TRANSPORTATION RESEARCH RECORD 835

Transportation System Analysis

TRANSPORTATION RESEARCH BOARD

COMMISSION ON SOCIOTECHNICAL SYSTEMS
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY OF SCIENCES
WASHINGTON, D.C. 1981

Transportation Research Record 835
Price $10.40
Edited for TRB by Brenda J. Vumbaco

modes
1 highway transportation
2 public transit

subject areas
12 planning
13 forecasting

Library of Congress Cataloging in Publication Data
National Research Council (U.S.). Transportation Research Board.
Meeting (60th: 1981: Washington, D.C.)
Transportation system analysis.

(Transportation research record; 835)
Reports prepared for the 60th annual meeting of the Transportation Research Board.
1. Transportation–Addresses, essays, lectures. 2. System analysis–Addresses, essays, lectures. I. Title. II. Series.
TE7.H5 no. 835 [TA1155] 380.5s [380.5] 82-6363
ISBN 0-309-03312-8 ISSN 0361-1981 AACR2

Sponsorship of the Papers in This Transportation Research Record

GROUP 1–TRANSPORTATION SYSTEMS PLANNING AND ADMINISTRATION
Leon M. Cole, Library of Congress, chairman

Transportation Forecasting Section
George V. Wickstrom, Metropolitan Washington Council of Governments, chairman

Committee on Transportation Systems Design
William G. Barker, North Central Texas Council of Governments, chairman
Virginia J. Ainslie, James H. Banks, Edward A. Beimborn, Melvyn Cheslow, Lewis P. Clopton, Gary R. Erenrich, Joel Ettinger, Leonard Goldstein, Robert E. Heightchew, Jr., Clinton L. Heimbach, William S. Herald, James Dan Jones, Robert L. Knight, Douglass B. Lee, Jr., Marvin L. Manheim, Charles William Ockert, Robert E. Paaswell, Robert L. Peskin, Darwin G. Stuart, Antti Talvitie, Yacov Zahavi

Committee on Passenger Travel Demand Forecasting
David S. Gendell, Federal Highway Administration, chairman
Moshe E. Ben-Akiva, Daniel Brand, David J. Dunlap, Robert T. Dunphy, Raymond H. Ellis, Robert E. Gall, Thomas F. Golob, Walter G. Hansen, David T. Hartgen, Thomas J. Hillegass, Joel L. Horowitz, Stephen M. Howe, Lidia P. Kostyniuk, T. Keith Lawton, Steven Richard Lerman, Eugene J. Lessieu, Frederick A. Reid, Martin G. Richards, Gordon W. Schultz, Gordon A. Shunk, Bruce D. Spear, Anthony R. Tomazinis, Edward Weiner, Yacov Zahavi

Committee on Traveler Behavior and Values
David T. Hartgen, New York State Department of Transportation, chairman
M.E. Beesley, Nils Anders Bruzelius, Katherine Patricia Burnett, Melvyn Cheslow, Ricardo DePaul Dobson, Michael Florian, Thomas F. Golob, David A. Hensher, Frank S. Koppelman, Peter S. Liou, Jordan J. Louviere, Daniel McFadden, Arnim H. Meyburg, Richard M. Michaels, Robert E. Paaswell, Shalom Reichman, Earl R. Ruiter, Peter R. Stopher, Antti Talvitie, Mary Lynn Tischer, Peter L. Watson

Committee on Computer Graphics and Interactive Computing
Dietrich R. Bergmann, General Motors Transportation Systems Division, chairman
Richard J. Bair, Carter W. Brown, Amir Eiger, Richard E. Esch, Gordon P. Fisher, Claus D. Gehner, David A. Green, Yehuda Gur, Hubert A. Henry, Everett S. Joline, Jerome M. Lutin, Harold Moellering, Granville E. Paules, Ludie G. Richard, Jerry B. Schneider, Robert L. Siegel, Robert A. White

Transportation Systems and Planning Innovations Section
Richard J. Bouchard, Daniel, Mann, Johnson and Mendenhall, chairman

Committee on Statewide Multimodal Transportation Planning
Kenneth W. Shiatte, New York State Department of Transportation, chairman
John A. Bivens, Jr., Malcolm F. Brenan, William John Cameron, Roger L. Creighton, Robert E. David, Clarence W. Friesen, Arne L. Gausmann, Antonio W. Gonzales, Charles H. Graves, Philip I. Hazen, Lester A. Hoel, Thomas F. Humphrey, Colin Ian MacGillivray, Michael I. Schneider, Isaac Shafran, Carl N. Swerdloff, William C. Taylor, Thomas J. Thompson, D.L. Wieman, Nobert Y. Zucker

James A. Scott, Transportation Research Board staff

Sponsorship is indicated by a footnote at the end of each report. The organizational units, officers, and members are as of December 31, 1980.

