular O/D markets. We plan therefore to develop a method that identifies the set of the paths used for a given product and O/D pair, given the solution to the multimodal multiproduct network assignment problem. This method will be used as an a posteriori analysis tool, using the path marginal cost information provided by the STAN assignment algorithm.

The main enhancements planned for user interface functionality consist of a generalized link selection facility and a network calculator. These developments will allow access to subnetworks by any link attribute, predicted flows, etc., and (in conjunction with the network calculator) will provide a powerful tool for calibration, data validation, and the implementation of various evaluation methods.

ACKNOWLEDGMENT

The authors would like to thank GEIPOT, Ministry of Transportation of Brazil, for the data that they have provided for the applications of STAN described in this paper. Their collaboration was essential to the success of this work.

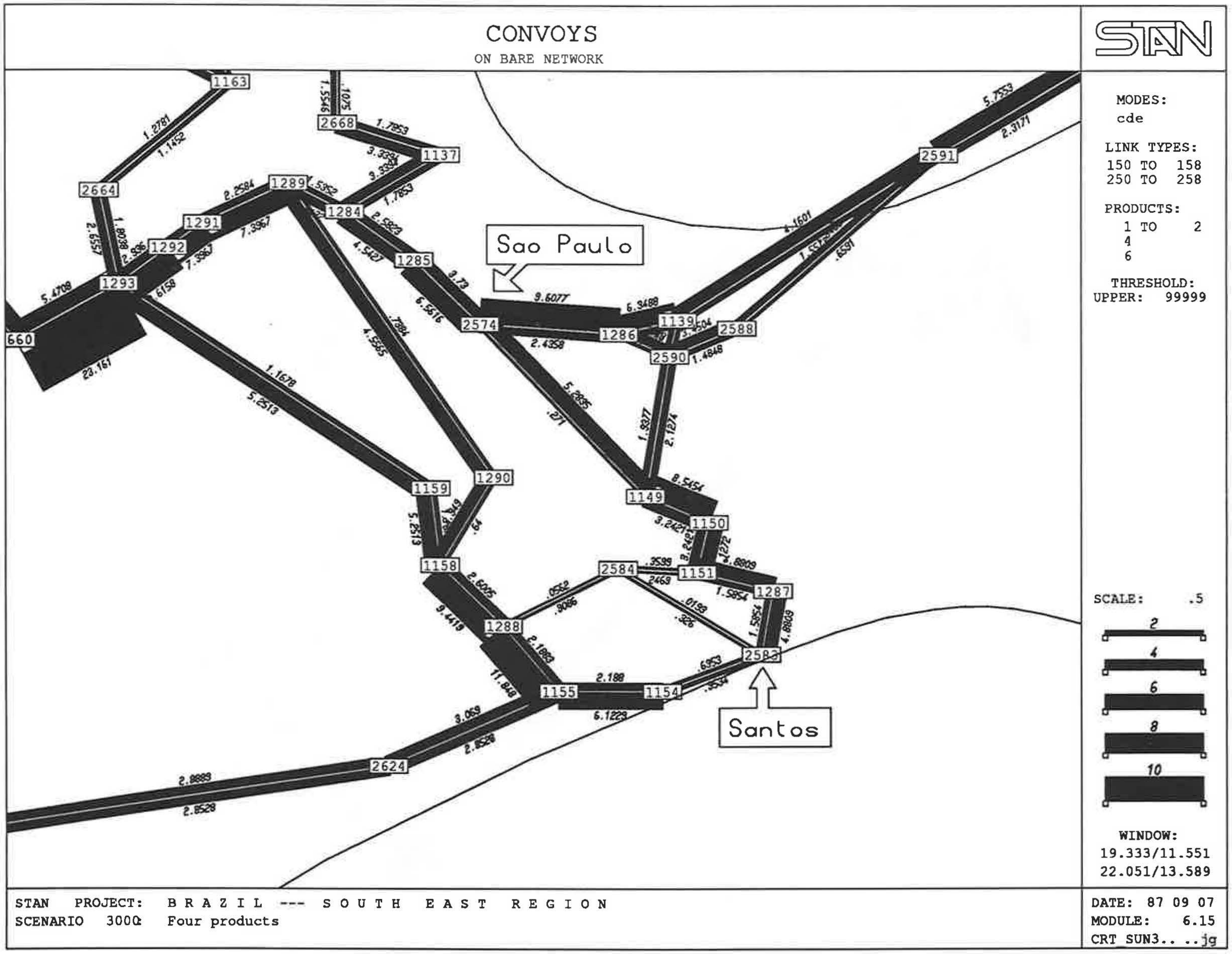


FIGURE 20 Total number of trains on a selected subnetwork.

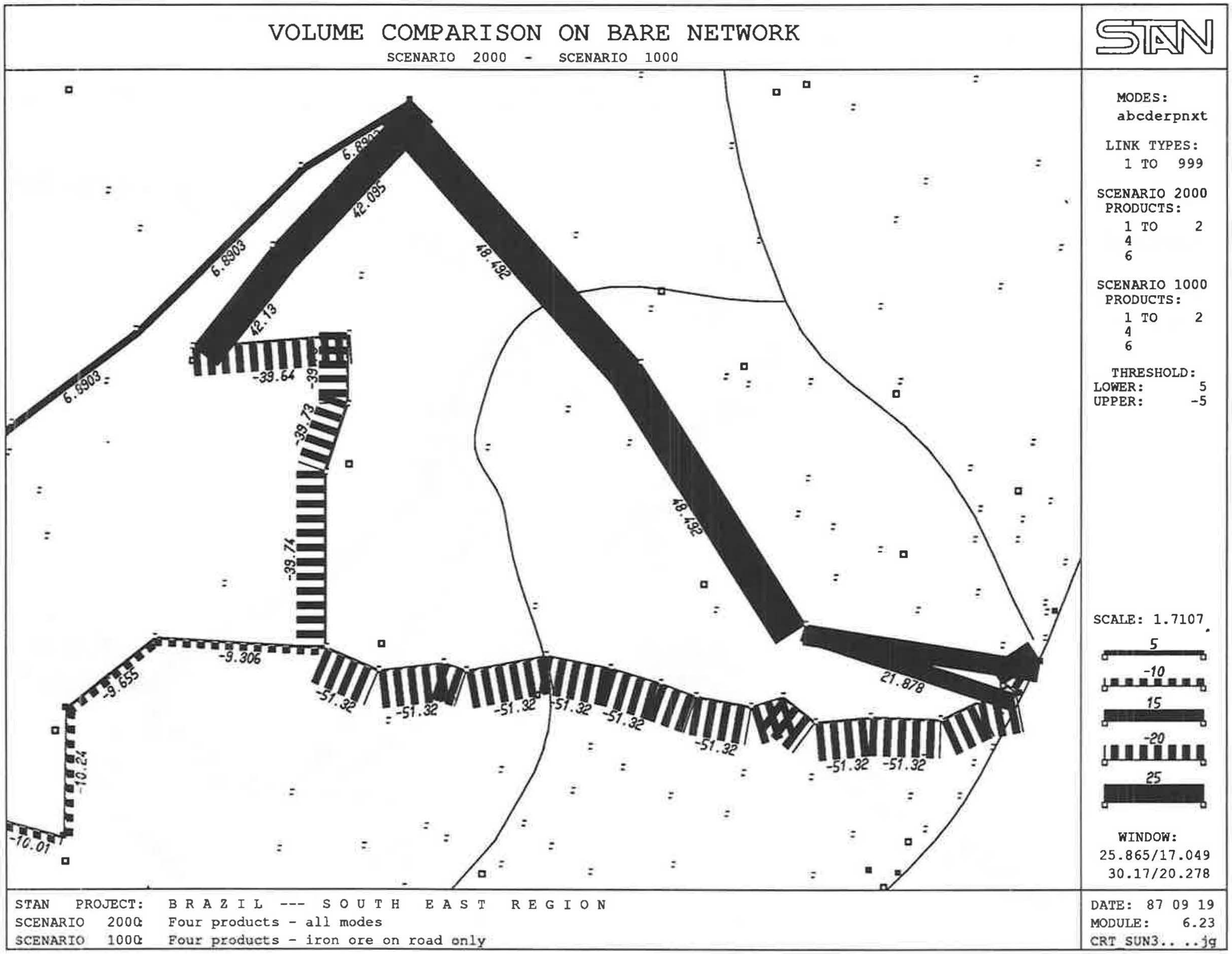


FIGURE 21 Volume comparison on bare network.

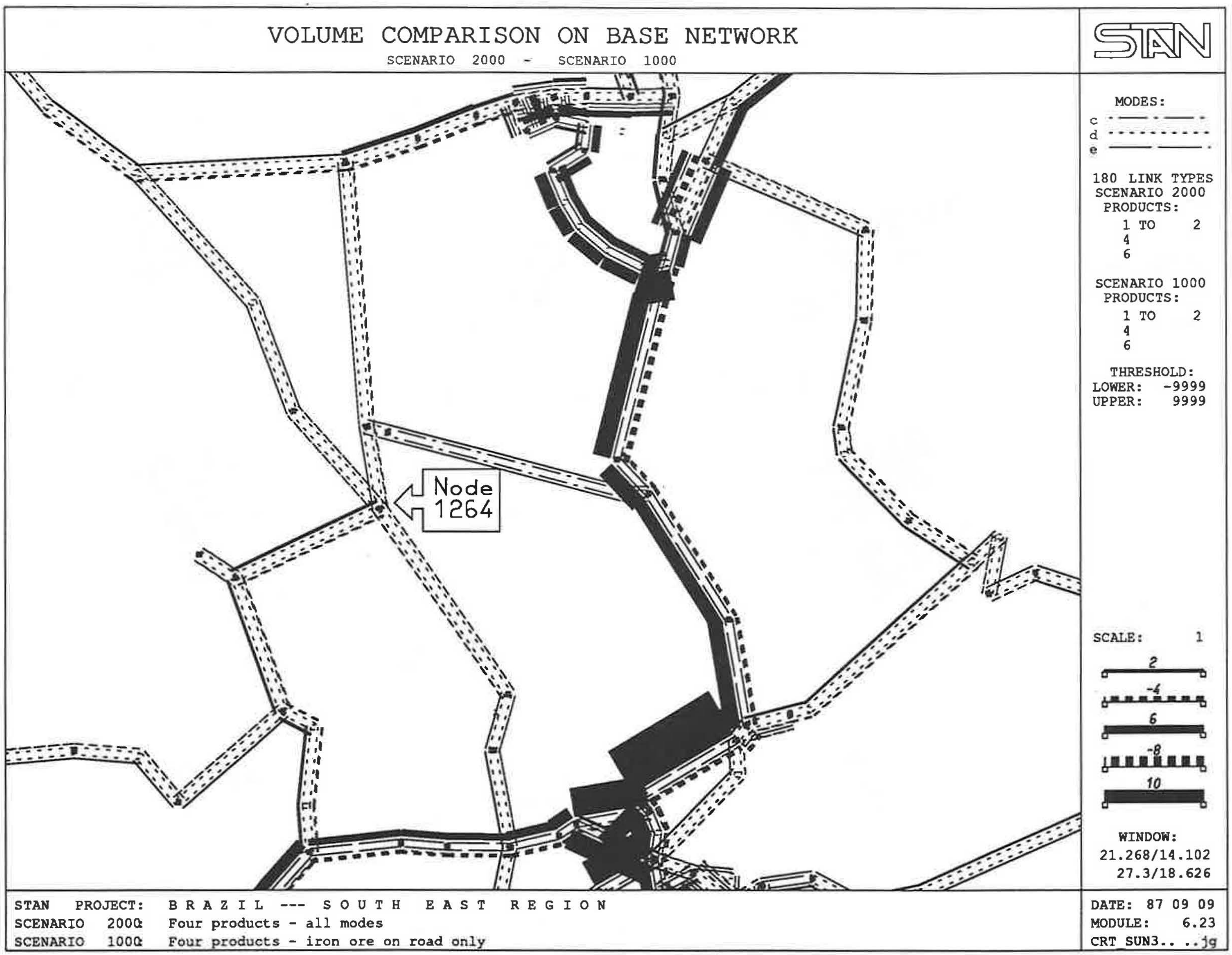


FIGURE 22 Volume comparison on base network.

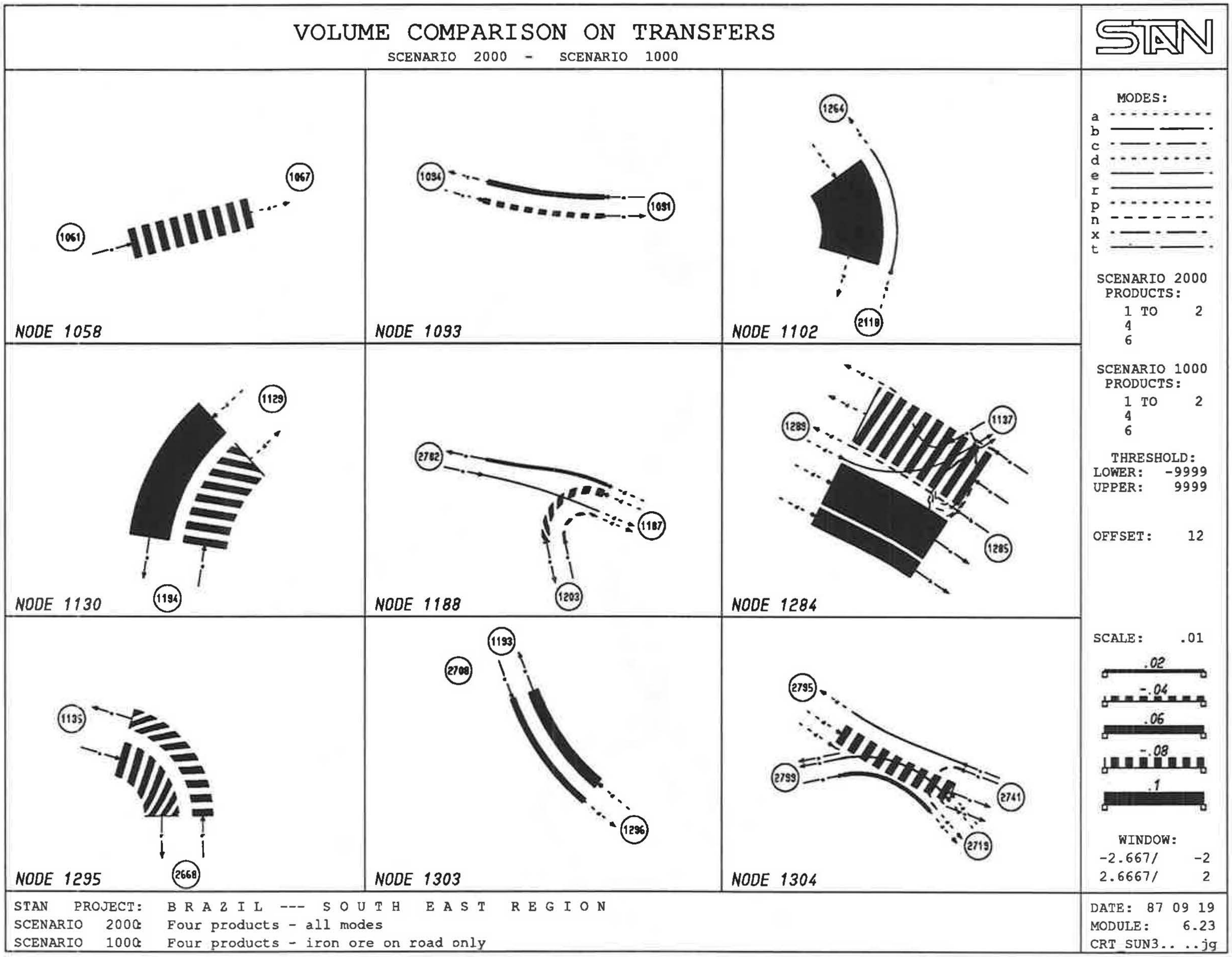


FIGURE 23 **Volume comparison on transfers.**

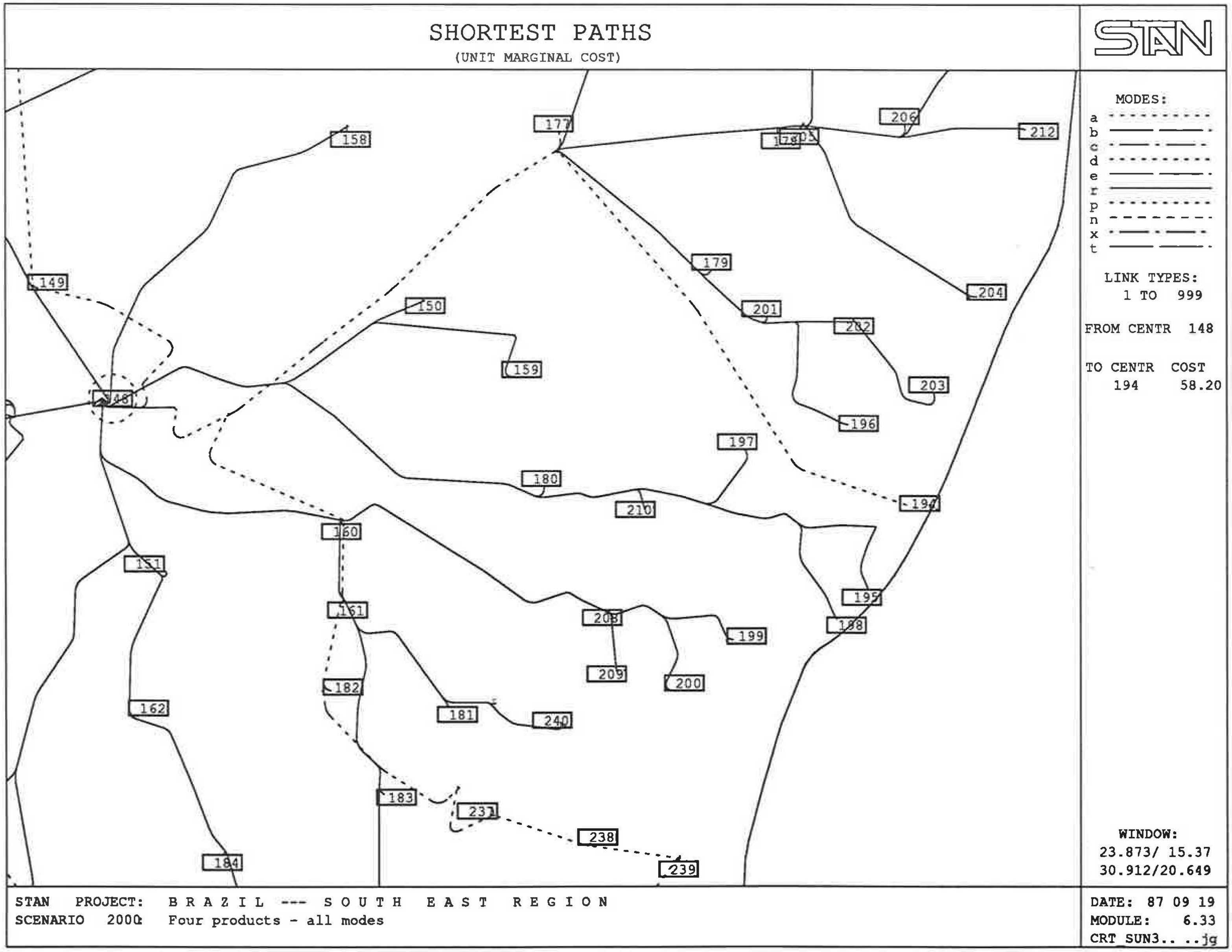


FIGURE 24 Shortest unit marginal cost routes.

REFERENCES

1. M. Florian and T. G. Crainic, eds. *Strategic Planning of Freight Transportation in Brazil: Methodology and Applications*, 5 vol. Reports 638–642. Centre de recherche sur les transports, Université de Montréal, Montréal, Canada, 1989.
2. T. G. Crainic, M. Florian, and J. E. Leal. A Model for the Strategic Planning of National Freight Transportation by Rail, *Transportation Science*, Vol. 24, No. 1, 1990, pp. 1–24.
3. T. G. Crainic. *An Interactive-Graphic System for the Analysis and Evaluation of Regional/National Transportation Networks*. Report 632. Centre de recherche sur les transports, Université de Montréal, Montréal, Canada, 1989.
4. H. Spiess. *Contributions à la théorie et aux outils de planification des réseaux de transport urbain*. Ph.D. thesis. Université de Montréal, Montréal, Canada, 1984.
5. A. Babin, M. Florian, L. James-Lefebvre, and H. Spiess. EMME/2: An Interactive Graphic Method for Road and Transit Planning. In *Transportation Research Record 866*, TRB, National Research Council, Washington, D.C., 1982, pp. 1–9.
6. J. Guélat, M. Florian, and T. G. Crainic. A Multimode Multiproduct Network Assignment Model for Strategic Planning of Freight Flows, *Transportation Science*, Vol. 24, No. 1, 1990, pp. 25–39.
7. INRO Consultants, Inc. *STAN User's Manual*. Montréal, Canada, 1991, 750 pp.
8. T. L. Friesz, J. A. Gottfried, and E. K. Morlok. A Sequential Shipper-Carrier Network Model for Predicting Freight Flows, *Transportation Science*, Vol. 20, 1986, pp. 80–91.
9. A. L. Kornhauser. *Examples of Corporate Distribution and Transportation Planning Using the Princeton Transportation Network Model and Graphic Information System*. Civil Engineering Department, Princeton University, Princeton, N.J., 1985.
10. A. L. Kornhauser. *Very Large Transportation Network Interactive Computer Graphic Information System and Problem Solver*. Civil Engineering Department, Princeton University, Princeton, N.J., 1985.

Publication of this paper sponsored by Committee on Transportation Supply Analysis.

TRUST: A LISP Program for the Analysis of Transit Route Configurations

M. HADI BAAJ AND HANI S. MAHMASSANI

TRUST is a program to analyze and evaluate a given set of bus transit routes and associated frequencies, in terms of several descriptors, including measures of user costs, service quality, and operator resources. The procedure assigns a known demand matrix to the transit network according to a path choice logic that explicitly considers transfers. As such, it calculates the percentages of the total demand trips that are able to reach their destination with no transfer, via one transfer, via two transfers, or simply cannot be satisified (with two or fewer transfers). Also computed are several node-level and route-level descriptors for use in the route network planning and design process. After the assignment is executed, the program determines the service frequency necessary on each route to maintain the passenger load factor below a specified maximum. The procedure can be used iteratively until the calculated frequencies are consistent with the input frequencies. TRUST is written in the LISP language because the latter's "list" data structure representation is particularly well suited to support the path search and enumeration activities inherent in the assignment logic and path choice rules appropriate in a transit network. The application of the program to the transit network of the Austin, Tex., urban area (with some simplifying assumptions) is presented, illustrating the program's capabilities and computational performance.

The purpose of this paper is to describe TRUST (Transit Routes Analyst), a LISP program developed to analyze a set of bus transit routes and associated service frequencies. It can be used either separately in a design-support function, or in the context of a formal solution procedure for the transit network design problem (TNDP). Such problems have been studied by several authors in the past (*1–8*). In the TNDP, one seeks to determine a configuration, consisting of a set of transit routes and associated frequencies, that achieves some desired objective, subject to the constraints of the problem. Mathematical formulations of the TNDP have been concerned primarily with the minimization of an overall cost measure, generally a combination of user costs and operator costs. The former is often captured by the total travel time incurred by users in the network, while a proxy for operator costs is the total number of buses required for a particular configuration. Feasibility constraints may include, but are not limited to: (1) minimum operating frequencies on all or selected routes (policy headways, where applicable); (2) a maximum load factor on any bus route; and (3) a maximum allowable bus fleet size.

Most existing formulations can be viewed as variants of the following mathematical program:

Department of Civil Engineering, University of Texas at Austin, Austin, Tex. 78712.

$$\text{Minimize} \left\{ c_1 \left[\sum_{j=1}^{n} \sum_{i=1}^{n} d_{ij} t_{ij} \right] + c_2 \left[\sum_{\text{all } k \in SR} f_k T_k \right] \right\} \quad (1)$$

subject to

$$\text{frequency feasibility } f_k \geq f_{\min} \quad \text{for all } k \in SR \quad (2)$$

$$\text{load factor constraint: } LF_k = \frac{(Q_k)_{\max}}{f_k CAP} \leq LF_{\max} \quad \text{for all } k \in SR \quad (3)$$

$$\text{fleet size constraint: } \sum_{\text{all } k \in SR} N_k = \left[\sum_{\text{all } k \in SR} f_k T_k \right] \leq W \quad \text{for all } k \in SR \quad (4)$$

where

d_{ij} = demand between nodes i and j,
t_{ij} = total travel time between i and j; $t_{ij} = t_{invtt,ij} + t_{wt,ij} + t_{tt,ij}$,
$t_{invttij}$ = in-vehicle travel time between nodes i and j,
$t_{wt,ij}$ = waiting time incurred while traveling between nodes i and j,
$t_{tt,ij}$ = transfer time (penalty per transfer) incurred on trip between nodes i and j,
N_k = number of buses operating on route k ($N_k = f_k T_k$);
f_k = frequency of buses operating on route k,
f_{min} = minimum frequency of buses operating on any route,
T_k = round trip time of route k,
W = fleet size available for operation on the route network,
LF_k = load factor of route k,
$(Q_k)_{max}$ = maximum flow occurring on any link of route k,
CAP = seating capacity of buses operating on the network's routes,
SR = set of transit routes, and
c_1, c_2 = weights reflecting the relative importance of the two cost components.

Note that by varying c_1 and c_2, one can generate different nondominated configurations that achieve a different tradeoff between user costs on one hand and operator costs (number of buses) on the other. In practice, other important tradeoffs need to be addressed in what is inherently a multiobjective problem. For example, should all demand be served, implying that some resources are allocated to low-density routes, or should service be more concentrated to provide high service levels on more productive routes, even if that means leaving some demand unmet? Similarly, a tradeoff needs to be made between directness of service (no transfers) and coverage. Such considerations have not typically been included in mathematical programming formulations, perhaps explaining the generally low degree of acceptance of such formulations in practice.

A central component in the solution of the network design problem is a procedure to evaluate the objective function components (in math programming formulations) and other measures of effectiveness or service quality that are of concern to the operator. This requires the assignment of the passenger trip demand matrix to the set of routes that define any particular network configuration. As indicated by Spiess and Florian (*9*), the transit assignment problem has been studied by several authors in the past, either as a separate problem (*10,11*) or as a subproblem of more complex models, such as transit network design (*1,4,7*) or multimodal network equilibrium (*12*).

Most transit assignment algorithms are variants of procedures used for private car traffic on road networks (such as shortest path, or stochastic multipath assignment), which are modified to reflect the waiting time phenomenon inherent to transit networks. However, path choice in a transit network, especially in large urban areas with overlapping routes, may not be well described by the assumptions of auto driver assignment. This was recognized by the transit route choice behavior model presented by Han and Wilson (*13*), which allowed for a lexicographic strategy in the choice among competing routes, with transfer avoidance and/or minimization acting as the primary choice criterion. This feature of their model is adopted in TRUST.

The transit route choice model at the core of TRUST considers two main criteria: the number of transfers necessary to reach the trip's destination, and the trip times incurred on different alternative choices. The tripmaker is assumed to always attempt to reach his/her destination by following the path that involves the fewest possible number of transfers. In case of a tie (i.e., when there is more than one path with the same least number of transfers), a decision is made after considering the competing paths' trip times. When there is one or more alternatives whose trip time is within a threshold of the minimum trip time, a "frequency share" rule is applied, as explained below.

The next section presents an overview of TRUST explaining what it does, how the input data is represented, the rationale for using LISP for this purpose, and the unique features and advantages that LISP offers relative to more conventional languages for this particular problem. Subsequent sections describe in detail the transit network representation and input data, the traveler's route choice model (which is at the core of the assignment procedures), and the procedures for computing the different network descriptors and performance measures, as well as other output information. An example case demonstrates the results of computer runs of TRUST on the Austin, Tex., transit network (with some simplifying assumptions on the demand side). The final section presents some concluding remarks and future directions for research and improvement.

OVERVIEW OF TRUST

TRUST is intended as an analytic tool that can support the design of transit routes: the transit planner generates a given set of routes (either manually or via a computerized route generation package) and assigns frequencies to these routes (based on either policy headways, past experience, or a formal frequency allocation procedure). TRUST then computes a variety of network descriptors and figures of merit, including the components of the objective function given earlier. In addition to the usual cost items, the program is notable for generating measures of service quality from the user's point of view, particularly with regard to transfers. The following five types of information are calculated:

1. The total travel time experienced by users in the network, and the respective percentages of in-vehicle travel time, waiting time, and transfer time (the latter reflecting a prespecified time penalty considered to be equivalent to a transfer);
2. The total number of demand trips, as well as the percentages of demand that are unsatisfied, or satisfied with 0, 1, or 2 transfers;
3. The number of trips originating at each node that could not be assigned, as well as the number of passenger trips that transfer at each node;
4. The link flows on each route's links and the route's maximum load factor; and
5. The frequency and number of buses required on each route to maintain its load factor under LF_{max} and the resulting number of buses required for the whole network.

If, for a given analysis run of a particular configuration, the output frequencies (i.e., the frequencies required to maintain all routes' load factors under a prespecified LF_{max}) are considered quite different from the input frequencies (say they differ by more than 5 percent), then the planner may reiterate the process by using the output frequencies of the previous run as the new input frequencies. For the example case of Austin's transit network presented below, the difference between input frequencies and output frequencies after two iterations was more than 10 percent for 9 out of 36 routes, between 5 percent and 10 percent for 3 other routes, and under 5 percent for the remaining 24 routes. After three iterations, this difference was still more than 10 percent for only 2 out of the 36 routes, between 5 percent and 10 percent in the case of 5 other routes, and under 5 percent in the case of the remaining 29 routes. The network descriptors recomputed in each new iteration differed from those of previous runs by small percentages, usually under 5 percent.

TRUST's input data may be grouped under four categories:

1. Network—the number of bus transit nodes, the names of routes operating in the network as well as the list of nodes defining each route, and a connectivity list specifying for each node its accessible neighboring nodes as well as the in-vehicle

travel time (i.e. over the road network) to each of the neighboring nodes;

2. Frequencies—the frequencies of bus service on each route;
3. Demand—a symmetric demand matrix representing the number of passenger trips between each pair of nodes (the symmetry requirement is not essential, but has been used so far for convenience); and
4. Design parameters—the transfer time (penalty) per transfer expressed in equivalent minutes of in-vehicle travel time, the bus seating capacity (assumed the same on all buses), and the maximum load factor allowed by the planner on any transit route.

TRUST is written in LISP (List Processing), a fifth generation computer language (*14*). The principal motivation for using LISP (or, more generally, a fifth generation language) lies in the nature of the computational activity taking place in TRUST, which consists of searching and screening paths in a graph. It has been common wisdom in transportation network applications to avoid any form of path enumeration. Thus, most existing assignment procedures are limited to shortest path constructs. However, other programming paradigms and advances in computing hardware and software can greatly facilitate some degree of path search and enumeration, which is justified by the added realism that it could allow into the resulting procedure. The present program is an attempt to explore these possibilities in the transit network analysis area.

LISP offers advantages over "conventional" languages such as FORTRAN, C, or Pascal both in terms of representation and search. The transit network data representation lends itself conveniently to the 'list' data structure representation of LISP, which in turn supports the kind of path search strategies of interest in this application. This can be illustrated by the following:

- The network connectivity can be conveniently represented in a descriptive language such as LISP: to each network node, one associates a set (or, in LISP, a list) of neighboring nodes as well as the trip time (cost) associated with the nodes. Thus, the list (2 ((1 11.4)(3 2.9)(6 8.0))) indicates that one can travel from node 2 to node 1 in 11.4 minutes, to node 3 in 2.9 minutes, and to node 6 in 8 minutes.
- A route can be represented as a list of nodes, thus route r25 is defined by the list of nodes (18 11 10 9 8 12 14).
- The search techniques that are specific to the transit network design problem can be readily programmed in LISP. In such techniques, a feasible path connecting two network nodes can be easily represented as a list. Thus, the list ((r1 9 16)(r8 16 21)) implies that one can travel from node 9 to node 21 by boarding route r1 from node 9 to node 16 and route r8 from node 16 to node 21 (i.e., node 16 is a transfer node).

Taylor (15) describes simple programs written in Prolog, another fifth generation language, to solve different route selection problems. His examples underscore the brevity of code as well as the relative ease of programming with fifth generation languages. At the basis of these programs are some general "predicates" (Prolog meta-statements) that test for set membership or append a new element to a set. Such meta-statements define the necessary condition for the required solution, thus isolating the programmer from worring about the elemental computing and house-keeping chores, as would be the case with conventional programming languages such as FORTRAN, Pascal, and C.

On the negative side, LISP, like most higher-level languages, may experience relatively slow computational performance when it comes to mathematical computations (as opposed to symbolic manipulations). However, tests with the program to date have shown reasonable execution times. Furthermore, when TRUST is eventually incorporated into a design procedure, it is envisioned that activities that are most efficiently and effectively handled by conventional languages (such as bookkeeping) would be programmed as such, while items such as the experts' design rules or path search strategies would be conveniently expressed in a symbolic language like LISP. Such use of multiple languages communicating in the execution of a particular program is an increasingly appealing approach to combine the advantages of artificial intelligence tools and standard scientific computing for the development of effective design procedures for engineering problems.

TRANSIT NETWORK DATA REPRESENTATION AND INPUT DATA

Nodes are defined to represent the demand points for transit trips in the network under study. A list of these nodes, called *NODE-LIST* (whose members are natural integers starting with 0 and ending with a number equal to 1 less than the number of nodes), is maintained internally. For each node (call it parent node), we specify the set of adjacent or neighboring nodes and the in-vehicle travel time in minutes between each neighboring node (call it child node) and the parent node. In list form, (2 ((1 11.4) (3 2.9)(6 8.0))) is the sublist associated with node 2. It indicates that one can travel between node 2 and nodes 1, 3, and 6 in 11.4, 2.9, and 8.0 minutes, respectively. The overall list (of all sublists) for all nodes is referred to as *CONNECTIVITY-LIST*.

Passenger trip demand is represented in the usual way as an origin-destination trip matrix, called *DEMAND-MATRIX*, containing the number of trips from each node to all others. In the present version, the matrix is assumed to be symmetric for convenience and with no loss of generality. A transit route is represented as a sequence of nodes, which are input by the user for each route, thereby forming the *ROUTE-LIST*. For example, a route called r25 would be represented as (r25 (18 11 10 9 8 12 14)).

In addition to the name, four properties are associated with each route: "list-of-nodes" (defined earlier), frequency, "round-trip-time", and "list-of-link-flows." The first two properties and their values are obtained from the input data. The frequency of service is assigned by the user for each route (an interactive capability in the program queries the user for this information). However, as explained below, the program has the capability to also set frequencies, in which case the entered values serve as initial values. The third and fourth properties are determined internally by the program.

The third property is assigned (via a call to the procedure "round-trip-times-of-routes") using the information contained in the above-mentioned *CONNECTIVITY-LIST*. The fourth property corresponds to one of the main objectives of the program and is determined as a result of the assignment procedure. Link flow values are initialized to zero (via a call to the procedure "create"). Thus for route r25, the "create" procedure assembles the list ((1-18-11 0.0) (1-11-10 0.0) (1-10-9 0.0) (1-9-8 0.0) (1-8-12 0.0)(1-12-14 0.0)). This list indicates that initially, before the assignment procedures are called, the flow on link 1-18-11 (which represents the link connecting nodes 18 and 11 of route r25) is 0.0. Now consider another route, r44, with the "list-of-link-flows" ((1-1-2 0.0)(1-2-6 0.0) (1-6-12 0.0)(1-12-14 0.0)). Both routes r25 and r44 utilize the same physical link joining nodes 12 and 14, but for the purpose of assignment and flow specification, (1-12-14) associated with route r25 is different from (1-12-14) associated with r44, as should be the case.

In addition to the above network and system descriptors, the user specifies values for the following parameters: *TRANSFER-PENALTY* (in equivalent minutes of network in-vehicle travel time), *BUS-SEATING-CAPACITY*, and *MAX-LOAD-FACTOR*.