Contents

Authors of Papers in This Record

Bauerlein, Joan W., Federal Highway Administration, U.S. Department of Transportation, 400 Seventh Street, S.W., Washington, DC 20590
Chin, Shih-Miao, Department of Civil Engineering, Rensselaer Polytechnic Institute, Troy, NY 12181
Daganzo, Carlos F., Department of Civil Engineering, University of California, 416 McLaughlin Hall, Berkeley, CA 94720
Dehghani, Youssef, Department of Civil Engineering, State University of New York at Buffalo, Buffalo, NY 14214
Eiger, Amir, Department of Civil Engineering, Rensselaer Polytechnic Institute, Troy, NY 12181
Gallagher, Marta J., Peat, Marwick, Mitchell and Co., 1990 K Street, N.W., Washington, DC 20006
Grayson, Alan, DTS-221, Transportation Systems Center, 55 Broadway, Cambridge, MA 02143
Hoffmeister, John F., Metropolitan Council of the Twin Cities Area, 300 Metro Square Building, 7th and Roberts Streets, St. Paul, MN 55101
Holec, James M., Jr., Peat, Marwick, Mitchell and Co., 1990 K Street, N.W., Washington, DC 20006
Kousheshi, Brenda, Department of Civil Engineering, State University of New York at Buffalo, Buffalo, NY 14214
Liotine, Matthew, Department of Civil Engineering, Princeton University, Princeton, NY 08544
Loudon, William R., Cambridge Systematics, 238 Main Street, Cambridge, MA 02142
Lutin, Jerome M., Parsons Brinckerhoff Quade and Douglas, Inc., One Penn Plaza, New York, NY 10119
Markowicz, Bernard P., Transportation Program, Princeton University, Princeton, NY 08544
May, Adolf D., Institute of Transportation Studies, University of California, 109 McLaughlin Hall, Berkeley, CA 94720
McKearney, Thomas, Department of Civil Engineering, State University of New York at Buffalo, Buffalo, NY 14214
Mertz, William L., Federal Highway Administration, U.S. Department of Transportation, 400 Seventh Street, S.W., Washington, DC 20590
Robertson, Richard B., Federal Highway Administration, U.S. Department of Transportation, 400 Seventh Street, S.W., Washington, DC 20590
Schwager, Dianne S., Peat, Marwick, Mitchell and Co., 1990 K Street, N.W., Washington, DC 20006
Sievert, Robert, Department of Civil Engineering, State University of New York at Buffalo, Buffalo, NY 14214
Skinner, Robert E., PRC Voorhees, Planning Research Corporation, 1500 Planning Research Drive, McLean, VA 22102
Stern, Wendy P., Transportation Policy Consultant, 77 Mason Terrace, Brookline, MA 02146
Westerman, Hans L., School of Town Planning, University of New South Wales, Sydney, Australia

Development of A National Highway Policy: An Interactive Process

WILLIAM L. MERTZ AND JOAN W. BAUERLEIN

The nation has spent the last 50 years building the best system of highways in the world. In the last decade, the federal government has dedicated $90 billion to this effort. This network of highways has become the backbone of the transportation system and has supported the economic growth and development of the United States. The highway system carries almost 90 percent of all personal and freight movements. Despite its significance, the highway system—particularly the national, interregional system—is deteriorating at an increasing rate as the federal and state governments grapple with decreasing revenues.

The nation has spent the last 50 years building the best system of highways in the world. In the last decade, the federal government has dedicated $90 billion to this effort. This network of highways has become the backbone of the transportation system and has supported the economic growth and development of the United States. The highway system carries almost 90 percent of all personal and freight movements. Despite its significance, the highway system--particularly the national, interregional system--is deteriorating at an increasing rate as the federal and state governments grapple with decreasing revenues.

Since 1970, travel has increased at an annual rate of 3 percent and, because of the growth in travel and the enormous impact of inflation on the highway capital dollar, conditions have deteriorated on the major highway system. Inflation has played a major role in the inability of governments to keep up with necessary improvements. Inflation in the country was about 10 percent in 1980. In the highway construction industry in 1979, the composite price index rose 13 percent. The bid price index, the indicator for construction costs, closed the fourth quarter of 1980 at 349.7, which means that what would have cost $100 in 1967, the base year, now costs $349.70.