THE ASSIGNMENT MODEL AND COMPUTATION OF TRANSIT NETWORK DESCRIPTORS

The assignment model is the core of the analysis program because it determines the passenger flows on each link, which are used to calculate the various costs and performance descriptors. The assignment process consists of a direct application of the underlying transit path choice logic. As noted earlier, we have adapted the lexicographic decision structure of Han and Wilson (*13*) because of its behavior realism and plausibility. The main feature that we have adopted from their model is the consideration of the number of transfers as the most important criterion, and one which exercises preemptive priority over other considerations. Thus the procedure first starts by identifying all transit paths between i and j that do not involve any tranfers. Only if none are found will paths involving one (and only one tranfer) be considered. If more than one zero-transfer path is available, then additional criteria come into play. It is in the details of this aspect that our procedure differs from Han and Wilson. When more than one path exists with the same number of (or no) transfers, the user's effective choice set is assumed to contain only those paths with respective travel times within a particular range. Trips are then assigned to these paths using an allocation formula reflecting the relative frequencies of service on the alternative paths. The details of this procedure are described below.

The assignment process considers each node pair separately. For a given node pair (i, j), the set of routes passing through node i, and those passing through node j (denoted by SR1 and SR2, respectively), are assembled by calling the "routes-passing-by" procedure for both nodes. If either SR1 or SR2 is empty, then at least one of these two nodes is not served by any transit route, and the demand d_{ij} cannot be assigned. In this case, the list ((i j) d_{ij}) is appended to the list *UNSATISFIED-DEMAND-LIST*, which was initially empty. If both SR1 and SR2 are not empty, then a call is made to procedure "assign-0-transfer?" that checks whether the demand can be assigned directly (without transfers). This is possible only if the intersection of SR1 and SR2, which is the subset of all routes that have both nodes i and j on their "list-of-nodes," is not empty. When this is the case, the intersection set is passed to the procedure "decide-0," whose function is to distribute the demand d_{ij} among the acceptable routes.

The procedure "decide-0" finds the route with the minimum in-vehicle travel time, then invokes a filtering process. Any route with in-vehicle travel time exceeding the minimum value by a specified threshold (say 50 percent, as selected in our implementation) is rejected. The demand d_{ij} is allocated to the routes surviving the filtering process using a simple "frequency-share" rule: a route carries a proportion of the flow equal to the ratio of its frequency to the sum of the frequencies of all acceptable routes. Thus, if there are three acceptable routes (after filtering) R1, R2, and R3 whose frequencies are f_1, f_2, and f_3 (in buses per hour), respectively, then R1 carries $\{[f_1/(f_1 + f_2 + f_3)] \times d_{ij}\}$ on all links between nodes i and j of this route. Such flow incurs an average waiting time (in minutes) of $\{60.0/[2(f_1 + f_2 + f_3)]\}$. Note that the corresponding flows on R2 and R3 also incur the same average waiting time. This rule is based on the assumption that uniformly arriving tripmakers board the first bus to arrive at their stop; the waiting time calculation further ignores stochasticity in bus headways.

If the "assign-0-transfer?" procedure is unable to assign the demand between i and j directly (with no transfers), a call to procedure "assign-1-transfer?" is made. The latter checks whether the trip can be completed with one transfer. This check is carried out by examining, for every possible combination of a route number of SR1 (i.e., that passes through node i, say R1) and another of SR2 (say R2), the intersection set of the "list-of-nodes" of both R1 and R2. If the intersection set is not empty, then its contents are possible transfer nodes between R1 and R2. For example, if the intersection set is (tf1 tf2), TRUST forms two possible paths for the assignment of demand between nodes i and j: ((R1 i tf1)(R2 tf1 j)) and ((R1 i tf2)(R2 tf2 j)). The first signifies that a possible path from i to j consists of boarding route R1's bus at i, and staying on it until node tf1 (transfer node 1), where the passenger should transfer to route R2 and travel on it until the destination j is attained. For each possible path involving one transfer, an estimate of the total travel time is calculated. Thus, for the first path above, the total travel time is computed as

$$t_{ij} = t_{invt,i\ tf1|R1} + t_{invt,tf1\ j|R2} + [60/(2f_1)] + [60/(2f_2)] + \text{*TRANSFER-PENALTY*} \quad (5)$$

where the third and fourth components are the average waiting times at nodes i and $tf1$ respectively. After all such paths between i and j are found, and the associated trip times calculated, a filtering process similar to the zero-transfer case is applied. All paths between i and j whose total travel time exceeds the minimum value offered by any path between that pair of nodes by more than a specified threshold (say 10 percent as selected here) are rejected.

Procedure "decide-1" subsequently distributes d_{ij} among the paths that have passed the filtering process. "Decide-1" utilizes the above-described "frequency-share" rule, except that it is now applied to classes of paths rather than to individual paths. Such modification is necessary in both "decide-1" and "decide-2." Paths that share the same starting route (say Rm) at the boarding node *i*, form one class of paths (say Cm). Demand is first allocated among alternative classes; within each class, demand is shared equally among the constituting paths. The fraction of the demand that the whole class carries is equal to the ratio of f_m (the frequency of the starting route Rm defining the class of paths Cm) to the summation of the frequencies of all the starting routes that define all the existing classes of paths.

To illustrate this process, Figure 1 shows five acceptable paths between nodes *i* and *j* that have survived the filtering process applied by "decide-1." Procedure "classify-paths" indicates that there are two classes of paths, C1 and C4, corresponding to routes R1 and R4, respectively, which can be boarded at node *i*. The class C1 contains three possible paths: ((R1 *i* tf3)(R3 tf3 *j*)), ((R1 *i* tf1)(R2 tf1 *j*)), and ((R1 *i* tf2)(R2 tf2 *j*)). The latter two indicate that the traveller can transfer to route R2 either at tf1 or tf2 since the link joining tf1 and tf2 is common to both routes R1 and R2. The flow allocated to C1 is $\{[f_1/(f_1 + f_4)] \times d_{ij}\}$, and a third of this flow is assigned to each of C1's three paths. The average waiting time incurred by all riders at node *i* reflects the frequencies of R1 and R4, and is given by $\{60/[2\ (f_1 + f_4)]\}$. An additional waiting time is incurred at the transfer node; it is equal to $[60/(2f_3)]$ for C1's first path and $[60/(2f_2)]$ for its second and third paths. Similarly, C4 in Figure 1 contains two different paths: ((R4 i tf4)(R5 tf4 j)) and ((R4 i tf5)(R6 tf5 j)). The flow allocated to C4 is $\{[f_4/(f_1 + f_4)] \times d_{ij}\}$, with half of it assigned to each of C4's two paths.

If it is determined that d_{ij} cannot be assigned with at most one transfer procedure "assign-2-transfer?" is called to search for paths involving two transfers. If no such path is found, the demand between *i* and *j* will remain unsatisfied. In other words, it is assumed that a passenger will simply not consider boarding the transit buses to accomplish a trip that requires three or more transfers. Hence, TRUST avoids searching for paths that reach destination *j* with three or more transfers, thereby avoiding an otherwise considerable amount of meaningless search and keeping the execution time within tolerable limits.

The "assign-2-transfer?" procedure searches for all paths with exactly two transfers between given nodes *i* and *j*. The search consists of finding a route that passes through neither *i* nor *j*, but which shares a node with a route passing through *i* (i.e., with a member of SR1) and another through node *j* (i.e., with a member of SR2). The set SR3 of routes that pass through neither node *i* nor node *j* is obtained as the complement (in *SR*, the set of all routes) of the union of the previously generated SR1 and SR2, (SR1 U SR2). For a trip to require exactly two transfers between origin *i* and destination *j*, the first route, R1, has to pass by node *i* (hence, R1 ∈ SR1); the second route, R3, has to pass by some node other than node *i* or node *j* (hence, R3 ∈ SR3); and the third route, R2, has to pass by node *j* (hence, R2 ∈ SR2). Thus, three "do" loops are executed: the outer one on SR3, the inner one on SR1, and the innermost one on SR2. A route from each of the above three sets is selected and the following test is performed: if the "list-of-nodes" of the route from SR3 (say R3) intersects both the "list-of-nodes" of the route from SR1 (say R1), and the "list-of-nodes" of the route from SR2 (say R2), then a possible two-transfer path is defined. For example, if the intersection set of the "list-of-nodes" of R3 and R1 is (tf1 tf2) and that of the "list-of-nodes" of R3 and

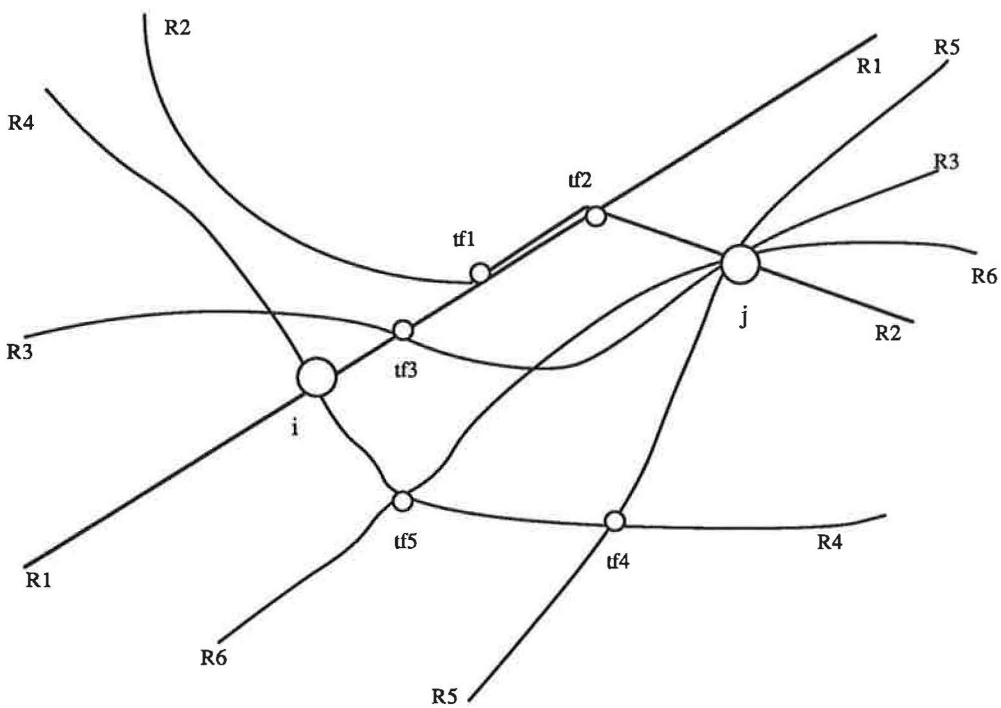


FIGURE 1 Five acceptable paths between nodes *i* and *j* involving one transfer.

R2 is (tf3), then TRUST defines two possible paths: ((R1 i tf1)(R3 tf1 tf3)(R2 tf3 j)) and ((R1 i tf2)(R3 tf2 tf3) (R2 tf3 j)).

After all possible combinations of triplets (R3 ∈ SR3, R1 ∈ SR1, R2 ∈ SR2) are checked and listed, the allocation procedure "decide-2" is called. Its logic is similar to that of the previously described "decide-1," in that it filters out paths with travel time exceeding the minimum possible total travel time by more than the prespecified threshold (10 percent here). The average total travel time of the above path is calculated as

$$t_{ij} = t_{invtt,i\ tf1|R1} + t_{invtt,tf1\ tf3|R3} + t_{invtt,tf3\ j|R2} + [60/(2f_1)] + [60/(2f_2)] + [60/(2f_3)] + [2\ (\text{*TRANSFER-PENALTY*})] \quad (6)$$

Following the filtering process, the remaining paths are grouped into different path classes, such that members of the same class share the same first route as the trip origin i. The demand d_{ij} is then assigned to the alternative paths in a manner similar to the previously described one-transfer case. Note that while "decide-2" could be rather time consuming computationally, it is usually executed a smaller number of times than "decide-0" or "decide-1," especially if the transit network allows the completion of most trips directly or via one transfer.

In summary, "allocate-demand" checks each node pair (one element of the demand matrix) to determine whether its trip can be assigned with zero, one, or two transfers. It appends the list ((i j) d_{ij}) to one of four possible demand lists: *UNSATISFIED-DEMAND-LIST*, *DEMAND-0-TRANSFER*, *DEMAND-1-TRANSFER*, and *DEMAND-2-TRANSFERS*. It also computes the contributions of this assignment to the networkwide measures of user costs: *NETWORK-INVEHICLE-TRAVEL-TIME*, *NETWORK-WAITING-TIME*, and *NETWORK-TRANSFER-TIME*. Once all elements of the demand matrix are assigned, the *NETWORK-TOTAL-TRAVEL-TIME* is calculated as the sum of the final values of the above three components. The program also considers each of the four demand lists above to compute the percentages of the total number of passenger trips that are unsatisfied, or satisfied via zero, one, or two transfers. The flow diagram of the assignment process is shown in Figure 2. The program output includes each of the four demand lists, so that one can identify how the demand associated with a selected pair of nodes was assigned, as well as the final "list-of-link-flows" associated with each transit route. This information provides the principal measures of service quality and user costs that are of interest in the evaluation of a particular transit route network configuration. The other type of measures consists of the resources required by the operator, primarily the number of buses necessary to serve that particular configuration. The next section discusses this aspect, as well as the use of the program in an iterative approach to calculate service frequencies consistent with a prespecified load factor constraint.

COMPUTATION OF ROUTE FREQUENCIES AND NUMBER OF BUSES

Once all elements of the demand matrix have been assigned, the final "list-of-link-flows" of each route k is used to locate the route's link carrying the highest volume, $(Q_k)_{max}$. The program then computes the number of buses required on route k to achieve the maximum allowed load factor:

$$(N_k)_{required} = \frac{(Q_k)_{max} \times T_k}{60.0 \times \text{*MAX-LOAD-FACTOR*} \times \text{*BUS-SEATING-CAPACITY*}} \quad \text{for all } k \in SR \quad (7)$$

The bus frequency required to maintain the route's load factor under LF_{max} is computed as:

$$(f_k)_{required} = \frac{(N_k)_{required} \times 60.0}{T_k} \quad \text{for all } k \in SR, \quad (8)$$

where all terms are as previously defined.

The above tasks are accomplished by a call to procedure "min-buses." The minimum fleet size necessary for the whole network, $(W)_{required}$, is also included in the program's output and is computed as:

$$(W)_{required} = \sum_{\text{all } k \in SR} (N_k)_{required} \quad \text{for all } k \in SR. \quad (9)$$

Note that the fleet size produced is a minimum value required to maintain set frequencies on the system's routes. It does not make allowances for reliability issues, which need to be taken into account separately by the operator, based on local experience and/or operating policies.

The calculation of route frequencies so as to achieve a preset peak load factor is a commonly used rule in the transit industry, especially in connection with congested networks. Of course, in order for it to be meaningful, the demand assignment must be performed over the peak period. Generally, one would expect to exercise TRUST for different time-of-day periods and separately for average weekdays and weekends. For less congested periods, the peak load factor method may come up with frequencies that are lower than what can be reasonably expected by riders. Minimum policy headways would then be used instead. From a route design standpoint, the number of required buses will be generally governed by the peak period. Because the focus of TRUST is on route design, its frequency-setting logic is limited to diagnosing a particular set of routes and frequencies against a maximum load factor constraint, so as to determine the minimum number of buses needed to achieve this constraint. It does not presently possess the capability to allocate frequencies given a fleet size constraint. For a comprehensive discussion of frequency allocation issues, the reader is referred to the review and model presented by Furth and Wilson (*16*).

As explained, the service frequencies computed are those that are consistent with the preset load factor. However, these frequencies are calculated after demand assignment on the basis of an initial set of frequencies. These two sets of frequency values may not be identical. The direction of the discrepancy indicates situations of higher or lower service than necessary. The program provides the capability to set frequencies in an iterative process where the last set of frequency values serves as the basis for the next assignment. While there is no guarantee of convergence, experience to date has indicated a converging pattern, as shown in the next case example.

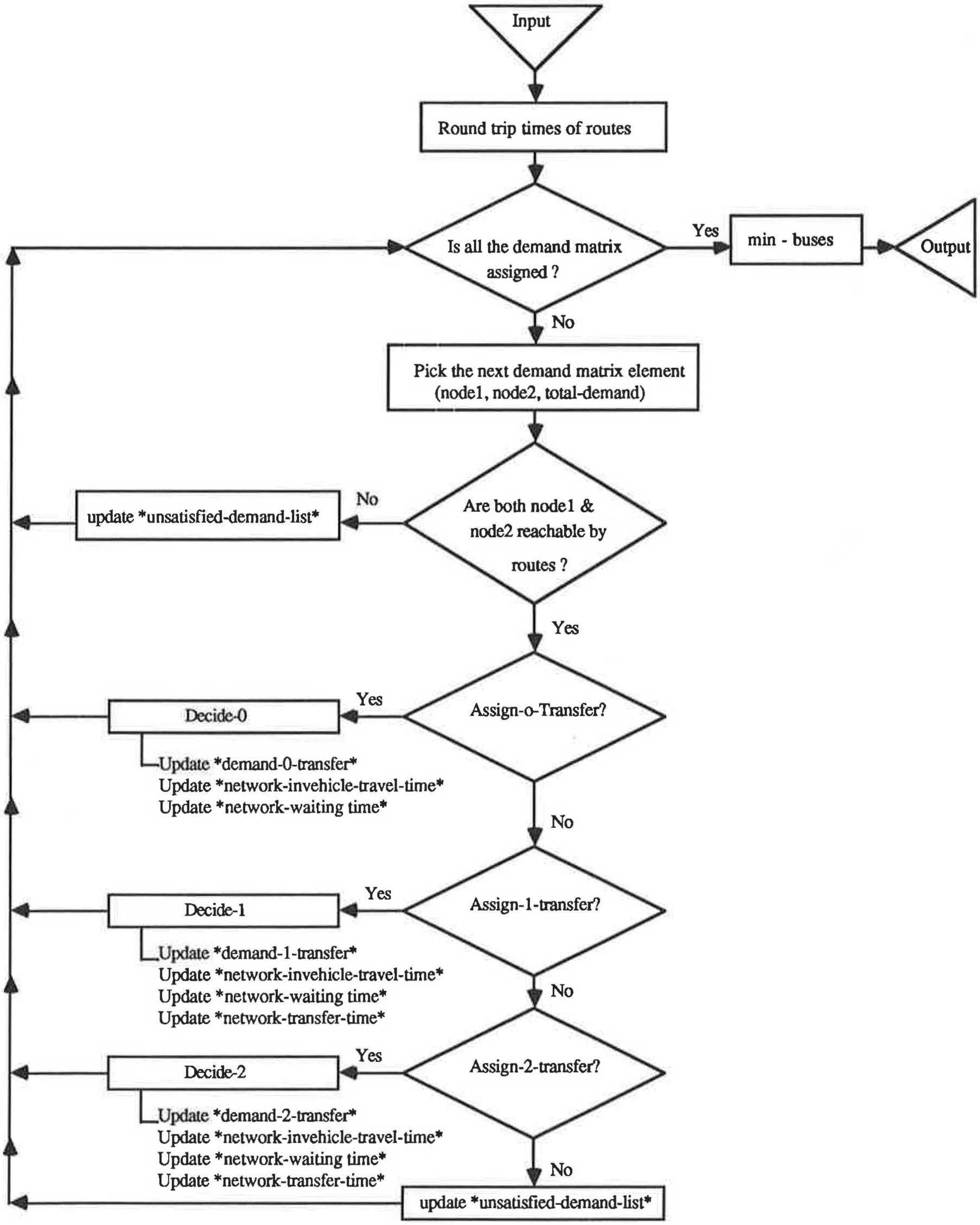


FIGURE 2 Flow diagram of TRUST.

APPLICATION TO AUSTIN NETWORK

The transit network of the Austin, Tex., urban area (metropolitan population approximately 500,000) was selected to illustrate TRUST and test its computational performance on a network of realistic size. Some simplifications were made in the process, due in part to data limitations, though they do not detract from the illustrative purpose of the application. The transit network consists of 36 routes with fixed schedules, operated by the Capital Metropolitan Transit Authority (Capital Metro, for short). We defined 140 nodes to describe the service area and associated network connectivity (the nodes defined include all intersection points of two or more routes). A trip demand matrix was not available in the desired format. In order to demonstrate the performance of TRUST, a symmetrical demand matrix was randomly generated. Each element of the matrix is a natural integer independently selected according to a discrete and linearly decreasing probability mass function between 0 and 4, inclusive. Thus, the number of trips between two nodes (a given demand matrix element) equals either 0, 1, 2, 3, or 4, with corresponding probabilities of 5/15, 4/15, 3/15, 2/15, and 1/15. For all runs, the penalty for a transfer was selected as 5 minutes of equivalent in-vehicle travel time, the bus seating capacity was taken as 40 passengers, and the maximum load factor was chosen to be 1.25 (i.e., up to 10 standing passengers are allowed at any time).

Figure 3 shows the contents of file "data.text," which lists the names of the network's routes and the node composition of each route. Figure 4 shows the different output quantities generated after one run (iteration) of the program. These include the total network aggregate values; thus, under the above assumptions, and for the fictitious demand matrix generated, only 9.6 percent can be served with no transfer, 69.1 percent require one transfer, 17.7 percent two transfers, and the rest cannot be served with less than three transfers. Also included in Figure 4 are results for two representative routes (R1 and R2) and nodes (0 and 69). Table 1 summarizes selected results of three successive iterations. For the first run, all input bus frequencies were arbitrarily set at 10 buses per hour; in subsequent runs, the input frequencies were selected as the output (required) frequencies of the corresponding previous run, as explained previously. It can be noted that, after two iterations, the difference between input frequences and output frequencies was more than 10 percent for 9 out of the 36 routes, between 5 percent and 10 percent for 3 other routes, and under 5 percent for the remaining 24 routes. After three runs, the difference between input frequencies and output frequencies was more than 10 percent for only 2 of 36 routes, between 5 percent and 10 percent for 5 other routes, and under 5 percent for the remaining 29 routes, revealing a generally converging pattern.

From the computational standpoint, all runs were executed on an Apple MAC-II workstation equipped with a special-purpose TI Explorer microchip that runs the LISP language compiler. Each run took about 14 minutes of real execution time, which is quite satisfactory given the extent of path search and enumeration required to implement the assignment logic and calculate the desired performance measures. This response time is certainly within tolerable limits for such applications.

```
r1  (134 9 16 26 43 48 139 60 69)
r2  (73 72 71 69)
r3  (14 19 31 47 139 59 69)
r4  (82 81 137 80 75 74 69)
r5  (14 15 20 31 43 49 60 69)
r6  (67 68 69)
r7  (4 11 10 17 16 21 25 61 69)
r8  (14 15 16 21 25 22 23 24 32 39 40 53 67 72 75 79 90)
r9  (27 28 30 45 44 57 69)
r10 (124 125 103 87 69)
r11 (51 50 41 42 131 136 64 132 63 62 61 69)
r12 (128 122 123 129 101 97 96 95 94 85 69)
r13 (115 116 112 104 105 93 133 69)
r14 (109 108 91 106 92 93 133 69)
r15 (25 41 50 62 66 69)
r16 (117 129 102 107 86 87 69)
r17 (77 137 76 75 79 78 69)
r18 (55 54 53 65 66 69)
r19 (27 28 29 30 46 58 59 69)
r20 (35 34 38 40 52 64 132 63 62 69)
r23 (13 14)
r25 (18 11 10 9 8 12 14)
r26 (89 82 90 88 69)
r27 (115 114 113 111 109 91 90 88 133 69)
r28 (99 100 101 102 103 104 105 109)
r29 (98 84 85 86 87 69)
r30 (99 130 83 84 85 86 87 133 69)
r33 (118 119 120 121 122 123 124 125 116 115 111 108 110)
r37 (37 36 35 34 33 32 42 131 51 63 66 69)
r38 (127 121 117 100 97 96 94 85 69)
r39 (138 18 24 23 22 25)
r40 (2 3 135 8 9)
r42 (0 3 5 8 9 16)
r44 (1 2 6 12 14)
r45 (126 125 116 115)
r46 (7 14)
EOF
```

FIGURE 3 Contents of the data file "Data.Text" for Austin transit network application.

CONCLUSIONS

TRUST is a program for the analysis and evaluation of alternative transit route network configurations consisting of a set of routes and associated frequencies. Its main function is to assign known demands between origin-destination pairs to the transit network, and compute a variety of performance measures reflecting the quality of service and costs experienced by the users, as well as the resources required by the operator. The program differs from existing assignment approaches in several respects, including: (1) the path choice mechanism that is the basis of the assignment procedure; (2) the use of LISP, a so-called fifth generation language associated with artificial intelligence (AI) applications, which greatly facilitates the implementation of the path search and enumeration inherent in the type of assignment procedure adopted; and (3) the broader range of performance measures and descriptors that are computed and displayed, particularly on the demand side. TRUST may be utilized in at least two different ways:

1. As a computerized tool for the analysis of a given set of transit routes, which might be examined by the transit planner in the process of short-range transit route design (modifications and service changes aimed at improving an existing transit network) or long-range transit system planning (whereby a whole new transit network is designed from scratch); and
2. As a tool for effective sensitivity analysis, whereby transit network descriptors are evaluated when variations in route configuration, route frequency, bus seating capacity, transfer penalty, and maximum allowable route load factors are considered jointly or separately. [In this regard, it can be noted that procedures "decide-0," "decide-1," and "decide-2" can easily be changed so as to model different travellers' route choice behavior mechanisms; furthermore, the various thresholds used, such as the 50 percent over the minimum in-vehicle travel time (as assumed by "decide-0") and 10 percent over the minimum total travel time (assumed in both "decide-1" and "decide-2") can be easily modified to better reflect local conditions.]

In addition to the substantive motivation in terms of providing effective procedures for transit planning and design problems, an additional motivation, from a methodological standpoint, has been to explore the possibilities of LISP and

(NETWORK TOTAL DEMAND = 25564.0 TRIPS)
(DEMAND SATISFIED WITH 0 TRANSFER = 2448.0 TRIPS) (FRACTION IS 0.09575966)
(DEMAND SATISFIED WITH 1 TRANSFER = 17664.0 TRIPS) (FRACTION IS 0.6909717)
(DEMAND SATISFIED WITH 2 TRANSFERS = 4528.0 TRIPS) (FRACTION IS 0.17712408)
(UNSATISFIED DEMAND = 924.0 TRIPS) (FRACTION IS 0.036144577)

(NETWORK TOTAL TRAVEL TIME = 1303556.7)
(NETWORK INVEHICLE TRAVEL TIME = 1025619.9) (FRACTION IS 0.7867857)
(NETWORK WAITING TIME = 144336.94) (FRACTION IS 0.11072547)
(NETWORK TRANSFER TIME = 133600.0) (FRACTION IS 0.10248882)

(FREQUENCY AVAILABLE ON ROUTE R1 IS 10)
(LOAD FACTOR ON ROUTE R1 IS 2.7416472)
(MINIMUM FREQUENCY REQUIRED ON ROUTE R1 TO KEEP LOAD FACTOR UNDER 1.25 IS 21.93318)
(NUMBER OF BUSES AVAILABLE ON ROUTE R1 IS 17.866667)
(MINIMUM NUMBER OF BUSES REQUIRED ON ROUTE R1 TO KEEP LOAD FACTOR UNDER 1.25 IS 39.18728)
(LINK FLOWS ON ROUTE R1 ARE ((L-134-9 182.99997) (L-9-16 670.4245) (L-16-26 848.2194) (L-26-43 925.2193) (L-43-48 962.66473) (L-48-139 1061.665) (L-139-60 1072.2091) (L-60-69 1096.6589)))

(FREQUENCY AVAILABLE ON ROUTE R2 IS 10)
(LOAD FACTOR ON ROUTE R2 IS 0.9984379)
(MINIMUM FREQUENCY REQUIRED ON ROUTE R2 TO KEEP LOAD FACTOR UNDER 1.25 IS 7.9875026)
(NUMBER OF BUSES AVAILABLE ON ROUTE R2 IS 6.8)
(MINIMUM NUMBER OF BUSES REQUIRED ON ROUTE R2 TO KEEP LOAD FACTOR UNDER 1.25 IS 5.431502)
(LINK FLOWS ON ROUTE R2 ARE ((L-73-72 172.0) (L-72-71 247.37497) (L-71-69 399.37515)))

(NUMBER OF BUSES AVAILABLE FOR THE SET OF ROUTES (R1 R2 R3 R4 R5 R6 R7 R8 R9 R10 R11 R12 R13 R14 R15 R16 R17 R18 R19 R20 R23 R25 R26 R27 R28 R29 R30 R33 R37 R38 R39 R40 R42 R44 R45 R46) IS 413.8667)

(MINIMUM NUMBER OF BUSES REQUIRED FOR THE SET OF ROUTES (R1 R2 R3 R4 R5 R6 R7 R8 R9 R10 R11 R12 R13 R14 R15 R16 R17 R18 R19 R20 R23 R25 R26 R27 R28 R29 R30 R33 R37 R38 R39 R40 R42 R44 R45 R46) IS 529.17017)

(160.0 TRIPS ORIGINATING-FROM-AND-ASSIGNED NODE-0)
(% OF TOTAL NETWORK DEMAND = 0.6258802)
(13.0 TRIPS ORIGINATING-FROM-AND-UNASSIGNED NODE-0)
(% OF TOTAL NETWORK DEMAND = 0.05085276)
(0.0 TRIPS TRANSFERING-AT NODE-0)
(% OF TOTAL NETWORK DEMAND = 0.0)

(92.0 TRIPS ORIGINATING-FROM-AND-ASSIGNED NODE-69)
(% OF TOTAL NETWORK DEMAND = 0.35988107)
(1.0 TRIPS ORIGINATING-FROM-AND-UNASSIGNED NODE-69)
(% OF TOTAL NETWORK DEMAND = 0.003911751)
(7988.1133 TRIPS TRANSFERING-AT NODE-69)
(% OF TOTAL NETWORK DEMAND = 31.247509)

FIGURE 4 Results of the first run of TRUST for two representative nodes and routes.

TABLE 1 PARTIAL RESULTS OF THREE ITERATIONS OF TRUST

	First Run			Second Run			Third Run		
route name	input freq.	output freq.	%change	input freq.	output freq.	%change	input freq.	output freq.	%change
r1	10	21.93	119.3	21.93	22.1	0.78	22.1	21.9	- 0.91
r2	10	7.99	- 20.1	7.99	7.96	- 0.38	7.96	7.93	- 0.38
r3	10	16.74	67.4	16.74	17.13	2.33	17.13	17.96	0.18
r4	10	13.54	35.4	13.54	13.93	2.88	13.93	13.96	0.22
r5	10	16.23	62.3	16.23	15.96	- 1.66	15.96	16.09	0.81
r6	10	5.21	- 47.9	5.21	5.03	- 3.45	5.03	5.02	0.20
r7	10	20.03	100.3	20.03	19.91	- 0.60	19.91	20.23	1.61
r8	10	14.85	48.5	14.85	15.27	2.83	15.27	15.36	0.59
r9	10	10.76	7.6	10.76	10.57	- 1.77	10.57	10.57	-
r10	10	8.97	- 10.3	8.97	8.78	- 2.12	8.78	8.78	-
r11	10	8.06	- 19.4	8.06	5.70	- 29.28	5.70	4.25	- 25.44
r12	10	23.54	135.4	23.54	24.38	3.57	24.38	24.66	1.15
r13	10	15.79	57.9	15.79	16.04	1.58	16.04	16.05	0.06
r14	10	12.88	28.8	12.88	12.66	- 1.71	12.66	12.63	- 0.24
r15	10	10.30	3.0	10.30	9.69	- 5.92	9.69	9.62	- 0.72
r16	10	8.07	- 19.3	8.07	7.13	- 11.65	7.13	6.84	- 4.07
r17	10	13.52	35.2	13.52	13.39	- 0.96	13.39	13.36	- 0.22
r18	10	10.6	6.0	10.6	10.48	- 1.13	10.48	10.49	0.10
r19	10	20.62	106.2	20.62	20.77	0.72	20.77	20.77	-
r20	10	13.15	31.5	13.15	13.83	5.17	13.83	15.07	8.97
r23	10	3.72	- 62.8	3.72	3.72	-	3.72	3.72	-
r25	10	3.11	- 68.9	3.11	2.03	- 34.73	2.03	1.69	- 16.75
r26	10	6.05	- 39.5	6.05	5.36	- 11.40	5.35	5.25	- 2.05
r27	10	14.77	47.7	14.77	14.88	0.74	14.88	14.89	0.07
r28	10	3.67	- 63.3	3.67	3.25	- 11.44	3.25	3.18	- 2.15
r29	10	5.74	- 42.6	5.74	5.12	- 10.80	5.12	4.78	- 6.64
r30	10	14.01	40.1	14.01	14.38	2.64	14.38	14.57	1.32
r33	10	9.27	- 7.3	9.27	9.31	0.43	9.31	9.32	0.11
r37	10	21.09	110.9	21.09	22.92	8.68	22.92	23.14	0.96
r38	10	15.70	57.0	15.70	16.08	2.42	16.08	16.14	0.37
r39	10	6.13	- 38.7	6.13	6.01	- 1.96	6.01	6.10	1.50
r40	10	5.66	- 43.4	5.66	4.73	- 16.43	4.73	4.29	- 9.30
r42	10	9.25	- 7.5	9.25	10.28	11.14	10.28	10.89	5.93
r44	10	9.10	- 9.0	9.10	10.98	20.66	10.98	11.59	5.56
r45	10	3.24	- 67.6	3.24	3.24	-	3.24	3.24	-
r46	10	3.50	- 65.0	3.50	3.50	-	3.50	3.50	-

Total demand : 25564.0 trips

	First Run	Second Run	Third Run
Demand-0-Transfer (%)	9.58	9.58	9.58
Demand-1-Transfer (%)	69.10	69.10	69.10
Demand-2-Transfers (%)	17.71	17.71	17.71
Unsatisfied-Demand (%]	3.61	3.61	3.61
Network total travel time	1.304×10^6	1.282×10^6	1.283×10^6
Network invehicle travel time (%)	78.68	79.85	79.73
Network waiting time (%)	11.07	9.72	9.86
Network transfer time (%)	10.25	10.43	10.41
Number of buses available	414	529	532
Number of buses required	529	532	534

other fifth generation languages for transportation network analysis and design problems. Such tools offer both opportunities and challenges, as described earlier. Our experience to date with the TRUST program has confirmed that the choice of LISP for this application has been an appropriate one, as it has allowed considerable flexibility in terms of the path choice rules and assignment logic that have been explored.

Clearly, the program is still in its developmental stages. As explained earlier, it is intended as part of an AI-based decision support system for transit network planning and design, which differs from existing approaches in several respects, including:

- use of heuristics and expert knowledge to guide search for design configuration;
- consideration of multiple criteria in the design process, as illustrated in the use of demand and transfer information in the present paper; and
- integration of AI software development tools and conventional languages in program execution.

TRUST is the core analysis part of the overall methodology, though it can be used independently, as shown in the present paper. Two principal additional components are contemplated in the overall design framework. The first is a heuristic route generation tool that produces "good" layouts, which would then be evaluated by TRUST. The second component identifies and suggests modifications to an existing configuration, given the descriptors computed by TRUST.