During the past 10 years, the costs of maintaining a mile of roadway have increased by more than 111 percent, while the costs of highway construction and reconstruction have increased by more than 190 percent. Since 1973, the average rate of inflation in highway construction has been 12.5 percent per year, a rate that doubles highway costs every six years.

While highway costs are rising sharply in the wake of inflation and expensive energy, the user revenues to pay for these costs are leveling off and declining. Due to the shift to smaller, more efficient automobiles and decreasing highway travel, the growth rate of the income into the Highway Trust Fund has dropped sharply. From 1970 to 1978, the income grew at a rate of 4.5 percent per year. From 1979 to 1984, the projected increase is 1.5 percent per year.

These factors have resulted in a reduction of capital expenditures by all levels of government on highways of 42 percent in constant dollars from 1970 to 1978. More than half of this drop was caused by the inability of federal and state spending to keep up with inflation. States have reduced spending on all highways while local governments managed to fund large real increases on their roads during this eight-year period. Part of the states' decline was due to the states shifting their resources into noncapital categories, such as maintenance, traffic control, administration, engineering, and law enforcement. The local government increase was due to increased funding from general revenue sharing and the U.S. Department of Housing and Urban Development Community Development Block Grants, which local governments chose to spend on local roads and streets.

As we enter the 1980s, the country is embarking on a national effort of refurbishment and recovery. There are many new challenges to add to the traditional objectives of the highway program, e.g., energy conservation, preservation of the environment, and economic recovery. The emphasis in the highway program will be on preserving the existing systems and increasing the productivity in the management of the highway network.

In this framework and based on data and issues that have been distilled from other inputs, the 1981 highway legislation was developed. For very practical reasons, it is important to enact highway legislation in 1981. First, the programs terminate with the 1982 authorizations. Second, it is critical that the Interstate completion program be modified as soon as possible so as to shift to the important work of rehabilitation and reconstruction. Finally, the taxes for the Highway Trust Fund end on September 30, 1984. The taxes and trust fund need to be extended, and, if authorizations are increased significantly, revenues must also increase.

The process of developing the legislation presents an excellent case study of how national highway policy is formulated and presented for consideration by the Administration and Congress. The following discussion outlines the use of data and inputs from constituent groups, the development of goals, options, and, finally, legislative recommendations.

SUPPORTING INFORMATION PROCESS AND DATA

Status of the Nation's Highways: Conditions and Performance

U.S. Senate Joint Resolution 81, enacted in 1965, requires that a report on the nation's highway needs be submitted by the U.S. Secretary of Transportation every two years to the Congress. The first highway needs report was compiled in 1968, followed by biennial reports in 1970, 1972, 1974, and 1977. Previous reports on highway needs were based on the determination of estimated costs of improving all roads so that by 1990 no road would have physical or traffic characteristics below certain uniform operating and physical standards. The total costs to meet such needs in the past were enormous so in 1974 the report introduced the concept of "performance" as a standard of measurement. A performance index scale was constructed to illustrate the relationship between investment and levels of performance.

About three years ago the Federal Highway Administration (FHWA) established the highway performance monitoring system (HPMS), a data-collection and analysis tool that has been used to develop the relationship between investment and performance and condition. The HPMS depends heavily on state participation and samples the conditions and performance on 100 000 federal-aid road segments. The HPMS also reports accident rates and bus usage and is being expanded to include a fuel consumption factor and air pollution emission rates. The findings from the 1980 report that were submitted to the Congress in January 1981 are discussed in the section on definition of the issues.

Cost-Allocation Study

The 1978 Surface Transportation Assistance Act called for a cost-allocation study to be submitted to Congress in January 1982. An interim report presenting preliminary findings was transmitted to Congress in January 1981.

Cost allocation is the process of dividing up the cost of program outlays among both the various classes of highway users and nonusers and translating such distributions into revenue sources, such as general revenue and user taxes.

The process normally involves assessing those costs that can be specifically attributed to certain vehicle classes, generally based on the size, weight, performance, and level of use of the vehicle class. The attributable costs are both governmental costs (construction, maintenance, operation), and/or user interference costs (congestion, accidents), and/or external costs (noise, air pollution). There are also common or joint costs that specific classes of vehicles cannot be shown to cause in an unambiguous way.

The process of cost allocation then covers both assessing those costs that can be clearly attributed to certain vehicle classes and allocating those remaining common or joint costs that cannot be clearly attributed to certain vehicle classes.

In addition to the cost-allocation study, there is a companion study to be completed by the U.S. Department of the Treasury that will evaluate various taxing methods, the burden of the taxes, administrative costs, etc.

Truck Size and Weight Study

Section 161 of the Surface Transportation Assistance Act directed the Secretary of Transportation to study and investigate the need for and desirability of nationally uniform truck size and weight limits, the effects on construction and maintenance of roads, and related topics.

The Federal-Aid Highway Act of 1956 set maximum limits for truck operation on the Interstate. The limitations were permissive in nature, that is, they did not preempt the states' rights to establish lower limits, and they applied only to the Interstate system. The oil embargo in 1973 and the subsequent enactment of the national 55-mph limit generated greatly increased pressure from the trucking industry to increase the existing weight limits and to make the limits uniform.

The results from the truck size and weight study are related to the cost allocation and the revenues necessary to support the highway program. This relationship will be apparent in the legislative recommendations that are submitted to Congress in the study and ultimately as a part of highway legislation. The Carter proposal coupled changes in motor vehicle size and weight limits to changes in the highway program structure and its funding. Specific proposals will be based on the following principles:

1. Interstate commerce should be fostered through increased uniformity in motor vehicle size and weight limits;
2. Local and regional economic conditions must be accounted for in the establishment of revised size and weight limits;
3. Changes in size and weight limits should consider other national goals, such as preservation of the nation's highways, energy conservation, safety, and environmental concerns; and
4. The revised highway user tax structure should be consistent with the revised size and weight limits in terms of the relative contribution of various classes of vehicles.

Interstate Cost Estimate

The Interstate Cost Estimate (ICE) is required by Section 104(b)(5), 23 U.S.C. The purpose of the estimate is to derive the ratio of the federal share of the estimated cost of completing the Interstate system in each state to the sum of such costs for all the states to serve as a basis of apportioning funds annually. The 1981 ICE is the tenth in a series of estimates submitted to the Congress. The remaining cost-to-complete the Interstate system as of January 1, 1980, is $53.8 billion, of which $48 billion is the federal share.

Each of the studies discussed above depends heavily on input and cooperation with the states. The data form the critical base that makes it possible to develop policy and legislation at the national level and relates policy and legislation realistically to what is actually happening in the states.

STEPS IN POLICY DEVELOPMENT PROCESS

Defining Issues

By necessity the national emphasis of the 1980s will be on stewardship and conservation of existing functional systems and resources. This becomes apparent when the findings about the current conditions and trends on the federal-aid systems are analyzed. A few details about each of the federal highway systems tells the story.

The primary system is 271 000 miles of major arterials that connect almost all of the nation's cities with more than 50 000 population. It carries 28 percent of all highway travel and is generally in good condition. However, 7 percent of the pavement requires immediate replacement, and it is anticipated that 50 percent of the pavement will have to be replaced during the 1980s. Some 20 percent of the bridges in rural areas have some deficiency, and 23 percent of the urban primary mileage experiences severe congestion during peak-hour periods.

The secondary system includes 390 000 miles that provide intracounty service. About 14 percent of the mileage is still unpaved, and 8 percent of the remaining pavement needs immediate resurfacing. Some 30 percent of all bridges on the secondary system have a serious deficiency.

The urban system is not a national system as are the other federal-aid systems. The program funds a broad range of projects, including public transportation capital projects. About 7 percent of the pavement needs immediate replacement, and 18 percent of the bridges have some deficiency. Congestion, which occurs on 20 percent of the system during peak periods, is probably a fact of urban life.

The Interstate system, however, is of the most concern from the point of interstate commerce and the economic well-being of the country. Although the Interstate makes up only 1 percent of the nation's highways, it carries 20 percent of the traffic. The goals of the Interstate program have nearly been achieved as 96 percent of the designated 42 500 miles is either serving traffic or under construction. Nevertheless, in the third decade of construction, some sections remain to be built and, in fact, may never be built because of skyrocketing costs, environmental objections, and more pressing priorities.

While originally estimated to cost about $27 billion, it is now estimated that the cost to complete the system will be $54 billion, or at least twice as much to finish the last 4 percent as it cost to construct the first 96 percent. Further, from 1975 to the present, the Interstate system has deteriorated to the point that 8 percent of the pavement needs resurfacing immediately, 13 percent of the bridges have some deficiency, and 23 percent of the urban Interstate is congested during peak-hour periods.