REFERENCES

1. W. Lampkin and P. D. Saalmans. The Design of Routes, Service Frequencies and Schedules for a Municipal Bus Undertaking: A

Case Study. *Operations Research Quarterly*, Vol. 18, 1967, pp. 375–397.
2. J. C. Rea. Designing Urban Transit Systems: An Approach to the Route-Technology Selection Problem. Report 204881. University of Washington, Seattle, Wash., 1971.
3. L. A. Silman, Z. Barzily, and U. Passy. Planning the Route System for Urban Buses. *Comp. Oper. Res.* Vol. 1, 1974, pp. 201–211.
4. C. E. Mandl. Evaluation and Optimization of Urban Public Transportation Networks. Presented at the 3rd European Congress on Operations Research, Amsterdam, Netherlands, 1979.
5. G. Newell. Some Issues Related to the Optimal Design of Bus Routes. *Transportation Science*, Vol. 13, 1979, pp. 20–35.
6. D. Dubois, G. Bell, and M. Llibre. A Set of Methods in Transportation Network Synthesis and Analysis. *Journal of the Operations Research Society*, Vol. 30, 1979, pp. 797–808.
7. D. Hasselstrom. *Public Transportation Planning—A Mathematical Programming Approach*. Ph.D. thesis. University of Gothenburg, Sweden, 1981.
8. A. Ceder and N. H. M. Wilson. Bus Network Design. *Transportation Research B*, Vol. 20B, No. 4, 1986, pp. 331–344.
9. H. Spiess and M. Florian. Optimal Strategies: A New Assignment Model for Transit Networks. *Transportation Research B*, Vol. 23B, No. 2, 1989, pp. 83–102.
10. R. B. Dial. Transit Pathfinder Algorithm. In *Highway Research Record* 205, TRB, National Research Council, Washington, D.C., 1967, pp. 67–85.
11. M. H. Rapp, P. Mattenberger, S. Piguet, and A. Robert-Grandpierre. Interactive Graphic System for Transit Route Optimization. In *Transportation Research Record 619*, TRB, National Research Council, Washington, D.C., 1976.
12. M. Florian and H. Spiess. On Two Mode Choice/Assignment Models. *Transportation Science*, Vol. 17, 1983, pp. 32–47.
13. A. F. Han and N. H. M. Wilson. The Allocation of Buses in Heavily Utilized Networks with Overlapping Routes. *Transportation Research B*, Vol. 16B, No. 3, 1982, pp. 221–232.
14. P. H. Winston and B. K. P. Horn. *LISP*, 2nd ed. Addison-Wesley, Reading, Mass., 1984.
15. M. A. P. Taylor. Knowledge-Based Systems for Transport Network Analysis: A Fifth Generation Perspective on Transport Network Problems. Department of Civil Engineering, Monash University, Victoria, Australia, 1989.
16. P. C. Furth and N. H. M. Wilson. Setting Frequencies on Bus Routes: Theory and Practice. In *Transportation Research Record 818*, TRB, National Research Council, Washington, D.C., 1981.

Publication of this paper sponsored by Committee on Transportation Supply Analysis.

Interactive-Graphic Interfaces for Planning the Transportation and Distribution of Freight

TEODOR GABRIEL CRAINIC AND JEAN-FRANÇOIS MONDOU

The development of interactive-graphic systems for the tactical/strategic planning of intercity freight transportation and distribution systems is discussed, with emphasis on the specification of the interactive-graphic interfaces. The requirements, difficulties, and possible ways of graphically representing the various components of the transportation system are analyzed. A number of general solutions to these problems are presented, solutions that are easily adaptable to other related problems. The discussion is illustrated with computer images obtained from an actual rail application.

Freight transportation is one of today's most important activities, not only as measured by its share of a nation's gross national product (which may be quite significant—it accounts for approximately 10 percent of the U.S. GNP), but also by the increasing influence that the transportation and distribution of goods have on the performance of other economic sectors (manufacturing, distribution, foreign trade, etc.). The current economic and regulatory environment "condemns" the transportation sector to economic efficiency, and it also increasingly places the emphasis on the quality of the service offered. Total delivery time and service reliability are increasingly important factors in the choice of a transportation or distribution firm. The forthcoming European and North American free trade zones will have additional impacts on the organization, operation policies, and competitivity conditions in the transportation industry. The constant evolution of this environment requires rapid, but comprehensive, assessments of actual or potentially new situations and of possible ways to adjust to these requirements. A "best" possible answer is normally sought.

Transportation systems, on the other hand, usually have quite complex organizations, involving a great deal of human and material resources and displaying intricate relationships and tradeoffs among the various policies affecting their different components. The planning process of such systems reflects their complexities and generally has an important component of scenario evaluation. Accurate, efficient, and flexible tools are therefore needed to help the planner in his task. A combination of operations research models and computer science techniques may be of considerable help by supplying them.

T. G. Crainic, Centre de recherche sur les transports, Université de Montréal and Université du Québec à Montreal, Montréal, Canada. J. F. Mondou, Centre de recherche sur les transports, Université de Montréal, Montreal, Canada.

The design of such tools, however, has changed recently. For example, systems traditionally required the user to squeeze his problem into predetermined forms, offering a limited range of results and drastically restricting the user's opportunities to intervene in the solution process. Instead of such "black boxes," the present trend is to develop integrated systems where the user is brought right in the middle of the process of defining and solving the problem. Furthermore, he is given the means to easily and efficiently control and direct it. These systems are comprehensive planning tools that integrate sophisticated operations research models and algorithms, large data banks and powerful interactive-graphic interfaces.

Not only has the design of the applications changed, but so has their immediate aim. Thus, these systems do not pretend to find the single, absolute, definitive solution to a given problem. Instead, the system is aimed at offering the planner a large array of means to enter, manipulate, visualize, and analyze vast amounts of data pertinent to a whole class of problems, plus the opportunity to merge the power of operations research methods with his special skills and inside knowledge of the problem. In addition, the planner does not have to worry about tasks that involve programming or data processing. Rather, he can focus on his prime function: investigate, analyze, and decide.

Two major factors have prompted this evolution. On one hand, the major advancements realized in computer science and technology have put tremendous computing power and sophisticated technology in a price range that facilitates their large penetration. We have also realized that, despite undeniable methodological advances, the problems facing us are becoming increasingly complex, requiring the consideration of a variety of policy rules, decision criteria, union conditions, etc., that may not be adequately addressed by traditional operations research or economic modeling techniques, yet are well known to the decision makers. Thus, the integration of operations research methodologies into comprehensive interactive-graphic planning tools offers a most interesting and promising development avenue, one which has already met with considerable success in numerous domains (*1–8*).

For our part, we are interested in the case of the intercity freight carrier, or distribution firm, which may possibly use several transportation modes to move freight of several commodity classes. We focus, in particular, on the issues related to the long and medium-term planning of operations of such a firm.

Models and tools aimed at this problem have been proposed, but the emphasis has been on the methodological, operations research modeling and algorithmic development, aspects of the problem. The objective of this paper is to complement this body of knowledge by looking into the requirements and difficulties involved in graphically representing and displaying the various components of the transportation system of an interurban multimode multiproduct freight carrier, and by presenting the solutions that we propose to these problems.

We start by recalling the main components of a freight transportation system and of the corresponding planning process. We also briefly review the general modeling framework that we have developed for this problem. We then present the main objectives and characteristics of a proposed interactive-graphic planning system. The description of the requirements, alternatives, and selected solutions for the graphical representation and display of the system constitutes the core of the paper. Our method is illustrated throughout the paper by plots drawn from a rail application. We conclude with a discussion on the perspectives and challenges that lie ahead.

FREIGHT TRANSPORTATION PLANNING

General Principles

The underlying structure of any large freight transportation system consists of a rather complex network of large terminals and small stations interconnected by physical (e.g., rail tracks or roads) or conceptual links (e.g., air or navigation lines). Transportation services are usually performed by a large number of vehicles of various types and belonging to several transportation modes that move the freight from one of its nodes (terminal) to another by following the links of this network. Vehicles may move either individually or grouped in convoys such as trains or multitrailer assemblies, and they may sometimes follow predetermined routes and schedules. During the journey, vehicles and convoys may interact with others, thus contributing to and being affected by congestion conditions on the network. These interactions may or may not be controlled by the firm (e.g., rail can be controlled, whereas modes using public routes cannot). In terminals, various operations may be performed on vehicles (classification, convoy formation, etc.) and on freight (classification, consolidation, etc.). The modeling of terminal operations and of their interactions with the rest of the system is a critical component of any comprehensive model of the transportation system. Freight is usually differentiated by commodity group according to physical characteristics, or value, or both.

To control these activities and to make the best possible use of the system's resources, several policies must be defined and implemented. These policies and the associated modeling problems are generally classified into three groups corresponding to the widely used three planning levels: strategic, tactical and operational (*9,10*):

1. *Strategic* (long-term) planning typically involves the highest level of management and usually requires large capital investments for resource acquisition over long-term horizons. Strategic decisions determine the general development policies of the company and broadly shape the operating strategies of the system by deciding upon the design and upgrading of the physical network, the location of facilities, the acquisition of resources, the general definition of service and tariff policies, and so on.

2. *Operational* (short-term) planning is performed by local management, in a highly dynamic environment where the time factor and detailed representations of vehicles, facilities, and activities are essential. Scheduling (timetable generation for service departures, maintenance activities, crews, etc.) and dispatching (vehicles and convoys into, out of, and inside terminals, as well as on the links of the network, when appropriate) are among the most important decisions at this level.

3. *Tactical* planning forms a vital link between the preceding two levels of decision making. It aims at ensuring, over a medium-term horizon, an efficient and rational allocation of existing resources to improve the performance of the whole system. Tactical decisions need to be made concerning the design of the service network (selecting the routes and characteristics of the transportation services that will be offered), the distribution of the traffic of each origin-destination pair in the network with a nonnegative demand, the general rules that specify for each terminal how freight and vehicles are to be sorted and grouped, and the distribution of empty vehicles to satisfy demand and to rectify imbalances in the vehicle availability due to uneven traffic. The output of tactical planning consists of a series of rules and policies that affect the whole system. These rules, which are collected in a transportation (load) plan, are then used to determine the day-to-day policies that guide the operations of the system. The same tactical transportation plan is also a privileged evaluation tool for "what-if" questions raised during strategic planning (*10,11*).

A number of efforts have been made toward the formulation of models aimed at tactical planning problems. The most successful have been the so-called "network models" that take advantage of the structure of the transportation system and integrate policies that affect several terminal and line operations. Important contributions to this field have been made in recent years (*10–19*).

Methodology

We now briefly recall the methodology we use; details may be found elsewhere (*10,20*).

Let $G = (N, L)$ represent the *physical network*, on which transportation activities are carried out. Then, the nodes of the set N mainly represent the terminals, stations and junctions of the network. We identify the set of terminals as Y, $Y \subseteq N$. The set L is the set of links representing the connections between the points in N.

On this network, the transportation demand is defined in terms of tonnage and number of vehicles of a certain commodity c to be moved from an origin terminal o, $o \in Y$, to a destination terminal d, $d \in Y$. To simplify, we refer to the market or traffic class $m = (o, d, c)$ with a positive demand of $g^m = g^m_{od}$ cars to be transported. The set M will identify the set of all markets.

Freight is transported by vehicles, possibly grouped into convoys, that move through the physical network following certain routes and transportation modes defined by the service structure S. A service, $s_i \in S$, has an origin and a destination in Y. Between these two points, it follows a route defined as a set of links that form a path in G and may eventually stop at some of the terminals or stations on this route to drop off or pick up traffic. It is convenient to identify the subpath between each two consecutive stops on the service route as the service leg. When the service has no intermediary stops, as in most trucking applications, the entire service route contains only one service leg. The service is also characterized by its service class, or mode, which defines the transportation mode (e.g., truck size on roads or truck-on-flat car), the speed/priority (e.g., slow- or rapid-moving train), the vehicle type, the vehicle and convoy capacities, etc. Capacities may depend on the physical characteristics of the physical links on its route. Last but not least, each service operates at a certain frequency f_t, specifying how often the service is run during the planning period.

Traffic moves following predefined paths that specify the physical route, the services to use and the terminals where operations have to be performed. For a given traffic-class, one or several such paths may be used, according to the level of congestion in the system and the service and cost criteria of the particular application. One may thus have a physical definition and a service definition of the various product paths. It is more conceptually convenient, however, to define an itinerary for a given traffic-class as a sequence of classification/consolidation terminals, where work is performed on the traffic of the traffic-class, and the respective sequences of services used to travel between two such consecutive terminals. For a given market, $m \in M$, there are several possible itineraries; the set I^m contains them. The volume of traffic that moves on each itinerary $k \in I^m$ is noted v_k^m.

The frequencies and itineraries are the central elements of the model. Fixing the values of the frequencies determines the design of the service network, the level of service and the feasible domain for the traffic distribution problem. On the other hand, the selection of the best itineraries for each traffic-class solves the traffic distribution problem and also determines the workloads and the general classification and makeup strategies for each yard of the system. The model may then be stated as follows:

Minimize $\Psi(v, f)$

$$= \sum_{m \in M} \sum_{k \in I^m} \Phi_k^m(v, f) + \sum_{s_t \in S} \Theta_t(f) + \Pi(v, f) \quad (1)$$

Subject to

$$\sum_{k \in I^m} v_k^m = g^m \qquad \forall\, m \in M \quad (2)$$

$$v_k^m \geq 0, \qquad \forall\, k \in I^m, \qquad \forall\, m \in M \quad (3)$$

$$f_t \geq 0 \text{ and an integer} \qquad \forall\, s_t \in S \quad (4)$$

where

v_k^m = quantity of flow from the traffic-class m traveling on its itinerary k;
f_t = frequency of service s_t;
v = vector of flows for the whole system;
f = vector of frequencies;
$\Phi_k^m(v, f)$ = total (operating and delay) cost of using the itinerary k for the traffic-class m; the sum represents the "variable" cost of transporting the flow v by a service network operated at level f;
$\Theta_t(f)$ = total (operation and delay) cost of operating the service s_t; the sum represents the "fixed" cost of offering transportation services at level f; and
$\Pi(v, f)$ = special terms modeling specific relations, such as train capacity, performance targets, etc., as "delays."

The objective function principally represents the total system cost: traffic- and itinerary-related costs plus the cost of offering the transportation service. It is a generalized cost, in the sense that both operating costs and service costs (delays, service targets, reliability, etc.) are included. It is at the level of this objective function that the relationships and tradeoffs among the various system and policy components are considered by the combination of the operating costs and the delays encountered by vehicles, convoys, and traffic in terminals and on the lines of the system due to congestion and the operating policies.

The model has the structure of a nonlinear, mixed integer, multimode, multicommodity flow problem, and it is solved by an efficient heuristic algorithm (*20*) based on decomposition, column generation, and "steepest descent" techniques. This approach has been successfully applied to problems from the Canadian (*21*) and French railways (*22*). It has also been adapted with considerable success to the multimode less-than-truckload trucking problem (*16*).

This model is intended as the methodological core of an integrated interactive-graphic planning system called TACTIC (*23,24*). The system combines a data bank together with procedures to access, analyze, and modify it; problem modeling and scenario defining facilities; optimization algorithms and post-optimal procedures; and, most importantly, interactive-graphic capabilities to display and help analyze data and results.

INTERACTIVE-GRAPHIC INTERFACES

Objectives and General Characteristics

TACTIC has a modular structure. Each module is an independent program that performs a specific task. Thus, all data transfers among modules occur exclusively via the data bank. The modular approach has several advantages. It lets the user concentrate on one specific aspect of the problem at a time. It also facilitates the maintenance and the enhancement of the code. The precise characteristics of the system are described elsewhere (*23,24*); here, we concentrate on discussing its graphical aspects. We want, however, to specify some of our major objectives. Specifically, the system aims to offer the decision maker the possibility to:

- Treat the problem globally, networkwide;
- Adjust the model to his problem, not the inverse [he may, for example, specify the cost (delays, reliability, etc.) functions for his particular application];
- Efficiently generate, evaluate, and compare a large number of global operating strategies (he may analyze and evaluate "what if" strategic questions);

• Manipulate a solution until it meets all the unwritten rules and criteria.

From the point of view of the computer implementation, the system offers the following features:

• A stand-alone system, easy to use, that may efficiently run on the current generation of microcomputers and workstations;

• A system where the user may completely define, build, and modify its problem and its network [this contrasts, for example, with the *Princeton Transportation Network Model and Graphic Information System* (*25,26*), which is linked to a representation of the North American transportation network; this implies that, on the modeling side, the user will have the choice among various terminal and line submodels and the corresponding terms in the objective function, while on the network side the system will offer a comprehensive interactive-graphic network editor]; and

• A representation of the data that respects and reflects the "path" structure of most of the elements that make up the transportation system, in particular the transportation services and the traffic distribution (this contrasts with STAN (*7*), for example).

While each of these properties may appear in some systems, to our knowledge they do not characterize any existing system for the strategic or tactical planning of operations for intercity freight carriers.

At present, the graphical procedures are implemented for use on any terminal of the Tektronix 4010 and 4110 series. Any output displayed on the screen may be copied by using the Tektronix 4631 hard copy unit. The output displayed on a Tektronix color terminal may be drawn by using the Tektronix 4696 ink jet color printer. We also make use of a plot file to store the image for further usage (e.g., sending it to a laser printer). In order to increase the portability of the system, drivers for EGA and VGA cards will also be developed.

The code is written in the C programming language. The choice of C is mainly due to the efficiency of the code it produces and to the control of complex data structures (in contrast to FORTRAN) and the manipulation of memory addresses (in contrast to Pascal) that it offers. Another advantage is its portability, since, although C exploits all the capabilities of its host computer, it is independent of any particular machine architecture. For example, we have successfully compiled our graphic procedures on three different computers: the Sun-3 system (operating under Unix), an IBM-AT compatible, and an IBM-XT equipped with a DSI-32 coprocessor board (operating under DOS) without having to make any modifications.

An important aspect of TACTIC is the graphic kernel it uses. According to the choice of the programming language, we wrote a new graphic library in C specifically aimed at network representations (*27*). This library also contributes to the portability of the system, since we control all the codes.

Another important decision in the development of a graphic system is the use of color. Color-enhanced computer images are very attractive, and color may help to distinguish objects. On the other hand, a large number of black and white terminals do exist and are in use. Moreover, the definition of the color set is specific to each terminal. Thus, in defining color tables some degree of portability may be lost if each color receives a particular signification. We therefore adopt a simple solution: when color is available, we use it only to distinguish among elements, without assigning any specific meaning to a particular color; in absence of color, only geometrical symbols and line patterns are used.

The remainder of this paper is dedicated to presenting the difficulties inherent in an efficient graphical representation and display of the various elements that make up the transportation system, and of the approach that we have developed in order to address them.

The Elements to Display

When building or analyzing a transportation plan, several different elements have to be considered simultaneously. These elements are related to (1) the physical network, (2) the service structure, and (3) the freight distribution. Each element has both a set of characteristics (e.g., the length of a link) and a set of results that describe the state of the system (e.g., the link flows). Furthermore, each result may be analyzed according to several different attributes (e.g., the link flow by commodity group or by traffic-class). Note that results may be obtained either from an optimization procedure, a manual solution, or from actual or historical data. The following is an enumeration of the various elements that make up the system, with their main characteristics and attributes that influence the requirements and possibilities of their graphical representation and display:

1. Physical network—

• *description*: nodes and links with their identification (numerical, alphanumerical, or both), type (e.g., single or double line, breakbulk, or end-of-line terminal), physical characteristics, user data, etc.;

• *utilization*: total flows of freight, vehicles, convoys, etc.;

• *result analysis*: see Table 1.

2. Service network—

• *description*: physical route, service segments, intermediary stops, mode or service type, service performance measures (in terms of service targets, delays, reliability, etc.);

• *representation*: physical (the actual routes through the network are displayed) or conceptual (only the service arcs are represented); the display may be by individual service, by origin/destination pair, or the whole network;

• *utilization*: frequency, total and marginal costs;

• *result analysis*: see Table 2.

3. Freight traffic—

• *representation*: demand by origin/destination matrices and aggregations; movements by itineraries, which may be drawn on the physical or service network and may identify working terminals, blocks, services;

• *utilization*: volumes, measured in tons or vehicles, and total and marginal costs (delays, reliability, etc.) for each itinerary or for the traffic-class;

• *selection*: by traffic-class, origin/destination pair, commodity group, etc.

This enumeration, although not exhaustive, covers the main possible applications. Note that not all combinations apply to all applications. We have looked, however, for general solu-

TABLE 1 RESULTS ANALYSIS—PHYSICAL NETWORK

Element	Results	Attributes
Links	Convoys	Mode or service type
	Tons Vehicles Blocks	Commodity group Mode or service type Traffic–class
Nodes	Convoys by mode or service type	Created Terminated Passing through
	Blocks, vehicles, tons by Origin, destination, commodity group, traffic–class, mode or service type.	Originating Terminating Classifying Transferring Passing through

TABLE 2 RESULTS ANALYSIS—SERVICE NETWORK

Element	Results	Attributes
Service arcs	Tons Vehicles Blocks	Traffic–class Origin/destination pair Commodity group

tions that may easily adapt to the particular requirements of each application.

The remainder of this section is dedicated to the analysis of the graphical representation of these elements. We illustrate our discussion with several computer images, originally drawn in color, from a real Canadian railway application (*20*).

Representation of the Physical Network

A drawing of the physical network is shown in Figure 1. The details on the central part of the network, representing the western provinces of Canada, are not distinguishable. By zooming in on this part of the image (this action is called making a "window"), we produced the more detailed representation in Figure 2. On this image, each type of node is represented by using rectangles and circles of different sizes and colors, the choice of which depends on the user. There are two link types in our application: simple and double rail tracks. In order to represent these, we use different colors and line patterns. Note that the terms used in the legend to indicate the node and link types (terminal, single line, etc.) may be customized, depending on the application, by interactively modifying the data bank. Graphical attributes (e.g., colors, line patterns, geometric symbols) used to display the network may be modified at any moment during an interactive session.

In the lower left corner of Figure 2 appears the "skeleton" of the network that is drawn. This feature has to be asked for by the user, and it indicates which part of the network is currently plotted. This feature may also be used to select a new window that is not adjacent to the current one. Physical characteristics of nodes and links can be interactively inspected by pointing (by using the crosshair cursor or a light pen) the desired element. Then, the information appears immediately in a small box. The box position is also determined by the user. This technique for examining data is elegant and very useful, and we use it in other contexts as well (see the following sections). To our knowledge, these two graphical features are unique to TACTIC.

Figure 3 illustrates the display of total freight on links through the use of bars and, possibly, numerical values. The display of similar information on nodes, by using circles, is shown in Figure 4. As usual, hatch width and scale may be interactively modified during the session, as well as the size and contents of the node boxes.

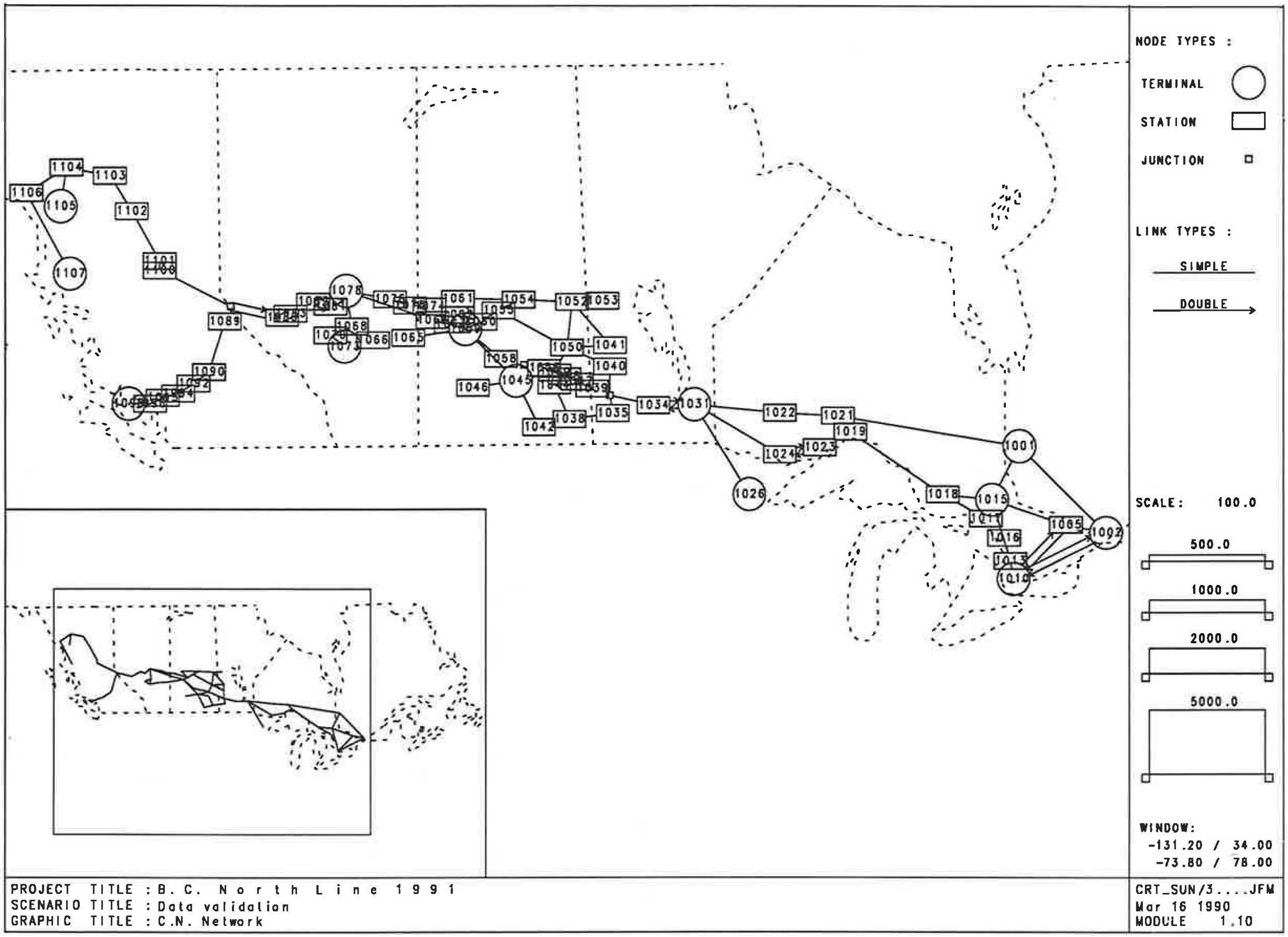


FIGURE 1 Representation of the physical network.

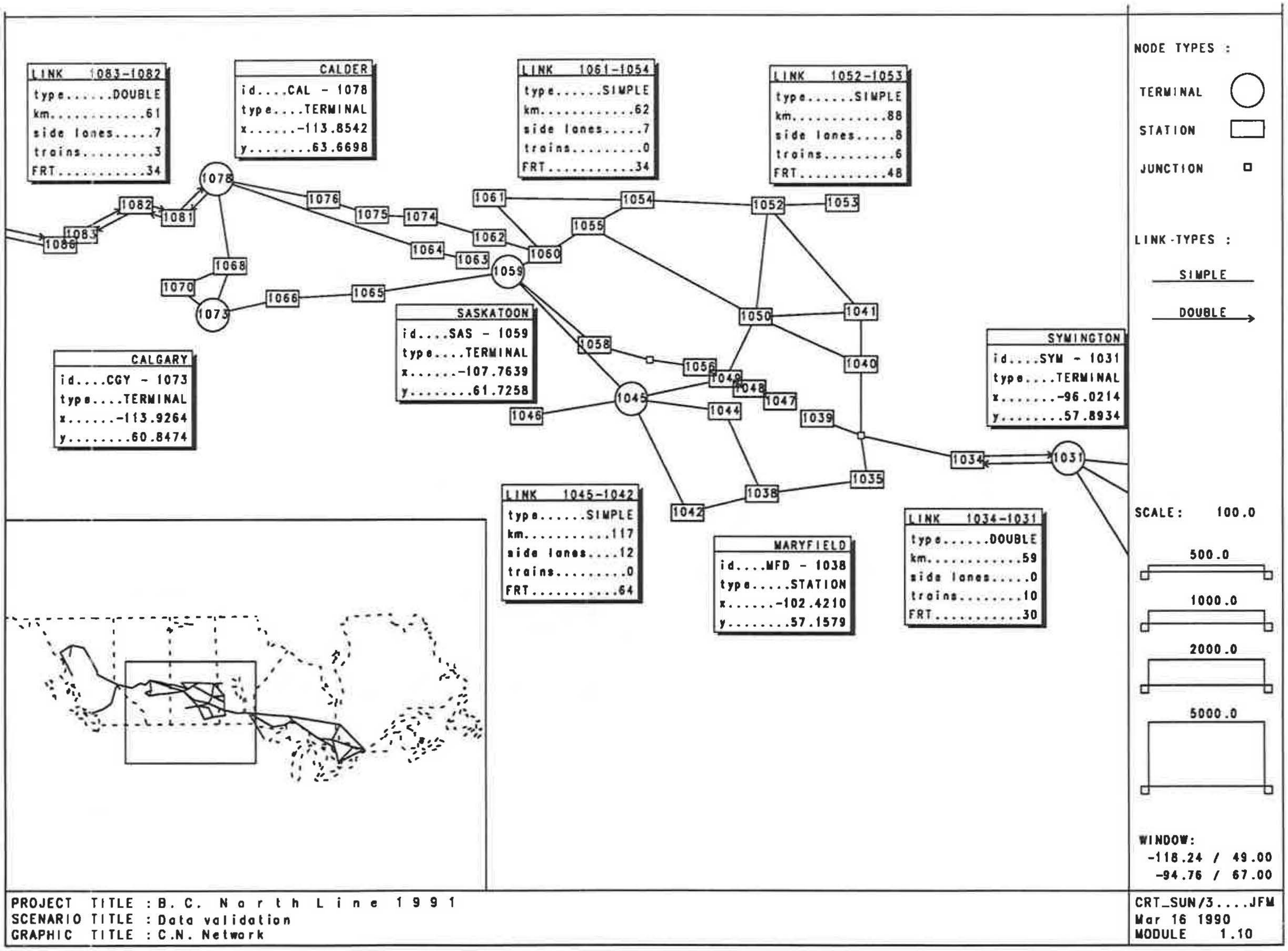


FIGURE 2 Physical network and characteristics.

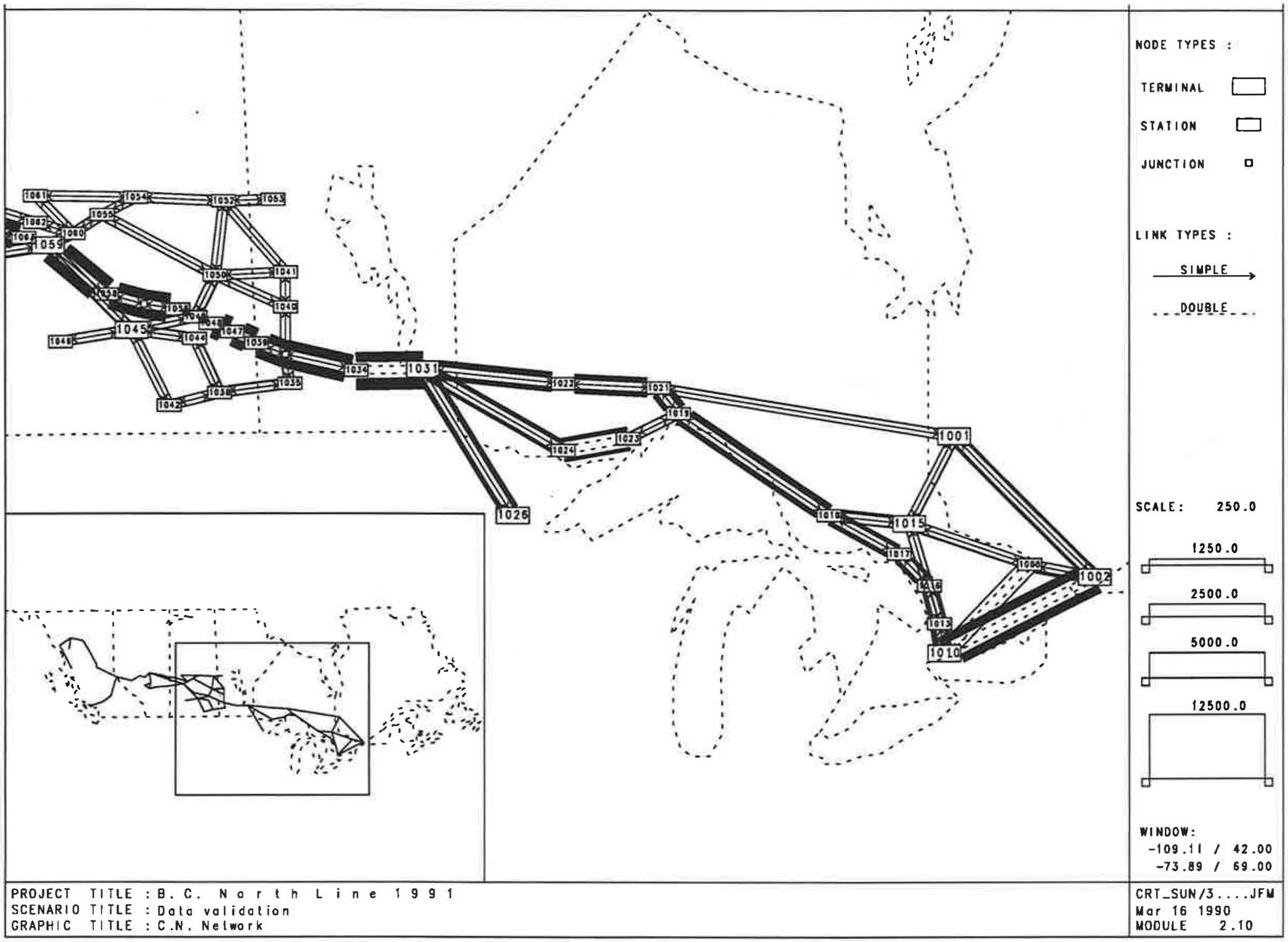


FIGURE 3 Volumes on rail links.

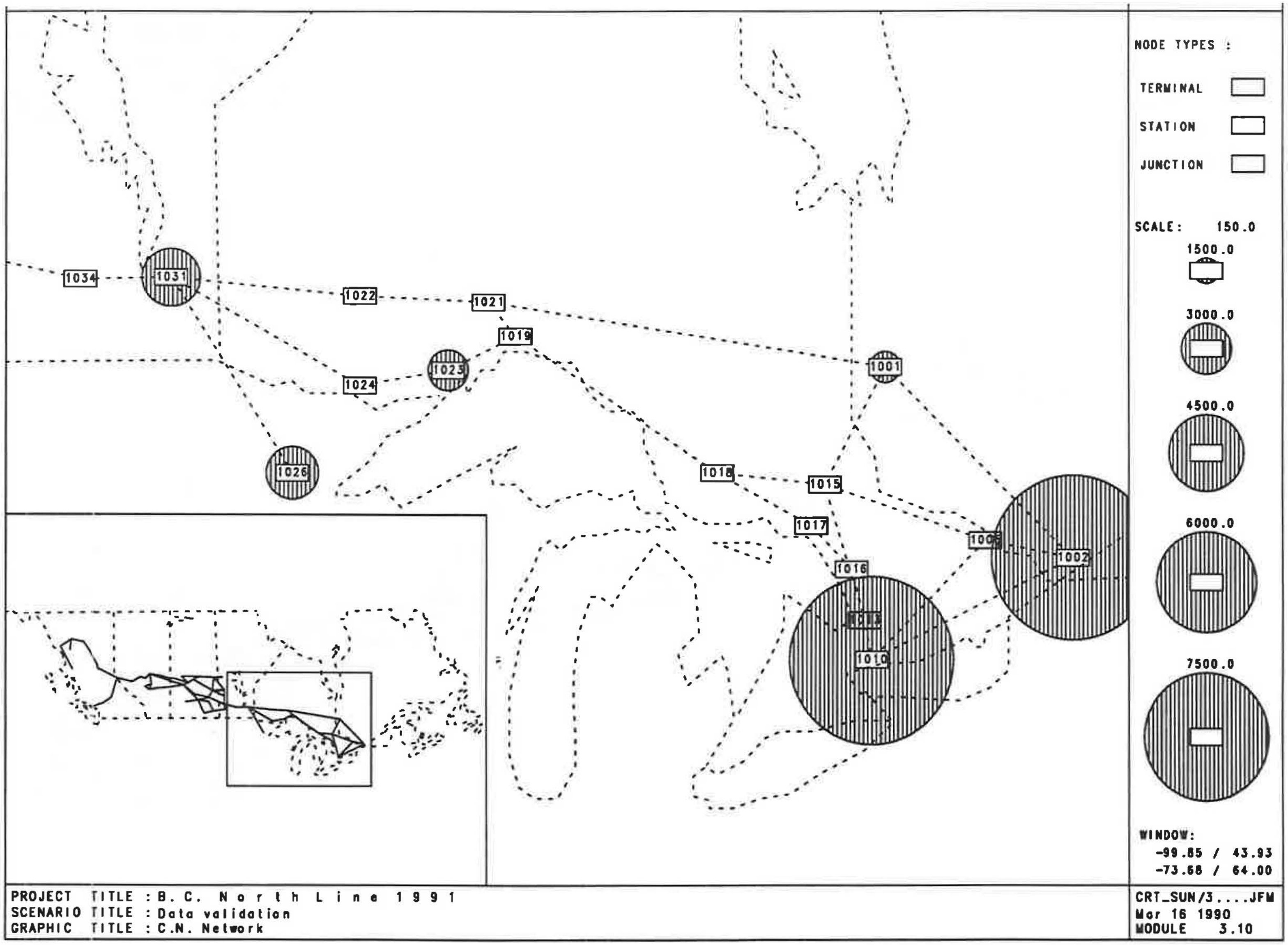


FIGURE 4 Volumes at nodes.

Result Analysis

One of the main problems one is faced with, when contemplating a way to display "results" on networks, is how to obtain a clear representation of the distribution on each element (link, node) of the network, of one or more results, each with several attributes. The problem is not trivial, because there may be a relatively high number of cases for an attribute (in our application, we had to take into account some 30 different commodity/vehicle groups), while networks may be quite dense. To our knowledge, no tactical or strategic planning tool currently allows the graphic display, on a network, of more than one measure at a time.

We first examined the graphical representation of one result-attribute combination on an element, namely a link. An example of such a requirement is the representation of the number of vehicles, per commodity group, using the link. Two solutions caught our attention. The first, illustrated in Figure 5, plots the histogram of the attribute on each link in both directions, if necessary. Unfortunately, the sharpness of the image deteriorates rapidly when the number of possible cases (values) that the attribute may take increases, or when there are several arcs adjacent to the same node, or when links are (almost) parallel.

The second approach, illustrated in Figure 6, uses color. We associate to each link a band, divided into as many sections as the number of cases of the attribute. Then, the color in each section indicates the intensity of the corresponding attribute value. This solution has the advantage of using less plotting space, due to the fact that each bar has the same small width. However, this technique requires a color monitor, since the utilization of filling patterns of variable grey shading, instead of color, is equivalent to the preceding solution, where the quality of the representation deteriorates with the number of possible values for an attribute. Thus, the use of color limits the portability of the system. Besides, it may be difficult to make the color intensity proportional to the attribute intensity, which is easy when using histograms. This may affect the precision of the display.

We also examined a number of ways to simultaneously plot two or more combinations of result-attributes on a network element. An example of such a case is the plotting of the

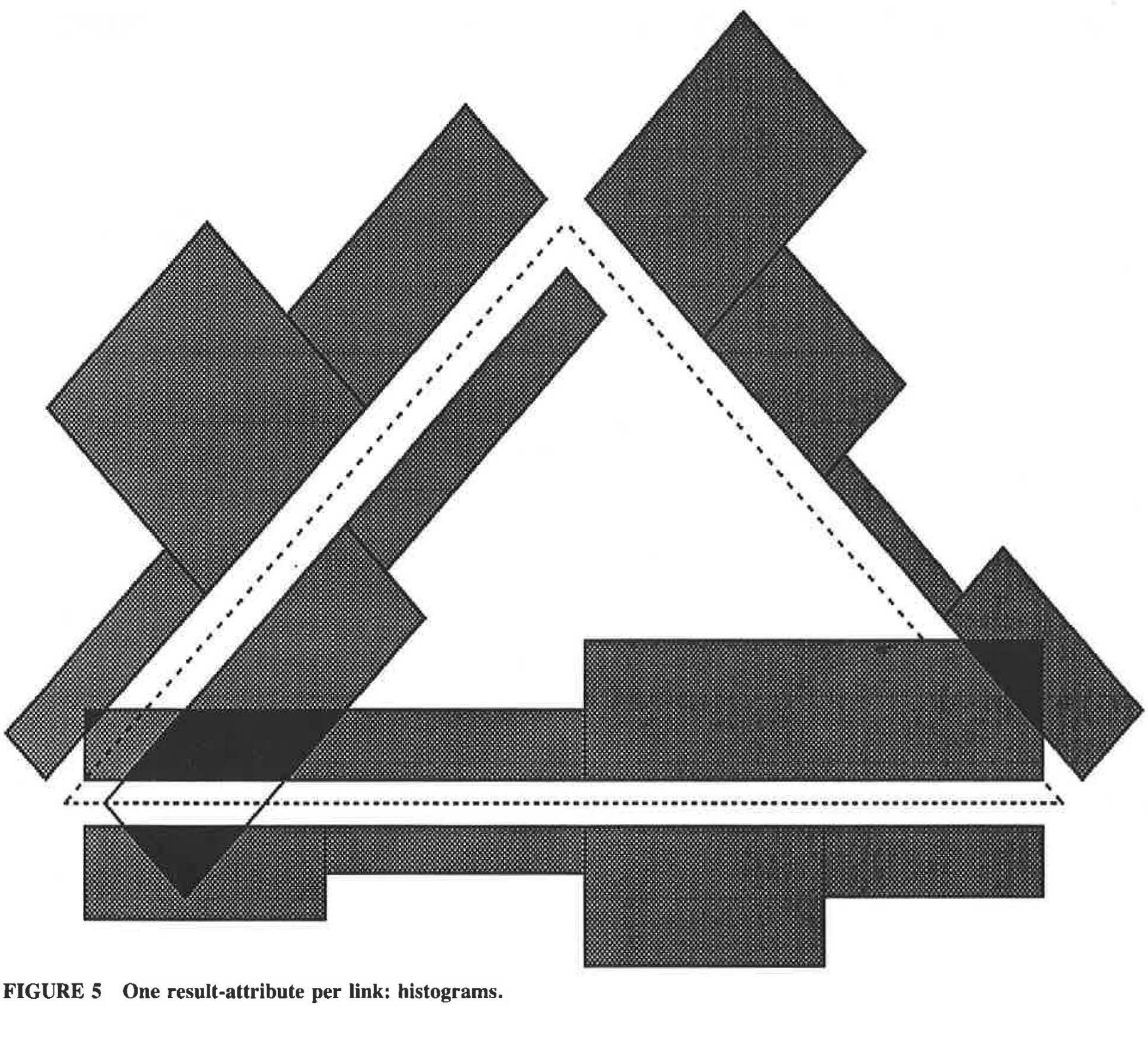


FIGURE 5 One result-attribute per link: histograms.

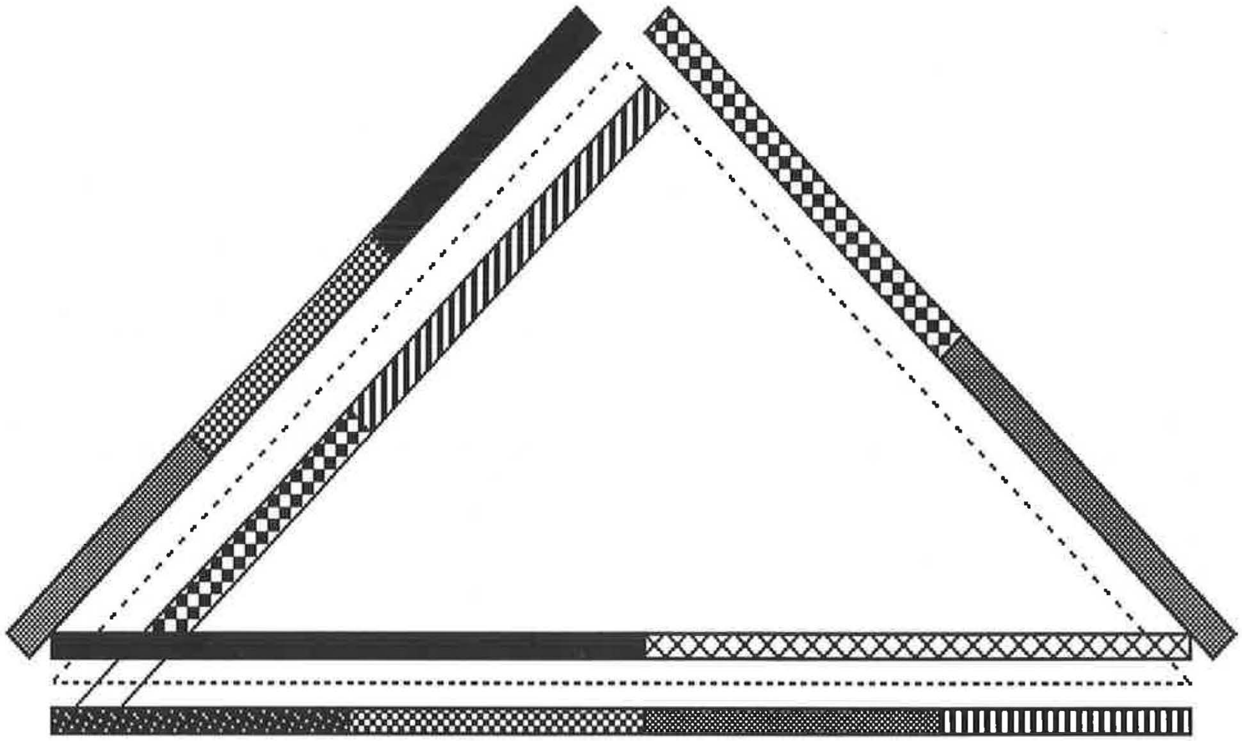


FIGURE 6 One result-attribute per link: color.

number of vehicles and of the number of blocks per service type that use the links of the network. Here are some of these possibilities:

- Three-dimensional histograms (appropriate only for two combinations at a time—see Figure 7);
- One color band for each combination; and
- Simultaneous screen copies of the network, each showing one combination.

Each of these alternatives presents a number of advantages, yet none is really acceptable. Most are not appropriate for the simultaneous plotting of several result-attribute combinations, and some restrict the portability of the system. They all result in unreadable screens when the quantity of data increases.

We therefore had to adopt the following solution, illustrated in Figures 8–10: represent only one combination of result-attribute per element at a time, by using histograms (and color when available), but plot several such combinations per screen or per page. The user chooses the elements and the combinations of results and attributes that he wants displayed, and the number of such combinations to be displayed on each page. Figure 8 illustrates one variant of this approach, where the maximum number of combinations that may be plotted on the same screen is fixed (here, for reasons of clarity, it has been fixed to four). Depending on the particular application and on the equipment used, this number may be easily modified. It is also quite easy to allow the user to determine this maximum number, (*3,7*), or even to completely control the composition of the screen (see Figures 12 and 13).

A second variant of this approach, similar to the method adopted for the display of the physical network, consists in working directly on the network. This approach is illustrated in Figures 9 and 10. Here, the user graphically selects an element, then chooses the result and attribute he wants plotted and, finally, points the position on the screen where the information has to be displayed. This second approach is very simple to implement and it is general, in the sense that it may successfully be used for the display and analysis of all the elements of the problem: physical network and results on network elements, services and itineraries.

Path Representation

In the physical network, a service is characterized by an origin, a destination, a link path connecting these two terminals, and possibly some intermediate stops on the route. The representation of services is similar to the display of bus lines (*3*).

Services often use the same physical links. Thus, it is necessary to use a rather complex mechanism to shift the representation of "parallel" services in order to clearly identify each one. Additionally, we have to identify intermediate stops and service legs and, to do so, we use special symbols to represent nodes. Figures 11–13 illustrate these features: stars represent intermediate stops, while small squares represent nodes that are bypassed on the route.

An image becomes rapidly incomprehensible when the number of services plotted on the same screen is too large. When this is the case, we let the user divide the screen into zones and plot one service per zone. The user is completely in charge of the image composition and is responsible for its quality. He may thus maximize the utilization of the visualization area. Figures 12 and 13 illustrate this feature. This approach is more efficient when the graphic terminal allows the selective erasing of the screen.

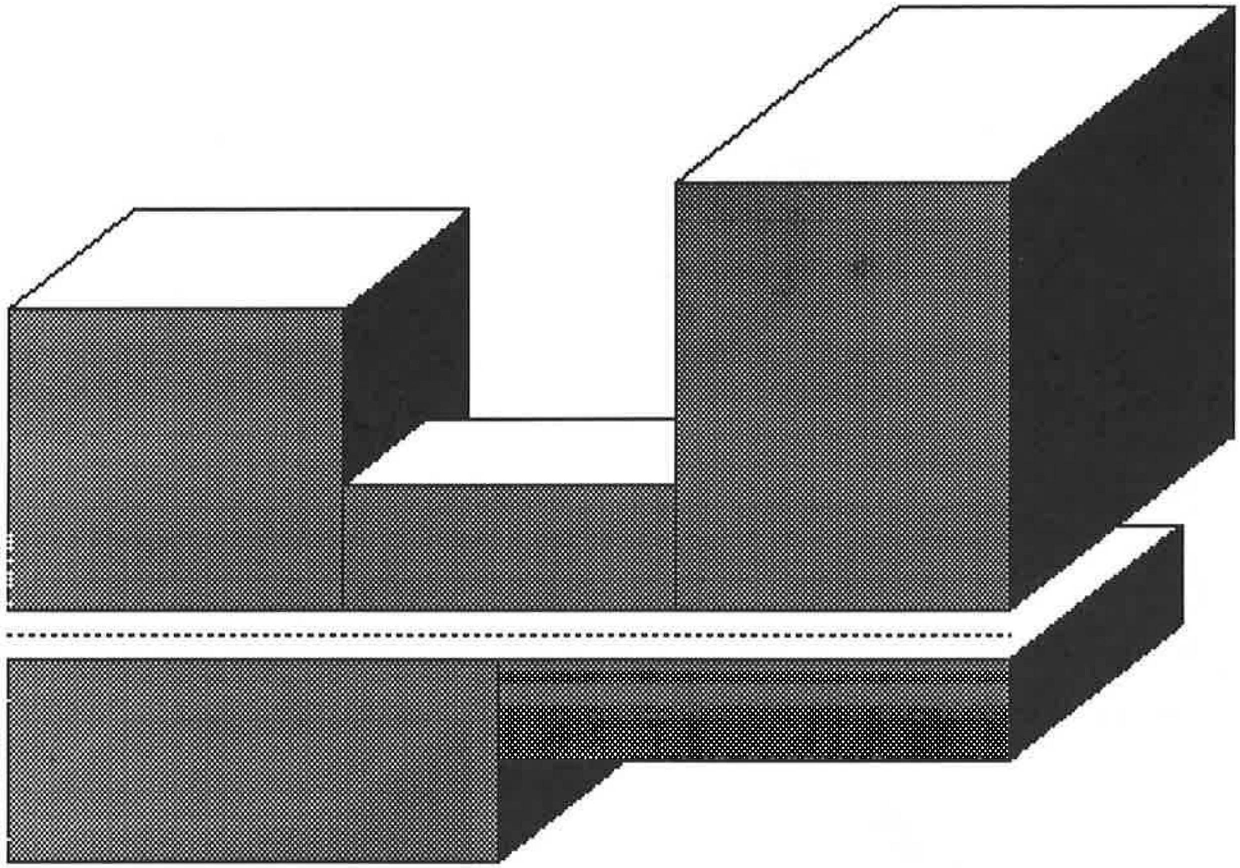


FIGURE 7 One result and two attributes per link: histograms.

Figure 14 illustrates the detailed representation of a service and of its utilization. Color and geometric symbols are used to identify elements of particular interest: circles for the origin and destination, dark-blue boxes for the intermediary stops, light-blue boxes for the other nodes on the service route, etc. All these characteristics are specified as module parameters that may be modified by the user. The descriptions of the service and its components are indicated in small boxes, interactively requested by the user and positioned on the screen by following his indications. Total flows on the service arcs complete the display. Detailed result-attribute analyses may also be added by using the same approach described above for the physical network. Finally, the origin and destination terminal of the segment are indicated on the skeleton network.

Itineraries used to move the freight of traffic-classes display structures very similar to those of services: an origin and a destination, both terminals in the network; a series of intermediate terminals where work is done on traffic, on vehicles, or on both; and service leg paths (doubled by link paths) connecting these terminals. The requirements for representation and display of itineraries are therefore similar, and the solutions we adopted use the same principles. Figures 15 and 16 illustrate the application of these ideas to the display of itineraries. Here, only the arcs of the service structure (i.e., the service legs) are identified, but their illustration follows the pattern of the physical network. Nodes that play a special role are identified by color and geometric symbols. In this application, special nodes are the origin, the destination, and the yards where classification or transfer operations occur. Arrows indicate the direction of movement, while the service number is identified on the service legs. Pop-up boxes and histograms, similar to those used for the physical and service networks, are used to describe the characteristics of the traffic-class, the itineraries, their composing elements (terminals, services, etc.), and the distribution of flows.

CONCLUSIONS

Interactive-graphic capabilities, interfaces, and procedures play an increasingly important role in the design and development of comprehensive planning systems for the transportation and distribution of freight. We have identified the requirements and explored the difficulties related to the development of such interfaces and have proposed general and elegant solutions to the problems of graphically representing and displaying the various elements, characteristics and results that make up the transportation system of an interurban freight carrier. These solutions may be used to display results obtained from optimization or simulation models, from actual operations or from historical sources. They are equally valid for representing freight carrier networks or logistic structures for freight distribution firms, at the tactical or strategic planning level.

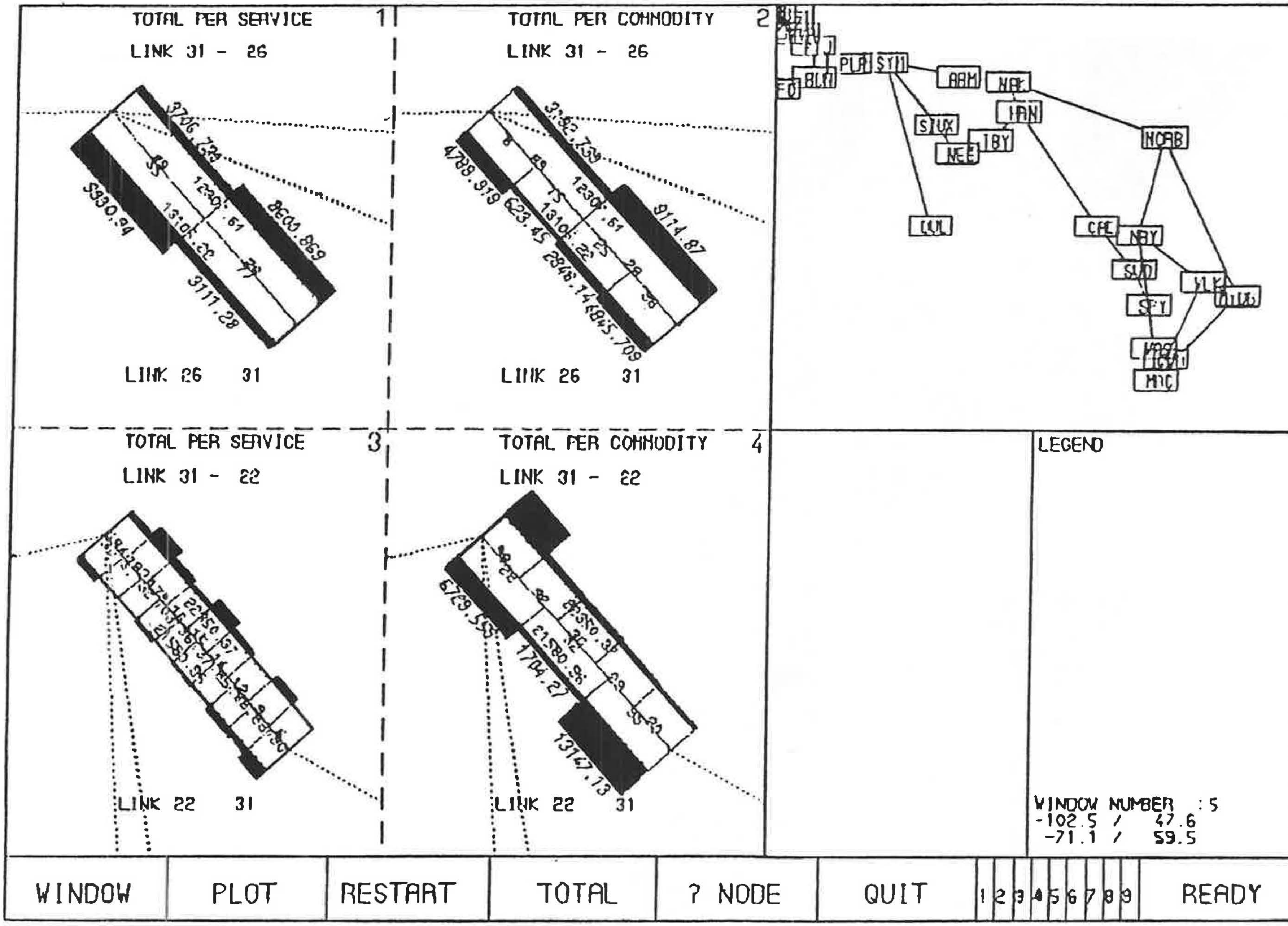


FIGURE 8 Link result analysis—fixed split screen.

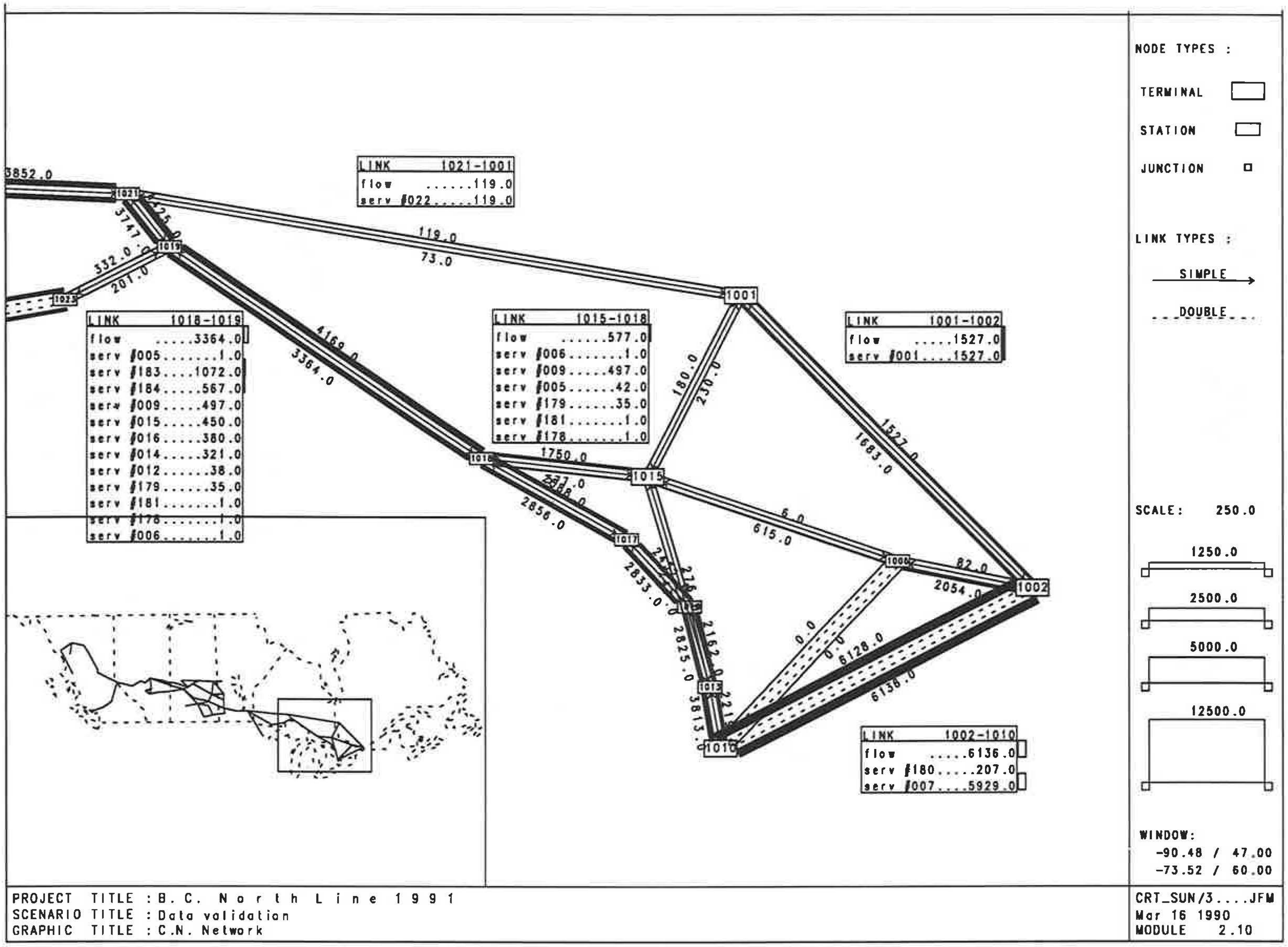


FIGURE 9 Link result analysis—moveable boxes and histograms.

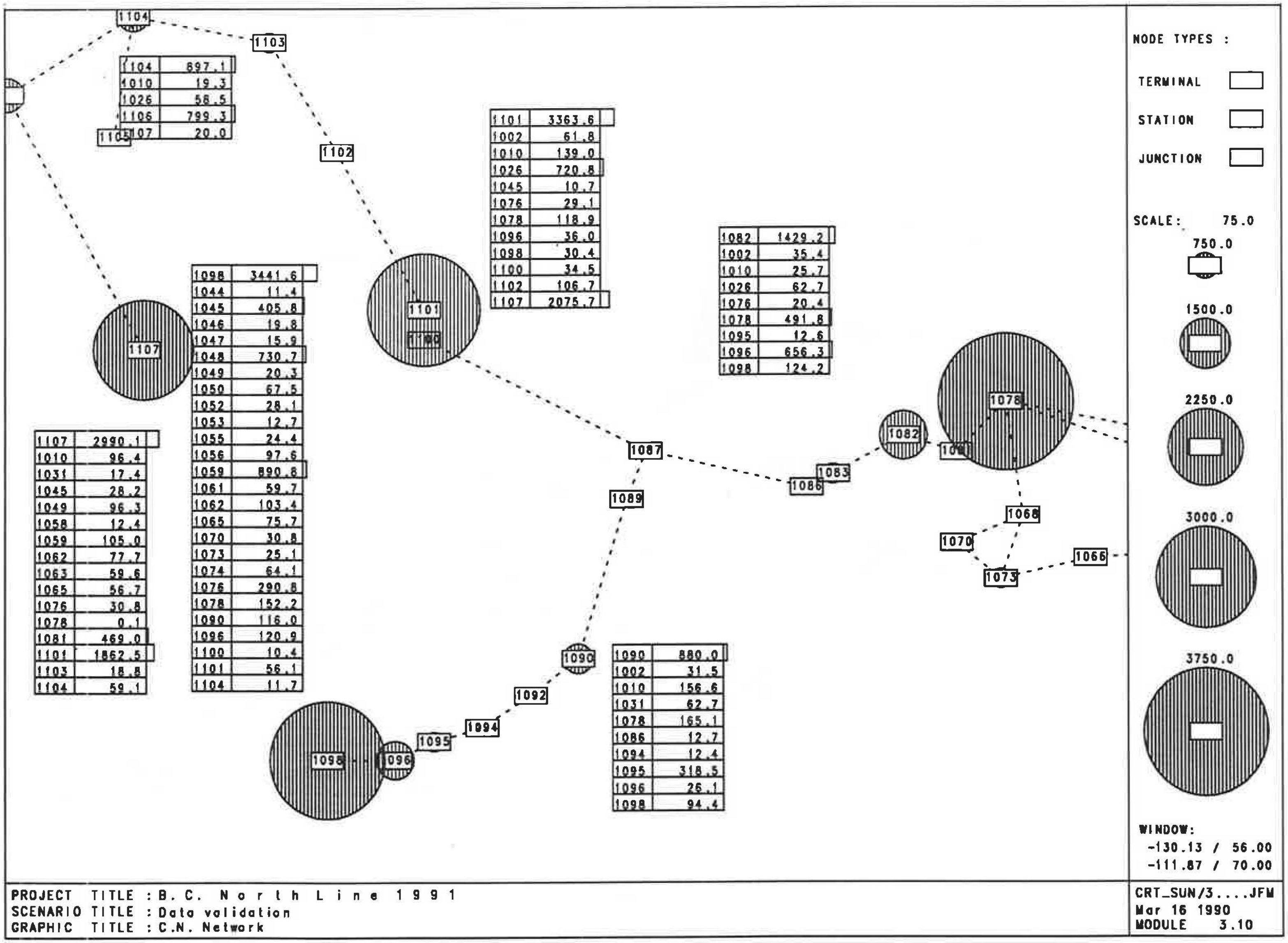


FIGURE 10 Node result analysis—moveable boxes and histograms.

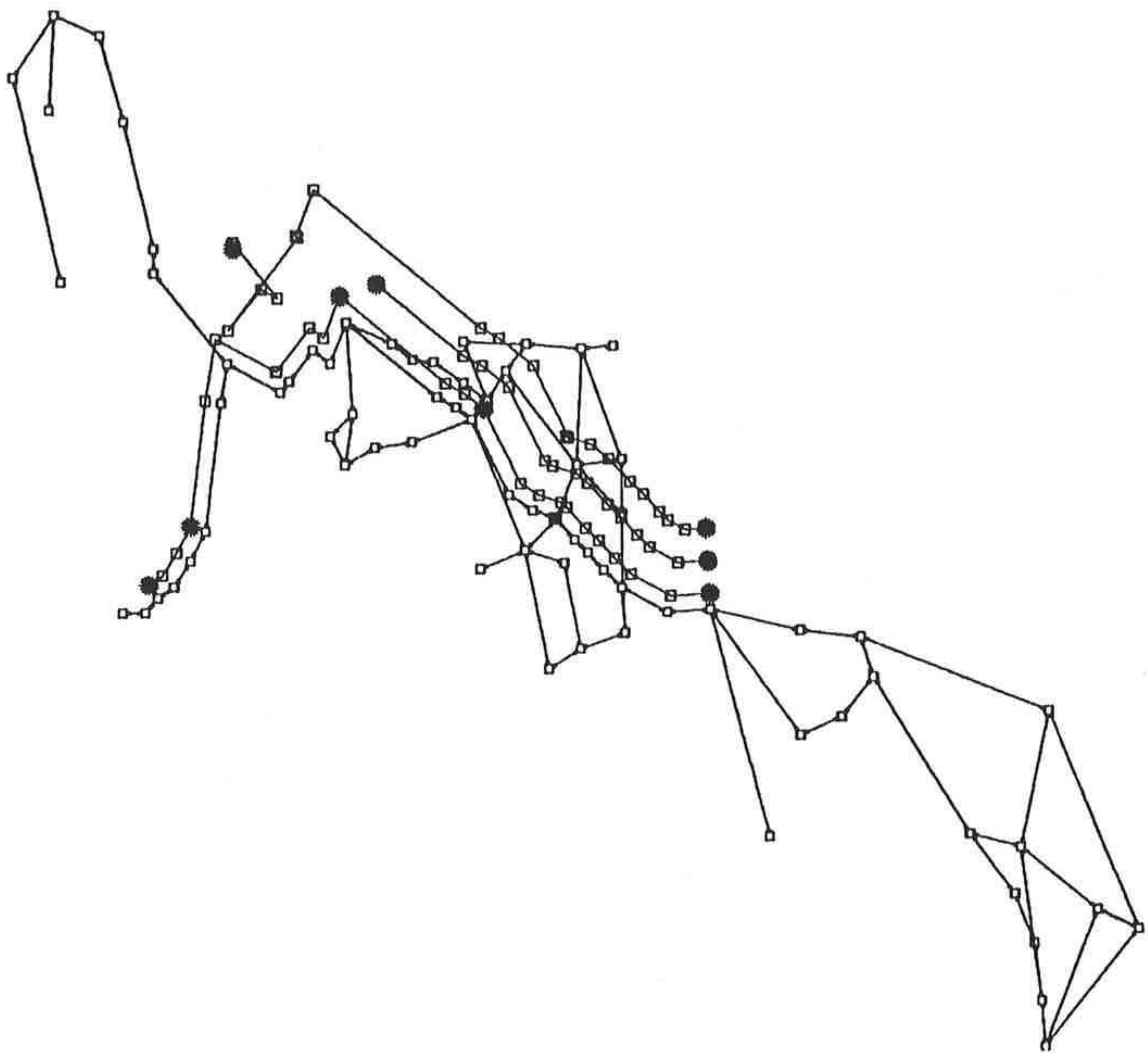


FIGURE 11 Plot of several services on the physical network.

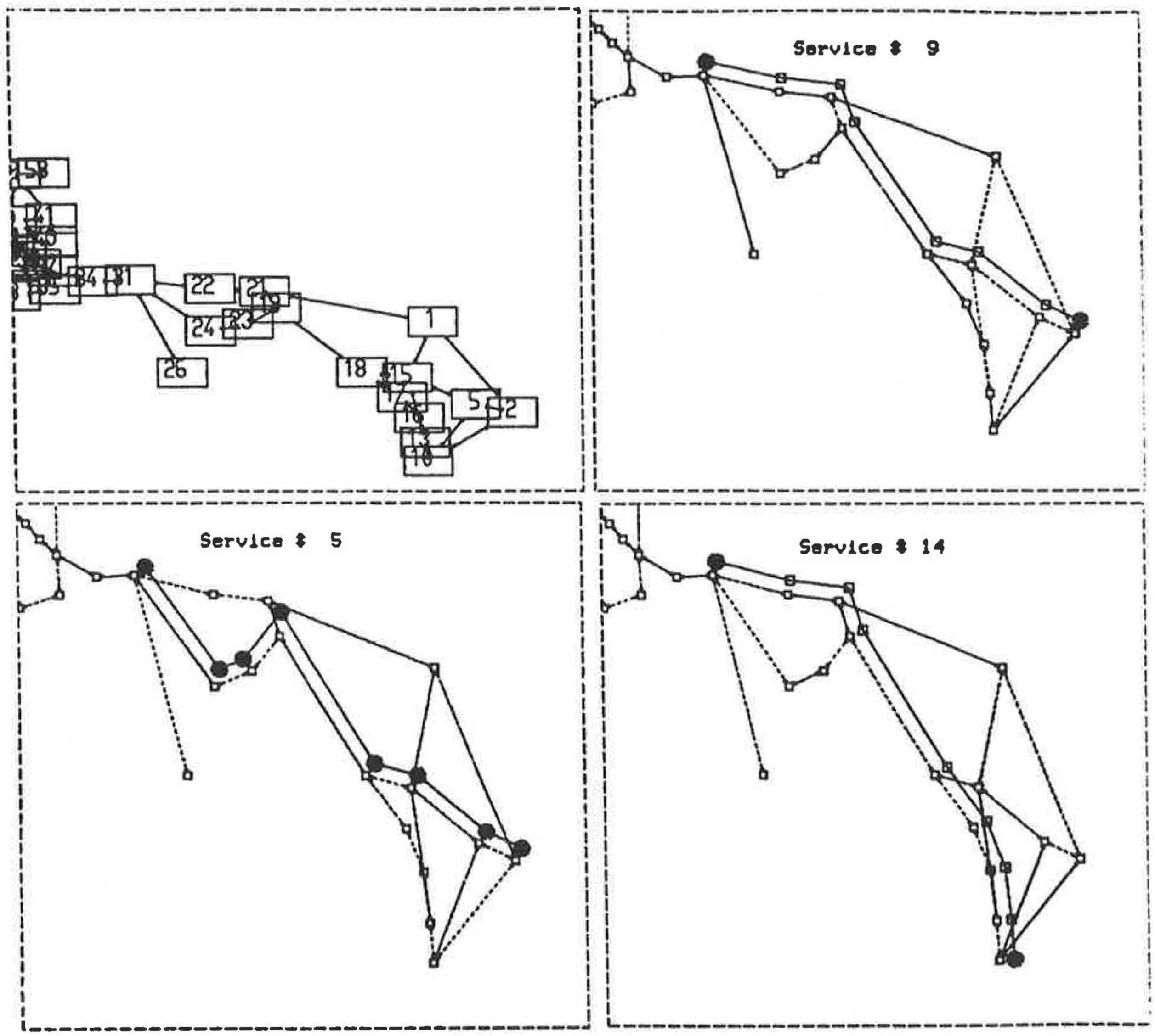


FIGURE 12 User-controlled screen composition—service representation.