Goals to Be Served

The traditional goals of the federal-aid highway program have been to provide for an Interstate highway network to serve national defense and commerce, to develop a balance among jurisdications in transportation, and to improve highway safety. In addition to serving the traditional goals, the highway program is now responding to the need to conserve energy; reduce inflation, minimize adverse social, economic, and environmental effects; and revitalize central cities. Balancing traditional goals and more recent national priorities involves two interrelated issues: (a) performance of the existing systems and the federal role in preserving the improving conditions, and (b) the source and amount of revenue that can be raised to finance the highway program. Given the interplay of these issues, it is likely that the Congress will consider the following:

1. Ensuring Adequate Highway Facilities--Real increases in authorizations may be considered in order to provide constant buying power over the life of the next highway bill because the real program level has decreased so dramatically over the last 10 years.
2. Completing the Interstate and Providing I-4R Funding--Provisions may be introduced to ensure rapid completion of remaining gaps on the Interstate and to resolve the problems surrounding the use of the ICE as the procedure for financing the Interstate. The intent would be to allow states to complete segments and then to focus on preserving and rehabilitating the existing system.
3. Providing Program Flexibility--States and others have expressed the need for a greater degree of program flexibility in selecting and implementing projects through consolidated program stucture and simplification of requirements.
4. Promoting Economic Revitalization and Recovery--Legislation will probably be considered in light of the need to revitalize the economy and to support other economic objectives.
5. Generating Sufficient Revenues--The legislation will reflect a consideration of whether the existing revenue structure will provide necessary financing to meet the goals of the program.

Development of Optional Program Structures

In March 1980, a legislative options paper was developed for presentation by FHWA as a part of the budget process and for review by the Office of the Secretary. This paper outlined some of the early information available on the conditions and performance of the federal-aid systems, the problems currently existing in the programs, and various alternative solutions. This included a discussion of the federal role in highway financing and individual program assessments. After the discussion paper was reviewed throughout the department and FHWA, it was distributed nationwide to states, state departments of transportation, legislators, local governments, and interested groups.

Optional Levels of Federal Involvement in Highway Programs

The federal role has become of particular interest in considering and formulating the 1981 federal-aid highway legislation due to the fundamental dilemma the program is facing, i.e., the increasing number of programs and national objectives being pursued with shrinking funds. The proliferation of programs and the multiplicity of goals have dissipated the leverage the federal government had through the program.

The federal role in the highway program has been asserted in many forms--the Constitution, legislation (Title 23, NEPA, etc.), policy statements, regulations, directives, program emphasis areas, etc. In the broadest sense, the highway program has grown out of the federal responsibilities under the commerce clause in the Constitution. Title 23 states that the objective of constructing federal-aid highway systems is "to meet the needs of local and interstate commerce, (and) for the national and civil defense."

The role of the federal government in the highway program has gradually expanded to assume responsibilities under the general welfare clause. The major changes in the last decade have been designed to assure that the highway program could accommodate national, state, and local goals and objectives. Thus, highways could not be built without considering the impact of the facility on the environment, without fully and fairly compensating individuals displaced, and without recognizing the amount of energy being consumed.

There are three options for continued federal involvement in the highway program. One is to continue to pursue the broad range of programs currently identified in Title 23 either with the current level of authorizations or with increased authorizations to maintain constant buying power. Second, the program could be substantially modified in order to target federal investments to more specific national objectives. This could include phasing out or reducing federal participation in one or several programs.

Another option would be fundamentally to change the federal role in financing the nation's highway and other surface transportation needs. This scenario would change the nature of the federal fuel tax from a user tax to a consumption tax or a conservation tax, and would include a significant increase in the federal revenues. Because the tax would no longer be a direct highway user tax, the federal-aid highway program neither could nor would be the lone beneficiary of the substantial receipts. These substantial receipts would provide an opportunity for a unified transportation trust fund, which has been proposed in the past or a tax turnback to the states.

Specific Program Options

Travel on the primary system during the 1980s is expected to grow at a slower rate than during the 1970s; however, even with a slow rate, approximately 20 percent more travel will occur on this system by the end of the decade. Also, the system will have to provide the capacity to carry coal and other resources for which there is an increasing domestic and foreign demand. As a result, 60 percent of rural and 80 percent of urban primary mileage will need repaving in the next 20 years.

The federal funding options over the next 10 years revolve around the issue of federal role in the primary program. These options are discussed below.

1. Increase Federal-Aid Primary Program Level to Maintain Today's Performance--This option would call for a major shift in the federal-state relationship in the primary program. It could put the federal government in the lead in determining how the system will perform over time and in standardizing, to some degree, 270 000 miles of arterial highway physical and operating characteristics. It would be extremely costly to assume responsibility for maintaining that level of performance. Assuming state-only funding would grow with federal program increases, the federal share of the 10-year needs would be more than three times the current program level.

2. Increase the Federal Program to Offset Inflationary Impacts--This option would continue to provide a stable financial base for the program yet would not provide the support necessary to offset the impacts of travel growth. The states would then have to determine individually how important the performance of the primary system is to their own development. A federal investment at today's constant dollar level would guarantee that no drastic deterioration would take place during the 1980s. If inflation averages 7 percent over the next 10 years, the federal program will have to increase to an average annual level of $2.6 billion, or a $0.8 billion increase over the 1981 authorized level of $1.8 billion.

3. Reduce the Federal Role in the Primary System--Under this option, which would reflect a major shift in the federal highway role, the primary system would no longer have the highest national significance. The states would bear more responsibility for improving their major arterials and ensuring connectivity and uniformity between states and regions of the country.

Travel on secondary routes is expected to increase at an average annual rate of 1.9 percent through 1990. This increase in travel will have little effect on the operational features of the system, because the capacity of the system far exceeds the demand placed on it. Exceptions to this will occur where rail branch lines are abandoned, and rural roads will be relied on to provide capacity for increased heavy truck traffic and increased demand for farm-to-harbor travel. The travel will, however, accelerate the deterioration of the pavement, and by 1990 nearly 90 percent of today's paved mileage will need replacing. In addition to the current deficiencies, future travel will create extensive problems on many sections that are currently adequately designed. By 1995, approximately 90 percent of all secondary system mileage will incur one or more deficiencies related to pavement, geometrics, roadway cross section, or operating performance.

From an operational standpoint the impact of the secondary routes on a nationwide highway network is marginal. The original purpose of the program, which was to develop paved farm-to-market routes, has certainly been accomplished since 85 percent of the routes are paved. Continued federal involvement in this program may no longer be so much related to system performance and needs as to revenue sharing and the distribution of federal revenues.

Federal options over the next 10 years are discussed below.

1. Increase the Federal Program Level to Maintain Current Level of Performance--The federal share of an annual cost to fund necessary improvements under this option would distort the importance of this program unless proportional increases were made to the remainder of the federal programs. It is likely that any significant increase in federal funding would signal a changing relationship in the current balance of responsibility for establishing priorities on this system. Increasing the federal funds available might force local priorities to be modified in order to meet the matching requirements.

2. Increase the Program Level to Compensate for Inflation--Growth of the federal-aid secondary program indexed to inflation would indicate that the federal role in this program remains directed toward achieving some level of system operating performance and that, from a national perspective, these routes have significance. It would also ensure that the federal government remains a major partner in the capital improvement program. However, if the states continue to draw away, as recent trends indicate, the program may become a federal-local partnership for the first time.

3. Consolidate into Block Grant Program--This option would consolidate several of the rural categorical programs into a more broadly based rural program, which would receive a single authorization.

4. Phase Out Federal Participation in This Program--It may be necessary at some point to choose between concentrating federal funding on certain federal systems such as the primary or Interstate and continuing federal funding for lower systems. Such a decision could be provoked if a decline in the federal program were to result in severe reductions in operating characteristics of the higher federal systems. In this case, the benefits derived from various investment options would have to be compared with the prevailing federal role in highways.

Two opposing forces are at work with regard to the federal role in urban highway programs; one is the higher national economic interest in the Interstate and primary system that drives the program away from local systems and local service decisions; and the other is the national concern with energy, air quality, urban development, service equity, and efficiency that drives the program into encouraging carpool/vanpool and transportation systems management actions. Two basic options would form the basis for the reassessment of the federal role:

1. Consolidate urban programs into block grant and focus effort on higher-order highway systems. Highway funds could also be consolidated with Section 3 and 5 Urban Mass Transporation (UMTA) funds.

2. Eliminate program and return the responsibility to the state and local governments.

FHWA has been analyzing new financing and program mechanisms that would accelerate completion of the initial construction of the system and provide a logical transition into a post-Interstate program that accommodates an expanded Interstate 3R program. The options define completion more narrowly than the term is currently defined. Cost elements that are deleted from the current definition, and, therefore, deleted from the ICE, would then be eligible for funding in an expanded Interstate 3R program, or a so-called "4R" program. The expanded 4R program would use an apportionment mechanism similar either to its current formula (based on mileage and travel), or would use a more traditional formula similar to that used for the federal-aid primary program (population, land area, and mileage).

Outreach Effort

While the states are involved in almost all of the data collected and actually manage the programs, it was felt that the program was at a turning point and in need of a fresh examination by all levels of government. Consequently, it was decided to conduct an open consultative process with various groups to receive their input directly into the legislative process.