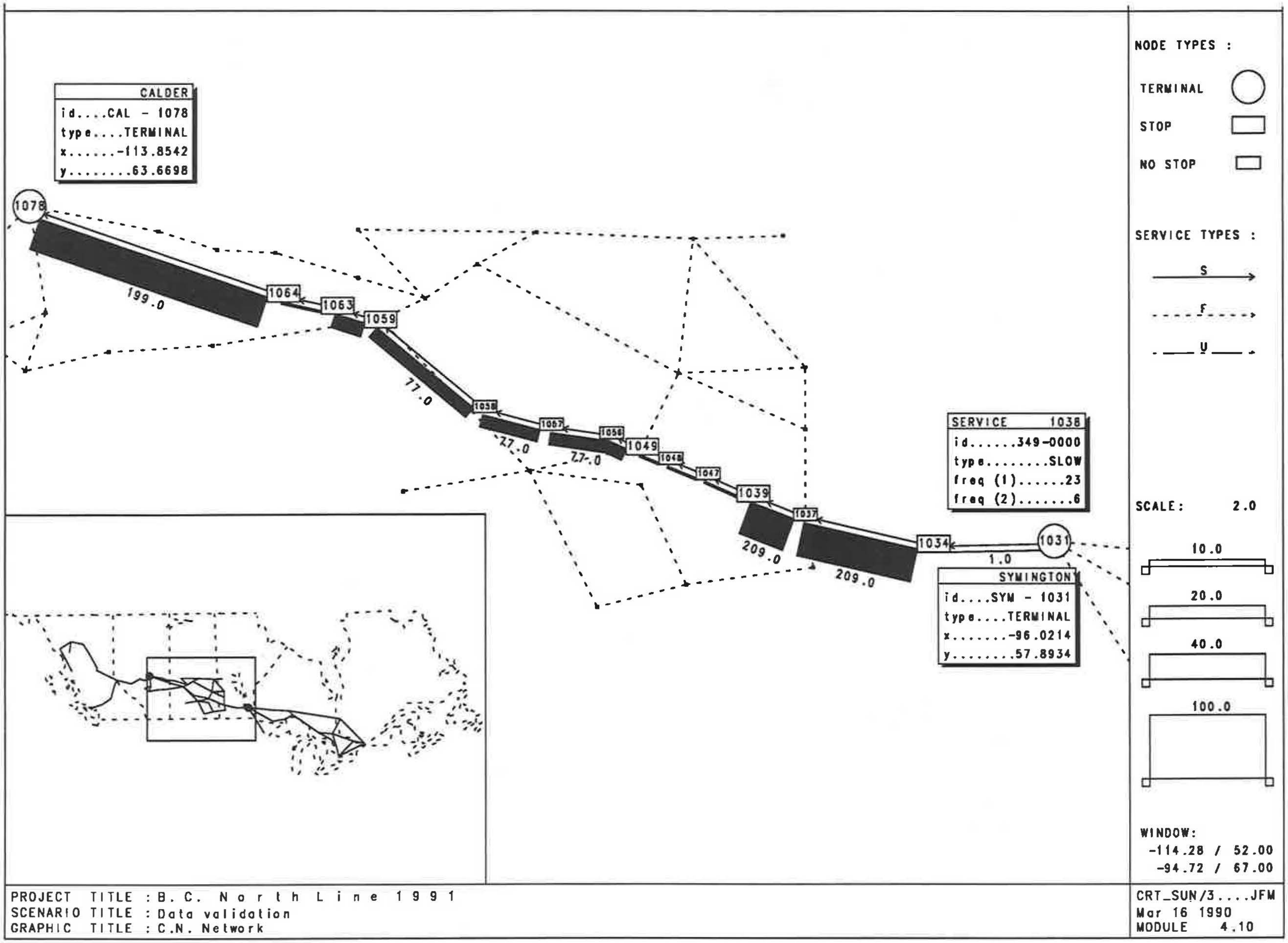


FIGURE 14 Detailed service analysis.

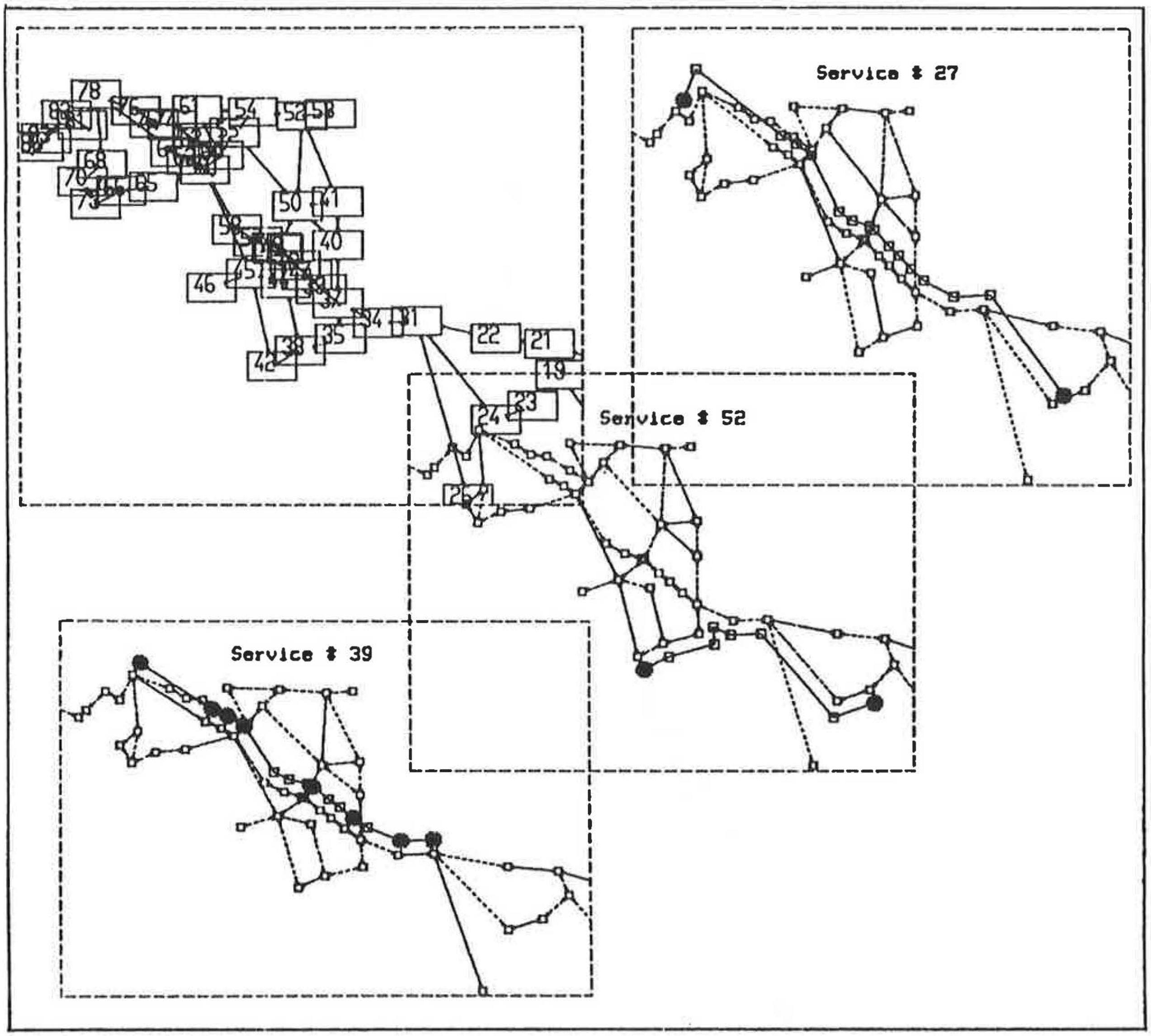


FIGURE 13 "Bad" image composition on screen without selective erasing.

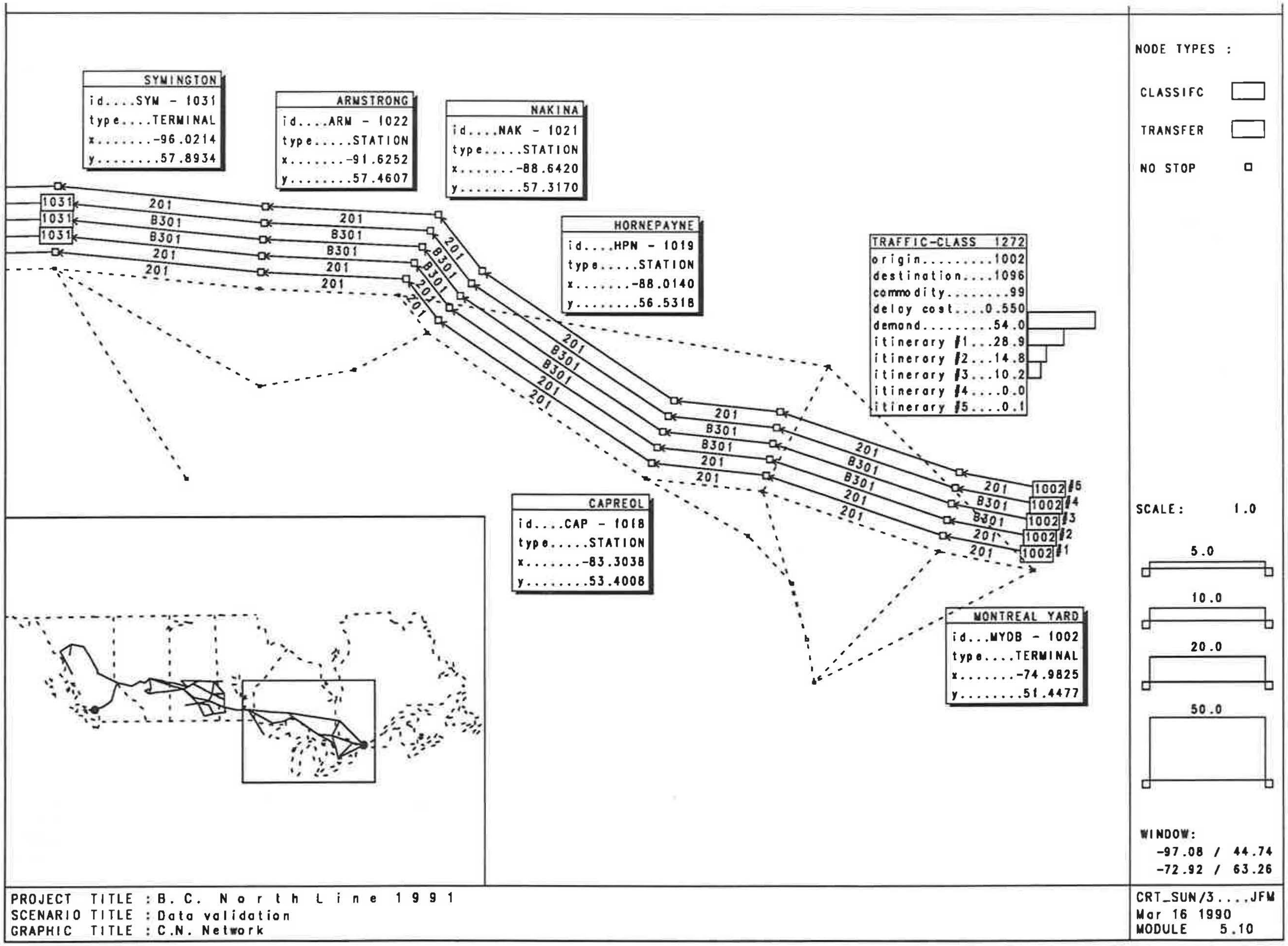


FIGURE 15 Plot of parallel itineraries for a traffic-class.

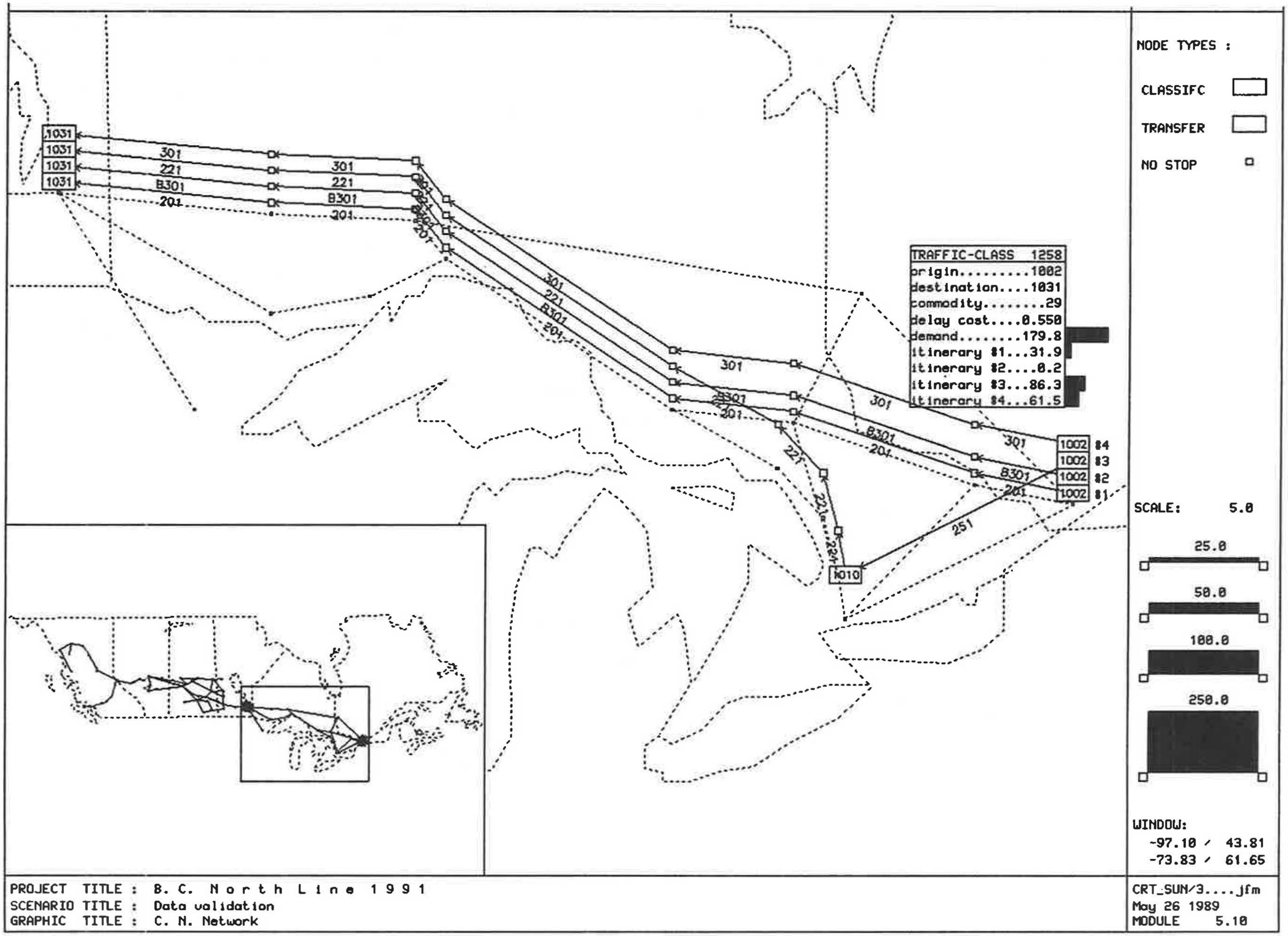


FIGURE 16 Plot of four itineraries for a traffic-class.

The graphic system we propose is particularly efficient for the representation of network features and operations. Moreover, it has been built to allow simple modifications and enhancements required by future developments in computer hardware or software. We made certain not to be linked to one particular graphic architecture, so that our procedures would not be restricted to a few machines.

For our part, we will continue to develop interactive-graphic planning systems that integrate powerful operations research models and algorithms. In this context, we are continually enriching the modeling framework and improving the performances of the solution algorithm. We are also working on the development of the graphical command language, the data bank structure, and all other operations research and computer procedures and modules required to complete a useful system.

REFERENCES

1. S. D. Brady, R. E. Rosenthal, and D. Young. Interactive Graphical Minimax Location of Multiple Facilities with General Constraints. *AIIE Transactions* Vol. 15, No. 3, 1980.
2. A. Babin, M. Florian, L. James-Lefebvre, and H. Spiess. EMME/2: An Interactive Graphic Method for Road and Transit Planning. In *Transportation Research Record 866*, TRB, National Research Council, Washington, D.C., 1982, pp. 1–9.
3. H. Spiess. *Contributions à la théorie et aux outils de planification des réseaux de transports urbains*. Report 382. Centre de recherche sur les transports, Université de Montréal, Montréal, Canada, 1984.
4. D. Lebeuf. *Allocation des appareils pour un horaire aérien: une approche interactive-graphique*. Report 415. Centre de recherche sur les transports, Université de Montréal, Montréal, Canada, 1985.
5. M. Fisher. Interactive Optimization. *Annals of Operations Research*, Vol. 5, 1986, pp. 541–556.
6. B. Sorenson. *Interactive Distribution Planning*. Ph.D thesis. Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, 1986.
7. T. G. Crainic, M. Florian, J. Guélat, and H. Spiess. *Strategic Planning of Freight Transportation: STAN, An Interactive-Graphic System*. Centre de recherche sur les transports, Université de Montréal, Montréal, Canada, 1989.
8. M. Florian and T. G. Crainic, eds., *Strategic Planning of Freight Transportation in Brazil: Methodology and Applications*, 5 volumes. Reports 638-642. Centre de recherche sur les transports, Université de Montréal, Montréal, Canada, 1989.
9. A. A. Assad. Models for Rail Transportation. *Transportation Research A*, Vol. 14A, 1980, pp. 205–220.
10. T. G. Crainic. Rail Tactical Planning: Issues, Models and Tools. In *Freight Transport Planning and Logistics* (L. Bianco and A. La Bella, eds.), Springer-Verlag, New York, 1988, pp. 463–509.
11. T. G. Crainic and J. Roy. O.R. Tools for Tactical Freight Transportation Planning. *European Journal of Operations Research*, Vol. 33, No. 3, 1988, pp. 290–299.
12. W. B. Powell and Y. Sheffi. Interactive Optimization for Motor Carrier Load Planning. *Journal of Business Logistics*, Vol. 7, No. 2, 1986, pp. 64–90.
13. W. B. Powell, Y. Sheffi, K. S. Nickerson, K. Butterbaugh, and S. Atherton. Maximizing Profits for North American Van Lines' Truckload Carriers: A New Framework for Pricing and Operations. *Interfaces*, Vol. 18, No. 1, 1988, pp. 21–41.
14. W. B. Powell and Y. Sheffi. Design and Implementation of an Interactive Optimization System for Network Design in the Motor Carrier Industry. *Operations Research*, Vol. 37, No. 1, 1989, pp. 12–29.
15. T. G. Crainic, J.-A. Ferland, and J.-M. Rousseau. A Tactical Planning Model for Rail Freight Transportation. *Transportation Science*, Vol. 18, No. 2, 1984, pp. 165–184.
16. J. Roy and L. Delorme. NETPLAN: A Network Optimization Model for Tactical Planning in the Less-than-Truckload Motor Carrier Industry. *INFOR*, Vol. 27, No. 1, 1989, pp. 22–35.
17. D. Jovanovic and P. T. Harker. SCAN: A Decision Support System for Railroad Scheduling. In *Impacts of Recent Computer Advances on Operations Research* (R. Sharda, B. Golden, E. Wasil, O. Balci and W. Stewart, eds.), North-Holland, Amsterdam, Netherlands, 1989.
18. L. Delorme, J. Roy, and J.-M. Rousseau. Motor-Carriers Operations Planning Models: A State of the Art. In *Freight Transportation Planning and Logistics* (L. Bianco and A. La Bella, eds.), Springer-Verlag, New York, 1988, pp. 510–545.
19. T. G. Crainic. Operations Research Models of Intercity Freight Transportation: The Current State and Future Research Issues. *Logistics and Transportation Review*, Vol. 23, No. 2, 1987, pp. 189–206.
20. T. G. Crainic and J.-M. Rousseau. Multicommodity, Multimode Freight Transportation: a General Modeling and Algorithmic Framework for the Service Network Design Problem. *Transportation Research B*, Vol. 20B, 1986, pp. 225–242.
21. T. G. Crainic. A Comparison of Two Methods for Tactical Planning in Rail Freight Transportation. In *Operational Research '84* (J. P. Brans, ed.). Elsevier, New York, 1984, pp. 707–720.
22. T. G. Crainic and M.-C. Nicolle. Planification tactique du transport ferroviaire de marchandises: quelques aspects de modélisation. *Actes du premier congrès international en France de génie industriel*, Tome 1, CEFI-AFCET-GGI, 1986, pp. 161–174.
23. T. G. Crainic, J.-F. Mondou, and J. Roy. TACTIC: A DSS for Tactical Planning of Freight Transportation. *International Journal of Physical Distribution and Materials Management*, 1991 (forthcoming).
24. J.-F. Mondou. *Mise au point d'un système interactif-graphique pour la planification tactique du transport des marchandises*. M.Sc. thesis. Département d'Informatique et de Recherche opérationnelle, Université de Montréal, Montréal, Canada, 1989.
25. A. L. Kornhauser. *Examples of Corporate Distribution and Transportation Planning Using the Princeton Transportation Network Model and Graphic Information System*. Civil Engineering Department, Princeton University, Princeton, N.J., 1985.
26. A. L. Kornhauser. *Very Large Transportation Network Interactive Computer Graphic Information System and Problem Solver*. Civil Engineering Department, Princeton University, Princeton, N.J., 1985.
27. J.-F. Mondou, T. G. Crainic, and S. Couture. *BGC: Bibliothèque de fonctions graphiques en language C*. Report 554. Centre de recherche sur les transports, Université de Montréal, Montréal, Canada, 1987.

Publication of this paper sponsored by Committee on Transportation Supply Analysis.

Roadway Automation Technology—Research Needs

STEVEN E. SHLADOVER

Although the concept of roadway automation has been in the public eye for nearly 50 years, and although it has now been 30 years since the first test track demonstrations of some of the requisite technologies, virtually no progress has been made toward implementation. This paper reviews the history of roadway automation development efforts and explains why it is worth re-examining roadway automation now in light of current transportation needs and technological progress. The main body of the paper outlines the technical questions that need to be answered in order to make roadway automation a reality, based on the nine functions of an automated roadway system: (a) intelligent traffic signalling; (b) traffic information systems; (c) driver warning and assistance systems; (d) automatic steering control; (e) automatic spacing control; (f) obstacle avoidance; (g) automatic trip routing and scheduling; (h) control of merging of strings of traffic; and (i) transitioning to and from automatic control.

The idea of roadway automation has been with us for a long time, although it has been manifested in many different guises. While the early futurists' visions of modern life (such as the General Motors "Futurama" of the 1939-40 World's Fair) may seem quaint to us now, much of what was contained in those visions has yet to be realized. It took nearly 20 years for the "driverless car" vision of the Futurama to appear on a test track for the first time, implemented with real hardware. In the 30 years since those first experiments, we have moved no closer toward the automated roadway in spite of the dramatic advances in many of the underlying technologies needed to make it work. Even if the automated roadway was ahead of its time 50 years ago and again 30 years ago, we owe it to ourselves to give it a serious look now, as we enter the final decade of the twentieth century.

Roadway automation technology can be interpreted to mean a variety of different things. For purposes of the discussion here, it is defined to be the application of communication and control technology to observe, guide, and/or control the movement of vehicles in a traffic system (or to assist in the performance of those functions). This definition is designed to be sufficiently broad to include advanced traffic signalling and monitoring systems used by drivers retaining control of their vehicles, as well as fully automated "driverless" vehicles, but not so broad as to include automation of special purpose exclusive guideway systems (BART, AIRTRANS, etc.). The terminology "roadway automation" is in itself controversial and has unfavorable connotations for some people. Other authors have chosen to refer to this subject as "intelligent vehicle-highway systems" or "road transport informatics." The ideal terminology has not been found.

Program on Advanced Technology for the Highway, University of California at Berkeley, 1301 S. 46th St., Bldg. 452, Richmond, Calif. 94804.

The potential benefits to be gained from roadway automation have grown rapidly in recent years, as the underlying technologies have been progressing (and therefore becoming more economical). The primary benefits fall into five categories:

1. Providing a cost-effective roadway capacity increase. By enabling vehicles to travel closer together (longitudinally and laterally), automation can increase the capacity of the existing roadway infrastructure at a lower cost (and with minimal disruption to the neighborhood) than the addition of more lane-miles of roadway. The increased capacity translates into reduced travel delays and, when combined with electrification of the vehicles, reduced environmental impacts as well.
2. Improving safety. By eliminating human error and foolhardiness from the control of vehicles on the roadway, the accident rate can be reduced, and with it the associated deaths and injuries, as well as the additional delays that accidents now impose on traffic flow.
3. Improving speed and reliability. Roadway automation should make it possible to travel safely at higher speeds than at present in urban areas. This speed increase can be obtained without sacrificing potential capacity increases, because fully automated vehicles are not governed by the same speed-volume relationships that apply to human drivers in normal traffic flow. By requiring stringent condition checking of all vehicles that enter an automated roadway, breakdowns should also be significantly reduced. Even when breakdowns and accidents do occur, their impacts on traffic flow should be greatly reduced by the elimination of "rubbernecking" by drivers in other lanes.
4. Increased convenience of travel. Roadway automation eliminates the tedium and stress of driving for travelers who are disabled, fatigued, too old, or too inexperienced to drive effectively.
5. Maintaining industrial competitiveness. If the United States does not proceed aggressively in the development of roadway automation technology, the rest of the world will not wait for us. Active, large-scale programs in Western Europe and Japan are developing roadway automation technology to ensure that their automotive industries will survive in the coming decades. U.S. automotive manufacturers can look forward to accelerated erosion in their shares of both the domestic and international markets unless this country remains competitive in the development of this class of technology.

In order to reduce a large subject to units of manageable scale, it is useful to subdivide the functions of an automated roadway system into nine categories:

1. Intelligent traffic signalling;
2. Traffic information systems;
3. Driver warning and assistance systems;
4. Automatic steering control;
5. Automatic spacing control;
6. Obstacle avoidance;
7. Automatic trip routing and scheduling;
8. Control of merging of strings of traffic; and
9. Transitioning to and from automatic control.

Each individual category represents a very useful function, and each will require considerable technology development before it can substantially benefit the traveling public. A system that serves any one of these functions could be highly beneficial, but truly dramatic improvements in our urban traffic systems will not be enjoyed until the "automatic vehicle control functions" (4–9 above) are available. However, these vehicle control functions cannot be provided until the "driver information functions" (1–3 above) are available.

The above listing, which serves as the framework for the main body of this paper, can also serve to indicate the sequence in which the automation functions are likely to become available. It is not a statement of research priorities, however, because the (later) automatic vehicle control functions are also the most difficult and will require the greatest investments of time and labor and the longest development lead times. Of course, they also offer the greatest potential capacity, safety, and travel time benefits and therefore deserve high current priority. The automatic vehicle control functions are essential if the capacity of the existing roadway infrastructure is to be doubled or tripled, if speeds are to be substantially increased or if accidents are to be reduced dramatically.

Consider, for example, the capacity increase that could be obtained by applying automatic steering and spacing control to a freeway or bridge having five 12-ft lanes. If steering control makes it possible to reduce lane width from 12 ft to 8.5 ft, the same pavement and structure could accommodate seven lanes instead of five. If automatic spacing control allows vehicles to operate at half-second average headways, each lane could accommodate 7,200 vehicles per hour (vph), rather than the approximately 2,000 vph achievable under driver control. The combined effect is an increase in capacity from 10,000 vph to more than 50,000 vph for the same structure and right of way.

With careful design of the control systems for safety and reliability, the incidence of accidents could be reduced significantly, particularly those caused by driver inattention and lapses of judgment. Furthermore, automatic control systems will completely eliminate the "rubbernecking" phenomenon, so that a breakdown or accident in one lane will not cause the vehicles in all the other lanes (in both directions) to slow down. The reduction of accidents and elimination of rubbernecking will substantially reduce the disruptions to traffic flow arising from incidents, enabling effective lane capacity to approach the theoretical limit the large majority of the time. Smooth flow, at constant speed, with the elimination of "stop and go" instabilities, also offers the potential for major improvements in energy efficiency and pollutant emissions.

The implementation concepts for roadway automation cover the range from completely centralized control of every vehicle in a network to fully autonomous vehicles with no centralized control (sometimes referred to as "automatic chauffeuring" in Europe). Each of these extremes has important limitations, so any realistic system is likely to represent a compromise between them. Such a compromise solution means that an automated roadway system will include significant elements that are both privately and publicly owned and controlled.

HISTORICAL BACKGROUND OF AUTOMATIC VEHICLE CONTROL SYSTEMS

Although we are no closer to driving on automated roadways now than we were 30 years ago, there has been much engineering activity devoted to roadway automation during this period. During the late 1950s, General Motors and RCA developed and demonstrated (on test tracks) automatic control of steering and longitudinal spacing of automobiles for what they called the "Electronic Highway" (*1,2*). By 1962, the first university research on automatic steering control of automobiles was reported from Ohio State University (OSU), under the sponsorship of the Ohio Department of Highways and the Bureau of Public Roads (*3*). This led to a later, long-term research program at OSU on both steering and longitudinal (spacing) control, under the sponsorship of the Ohio Department of Transportation and the Federal Highway Administration (FHWA), from 1965 to 1980 (*4,5*). The first broad-scale investigation of the application of automation technologies to urban transportation problems appears to have been in the MIT Project METRAN, in the spring of 1966(*6*).

Major federal government involvement in the pursuit of roadway automation began with the New Transportation Systems Research Act of 1966, which instructed the U.S. Departments of Housing and Urban Development and Commerce to study new systems for urban transportation [since the Department of Transportation (DOT) did not yet exist]. This "New Systems Study" led to the creation of series of reports from the Stanford Research Institute and General Research Corporation, culminating in the summary report, *Tomorrow's Transportation: New Systems for the Urban Future,* in 1968 (*7*). Six new urban transportation systems were recommended for development in that report, one of which, called "Dual Mode," was essentially the automated roadway using vehicles that could operate in either a manual or an automated mode (hence the name).

Although several research programs grew out of the recommendations in *Tomorrow's Transportation*, the progress in the dual-mode area was particularly slow. The DOT Transportation Systems Center (TSC) conducted a major economic evaluation of dual mode in 1971-73, leading to the recommendation that dual-mode systems be developed for both transit and private (automotive) vehicles (*7*). UMTA initiated studies of three different dual-mode transit (bus and pallet) systems in 1973, but the program never advanced beyond the initial concept development phase, and "dual mode" soon became a dirty word within the transportation establishment.

The work on dual mode and automated highway technologies was reported in a Transportation Research Board conference in 1974, although the proceedings unfortunately did not contain full texts of many of the papers (*9*).

In addition to the ongoing work at OSU, FHWA sponsored a study of the feasibility of an automated highway system, leading to a report in 1977 that recommended several areas needing further research (*10*). A second FHWA study of automated highways by General Motors in 1980-82 examined the technical potential of the automated highway in the context of realistic implementation problems (*11*). The most recent thinking of the FHWA in this area is represented in Working Paper No. 7 of the FHWA's "Future National Highway Program—1991 and Beyond" (*12*).

Privately funded work on the development of roadway automation technologies is not generally reported because of proprietary restrictions on release of information. This makes it difficult to tell how much activity there actually has been in this area. Certainly General Motors has been active in work on roadway automation technology since the time of their original work, although only limited glimpses of this have been available to the research community in general (*13*).

A large portion of the technology development work relevant to roadway automation has actually been performed in pursuit of automated guideway transit (AGT) systems, formerly referred to as personal rapid transit systems. This activity was primarily sponsored by the Urban Mass Transit Adminsitration (UMTA) during most of the 1970s and produced a large body of research results, which must be absorbed and understood by anyone hoping to advance the technologies for roadway automation. AGT researchers had to address most of the issues that must be solved for roadway automation, except for the problems associated with transitions to and from automatic control and with developing vehicles that can operate under both manual and automatic control. The results of the AGT research are scattered across an extensive range of reports, journal papers, and conferences. The central references, from which a more comprehensive literature search can be started, are the four conference proceedings (*14–17*).

The development of the roadway automation technologies never achieved a high enough priority on the national agenda to proceed very rapidly. In the late 1970s its priority declined, leading to a virtual disappearance in the 1980s. Virtually no new domestic work on vehicle automation has propagated into the literature since 1980.

Overseas, however, roadway automation work has recently seen a dramatic revival (*18*). There are hopes for a domestic revival with the advent of the PATH program in California and Mobility 2000 on the national level. Indeed, the climate should be more favorable for roadway automation now than it has ever been in the past, because of many of the changes that have occurred within the past decade:

- Significant growth in urban and suburban congestion;
- Further increases in the real costs of contruction of new roads;
- Public opposition to construction of new roads;
- Environmental problems created by the use of fossil fuels for transportation (pollution, greenhouse effect, petroleum dependency);
- Dramatic technological advances in sensing technologies, computer hardware and software, communication technologies, and control theory;
- Increased foreign competition for the automotive industry; and
- The threat of losing U.S. technological leadership in the world.

All of these factors provide encouragement for a concentration of effort on the development of roadway automation technologies. Such development efforts must not go about reinventing the wheel, but must build on the generations of work that have gone before, and must be carefully planned in order to avoid wasting time and effort. The development work must be based on clear statements of goals and a clear understanding of how to go about achieving those goals. The remainder of this paper outlines an agenda for such development efforts, indicating the technical problems that need to be solved before we can enjoy the benefits of automated roadways. More detailed statements of the research needs outlined in the next section of this paper may be found elsewhere (*19*).

CRITICAL TECHNICAL ISSUES

The operations embedded within a roadway automation system primarily consist of gathering data, processing the data, and then communicating the results elsewhere for action. These operations can be understood best by illustrating them schematically, showing the flow of data from one component of the system to another (see Figure 1). The components to be considered are:

- The vehicle being controlled (or more specifically its sensors and its onboard computer);
- The driver of the vehicle;
- The roadway;
- External objects;
- Other vehicles and their drivers;
- Wayside computer(s); and
- Wayside traffic signals.

Each of the nine functions of roadway automation can be described in terms of the data flows among these components. Once those data flows are specified, it becomes easier to focus on the technical issues that must be resolved in order to implement each automation function.

Intelligent Traffic Signalling

Modern traffic signals already incorporate some elements of "intelligence," whether it be programming that varies with the time of day or some responsiveness to measured traffic conditions (generally vehicle presence detection). This function assumes that the vehicles of today continue to operate entirely under driver manual control, making it the least advanced and nearest-term of the nine automation functions. All the technology changes would be on the wayside, in the collection of real time traffic data, processing of the data, and displaying it to drivers. Considerable advances have been made toward implementing this function in demonstration systems in recent years, such as the Los Angeles "Smart Corridor," the Australian SCATS, and the British SCOOT and CITRAC systems (*12*).

The data flows needed to implement intelligent traffic signalling [also referred to as advanced traffic management system—(ATMS)] are illustrated in Figure 1. Most of these data flows already exist in current traffic systems and involve no automation functions. The new development work will need to be focused on collecting data about the vehicles as they travel, processing that data, and displaying it to the drivers so they can take appropriate actions. The technical issues that must be addressed here include:

- Selection of performance measures by which to judge success;
- Selection of the information to be collected;
- Selection of monitoring and communication methods;
- Development of control algorithms;
- Application to an urban corridor; and
- Application to a complete urban region.

Traffic Information Systems

Traffic signal systems are, of course, traffic information systems of one limited type. However, this function also includes advanced driver information systems (ADIS): two-way communication of information between the wayside computer system and the vehicle's onboard computer, which serves as the interface to the driver. This requires that the vehicle owner invest in additional equipment on the vehicle, but it continues to leave the driver in complete control of the vehicle. As with the previous function, much of this functionality is starting to enter some urban demonstration projects. These include the "Pathfinder" demonstration planned for the Los Angeles "Smart Corridor," the AMTICS (Advanced Mobile Traffic Information and Communication Systems) demonstration in Tokyo (*20*), and the ALI-SCOUT demonstration in Berlin (*21*).

The data flows needed to implement a comprehensive traffic information system are illustrated in Figure 2. As before, many of these flows already exist as a natural part of driving. The new developments involve the communications between the wayside computer and the computers on the individual vehicles (as well as the processing of the data at both ends). The technical issues to be addressed here include:

- Selection of performance measures;
- Selection of the information to be collected;
- Selection of monitoring and communication methods;
- Selection of information to display to the driver;
- Selection of how to convey information to driver;
- Development of the technologies to convey the information to the driver;
- Development of control algorithms; and
- Selection of centralized functions (AVM, etc.).