Distribution of Discussion Paper for Comment

As mentioned earlier the discussion paper developed by FHWA and presented to the U.S. Department of Transportation (DOT) for review was circulated to nearly 2000 groups and individuals for comment. The discussion paper generated written responses from about 90 local governments, 45 states, and 14 special interest groups. The comments were compiled and integrated into a second discussion paper, which is available on request. The comments received are summarized below.

Federal Requirements and the Federal Role

There were repeated references to the multiplicity of goals being addressed by the federal highway program. In general, it was felt that the highway program was being distracted from its primary purpose. State and local governments would like to see the program streamlined, categories of funding reduced, liberal transfer provisions between programs adopted, and reporting requirements reduced. Environmental, equal employment opportunity, handicapped, inflation, and other cross-cutting requirements are real thorns in the side of state and local governments. More federal money and less federal control were desired by the respondents.

Interstate and Primary Systems

The strongest message from state departments of transportation was to complete all gaps in the Interstate system and to provide substantial financial support for restoring, resurfacing, and rehabilitating the Interstate and primary systems. Western states were especially concerned about losing funding to complete their gaps while states whose segments are virtually complete fear that their 3R needs will be overlooked in an attempt to concentrate on completion. Most respondents agreed that maintenance (correction of minor deficiencies and routine upkeep activities) should be funded at the state and local levels.

Urban (FAUS) Program

Most of the comments regarding this part of the paper came from local governments. These comments reflected a desire to continue funding, preferably through a block grant approach similar to that of the Community Development Block Grant or General Revenue Sharing Program. There was some hesitancy to combine this program with UMTA programs. While more federal money was requested, decreased federal and state control was recommended. Most aggravating to state and local governments are the federal design standards, environmental, and other cross-cutting requirements, and the obstacles and delays to which the program is subject. As might be expected, state-level officials were opposed to direct federal funding of urban projects although they would like to see the program continue and possibly expand.

Secondary Program

Fewer respondents addressed the issue of the secondary program, but those who did were adamant in their support for its continuance. These responses came primarily from southern and western agricultural states and from the Association of General Contractors. A block grant approach was favored.

Safety Issues

A number of states indicated support for consolidating all existing categorical safety construction programs. Colorado called for a reorganization/restructuring of safety programs under a single DOT agency. A similar sentiment was voiced by the American Society of Civil Engineers.

Trust Fund

There was tremendous support from all sectors for extending the life of the Trust Fund and continuing to rely on user taxes. Most respondents also agreed that the user taxes should be indexed to keep up with inflation. While an ad valorem tax was acceptable to many, tying the charges to the gross national product or the Consumer Price Index seemed to have greater support. There was strong sentiment that a general transportation Trust Fund not be established, but that a fund similar to the Highway Trust Fund be set up for transit.

Field Trips

Four regional trips were conducted by officials from FHWA, the Office of the Secretary, and congressional staffs. About 14 states were visited during the trips, and officials from another 10 or 12 states participated in the meetings. Further, interest groups representing environmental concerns, contractors, and various construction industries also participated. The basis for the discussion was the options paper circulated earlier. Useful comments were gathered during the discussions and based on the experiences, philosophies, and policies of the different states, regions, and groups represented. These views were recorded in trip reports, which are available from the Office of Congressional and Intergovernmental Affairs in the Office of the Secretary.

Legislative Initiatives

The legislative initiatives were assembled, reviewed at DOT, and transmitted to the Office of Management and Budget (OMB). The negotiations between OMB and DOT resulted in the following legislative recommendations submitted to the Congress by the Carter Administration.

Interstate Completion and 4R Program

The Carter legislation proposed to redefine completion by limiting the elements included in he ICE to upgrading all segments and building all gaps to a uniform and minimal level of service. This definition includes access control, pavement design for 20 years from the time of initial construction, and maximum lanes based on population (four lanes in rural areas, six lanes in areas with more than 400 000 population). With the redefinition of completion, there was also expansion of the 3R program to include all items deleted from the ICE and the reconstruction items not currently eligible under 3R. The authorization for the new 4R program would be about five times the size of the existing 3R program. The new cost to complete reduced the overall cost to $31 billion from $53 billion. (The Reagan Administration has proposed a similar approach in its highway bill.)

Consolidation of Programs

In the Carter bill, many narrow categories were combined into three larger categories; they are the federal-aid rural, federal-aid urban, and the federal-aid safety programs. The rural program was broadened to include capital expenditures for public transportation and rail branch lines. Both the urban and rural programs eliminated the concept of federal-aid system for project eligibility. Funds could be spent on any project on any public road. The consolidation of the safety programs included the rail-highway crossing safety programs, but would not change the congressional intent of the categorical programs, i.e., safety funds must be spent on safety projects, and they were not transferrable to any other programs. The Bridge Replacement and Rehabilitation Program was retained as a separate program. (While the Reagan bill retains the bridge program, it phases out the urban and secondary programs in two years and eliminates most of the safety and other small categorical programs in 1982.)