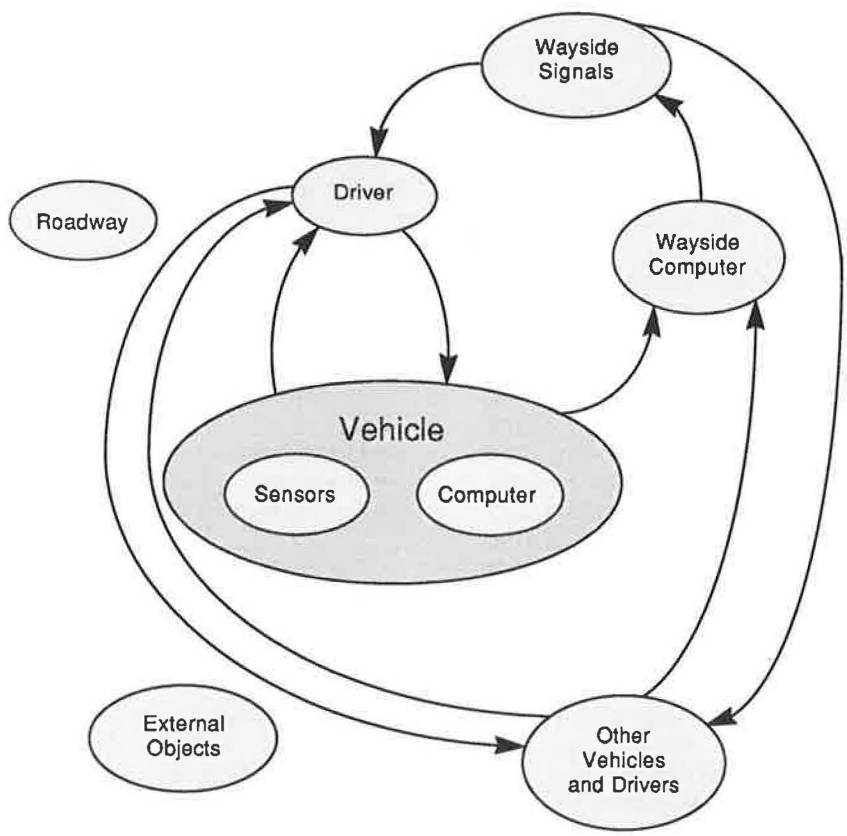


FIGURE 1 Information flows in intelligent traffic signalling system.

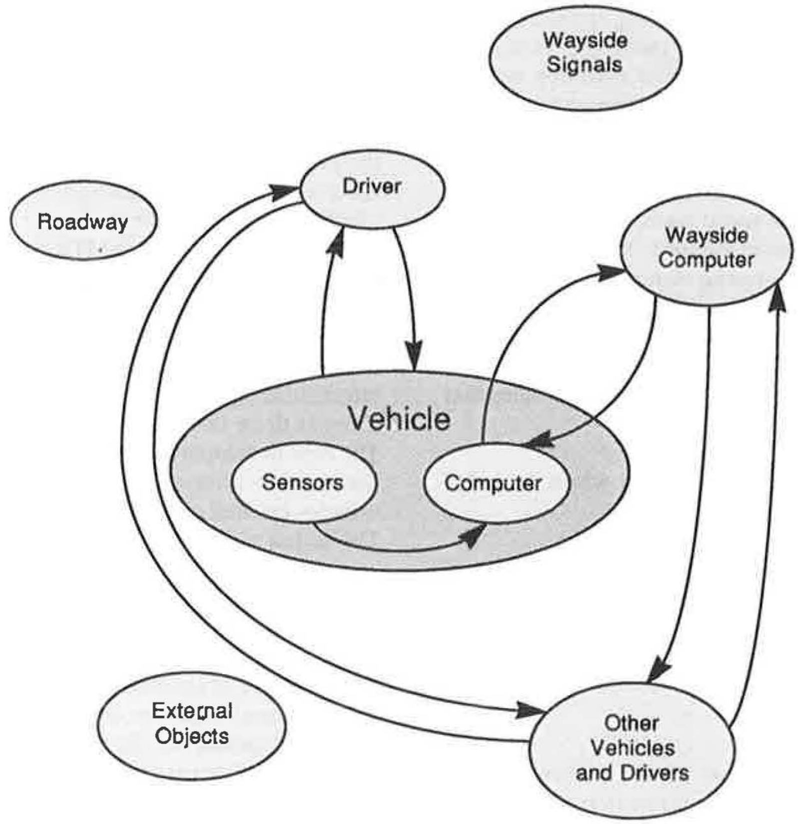


FIGURE 2 Information flows in traffic information system.

Driver Warning and Assistance Systems

Modern electronics can make automobiles safer and easier to drive in a variety of ways (antilock braking systems and cruise control, for example). However, there are realistic near-term opportunities available for significantly more dramatic electronic driver aids, performing higher-level functions than those now available on automobiles. These can be implemented entirely onboard the vehicle, without requiring the involvement of any public agency or wayside facility.

The data flows needed to implement driver warning and assistance functions are shown in Figure 3. The driver interacts with the automobile as he would under any normal circumstances. However, his vehicle would also be equipped with sensors to detect the vehicle's status relative to the roadway, as well as other vehicles and indeterminate external objects. The sensors would supply information to an onboard computer, which would in turn signal the appropriate information to the driver or the vehicle's other systems. Examples of this include collision-warning radar, near-obstacle detection, lane-center detection for driving assistance, and automatic steering compensation for external force disturbances.

The collision-warning radar and near-obstacle detection systems would issue a warning to the driver of impending danger, in effect supplementing his own eyes and ears. These systems could begin the evolution to the obstacle avoidance function, described below, in which the vehicle would take the corrective action without the driver's intervention. The lane-center detection system could help the driver steer in conditions of bad visibility (rain, snow, fog) or when he is fatigued, by supplying a dashboard or heads-up display of the vehicle's position relative to lane center. The steering compensation system could estimate the effect on the vehicle's steering of sudden external force disturbances (wind gusts or wakes of passing vehicles) in real time and compensate for them by adjusting the steering system, without requiring the driver to make abrupt and stressful corrective maneuvers.

The technical issues that must be addressed to lead to the development of such systems tend to be quite system-specific, but they can, in general, be related to the sensing, computation, and display functions:

- Development or selection of sensors;
- Development of computational capability; and
- Development of display capability.

Automatic Steering Control

With this function, we clearly cross the threshold from driver control of the vehicle to automatic control of one of the key driving operations. This will require extensive modifications to the vehicles, as well as some equipment very carefully installed within the roadway. This function therefore requires cooperation between the vehicle and roadway (and hence between the public and private sectors) in order to work. Fully automatic steering control is not likely to be implemented on a large scale independent of the other automatic vehicle con-

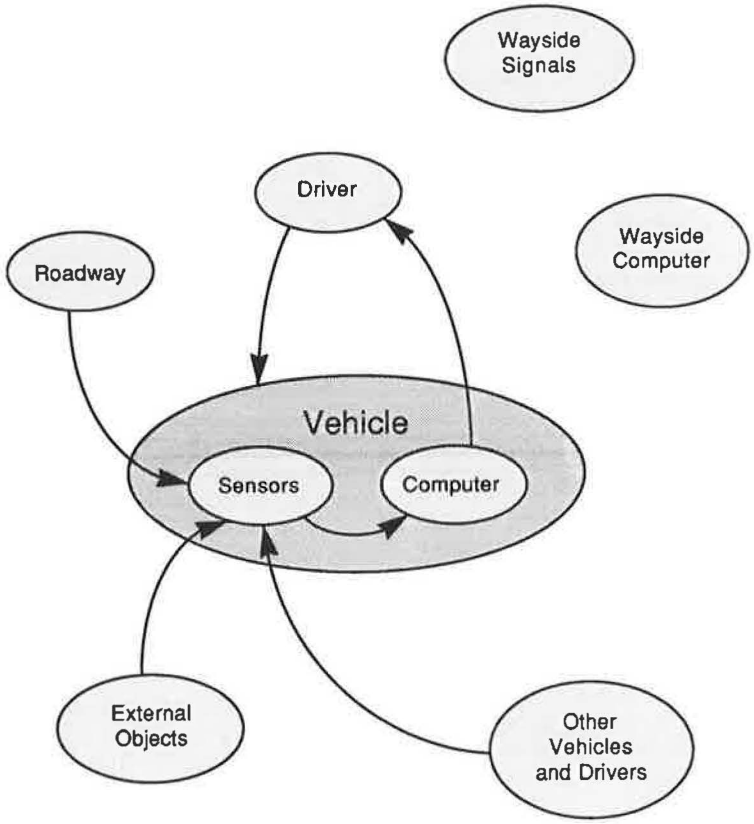


FIGURE 3 Information flows in driver warning and assistance systems.

trol functions (automatic spacing control, obstacle avoidance, etc.) because of the safety implications, particularly in terms of driver alertness. There does not appear to be a safe way to implement partial automation, because a vehicle must be under the control of either the driver or the automatic system, but not a mixture of the two.

The data flow needed to implement automatic steering control is simply from the roadway to the on-vehicle sensors to the computer. The complexity in automatic steering control is found at a lower level in the hierarchy of data flows. More than any of the other nine functions, the automatic steering function can be decoupled from the operations of the other vehicles in the traffic system and from the wayside computer system(s). The technical issues to be addressed here are:

- Selection of steering reference and sensors;
- Adaptation to varying conditions;
- Tradeoff between accuracy and comfort;
- Design of feedback control law;
- Development of actuation system;
- Provision of redundancy, backups, and fault accommodation; and
- Tradeoff between vehicle and roadway costs.

Automatic Spacing Control

This function represents another example of a driver ceding the control of his vehicle to an automatic system, with all of the reliability and safety implications that entails. The data flows are more complicated than they were for the automatic steering because of the strong interactions with the other vehicles on the road and with the wayside computer. The vehicle being controlled must detect its proximity to adjacent vehicles, using special sensing equipment, and the wayside computer needs to keep track of all of the nearby vehicles so that it can provide the appropriate instructions to the controlled vehicle's onboard computer. Although it is in theory possible to implement some spacing control without involving the wayside computer, that would require each vehicle to obtain information about vehicles other than those immediately adjacent to it, which would introduce other problems of comparable complexity. The technical issues to be addressed here are:

- Selection of safety criteria;
- Defining performance requirements;
- Defining permissible demands on vehicle performance;
- Tradeoffs between vehicle and wayside intelligence;
- Tradeoffs between vehicle following and point following;
- Selection of references and sensors;
- Definition of communication requirements;
- Definition of feedback control laws for asymptotic stability;
- Defining reliability requirements;
- Provision of redundancy, backups, and fault accommodation; and
- Deciding whether vehicles should be operated individually or in platoons.

Obstacle Avoidance

The automatic steering and spacing control functions described previously are intended to address vehicle operations in the presence of other automated vehicles, but not in the presence of other (unanticipated) obstacles. The appearance of a foreign object in the path of an automated vehicle has serious safety implications. If the object should be a person or a domestic animal, one would want to have the vehicle stop or swerve to avoid running over and harming it. If the object should be large or heavy, the vehicle should avoid running into it to save both passengers and vehicle from harm. On the other hand, if the object is inconsequential (leaves, newspaper, squirrel) the vehicle should continue on its way rather than changing its course and producing delays.

The most important flows of information are from the external objects to the onboard sensors and then from the sensors to the computer. The other data flows are needed to warn adjacent vehicles of the evasive measure about to be taken by the vehicle that detects the foreign object. This kind of direct advance warning is needed to ensure high performance and safety in a short-headway, high capacity system.

The obstacle detection and avoidance function described here differs from the driver warning and assistance function, in that it extends beyond providing a warning to the driver and includes directing the vehicle to take corrective actions as well. The technical issues that must be addressed here include:

- Development of sensing technology;
- Development of data processing capability; and
- Development of coupling to vehicle steering, speed, and spacing control systems.

Automatic Trip Routing and Scheduling

This function enters a new realm of scope and complexity, considering the number of vehicles it involves and the volume of the information that must be handled. Automatic trip routing and scheduling can assume a wide range of forms, with different degrees of centralization. On the one extreme, it can consist of complete prescheduling of every vehicle trip from origin to destination at the time the vehicle enters the automated system. On the opposite extreme, it can consist of en route rerouting of vehicles to obtain smoother flows and better utilization of line-haul capacity throughout the system, plus coordination of merge junctions and vehicle entries and exits. Many intermediate concepts are possible as well, offering a variety of tradeoffs between complexity and performance.

The data flows required for automatic trip routing and scheduling involve two-way communication between each vehicle and the wayside computer(s). The quantity of information to be passed to and from each vehicle is likely to be relatively limited (origin, destination and present locations, plus the command to turn or not turn at each demerge junction), but because of the number of vehicles that could be operating at once in a large-scale system, the total communication burden could be very large. The heaviest burden is in the computations that must be performed by the wayside computer(s) in real time. The technical issues that must be addressed here include:

- Selection of performance measures,
- Design of network for desired capacity,
- Tradeoff between prescheduling and en route rescheduling,
- Selection of degree of centralization of control,
- Development of routing and scheduling algorithms, and
- Detection of capacity limit in real time.

Control of Merging of Strings of Traffic

An essential feature of an automated roadway that consists of more than a single network link is automatic control of merges, which include lane changes, entries, and exits. Consecutive vehicles in a string of traffic may need to branch off to different destinations, or two strings of traffic may need to merge to form a single string (which poses more difficult problems). The wayside computer serves a central role in exchanging information to and from all the vehicles involved. In addition, the vehicles may communicate directly with each other, sensing the positions of their neighbors and possibly even commanding their neighbors to make minor course corrections. The technical issues to be addressed here include:

- Definition of safety requirements;
- Definition of priorities at merges;
- Definition of sensing requirements;
- Definition of communication requirements;
- Definition of reliability requirements;
- Provision of redundancy, backups, and fault accommodation; and
- Definition of merging for platoons.

Transitioning to and from Automatic Control

The final function of roadway automation is unique to the roadway automation technology, because it is not an issue for manually driven vehicles or for fully automated captive guideway vehicles. This is the function of transitioning between the manual and automatic modes of operation, where the driver cedes control of the vehicle to the automatic system upon entry to the automated roadway and then recovers control upon leaving the automated roadway.

The driver communicates with his vehicle and the vehicle's computer communicates with the wayside computer to effect the handoffs between the two modes of control. Safety and reliability issues are of prime importance here, in order to ensure that a vehicle is in good working order before it is admitted to the automated roadway and to ensure that the driver is awake, alert, and ready to take command when it returns to the normal street system. The technical issues to be addressed here include:

- Integration of entrances and exits with local street system;
- Entrance of a vehicle to the automated system; and
- Exit of a vehicle from the automated system.

The Fully Automated Roadway System

The fully automated roadway would incorporate all of the functions from category 4 to 9 described above, employing the information flows shown in Figure 4. It would obviously be quite a complicated system, and one that would take years to develop and implement. There is no single "best" way to implement the research and development plan for such a system. Each of the research groups working in this field will be devoting considerable effort to developing its own plan. The definition of needed elements or capabilities, as outlined above, is a necessary first step. However, development of a research plan is well beyond the scope of this paper.

The most important issues in the development of a research plan are ensuring that the plan has a coherent focus and that it follows good system engineering principles. This means that it should proceed from the statement of needs to mathematical modeling and analysis, design, and implementation of limited-scale experiments, and then multiple cycles of design refinement and testing, before proceeding to public demonstrations of any substantial scale. Politically inspired near-term demonstrations are very risky, as proven by the damage done to the entire AGT activity by the Morgantown demonstration of the early 1970s.

Many of the technologies needed for the automated roadway have been developed and demonstrated in experimental form during the past 30 years, but very little of that has been in evidence during the past decade. The first priority should therefore be in rebuilding the knowledge base of that prior work and then applying the new technological developments of today to that foundation in carefully conceived experiments. We should expect a major application of effort to providing the needed reliability and safety for each of the functions and to reducing the costs of the needed equipment. The most difficult topic is likely to be the development of generally applicable automatic routing and scheduling methods for complicated networks. The routing and scheduling issues have only been addressed for small, topologically simple systems to date, and even these have proven to be very difficult.

POLICY ISSUES AFFECTING ROADWAY AUTOMATION

In spite of the very large potential benefits that could be gained from large-scale implementation of the more advanced roadway automation functions (i.e., those involving automatic vehicle control), there has been virtually no movement in that direction by any public agencies. It is apparent that a similarly large set of potential problems stands in the way of progress toward complete roadway automation. The solutions to these problems will be challenges of a very different type from the solutions to the technical problems described above.

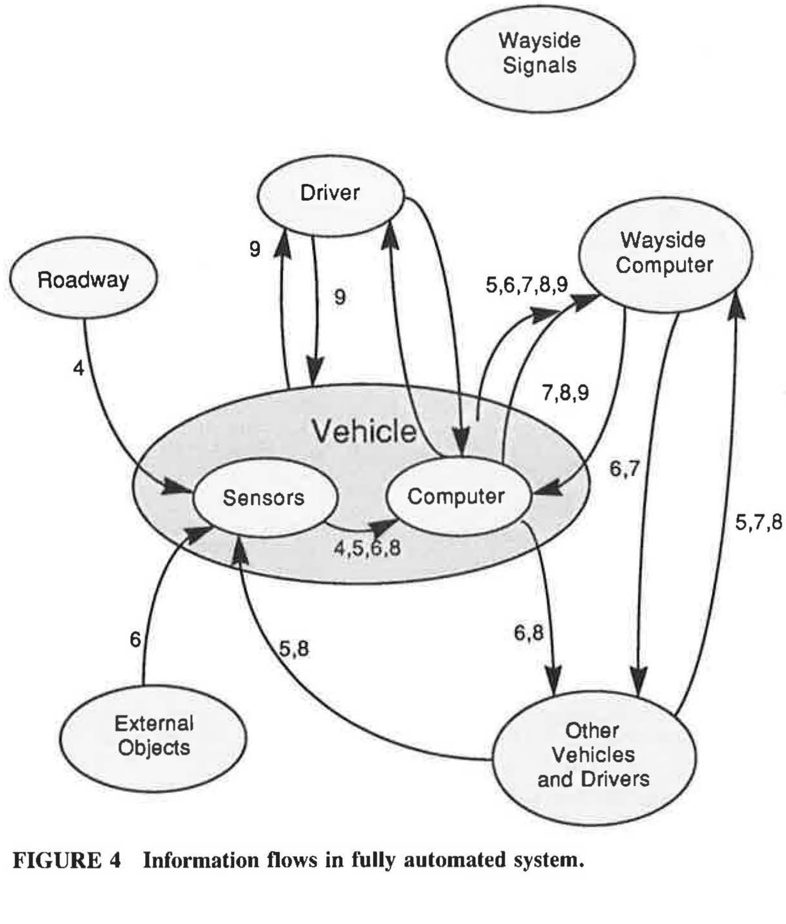


FIGURE 4 Information flows in fully automated system.

Political Need for Near-Term Results

Politicians need near-term results, which can be demonstrated to their constituents before the next election, in order to develop enthusiasm for new programs. Roadway automation is not well suited to this need because of the lengthy, laborious research needed to bring its most attractive features to the point that they can be demonstrated. Some incremental progress can be made towards short-term goals, but only at the expense of sacrificing the much greater long-term benefits of solving the harder problems.

Scale Effects

The benefits of roadway automation have very strong scale effects. A small-scale automation project or a project in a small city will show relatively small benefits, and quite possibly it will not show benefits that exceed its costs. The dramatic benefits of this technology can really be gained only with large-scale implementation in the largest and most congested metropolitan areas, where it can most significantly reduce delays and increase capacity. It is obviously not practical to make such a large and risky investment until the technology has been proven to work on a smaller scale. By the same token, it is difficult to justify the small-scale demonstration on its own merits if its benefits will not exceed its costs. Some strong political leadership, imbued with a long-term perspective, will be needed to overcome this problem.

The Chicken and the Egg

Apart from the scale effect problem cited above, there is an additional "chicken and egg" problem to confront. It is difficult to gain public support for the infrastructure investment needed to install a roadway automation system if very few people have vehicles equipped to use the system. At the same time, people will not be willing to invest in the additional equipment needed to automate their private vehicle if there is no roadway available on which to use it. The cycle must be initiated by somebody willing to make an investment for the long term, without hope of a near-term payback.

Competing Interests of Public and Private Sectors

The public and private sectors must cooperate with each other to achieve an automated roadway system. The public sector must furnish the roadway and wayside facilities, while the private sector must equip its vehicles with the necessary onboard equipment. The equipment and facilities must be standardized for compatibility with each other, and these standards must at least apply on a nationwide basis (if not internationally) so that vehicles can use the same onboard equipment everywhere. There are many tradeoffs in the system design which affect the distribution of costs and risks between the vehicles and the roadway, but there is no impartial, unbiased entity to decide how to balance those competing interests. This raises the possibility of battles between self-interested entities over the very nature of a roadway automation system.

Liability

In our litigious modern society, every accident has the potential to become a major tort liability case. An injury-producing accident on an automated roadway could be a trial lawyer's dream (and the nightmare of everybody else) unless policies are implemented in advance to preclude that. Indeed, roadway automation will not be able to advance beyond the test track until the liability issues are resolved so that all of the interested parties (designers and builders of systems and components, public agencies, testing laboratories, consultants, insurance companies, etc.) can be assured that they will not be liable for unlimited damages when failures occur. Creative legislation is likely to be needed to reduce the liability threat to manageable proportions.

Hope

In spite of the technological and institutional problems that need to be solved in order to make roadway automation a reality, the benefits to be derived from solving those problems appear to be so great that it is still worth doing. If we in the United States persist in applying short-term decision criteria to discourage attempts to solve our large-scale transportation problems, the solutions will eventually be imported from Europe or Japan. If the implications of that for the health of our automobile and electronics industries, as well as for our national balance of payments, can be raised in the public consciousness, perhaps we will be able to generate the interest needed to enable us to get to work solving the problems.

REFERENCES

1. K. Gardels. *Automatic Car Controls for Electronic Highways.* Report GMR-276. General Motors Research Laboratories, Warren, Mich., 1960.
2. V. K. Zworykin and L. E. Flory. Electronic Control of Motor Vehicles on the Highways. *HRB Proceedings*, Vol. 37, 1958.
3. D. Barrick. Automatic Steering Techniques. *IRE International Convention Record*, Vol. 10, Part 2, 1962, pp. 166–178.
4. An Investigation of Highway Automation. Reports EES 276A-6 and EES 276A-15. Ohio State University, Columbus, Ohio, 1969 and 1974.
5. *Fundamental Studies in the Longitudinal Control of Automated Ground Vehicles.* Report FHWA-RD-77-28, Ohio State University, Columbus, Ohio, 1976.
6. M. E. Hanson, ed., *Project Metran: An Integrated, Evolutionary Transportation System for Urban Areas.* MIT Press, Cambridge, Mass., 1966.
7. *Tomorrow's Transportation: New Systems for the Urban Future.* U.S. Department of Housing and Urban Development, Washington, D.C., 1968.
8. P. Benjamin, et al. Analysis of Dual Mode Systems in an Urban Area. Report DOT-TSC-OST-73-16A. Transportation Systems Center, U.S. Department of Transportation, Cambridge, Mass., 1973.
9. *Special Report 170: Dual Mode Transportation.* TRB, National Research Council, Washington, D.C., 1976.
10. Calspan Corp. *Practicability of Automated Highway Systems.* Report DOT-FH-11-8903. FHWA, U.S. Department of Transportation, Washington, D.C., November 1977.
11. J. G. Bender, et al. *System Studies of Automated Highway Systems.* Report FHWA/RD-82/003. FHWA, U.S. Department of Transportation, Washington, D.C., 1982.

12. *Advances in Vehicle and Traffic Control Technology*. The Future National Highway Program 1991 and Beyond, Working Paper No. 7. FHWA, U.S. Department of Transportation, Washington, D.C., 1988.
13. R. R. Smisek and G. A. Harder. Automated Control of Guideway Transit Vehicles. *Proc., 29th IEEE Vehicular Technology Conference*, Arlington Heights, Ill., March 1979.
14. J. E. Anderson, et al. eds., *Personal Rapid Transit*. Institute of Technology, University of Minnesota, Minneapolis, Minn., 1972.
15. J. E. Anderson, ed., *Personal Rapid Transit II*. Institute of Technology, University of Minnesota, Minneapolis, Minn., 1974.
16. D. A. Gary, ed., *Personal Rapid Transit III*. Institute of Technology, University of Minnesota, Minneapolis, Minn., 1976.
17. *Proc., International Conference on Advanced Transportation and Urban Revitalization*, Advanced Transit Association, Indianapolis, Ind., May 1978.
18. T. Karlsson. PROMETHEUS: The European Research Program. Presented at 67th Annual Meeting of the Transportation Research Board, Washington, D.C., 1988.
19. S. E. Shladover. Research Needs in Roadway Automation Technology. Presented at SAE Conference on Future Transportation Technology, Vancouver, Canada, August 1989.
20. H. Okamoto. Advanced Mobile Information and Communication Systems (AMTICS). Presented at Future Transportation Technology Conference and Exposition, San Francisco, Calif., August 1988.
21. R. Von Tomkewitsch. The Traffic Guidance System ALI-SCOUT. *Proc., Conference on Technology Options for Highway Transportation Operations*, Sacramento, Calif., October 1986, pp. 130–142.

Publication of this paper sponsored by Task Force on Advanced Vehicle and Highway Technologies.

Multipath Capacity-Limited Transit Assignment

JOSEPH N. PRASHKER

At present most patronage predictions of transit systems are performed using UMTA Transportation Planning System (UTPS) package or some adaptation of it. The transit assignment produced by a typical UTPS system can be classified as an all-or-nothing limited equilibrium assignment. However, passenger loads assigned to a transit line can far exceed the line capacity. In such a case line headway has to be reduced to provide enough capacity to accommodate transit demand. If the increase in frequency is not accounted for by iterating again through the mode choice and assignment models, the equilibrium assumptions are violated. If equilibrium between demand and supply is achieved, it might occur at a point that requires transit capacity much beyond economically feasible or engineering practical levels. Thus the present transit assignment procedure suffers from two problems. First, trips are assigned to transit lines without regard to their actual capacity. Second, while some lines are assigned passenger loads beyond capacity, there might be other lines with just slightly longer travel times that are greatly underutilized. A realistic assignment should take into account and not exceed the actual capacity of every transit line. Furthermore, it should consider lines' capacities while rationally simulating people's travel behavior. A transit assignment algorithm is presented that takes into account the actual capacity of transit lines and assigns trips to more than a single path when the shortest path reaches its capacity. This procedure produces a practical Multipath Capacity-Limited Transit Assignment (McLAT). The procedure was implemented on an IBM mainframe computer using the standard UTPS package with the addition of only one FORTRAN program.

At present most patronage predictions of transit systems are performed using the UMTA Transportation Planning System (UTPS) package (*1*) or some adaptation of it. This set of programs is typically applied once in a customary sequence of mode choice and assignment programs to produce ridership predictions for the various components of a transit system during typical periods of the day. The transit assignment produced by a typical UTPS application can be classified as an all-or-nothing limited equilibrium assignment. The equilibrium achieved by this type of assignment procedure under the usual assumptions of constant travel times and headways has the property that no individual using the system can improve his utility by using a different transit line or switching to a different mode.

However, passenger loads assigned to a transit line can far exceed the line capacity. In such a case line headway has to be reduced to provide enough capacity to accommodate transit demand. If the increase in frequency is not accounted for by iterating again through the mode choice and assignment models, the equilibrium assumptions are violated. If this iteration is performed, on the other hand, the new demand will be even higher, requiring more transit capacity. This process may or may not converge, but even if it does it might occur at an equilibrium point that requires transit capacity much beyond economically feasible levels. The equilibrium may even occur at a point that violates engineering constraints, such as street capacities or minimum headway separation between vehicles. Thus the presently used transit assignment program suffers from the following two undesirable and unrealistic characteristics:

1. Trips are assigned to transit lines with disregard to their actual capacity. Thus, some lines might be loaded with passengers much beyond their ability to carry those loads.
2. While some lines are assigned passenger loads beyond capacity, there might be other lines serving the same origin/destination (O/D) pairs with just slightly longer travel times that are greatly underutilized.

These two problems occur because of the simple all-or-nothing procedure used for transit assignment, and they have very serious practical implications on the validity of the transportation planning process. From the point of view of a transit agency, the amount of service that it can provide at a given future year is dictated by economic considerations and budget limitations. Thus, a clear planning objective for a transit agency is to achieve a realistic transit assignment for a given level of service. In this context, transit level of service should be treated as a predetermined policy decision, if not throughout the whole planning process, than at least in its final stages. Thus a realistic assignment should take into account and not exceed the actual capacity of every transit line. Furthermore, it should consider lines' capacity while rationally simulating people's travel behavior. A rational transit assignment model should take into account not only the fastest transit route serving an O/D pair, but should also consider second- or even third-best transit alternative options. The second- or third-best transit alternatives should be considered as long as the best option is overcrowded and the alternatives' travel utility is higher than the nontransit alternative.

In this paper we present a transit assignment algorithm that takes into account the actual capacity of transit lines and assigns trips to more than a single path when the shortest path reaches its capacity. This procedure produces a practical Multipath Capacity Limited Transit Assignment (McLAT). It was implemented on an IBM mainframe computer using the standard UTPS package with the addition of a single FORTRAN program and a minute modification of one existing program.

Faculty of Civil Engineering and Transportation Research Institute, Technion, Israel Institute of Technology. Former affiliation: Institute of Transportation Studies, University of California at Irvine, Calif. 92717.

In the following sections we discuss the theoretical background of the proposed algorithm within the UTPS framework and formally present the algorithm itself. Later sections provide an outline of programming considerations for the implementation of the algorithm and a comparison between proposed transit assignment and a standard UTPS procedure applied to the Los Angeles metropolitan area.

THEORETICAL BACKGROUND OF THE PROPOSED PROCEDURE

At present the typical UTPS process performs transit assignment assuming no limits on the capacity of transit lines. Thus the present transit assignment procedure is an all-or-nothing assignment, in the sense that for every O/D pair, all trips are assigned to a single transit path. This will happen even if the assigned passenger volumes far exceed the shortest line capacity and even if an alternative underutilized path exists between the same O/D pairs with only slightly longer travel times. The underlying behavioral assumption of the proposed McLAT procedure is that people will chose to use second- or third-best transit alternatives as long as these alternatives possess a higher utility than all other nontransit alternatives.

This assignment does not reach equilibrium in the transit network, because there are people using transit between the same O/D pairs who could improve their utility by switching to another transit path. However, in spite of the fact that transit equilibrium is not maintained, it will be shown next that equilibrium conditions exist between total demand for travel and supply. We argue that the proposed procedure is more realistic than the present ones, which neglect to realize economic or engineering capacity limits. On moderately crowded transit systems, the present assignment procedure might produce erroneous results. On very crowded transit systems, which are typical of rush-hour periods in large metropolitan areas, the results of existing transit assignment procedures might produce completely unrealistic patronage forecasts that greatly overestimate actual transit usage. A short discussion of the proposed McLAT algorithm with respect to equilibrium in the urban transportation system follows.

To begin the discussion, we adopt two basic assumptions that are customary within the framework of UTPS models system:

1. The characteristics of demand and supply are stationary for the simulated time period. This time period can be a whole day or any typical part of it.
2. The total travel demand on all modes for each O/D pair is fixed. Thus the O/D matrix is exogenous to the transportation modeling system.

These two assumptions narrow the equilibrium problem to the distribution of trips between the various network modes and routes. The demand function is a standard Logit function, while the supply function is determined by the transit and highway characteristics. The supply curve of a single transit line is not influenced by travel volumes, whereas travel time on the highway monotonically increases with volume. In the present application of the UTPS system it is assumed that the transit line does not have any capacity limits. However, it seems very unrealistic to assume no capacity limits on heavily crowded transit systems. If the shortest transit route serving a pair of zones has a specific capacity of passengers per hour, all excess demand has to use the highway network or some other transit path.

In the proposed McLAT procedure, we assume that any transit path that has excess capacity and provides the users with a higher travel utility than the highway system constitutes a feasible transit alternative and will be used for travel. This hierarchical choice process can be stated as follows: a zone pair is serviced by two or more transit lines with limited capacity. The supply functions of these lines are:

$$TT1 = \begin{Bmatrix} A & VT1 \le CT1 \\ 0 & VT1 > CT1 \end{Bmatrix} \text{ and}$$

$$TT2 = \begin{Bmatrix} B & VT2 \le CT2 \\ 0 & VT2 > CT2 \end{Bmatrix} \tag{1}$$

where

$CT1$ and $CT2$ = capacity limits of Lines 1 and 2;
$VT1$ and $VT2$ = passengers loads of Lines 1 and 2;
$TT1$ and $TT2$ = travel times of Lines 1 and 2, with $TT2 > TT1$.

The supply function for the highway network can be stated as:

$$TC = T0 + F(VC) \tag{2}$$

where $T0$ equals the free-flow travel time on the highway path and $F(VC)$ equals an increasing function of volume VC.

The equilibrium state for this system is presented in Figure 1. In the figure the demand function (D-D) is a simple Logit function, and the supply function (S-S) is defined as:

$$TC - TT1 = T0 + F(V - VT) - TT1 \tag{3}$$

where V is the total travel demand between a O/D pair. Since Line 1 is superior to Line 2, providing a shorter travel time, its travel time ($TT1$) is used in the supply equation. The potential demand for Line 1 is (VP), but its capacity is ($CT1$), and $VP > CT1$. Thus, the actual number of passengers who can use Line 1 is ($VC1$), producing unassigned demand of the magnitude of ($VP - CT1$). This demand cannot use Line 1 and has to choose between the highway and transit Line 2. This leftover demand can be split between the two modes using the original Logit function. The unassigned volume ($VP - CT1$) should be distributed between the highway and transit Line 2 as shown in Figure 2. The proposed procedure can be expanded to any number of lines serving the same O/D pair.

This iterative process produces a multipath capacity-limited transit assignment by assuming a hierarchical choice process in which transit lines are considered consecutively in order of their level of service and compared to the highway alternative. The best line is considered first, then the second best, and so on until all demand is exhausted. In this two transit line example, the total transit volume will be $VT12 = CT1 + VT2$ and the total highway volume will be $VC = V - VT12$.

This assignment procedure can be implemented as a simple extension of the customary UTPS procedure. It was applied

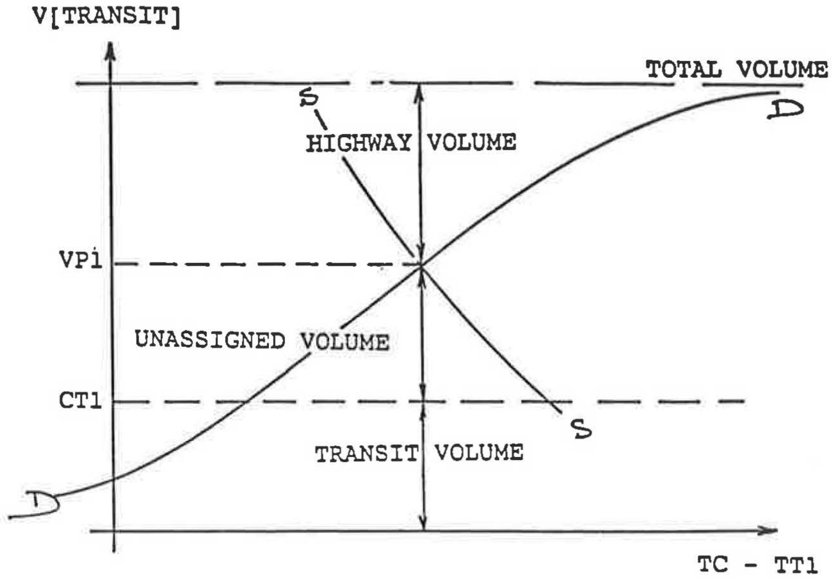


FIGURE 1 Equilibrium between demand and limited-capacity transit line.

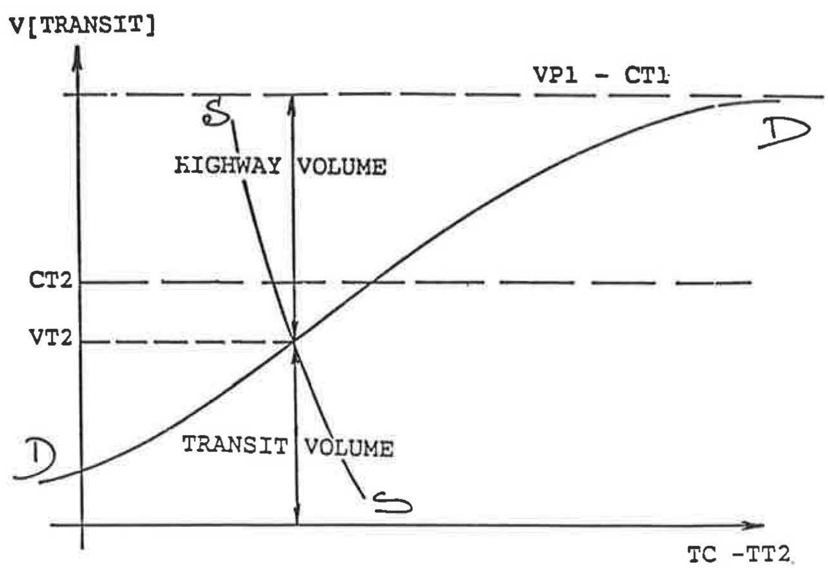


FIGURE 2 Equilibrium between unassigned demand and line 2.

in an almost completely automated way using standard UTPS programs and one additional FORTRAN program. The new proposed method is an iterative procedure, as defined below.

GENERAL DESCRIPTION OF THE PROPOSED ALGORITHM

The proposed algorithm was developed and used for transit alternative investigation in Los Angeles. The need for a capacity-limited assignment arose from the fact that the operating costs of the transit system had to be kept below a predetermined level to meet the agency's budget constraints. The McLAT procedure was used in the final stage of the transit planning process, after an in-depth analysis of various transit alternatives was performed using standard UTPS process. Because of the size of the area and complexity of the transit network, the only practical way to implement the McLAT algorithm was to maximize the use of standard UTPS programs. An iterative procedure was used, consisting of standard UTPS programs such as UPATH, USTOS, UMATRIX, ULOAD(UPRAS), and UMODEL for Logit predictions (*2*), as well as a special FORTRAN program called Overload Line Identification and Network Manipulation (UOLIM) (*3*).

The notation convention for this discussion is as follows:

Notation	*Meaning*
[OD]	original O/D matrix
od	a cell in [OD]
[OD]*xx*	xx% of the original O/D matrix
[OD]*xx*(*i*)	ODxx matrix used in iteration *i*
{*ST*}(*i*)	a set of all stops which a vehicle passes while it carries passengers at or above capacity level at iteration *i*
*NT*0	original (unmodified) transit network

Notation	*Meaning*
$\{LNovl\}(i)$	a set of transit lines which contain stops included in $\{ST\}(i)$
$NT(i)$	modified transit network at iteration i
[ODovl](i)	a matrix containing all O/D pairs whose in-vehicle part of their minimum path starts at a stop included in $\{ST\}(i)$
[ODfre](i)	a matrix containing all O/D pairs whose in-vehicle part of their minimum path *does not* start at a stop included in $\{ST\}(i)$
[LD](i)	Passenger loads on the transit network at iteration i

The proposed algorithm involves the following steps:

1. Take [OD]xx and $NT0$.
2. Perform a standard UTPS simulation run to produce [LD](i).
3. Apply UOLIM to identify overloaded stops in $\{ST\}(i)$, to identify overloaded lines $\{LNovl\}(i)$, and to create a new network $NT(i + 1) = NT0 - \{LNovl\}(i)$.
4. Identify [ODovl](i) for all stops in $\{ST\}(i)$ using USTOS.
5. Using UMATRIX, create a new yy fraction of [ODfre]$yy(i + 1)$ and [ODovl]$yy(i + 1)$, as follows:

$$[OD]yy(i + 1) = yy*[OD] \tag{4}$$

$$[ODovl]yy(i + 1) = [OD]yy(i + 1) * \begin{Bmatrix} 1 \text{ if od} \ni [ODovl](i) \\ 0 \text{ otherwise} \end{Bmatrix} \tag{5}$$

$$[ODfre]yy(i + 1) = [OD]yy(i + 1) - [ODovl]yy(i + 1). \tag{6}$$

6. Perform a full UTPS simulation run using network $NT(i + 1)$ and O/D matrix [ODovl]$yy(i + 1)$ to produce [LDovl]$(i + 1)$.
7. Perform a full UTPS simulation run using network $NT0$ and O/D matrix [ODfre]$yy(i + 1)$ to produce [LDfre]$(i + 1)$.
8. Using ULOAD(UPRAS), combine transit loads:

$$[LD](i + 1) = [LD](i) + [LDfre](i + 1) + [LDovl](i + 1). \tag{7}$$

9. Go to Step 3 and repeat through Step 8 with decreasing increments of the OD matrix until all demand is exhausted.

REMARKS ON THE ALGORITHM

The basic idea behind the McLAT algorithm is to assign to transit an increment of the total demand and test the transit lines for overcrowding. In the next step, partition the transit network and another increment of demand into two subsets. One subset includes transit lines that did not reach capacity and all passenger loads belonging to O/D pairs that, in the previous iteration, boarded transit at stops that reached line capacity. Thus those lines are not able to carry additional passengers. Using this set of O/D volumes and subset of the original network, new transit paths in the system are created. The second subset includes the original network and an increment of passenger loads belonging to O/D pairs that boarded the transit system in the previous iteration at stops which had excess capacity. It is these lines that are able to carry additional passenger loads. More details of the algorithm are discussed below.

The algorithm presented above consists mainly of iterations of the customary UTPS process and additional simple manipulation of O/D matrices and transit network coding. The O/D matrix is manipulated using standard UTPS programs, while the network is modified by the UOLIM program. The whole process is automated through a special feature in the UOLIM program, as discussed below. Each iteration except the first consists of applying twice the full sequence of UTPS programs necessary to generate the customary transit assignment. Practical considerations, and the level of overcrowding of the origin network, define the number of iterations and the fraction size of the O/D matrix at each iteration. The fraction size of the O/D matrix used in each successive iteration should be no larger than the one used previously. Furthermore, it makes practical sense to start the iterative process with a relatively high fraction, which can be predetermined by the ratio of unconstrained passenger loads to capacity on the most loaded lines. In most networks, therefore, three to four iterations of the proposed algorithm will produce assignment results that will not overload the transit network by a significant amount.

Most of the algorithm steps are straightforward and do not present any computational problem. All steps except Step 3 are performed using standard UTPS programs. Step 3, however, deserves special explanation regarding the way it operates. This step, which identifies overloaded stops along the transit line, can be implemented in two ways. Assume that at a point in the iterative process zz percent of the total demand was already assigned to the transit network. Then the following two alternatives exist to identify overloaded stops:

1. a stop can be identified as overloaded if passenger loads in the vehicles exceed total capacity of the line; or
2. a stop can be identified as overloaded if passenger loads in the vehicle exceed zz percent of total capacity of the line.

Assume that the first definition is adopted and the overloaded stops occur somewhere downstream along the line. In the next iteration, passengers who board the transit line at stops before the overloaded ones and travel through them will cause loads on the vehicles beyond their stated capacity. This method of simulation might produce unrealistic transit assignment. Trips originating from zones that board the transit line at overloaded stops are overestimated, while trips from zones that board transit at stops where capacity exists are estimated correctly. On the other hand, if the second definition of capacity is adopted, the proposed transit assignment might significantly underestimate passenger loads.

To remedy this problem, Step 3 of the algorithm should be improved. An exact enumeration of all transit paths, while testing for stop's capacity in their order along the transit route, from first to last, cannot be done efficiently given the size of the transit network. A rigorous and efficient mathematical method to solve the problem is not available at present. Given the size of the transit network at hand, the only practical solution was to modify the heuristic approach presented above.

An alternative definition to those presented for testing capacity can be defined as follows. Assume that *zz* percent of total demand was already allocated to the transit network. Let *CAPzz* be a capacity level after *ZZ* percent of total demand was allocated, and let *CAPtot* be the absolute capacity level of the line. Define capacity testing level as:

$$CAPzz = CAPtot*[ZZ + 0.5*(1.0 - ZZ)] \quad (8)$$

This definition for capacity level alleviates most of the problems associated with the two capacity definitions stated above. It will mark a transit stop as overloaded when loads inside the vehicle are above a certain percentage of capacity level. Because this capacity level is still below total capacity, however, additional passengers in the next iteration can board the transit line at upstream stops without violating capacity restrictions. Note that the value of *CAPzz* cannot exceed total capacity of the transit line. This definition represents an intermediate value compared to the two previous capacity definitions.

PROGRAM IMPLEMENTATION AND PROCEDURE AUTOMATION

The main programming effort in the process of implementing the proposed McLAT procedure was in the development of UOLIM program (Overload Line Identification and Network Manipulation). All inputs to this program, except for one, exist as standard UTPS files. There was a need to create an additional file that contains all transit lines and passenger loads at stops along their routes. Although such a file is not created by any of the existing UTPS program, "printed report no. 3" produced by UPRAS module of the ULOAD program (*4*) contains precisely the necessary data. A very simple modification of this program was implemented to create the necessary file optionally. This file contains line identification information and, for each direction and each stop, the number of passengers boarding embarking and travelling in the vehicle. This file is created at Step 8 of the algorithm and is used as input to UOLIM program.

The UOLIM program was coded in FORTRAN following UTPS programming conventions and using its service subroutines. Detailed description of the program is presented in UOLIM user's manual (*3*), and only the main features of the program will be discussed here. The UOLIM program performs mainly the following three tasks:

1. Identify overloaded stops along transit lines;
2. Generate a new transit lines file containing only non-overloaded lines; and
3. Automatically update the JCL stream and control cards for next iteration of the McLAT procedure.

The last two features of the program require further explanation. The generation of the updated transit network is performed as follows. The original lines file is a standard input to the UOLIM program. Each transit line in this file is separately tested for overloading in each direction of travel. If the line is overloaded in both directions, it is completely removed from the network. If it is overloaded in a single direction, only one direction is removed. If the transit line is coded in the correct stops order, only the direction code is changed to indicate a single directional line. If the line is overloaded in the direction opposite to the coded stops, the order of stops is reversed and the directional code is changed to indicate a single directional line. This new lines file is used to perform Step 6 of the algorithm.

The third function of UOLIM program is to automate as much as possible the execution of the proposed McLAT procedure. The program produces very extensive reports of overloaded lines and associated stops, as well as a very extensive statistical summary. However, to run the proposed McLAT procedure many control cards must be prepared after each iteration for USTOS and UMATRIX programs. Manually preparing control cards after each iteration from computer printouts is a lengthy, tedious, and error-prone task. To overcome this problem, UOLIM program accepts as input a JCL file that includes a generic setup of control cards to perform all the USTOS and UMATRIX runs defined in Steps 4 and 5 of the algorithm. The program updates USTOS PARAMETER cards, which are contained in the JCL stream with the list of new overloaded stops. The new updated JCL file can be submitted as is to perform the UTPS runs that execute Steps 4 and 5 of the McLAT procedure. Steps 6 and 7 of the McLAT procedure are standard transit assignment runs and are executed using a standard JCL setup with very little manual intervention. Step 8 of the McLAT procedure is performed by a simple modification of the standard JCL setup for UPRAS module of ULOAD program. Thus the whole McLAT procedure can be run almost automatically without any appreciable additional manual work compared with the manual tasks which are needed to perform a conventional UTPS transit assignment.

The performance of the UOLIM program is extremely flexible. The user can specify capacity levels of transit vehicles by mode and line; transit lines to be excluded from capacity checking by line, mode, or transit company. The choice of capacity check (from the three defined above) can also be specified by the user. Other control parameters of the program dictate various options of network manipulations and the JCL stream updates.

COMPARISON OF TRANSIT ASSIGNMENT RESULTS

To test the performance of the McLAT procedure, it was applied to the Los Angeles metropolitan area in the final analysis stage of one of the proposed alternatives for the Metro Rail system. The alternative chosen for analysis was first evaluated using a standard UTPS procedure, the outline of which is presented in Figure 3. The size of this transportation planning problem is extremely large, consisting of over 1,600 zones, 500 transit lines operated by 12 transit companies, a network of 7,000 nodes and 14,000 links, and four trip purposes (*2*). The motivation for developing and using McLAT was a policy decision by the Southern California Rapid Transit District (SCRTD) to restrict the annual operating costs of its bus system to $525 million.

The operating costs are calculated by the following equation:

$$OPCOST = K0 + K1*[PK\text{-}VEH] + K2*[VHT] + K3*[VMT] + K4*[PSGRS] \quad (9)$$

where

$K0$ through $K4$ = constants,
PK-VEH = the number of peak vehicles necessary to provide the service,
VHT, VMT = the annual hours and maintenance costs of the bus fleet, and
PSGRS = the number of passengers expressed on annual basis.

When the standard UTPS procedure was applied to the network, the demand for transit exceeded the coded bus system capacity. ULOAD and URAP modules (see Figure 3) adjust the line frequencies, up and down, to accommodate demand. The operating costs of the SCRTD bus system, after its service frequencies were adjusted, were about $608 million annually. This exceeded the stated goal of the agency by $83 million annually, or about 15.8 percent. At the same time, the adjusted bus frequencies violated the equilibrium conditions between demand and supply as assumed by the mode choice model.

The McLAT procedure was applied to the same network in the following way:

1. Of the four trip purposes, only HBW trips (home to work and back) were assigned by the McLAT procedure, since most of the overcrowding on the network occurred during the morning and evening peak periods, which are mainly loaded with HBW trips. The other three trip purposes were assigned in one step using the standard UTPS procedure.
2. Three full iterations of the McLAT procedure were performed, assigning successively 40, 30, and 30 percent of total demand.

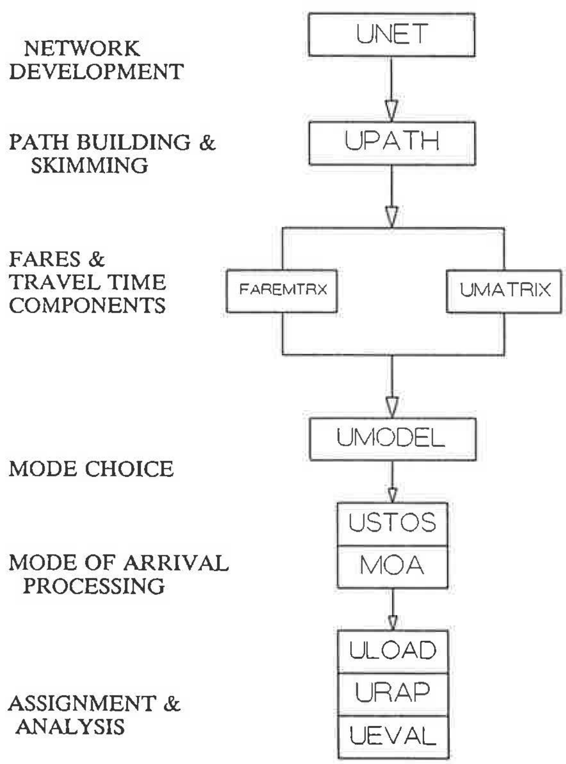


FIGURE 3 SCRTD patronage forecasting process.

Comparison between the results of the standard UTPS procedure and the final outcome of the McLAT procedure are presented in Table 1. The annual operating costs of the McLAT transit assignment are $563 million, saving $45 million relative to the standard UTPS transit assignment. The reduction in operating costs was achieved mainly due to the reduction in peak fleet requirements by 270 buses (there is also a small reduction of 56 vehicles during off-peak periods), thereby reducing the annual fleet hours and maintenance. However, these savings were achieved at the expense of the passenger loads carried: the number of HBW person-trips carried by transit declined by 7.7 percent, from 650,000 to 600,000.

The total reduction of 270 peak buses is the net of eliminating 308 buses operating along 78 transit routes and adding 38 buses operating along 13 transit routes. The number of transit lines along which the number of peak buses increased is a proxy measure for the number of multiple paths created by McLAT procedure, relative to the standard UTPS assignment. In the present simulation this number was not very high: only 13 transit routes, and along 9 of the 13 routes line capacity was reached, indicating that there was not much excess capacity in the alternative transit routes. The relatively small increase in the number of peak buses (38) also points to high crowding on the network.

The number of overloaded transit lines declined from 87 to 76, and the number of overloaded stops from 439 to 320. At first glance this reduction appears to be relative low. However, these figures represent the number of overloaded lines and stops, not the magnitude of overloading. Out of the 76 overloaded bus lines, 31 carried loads in excess of 20 percent of capacity. This figure is much lower, both in the number of overloaded lines as well as load levels, compared to standard UTPS assignment. Fine tuning of the McLAT procedure applying four increments of 40, 30, 20, and 10 percent would have significantly reduced overloading.

Given the complexity of the network and the high level of crowding, it seems that the McLAT procedure performed very well. The stated goal of the agency, to lower operating costs of the bus system to $525 million, was not reached. Analyzing assignment results, it seems that the coded line frequencies and demand patterns were too high. To reach the stated budgetary constraint it would be necessary to further reduce frequencies of the transit service.

CONCLUSIONS

The McLAT procedure presented in this paper served the purpose it was developed for—i.e., to perform transit assignment under capacity restrictions. It is believed based on the experience at hand that McLAT provides more realistic results than the standard UTPS transit assignment. It was developed as a heuristic procedure in the framework of UTPS and can be easily adapted to similar micro and mainframe transportation planning packages. McLAT is an iterative procedure in which almost all manual, error-prone tasks were eliminated. The whole iterative process can be executed in an almost automatic way. Each iteration, however, is a lengthy task. For a network of the size of Los Angeles, under the best circumstances, the turnaround time needed to perform one iteration is one day. Thus to perform the full McLAT

TABLE 1 COMPARISON OF RESULTS OF STANDARD UTPS AND MCLAT PROCEDURE

SYSTEM CHARACTERISTICS	STANDARD UTPS	McLAT PROCEDURE
SCRTD bus lines	287	287
Other bus lines	224	224
Heavy rail	1	1
Light rail	4	4
SCRTD overloaded lines		
Both directions	21	18
One direction	66	58
Overloaded stops	439	320
Daily HBW modal split (persons)		
Auto modes	8,600,000	8,650,000
Transit	650,000	600,000
Transit share (%)	7.0	6.5
Daily non HBW modal split (persons)		
Auto mode	39,000,000	39,000,000
Transit	1,100,000	1,100,000
Transit share (%)	2.7	2.7
SCRTD daily vehicles requirements		
coded pk veh	2,234	2,234
op veh	1,034	1,034
Modified pk veh	2,382	2,112
op veh	1,248	1,192
SCRTD Peak vehicles changes McLAT vs. UTPS		
Reduced pk veh requirements	308	
Increased pk veh requirements	38	
Net savings		270
SCRTD line changes McLAT vs. UTPS		
Reduced pk veh requirements	78	
Increased pk veh requirements	13	
SCRTD annual (1986 $)		
Bus operating costs	608,000,000	563,000,000
Net savings		45,000,000

procedure, while fine tuning it, might require four to five days. This time frame is acceptable only when used as the last stages of the transportation planning process and applied to few prescreened alternatives. At present the McLAT procedure seems a practical and acceptable method to overcome the deficiencies of standard UTPS and similar transit assignment procedures.

However, transit assignment deserves a better fate. From the point of view of software technology and mathematical sophistication, all available transit assignment procedures are at best slight improvements over procedures that are at least 15 years old. We cannot expect to improve the performance and financial integrity of the transit industry if we are unable to provide it with decent planning tools. The McLAT procedure is a very modest step in the right direction. If it does nothing more than stimulate the development of mathematically rigorous multipath capacity-limited assignment, its contribution will be significant.

ACKNOWLEDGMENTS

During the development and implementation of the McLAT procedure the author benefited from remarks and discussions with Peter Stopher, Ken Kaltenbach, and Chaushie Chu of Schimpeler Corradino. Gary Spivak and Keith Killough of SCRTD also provided significant encouragement and valuable remarks. Thanks are due to Suresh Rajagopalan of SCA for performing the actual transit assignment runs and analyzing their results. Last but not least, the comfortable and stimulating atmosphere provided by the Institute of Transportation at the University of California, Irvine, while writing this paper is greatly appreciated.

REFERENCES

1. *UMTA Transportation Planning System Reference Manual* (*UTPS*). UMTA, U.S. Department of Transportation, Washington, D.C., 1975.
2. *Patronage Forecasting Procedures*. Technical Memorandum 86.1.4. General Planning Consultants to Southern California Rapid Transit District, Los Angeles, Calif., April 1987.
3. *Overload Line Identification and Network Manipulation, Program Documentation—UOLIM*. Technical Memorandum 88.3.3. General Planning Consultant to Southern California Rapid Transit District, Los Angeles, Calif., February 1988.
4. R. B. Dial and K. K. Rebibo. *ULOUD—Transit Assignment Program. UTPS—Manual*. UMTA, U.S. Department of Transportation, Washington, D.C., 1975.

Publication of this paper sponsored by Committee on Transportation Data and Information Systems.

Network Representation of Performance Analysis for Intersections

CHRIS HENDRICKSON, CARLOS ZOZAYA-GOROSTIZA, AND SUE MCNEIL

Network representations of performance analysis parameters and relationships are described. In these models, nodes represent variables or other information and links represent relationships. The intersection performance analysis procedures implicit in the *Highway Capacity Manual* are formalized as an example of a network representation. These representations make use of semantic net and frame representations originally developed in the field of artificial intelligence. The purposes of network representations of analysis procedures are (1) to aid flexible, computer-based implementations; (2) to facilitate interpretation of results; and (3) to provide a check on the completeness and uniqueness of particular analysis procedures. For computer-aided analysis and design, these representations provide a means to provide localized or variable specific representation of "intelligent" solution schemes.

Performance analysis for many transportation facilities is complicated by the numerous design characteristics, facility attributes, and measures of effectiveness involved. For example, the evaluation of the performance of an intersection may involve over a hundred separate parameters representing input data, intermediate analysis results, and desired final results. In this paper, we discuss network representations of analytical relationships and procedures. These network representations can provide procedural information for analysis, aid in the interpretation of performance, and serve as a check on analysis aids. They can also serve as representational and analytical building blocks for knowledge-based expert systems. As an example of network representation, we formalize the intersection performance analysis procedures implicit in the *Highway Capacity Manual* (*1*).

We employ two types of network representations of analysis procedures. First, semantic networks or nets represent problem parameters as nodes and relationships as links among these parameters (*2*). In this representation, additional information about link relationships may also be recorded, such as restrictions on the existence of more than one relationship at a time. Second, frame representations provide facilities for storing information about a particular entity as well as links among the different entities (*3*). In our representations, frames can represent particular variables, equations, or higher-level abstractions about the facility performance. The two representations are similar in that frames represent a compact summary of portions of semantic nets. Thus, a frame representation can be generalized into a semantic net if desired. Our frame representation also contains solution information specific to each variable. A third possible representation scheme would be to apply object oriented programming, but this representation is not pursued here (*4*).

Of course, the representation of analysis procedures does not add any new information to the process. However, existing representations typically do not explicitly provide all relevant information. In mathematical or algebraic summaries of relationships, solution procedures are often not formally recorded. Furthermore, constraints to assure meaningful results are often not formally applied; for example, consistency in units of measure can be imposed. Moreover, appropriate analysis procedures may be difficult to represent algebraically. In procedural summaries, such as the tables contained in the *Highway Capacity Manual*, only one set of inputs and outputs is typically described. In many practical analysis cases, a designer may wish to alter the standard analysis procedures to constrain a normal output value or to allow a normal output parameter to be an input. In a network representation, these variations can be accommodated by altering the fashion in which network traversal occurs. As a result, computer aids based on network representations are more flexible in application. Finally, construction of analysis networks serves to insure completeness and uniqueness in the problem and solution description.

Signalized intersections were chosen as an example application because of the complexity of the problem, its considerable practical significance, and the existence of extensive knowledge about intersection analysis. Signalized intersection analysis considers a wide variety of factors, such as geometric characteristics, levels of service, traffic flow volumes and compositions, and signal timing. Analysis procedures typically require a hierarchical application of design decisions and calculations (*5*). Although detailed methodologies for capacity analysis and levels of service determination have been developed (*1*), intersection design is still a complex task because of the many interactions among the design variables.

In the next section, the use of network representations is introduced. Following this, the analysis procedures for signalized intersections contained in Chapter 9 of the *Highway Capacity Manual* are formalized as an example. The following section illustrates the use of the network representation in several example analyses. The final section discusses the merits of alternative procedural representations. The notation used in this paper is listed at its end.

SEMANTIC NETS AND FRAME REPRESENTATIONS

Semantic nets provide a useful model to represent relations among variables of any kind (*6*). In semantic nets, nodes are

C. Hendrickson and C. Zozaya-Gorostiza, Department of Civil Engineering, Carnegie-Mellon University, Pittsburgh, Pa. 15213. S. McNeil, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Mass. 02139.

used to represent variables or values, and links are used to represent relationships. Figure 1 shows some possible node-to-node relations that may occur in the semantic net of an analysis problem. Another type of relation in the semantic net is link-to-link relations. Figure 2 shows some possible link-to-link relations that may occur in a semantic net. Node-to-node and link-to-link relations may vary during the solution process. In particular, the direction of a link in the semantic net depends on the methodology used to solve a specific problem. For example, in certain types of problems V2 has to be determined before V1, and in other problems the order is reversed.

Finally, constraints are relations bound to groups of variables. A constraint might represent a functional relationship among the different variables involved in an equation. Figure 3 shows some alternative semantic net representations for the following simple equation:

$$N*W = TW$$

where

N = number of lanes in an approach,
W = width of each lane in the approach, and
TW = total width of the approach.

Instances 1 and 2 represent the cases where the total width of the approach is a fixed input variable. In this case, design variables W and N are inversely proportional to each other. Instance 3 represents the case where TW is calculated directly from N and W.

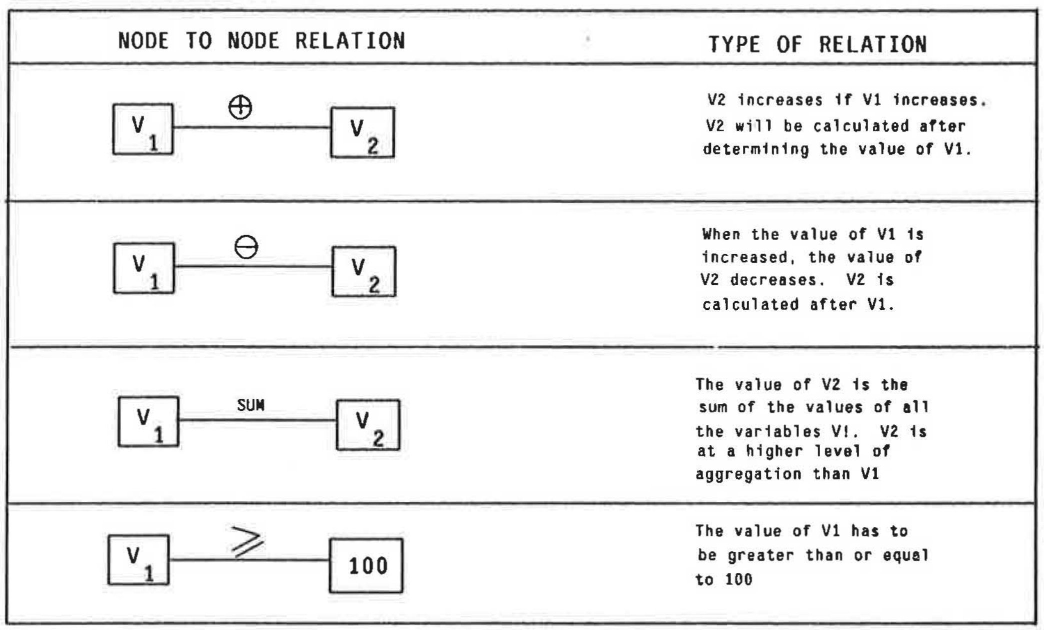


FIGURE 1 Example of node-to-node relations in a semantic net.

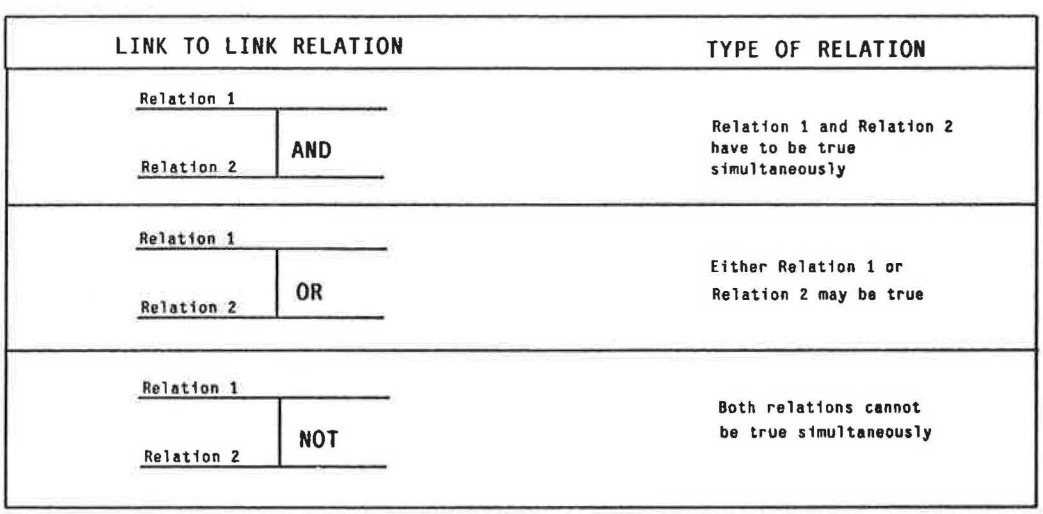


FIGURE 2 Example of link-to-link relations in a semantic net.

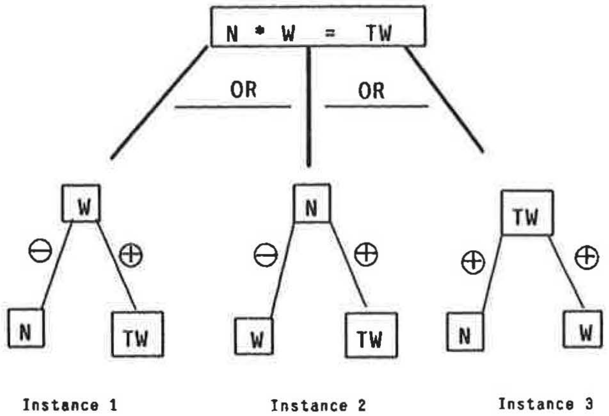


FIGURE 3 Example of a functional relationship involving three variables.

With a semantic network representation scheme, a particular network can represent the general possible relationships or a specific realization, as illustrated in Figure 3. It would be desirable to have a design system capable of generating any of the possible instances of a functional relationship depending on the characteristics of a specific design problem. Figure 4 shows a generic semantic net for the problem in consideration. Instance 1 would be generated as follows. First, the system identifies *W* as a calculated output variable. Then it backtracks to see what variables may be used to obtain *W*. The only possibility is to obtain it using *N* and *TW*. Both of these are input variables, so instance 1 has been fully created.

Constraints may also be used to represent relations among variables at different levels of aggregation. For example, in single-lane approaches the following condition is true: *Flow turns and lane turns must be compatible.* In other words, a flow may only have turn movements that are included in the allowable turn movements for the lane. A left-turn lane cannot have through or right-turn flows. Figure 5 represents the generic semantic net for this constraint. In this example, the object *flow* is more specific than the object *lane*. A lane may have more than one flow, but a flow is in only one lane.

In a complicated analytical problem such as intersection analysis, flexibility in the representation of relationships and equations is very important. As a result, simple types of link relationships such as those shown in Figure 4 will be insufficient. Several alternatives for explicit representation of equations exist. As one possibility, the software package TK!Solver™ uses "R-graph" representations in which the

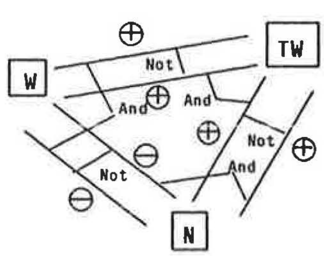


FIGURE 4 Generic semantic net for an equation with three variables.

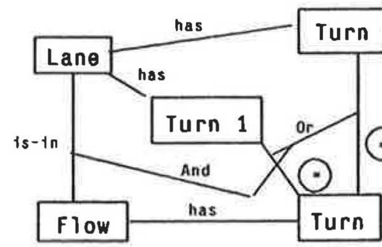


FIGURE 5 Generic semantic net for the constraint relating flow turns and lane turns.

polygonal area of a semantic network surrounding the variables in an equation is associated with additional information (7). With this representation, the program TK!Solver has the capability for automatically generating the solution trees shown in Figure 3. More generally, frames can be used to represent both variables and relationships among variables.

Frames are a general purpose data representation scheme in which information is arranged in "slots" within a named frame. Slots may contain lists, values, text, procedural statements (such as calculation rules), pointers, or other entities. Inheritance and reference to values in other frames is permissible within any slot. Slot functions called "demons" (or daemons) can be used to compute the value of the unknown variables. In addition, full rule sets can be added in slots to provide a local expert system specifically for solution of particular variable values. For example, separate rules could be added in a variable frame slot to indicate when and how the variable could be calculated. Constraints on calculations can also be imposed in these rulesets to insure appropriate ranges of variables, consistent units, etc. Particular calculation rules are only instantiated (or "fired") when the necessary other variable values are available. As a result, intelligent variable frames can be developed that include solution information within the local variable frames. The result would be an extremely flexible analysis system. This representation system should also be amenable to parallel processing on a network of computers.

As a small example of this representation, Figure 6 shows a frame representation of the relationships among the variables *N*, *W*, and *TW* presented earlier. In this representation, each equation's variables is represented as a frame with the following slots:

- (*is-a*) and (*has-children*). These slots are used to specify the other variables that are related with the variable of the frame. Inheritance of values and attributes can be accomplished via such slots, although it is not used here.
- (*current-value*). This slot contains the current value of the variable. The variable may have been calculated or given by the user. This slot also contains the facet (*if-needed*), which has a demon that is fired when a value for the variable is requested. This function calculates the value of each variable if the other defined variables are known.
- (*source-value*). This slot specifies how the value of the variable was obtained. Two values for this slot are possible: input or calculated.

Whenever two variable values are input into this system, the (*if-needed*) demon is invoked to calculate the third. As a

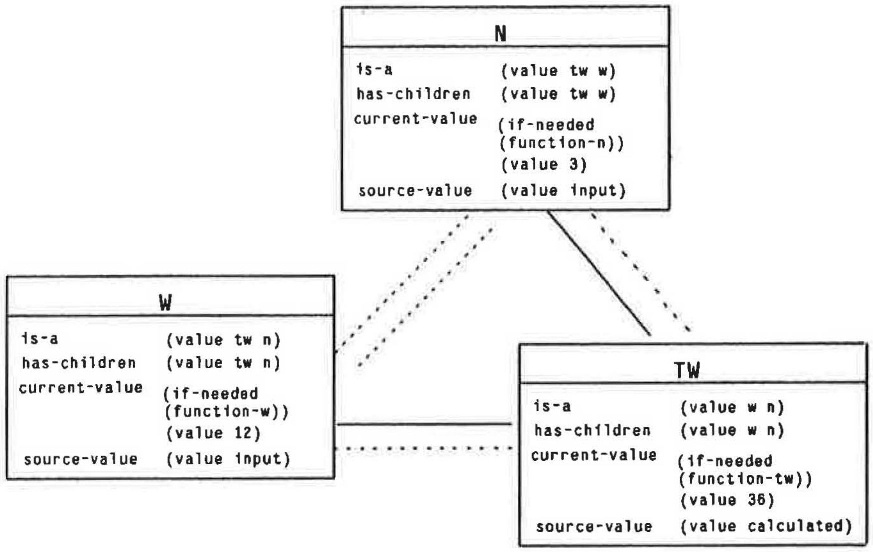


FIGURE 6 Example of a simple frame representation.

result, no control strategy is required to perform analysis in this case since the solution calculation will propagate through the frames automatically.

As noted above, a more general representation would replace the single calculation demon contained in the (*current-value*) slot with a series of calculation or solution rules, with one rule available for each equation in which the variable appears. Rather than simply cycling through all the rulesets to identify available calculation rules, it is also possible to impose a control strategy to search the most likely calculations first. A procedure for guessing at likely values and iteratively solving simultaneous equation sets might also be imposed, as in TK!Solver.

USING SEMANTIC NETS FOR INTERSECTION ANALYSIS

Typical Analysis Problems

The *Highway Capacity Manual* considers four types of components in any design problem of a signalized intersection: service flow rates, geometric design characteristics, signalization conditions, and levels of service. Operational analysis methodologies are used for determining the value of any of these four components when the other three components are known. Figure 7 presents the four typical intersection design problems:

1. Solve for levels of service. Service flow rates, geometric characteristics, and signalization conditions are given to the system. The system directly computes stopped delays for each lane group. Levels of service are obtained using Table 9-1 of the manual, where the level of service criterion is defined as a function of the stopped delay.
2. Solve for allowable service flow rates. Geometric characteristics, signalization conditions, and desired levels of service are given to the system. The system performs an iterative procedure for determining service flow rates that accomplish the desired levels of service. The system may have to try alternative lane group definitions in order to satisfy the design requirements.
3. Solve for signal timing. Geometric characteristics, service flow rates, and desired levels of service are given to the system. Alternative phase distributions and cycle lengths are tested in order to accomplish the desired levels of service.
4. Solve for geometric characteristics. Service flow rates, signal timing, and desired levels of service are given to the system. The system tests alternative geometric characteristics by changing the number or the allocation of lanes in the approaches, in order to satisfy the desired levels of service.

Identifying these four typical intersection design problems is useful to conceptualize mutual relationships among the different design variables. However, implementation of a flexible system capable of applying operational methodologies in any of the four cases is not straightforward. Moreover, a practical intersection design problem may have combinations of the typical design problems described above. For example, a common design problem is to solve for signal timings and geometric characteristics. Alternatively, the user may want to specify a desired level of service for the whole intersection and let the system compute the levels of service of the different approaches. A similar situation would occur when the number and width of the lanes in only one of the approaches may be changed. For both cases, classification of the design variables according to different levels of aggregation is required for solving the problem. In this paper, we shall restrict our attention to the more straightforward analysis problems associated with identifying levels of service.

During the design process, interactions among different subproblems may change dynamically. It is important to identify each subproblem by classifying the design variables related with it. A design variable may be classified with respect to any of these criteria: aggregation level, type of variable, and design condition. The three criteria are required in order to

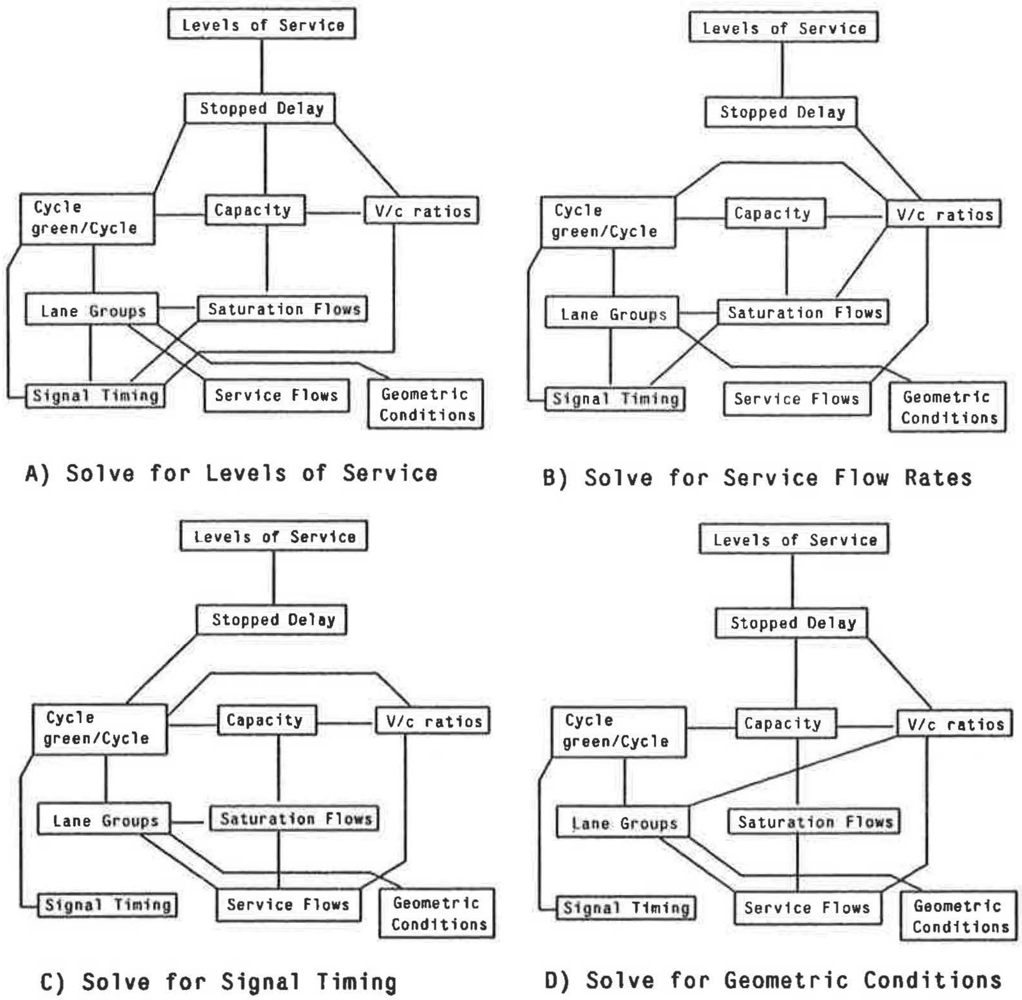


FIGURE 7 Four typical intersection design problems.

have a full classification of a specific design variable. Classification of a design variable may change during the design process.

With regard to aggregation, each design variable may appear at a particular level:

- Intersection.—The design variable has a unique value for the whole intersection. Examples of these variables are the total delay at the intersection, the critical volume/capacity ratio and the type of area (CBD or other) where the intersection is located.
- Approach.—The design variable has a value for each approach of the intersection. Examples of these variables are the number of lanes, the width of these lanes, and the total approach volumes.
- Phase.—The design variable has a value for each phase of the signal cycle. Examples of these variables are the available green time and the red plus yellow time for each phase.
- Lane group.—The design variable has a value for each lane group of the intersection. Examples of these variables are the saturation flow, the average delay, and the opposing volume for each lane group. Lane groups are used to join one or more flow movements of the same approach. A lane group may be in one or more phases. The latter case is called a "phase overlapping" situation.

In addition to the aggregation level, a design variable may be of any of the following types:

- Geometric.—The variable is used to describe a geometric characteristic of the intersection. Examples of these variables are the lane widths, the approach angles and the number of lanes in each approach.
- Signalization.—The variable is used to describe the operation of the traffic light signal. Examples of these variables are the cycle length and the total lost time.
- Traffic condition.—The variable is used to describe a characteristic of a traffic movement. Examples of these variables are the percentage of heavy vehicles and the flow volumes.
- Levels of service.—The variable is used to describe the service that a design alternative provides to a group of flows. Examples of these variables are the average delay for each lane group and the average delay for the intersection.

Finally, design conditions can be used to describe the status of a design variable during the design process. A variable may be any of these design conditions:

- Fixed input.—The value of the variable may not be altered during the design process.
- Design input.—The user has given explicitly a value to the variable.
- Assumed input.—A value has been assumed for the design variable due to lack of additional information.
- Desired output.—The user has expressed explicitly bounds for the value that he or she wants for the design variable.
- Calculated Output.—The value of the variable is calculated using formulas or tables.

Table 1 shows the classification of some design variables for a specific design problem to determine the intersection level of service. Design conditions vary depending on the type of design problem under consideration.

Functional Relationships for Intersection Analysis

A detailed methodology for determining capacities and levels of service in signalized intersections is given in Chapter 9 of the *Highway Capacity Manual*. The methodology corresponds to Case A of Figure 7, in which geometric characteristics, traffic conditions, and signalization timing are known and the goal is to determine levels of service. The portions of the semantic net that will be shown correspond to this particular problem. We shall summarize these relationships with respect to the four levels of aggregation, corresponding to variables and relationships at the lane group level, the approach level, the phase level, and finally at the intersection level. Notation used in the representation is summarized below.

Lane Group Level

Figure 8 presents the functional relationships among different lane group variables and the corresponding semantic nets that may be used to represent them. These semantic nets are particular instances of generic semantic nets, similar to those shown in Figure 3. In the figure, boundaries are used to separate variables at different levels of aggregation. For example, in the first equation, the capacity (C) and the saturation flow (S) are both at the lane group level of aggregation, while the fraction of green time ($g/Cycle$) is at the phase level of aggregation.

Determination of f_{rt} and f_{lt} is done using tables for each of the eight possible types of lane groups: (1) exclusive lane with protected phase, (2) exclusive lane with permitted phase, (3) exclusive lane with protected and permitted phases, (4) shared lane with protected phase, (5) shared lane with permitted phase, (6) shared lane with protected and permitted phases, (7) single lane, and (8) double exclusive lane with protected phase. However, in all tables f_{rt} is inversely proportional to the pedestrian volume ($PEDS$) and to the percentage of right turns (P_{rt}). Similarly, f_{lt} is inversely proportional to the opposing volume (V_o) and to the percentage of left turns (P_{lt}).

Approach Level

Figure 9 shows some functional relationships that occur among design variables at the approach level of aggregation.

Phase Level

Figure 10 shows the functional relationship among the minimum green time in a phase and the green time, walking distance, total red time, and total lost time.

Intersection Level

Figure 11 shows some functional relationships occurring among variables at the intersection level of aggregation.

Example of a Semantic Net

In the previous section we gave representations of some functional relationships among design variables of signalized inter-

TABLE 1 CLASSIFICATION OF SOME DESIGN VARIABLES

Design Variable	Symbol	Aggregation Level	Type	Design Condition
Lane Width	W	Lane Group	Geometric	Design Input
Level of Service	LOS	Intersection	Level of Service	Desired Output
Level of Service	LOS	Lane Group	Level of Service	Calculated Output
% Heavy Vehicles	% HV	Approach	Traffic Condition	State of Nature

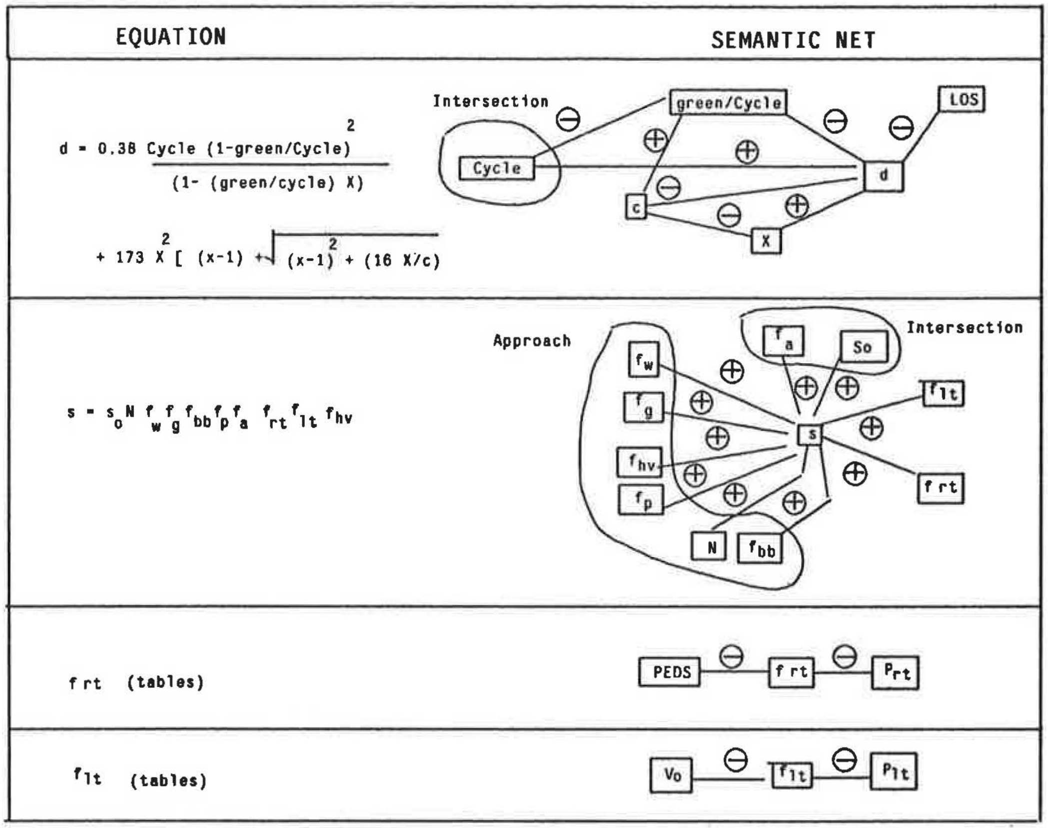


FIGURE 8 Functional relationships at the lane group level.

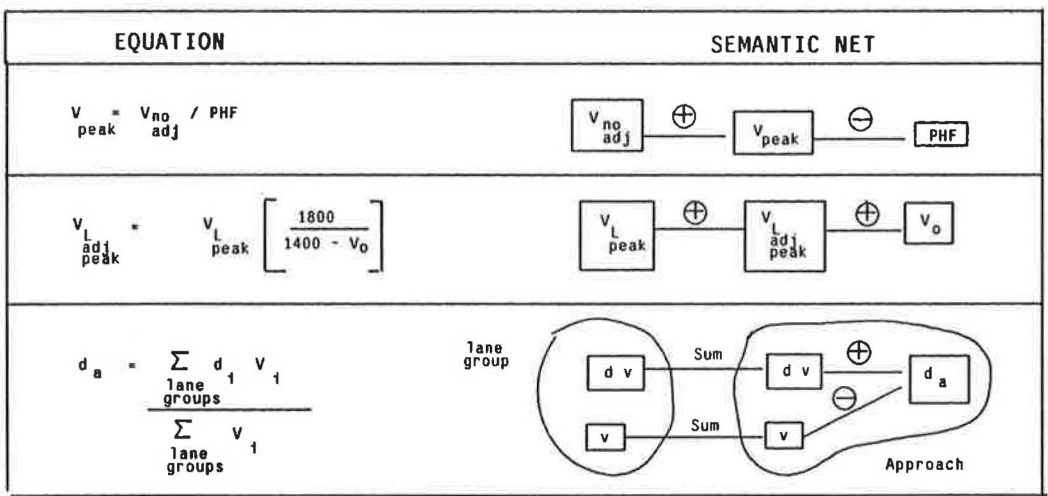


FIGURE 9 Functional relationships at the approach level.

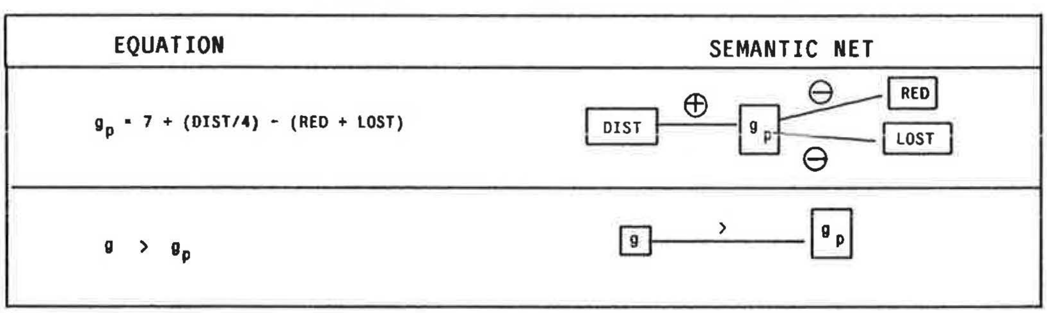


FIGURE 10 Functional relationships at the phase level.

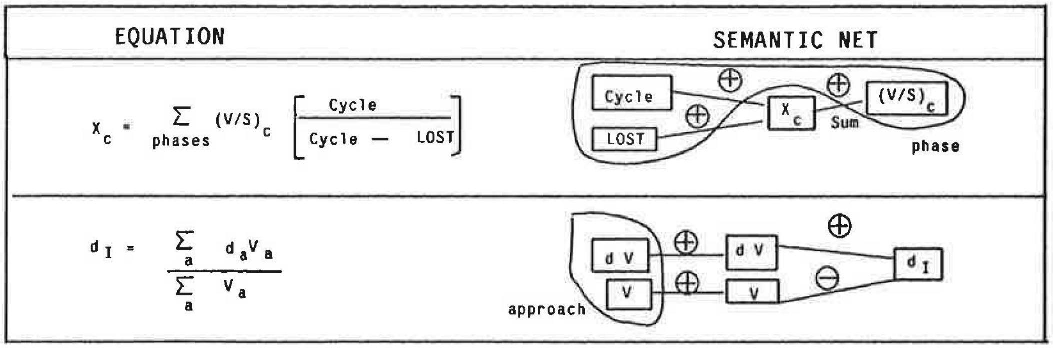


FIGURE 11 Functional relationships at the intersection level.

sections. These representations corresponded to the case where levels of service are unknown. Joining these semantic nets into a common network, we obtain the semantic net shown in Figure 12. Link-to-link relations are not included in the semantic net. However, they were used to create the instances of node-to-node relations shown in the network. The global network is one of the many possible networks that may result when individual instances of functional relationships are changed.

Figure 12 is a simplified version of the complete network. For example, nodes at the approach level of aggregation should be created for each approach of the intersection. Similarly, nodes of the phase level of aggregation should be created for each phase of the signal, and several lane group nodes are

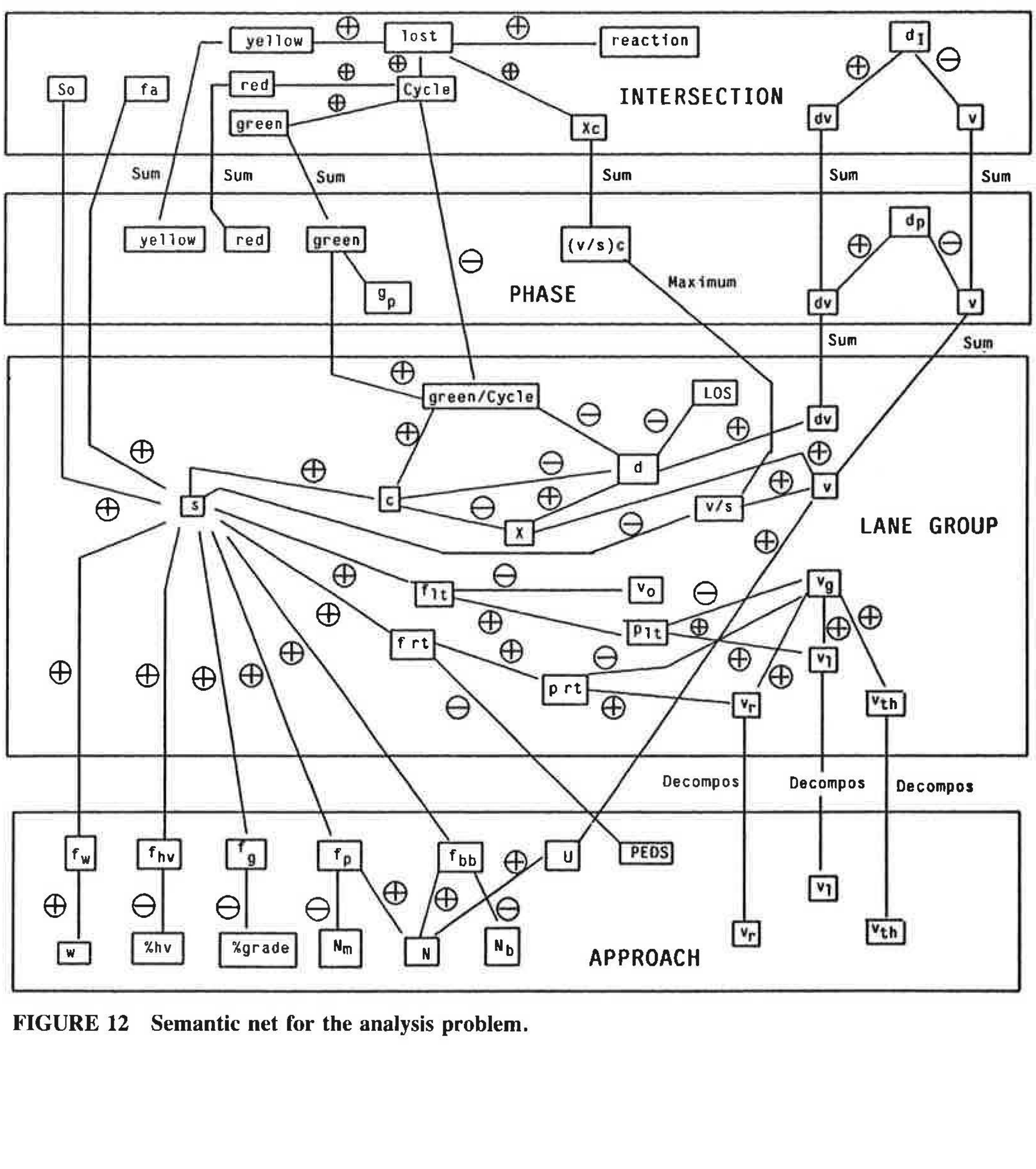


FIGURE 12 Semantic net for the analysis problem.

created for each approach. An expert system example of such groupings appears in the TRALI traffic signal setting system (*5*).

EXAMPLE APPLICATIONS

The following examples illustrate the use of the semantic network shown in Figure 12. The first example is based on a hypothetical intersection and uses the network to aid in assessing the effect of varying an input. The second example is based on volume data from an intersection in Cambridge, Mass., and uses the network to evaluate the impact of changing the phasing.

The input data for the first example are shown in Figure 13, the input worksheets as used in the *Highway Capacity Manual*. The intersection has four lanes on each approach, with large traffic volumes on Main Street and somewhat smaller volumes on First Street. Five phases are used with a total cycle length of 137.7 seconds. The lane group, approach, and intersection level of service are shown in Table 2. Suppose that the local transit authority is considering making a bus stop at the intersection, with a headway of six minutes between buses. The impact of these buses is evaluated with the aid of the network.

With ten buses per hour at the intersection, the input variable N_B at the lane group level of aggregation changes from 0 to 10 for both the eastbound and westbound right turn (EB and WB RT) lanes. Using the semantic net of Figure 12, the impact of the change can be assessed as follows:

- increasing N_B decreases f_{bb}, the factor for buses; and
- decreasing f_{bb} decreases s, the saturated flow for the lane group.

The change in s has two impacts:

- increasing v/s, which affects the critical volume capacity ratio for the intersection if the lane group including the right lane is critical, or becomes critical when the buses are added; and
- decreasing the capacity, which increases the delay for the lane group, approach, and intersection.

After tracing these changes, the level of service associated with the EB approach dropped from category D to E, although the entire intersection remained in level of service category E.

Geometry, volume data, vehicle characteristics, signal phasing, and signal timing for the second example are shown in Figure 14; this intersection can be found in Cambridge, Mass. The existing levels of service are shown in Table 3. As in the previous example, the semantic network can be used to explore the effect of changing inputs. In this case, we wish to assess the impact on the intersection performance of changing the signal phasing. The existing phasing includes an all-red phase for pedestrians, for safety reasons.

Eliminating this pedestrian phase may have two impacts on performance:

- pedestrians may not have sufficient time to cross; and
- the reduced cycle time may affect delay.

Using the semantic network in Figure 12 allows us to trace the impact of these changes. First, with the pedestrian phase present, green time (g) was always greater than the minimum time (g_p) required for pedestrians during that phase to cross the crosswalk that conflicts with right turns. With the pedestrian phase eliminated, g_p is calculated and compared to g. In this example, g is greater than g_p in all cases. Similarly, for the existing phasing the pedestrians do not affect the right turn movements. The presence of pedestrians during the other phases will decrease the factor f_{rt} and decrease the saturation flow. Ultimately, the delay increases for that lane group. This effect is similar to that traced in the first example above. However, in this example, the reduced lost time resulting in a shorter cycle will yield a larger green-to-cycle ratio, thereby decreasing delays. This is demonstrated by tracing the semantic network from the red time input for each phase. Eliminating the red time for one phase has the following effects:

- the total red time for the intersection decreases;
- the cycle length decreases;
- the g/cycle ratio increases for each lane group; and therefore
- the total delay decreases for all lane groups.

This tradeoff between increased delays due to conflicts and decreased delay due to small cycles is not clear in the worksheet analysis of the *Highway Capacity Manual*. When pedestrian volumes are low and the red phase is long, the reduced delay is likely to be greater due to increased g/cycle than increased delay due to the presence of conflicting pedestrians. For this example, the level of service is improved for most lane groups and for the intersection as a whole, as shown in Table 3.

CONCLUSIONS

We have suggested new representations of analysis procedures based on semantic networks and frames. In addition to their usefulness as manual aids to understanding and analysis, these representation schemes could be very useful in the development of computer-aided analysis and design tools. For example, the representation used here could be readily adopted to expert systems for intersection analysis or design (*8*). By localizing analysis knowledge in frames, the network representation used here might also support parallel processing of a network of signals over a geographic area. Further research in this area may be beneficial.

NOTATION

c	capacity (vph)
d	average delay for a lane group (sec)
d_a	average delay for an approach (sec)
d_I	average delay for the intersection (sec)
Cycle	cycle length time (sec)
f_a	adjustment factor for area type

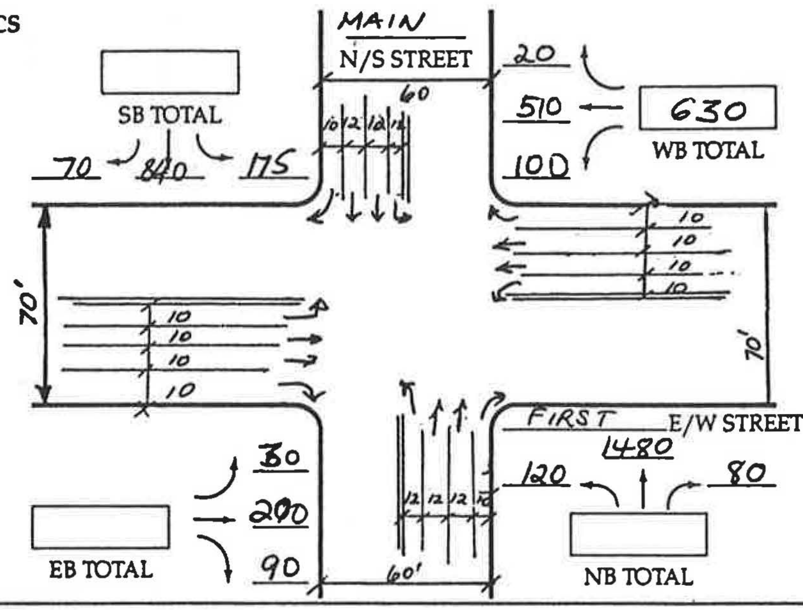
Intersection: MAIN ST & FIRST ST Date: 7/24/86.

Analyst: SM Time Period Analyzed: 8:00-9:00 Area Type: ☒ CBD ☐ Other

Project No.: TRB City/State: WASHINGTON D.C.

VOLUME AND GEOMETRICS

NORTH



IDENTIFY IN DIAGRAM:

1. Volumes
2. Lanes, lane widths
3. Movements by lane
4. Parking (PKG) locations
5. Bay storage lengths
6. Islands (physical or painted)
7. Bus stops

TRAFFIC AND ROADWAY CONDITIONS

Approach	Grade (%)	% HV	Adj. Pkg. Lane Y or N	Adj. Pkg. Lane N_m	Buses (N_B)	PHF	Conf. Peds. (peds./hr)	Pedestrian Button Y or N	Pedestrian Button Min. Timing	Arr. Type
EB	0	5	Y	5	0	0.85	200	Y	19.0	3
WB	0	5	Y	5	0	0.85	200	Y	19.0	3
NB	0	2	N	0	0	0.90	50	Y	21.5	3
SB	0	2	N	0	0	0.90	50	Y	21.3	3

Grade: + up, − down
HV: veh. with more than 4 wheels
N_m: pkg. maneuvers/hr
N_B: buses stopping/hr
PHF: peak-hour factor
Conf. Peds: Conflicting peds./hr
Min. Timing: min. green for pedestrian crossing
Arr. Type: Type 1-5

PHASING

DIAGRAM

Timing	G= 6 Y+R=3	G= Y+R=	G=2' Y+R=3	G=71.7 Y+R=3	G=5 Y+R=3	G=5 Y+R=3	G=20 Y+R=3	G=10 Y+R=10
Pretimed or Actuated	A							

Protected turns — - - Permitted turns - - - - - - - Pedestrian Cycle Length 137.7 Sec

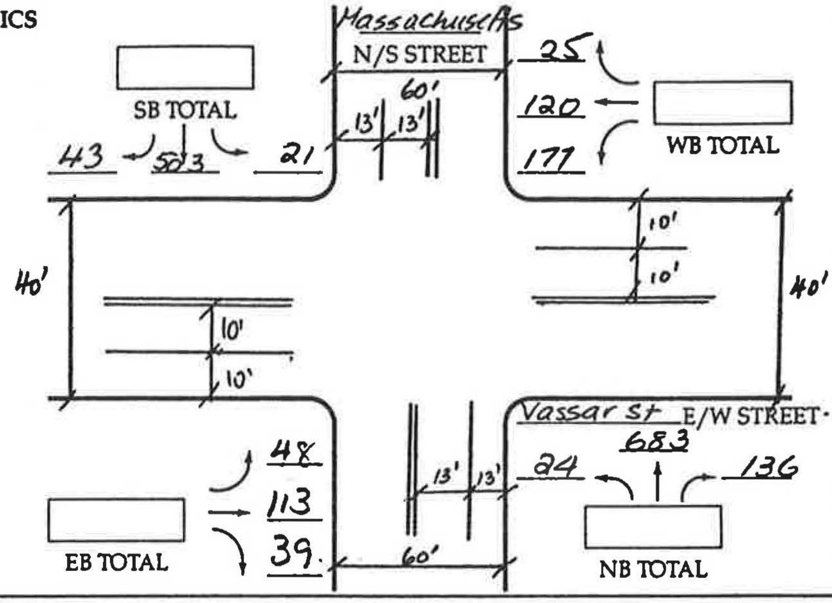
FIGURE 13 *Highway Capacity Manual* **input worksheet for Example 1.**

Intersection: *Massachusetts Ave and Vassar St* Date: *Nov 85*

Analyst: ______ Time Period Analyzed: *4:30-5:30pm* Area Type: ☒ CBD ☐ Other

Project No.: ______ City/State: *CAMBRIDGE, MA*

VOLUME AND GEOMETRICS



IDENTIFY IN DIAGRAM:

1. Volumes
2. Lanes, lane widths
3. Movements by lane
4. Parking (PKG) locations
5. Bay storage lengths
6. Islands (physical or painted)
7. Bus stops

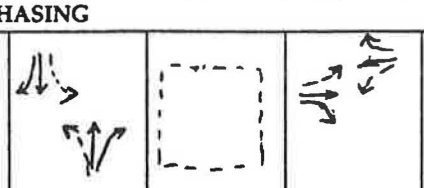
TRAFFIC AND ROADWAY CONDITIONS

Approach	Grade (%)	% HV	Adj. Pkg. Lane Y or N	Adj. Pkg. Lane N_m	Buses (N_B)	PHF	Conf. Peds. (peds./hr)	Pedestrian Button Y or N	Pedestrian Button Min. Timing	Arr. Type
EB	0	1	Y	11	1	0.75	162	N		4
WB	0	1	Y	3	1	0.89	141	N		4
NB	0	1	Y	14	6	0.83	301	N		1
SB	0	–	Y	5	14	0.87	157	N		3

Grade: + up, − down
HV: veh. with more than 4 wheels
N_m: pkg. maneuvers/hr
N_B: buses stopping/hr
PHF: peak-hour factor
Conf. Peds: Conflicting peds./hr
Min. Timing: min. green for pedestrian crossing
Arr. Type: Type 1-5

PHASING

DIAGRAM



Timing	G = 42 Y + R = 3	G = 0 Y + R = 23	G = 29 Y + R = 3	G = Y + R =	G = Y + R =	G = Y + R =	G = Y + R =	G = Y + R =
Pretimed or Actuated	P	P	P					

Protected turns — Permitted turns — ------- Pedestrian — Cycle Length *100* Sec

FIGURE 14 Input worksheet for Example 2.

TABLE 2 INTERSECTION LEVEL OF SERVICE FOR EXAMPLE 1

Lane Group Level of Service		Approach Level of Service		Intersection Level of Service
EB LT	D			
EB TH	D	EB	D	
EB RT	E			
WB LT	D			
WB TH	F	WB	F	
WB RT	D			E
NB LT	C	NB	E	
NB TH	E			
NB RT	B			
SB LT	F			
SB TH	B	SB	D	
SB RT	B			

TABLE 3 INTERSECTION LEVEL OF SERVICE FOR EXAMPLE 2

Approach and Lane Group	Existing Level of Service	Level of Service Without Red Phase
EB LT-TH-RT	B	B
WB LT-TH-RT	C	B
NB LT-TH-RT	E	C
SB LT-TH-RT	B	B
INTERSECTION	D	B

f_{bb}	adjustment factor for blocking effect of local buses
f_g	adjustment factor for approach grade
f_{hv}	adjustment factor for heavy vehicles
f_{lt}	adjustment factor for left turn
f_p	adjustment factor for the existence of a parking lane
f_{rt}	adjustment factor for right turn
green	green time (sec)
g_p	minimum green time (sec)
LOS	level of service
N	number of lanes
N_b	number of buses (buses/hr)
N_m	number of parking maneuvers (maneuvers/hr)
PEDS	pedestrian volume (peds/hr)
PHF	peak-hour factor
P_{lt}	percent of left turns
P_{rt}	percent of right turns
reaction	average reaction time for a driver (sec)
red	red time (sec)
s	saturation flow rate for lane group (pcphg)
s_0	ideal saturation flow rate per lane (usually 1,800 pcphgpl)
U	lane utilization factor
V	flow rate (vph)
V_{adj}	adjusted flow rate (vph)
$V_{adj,peak}$	adjusted peak flow rate (vph)
$V_{no\ adj}$	unadjusted flow rate (vph)
V_o	opposing volume (vph)
V_{peak}	unadjusted peak flow rate (vph)
X	V/C ratio for a lane group
X_c	critical V/C ratio for the intersection
W	lane width
yellow	yellow time (sec)
% grade	percent of grade
% HV	percent heavy vehicles

REFERENCES

1. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
2. N. V. Findler. *Associative Networks: The Representation and Use of Knowledge by Computers.* Academic Press, New York, N.Y., 1979.

3. M. Minsky, ed. *Semantic Information Processing*. MIT Press, Cambridge, Mass., 1968.
4. M. Stefik and D. G. Bobrow. Object-Oriented Programming: Themes and Variations. *AI Magazine* Vol. 6, No. 4, 1986, pp. 40–62.
5. C. Zozaya-Gorostiza and C. Hendrickson. Expert System for Traffic Signal Setting Assistance. *ASCE Journal of Transportation Engineering*. Vol. 113, No. 2, 1987, pp. 108–126.
6. P. Harmon and D. King. *Expert Systems: AI in Business*. Wiley, New York, N.Y., 1985.
7. M. Konopasek and S. Jayaraman. *The TK!Solver Book*. Osborne/McGraw-Hill, Berkeley, Calif., 1984.
8. C. T. Hendrickson, C. Zozaya-Gorostiza, and S. McNeil. *A Knowledge Based Expert System Architecture for Computer Aided Analysis and Design of Intersections*. Department of Civil Engineering, Carnegie-Mellon University, Pittsburgh, Pa., 1986.

Publication of this paper sponsored by Committee on Transportation Supply Analysis.