Highway Trust Fund

The Carter proposal recommended retaining the Highway Trust Fund as the main vehicle for financing highways and increasing the fuel tax from 4 to 6 cents. Other taxes were increased for heavy trucks, and all exemptions would be subject to sunset requirements in 1987. (The Reagan Administration bill extends the existing taxes through 1989 and the Trust Fund until 1990.)

CONCLUSION

The federal-aid highway programs and initiatives that will be developed in 1981 will be closely intertwined with national issues in transportation and other major issues facing the country, such as the need to control inflation and government spending. Nevertheless, the contribution that a well-functioning national system of highways makes to the growth of the national economy is significant, and the maintenance of the system is an important goal. The systems' conditions are not a surprise to highway officials who have been trying for years to solve many of these problems with declining revenues. In 1981 we have another opportunity to establish effective policies and to set realistic priorities to address these problems.

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

Role of Multistate Regions in Development of National Transportation Policy

RICHARD B. ROBERTSON

The experience of the Appalachian Regional Commission is used as an example of the role multistate regions can play in the development of national transportation policy. Most initiatives come from the states rather than federal agencies, in part because federal agencies do not need or want such assistance, or because they feel the states should decide such matters. Work done by regional commissions is generally welcomed by the states, but the reception by federal agencies is less enthusiastic. Conclusions and recommendations deal with national policy and agency regulations while calling for significant additional transportation investments in a particular region as opposed to the nation.

Is there a role for multistate regions in the development of national transportation policy? If so, how should a multistate area organize to make an input into such development? What are some examples of what has been tried and where have efforts succeeded and failed and for what reasons? This paper will address these points to some degree by using the Appalachian Regional Commission (ARC) as an example. It is not an attempt to settle the issue once and for all.

There is a valid role for an organized group of states in the development of national transportation policy. Some basic reasons are (a) recognition that many national transportation policies are interstate (or international) in nature, (b) to bring greater resources to bear on the identification of critical issues for a particular area, and (c) to apply these multistate resources to the resolution of such problems, with particular emphasis on consideration by the Congress and the Administration.

No single organizational arrangement is best for every issue, and several multistate organizations may often work toward resolution of the same problem. The American Association of State Highway and Transportation Officials (AASHTO), the National Governors' Association, the National League of Cities, and others assist groups of states on special interests, but they are national organizations usually trying to develop a national consensus. On the other hand, there are many multistate organizations such as the ARC (an independent agency), the Title V commissions (agencies within the U.S. Department of Commerce), the Tennessee-Tombigbee Waterway Authority (created by a compact of five states), etc., which normally seek special legislation favoring certain projects or geographic areas.

The ARC is an excellent example of how a multistate organization was created for certain reasons. One of the most important was a need to construct a highway system that, in conjunction with the Interstate system, would open up areas with a developmental potential. Perhaps the most important contribution made by the ARC is its way of making decisions. For that reason this paper will begin with a brief explanation of how the ARC is organized and then move to specific examples of activities in transportation areas.

The Appalachian region includes all of West Virginia and portions of 12 other states from New York to Mississippi. The ARC was created in 1965 by the U.S. Congress in response to the recommendation of a presidential advisory commission and the Conference of Appalachian Governors. Its overall purpose was to assist Appalachia in meeting its special problems, to promote its economic development, and to establish a framework for joint federal and state efforts on a coordinated and concerted regional basis. Some key elements of its structure and operation are

1. Independent commission comprised of a presidential appointee and the 13 Appalachian governors,
2. Decisions by the commission require the affirmative vote of the federal cochairman and of a majority of the state members,
3. A staff of 120 whose salaries are paid one-half by the federal government and one-half by the 13 states, and
4. Projects and programs funded from federal general funds and matching state/local funds.

Since the ARC is an independent agency, it can present its views directly to federal agencies, the Administration, and the Congress. Since the staff are paid equally by the states and the federal government and the executive director of the staff is appointed by the Commission, the staff is generally responsive on an equal basis to federal and state interests. The fact that Commission funds come from the general fund of the United States is significant for the Commission's highway program, because highway allocations to the states are in addition to federal Highway Trust Funds.

How, then, has the ARC been involved in the development of national transportation policy? There are a number of areas that deserve comment.

RURAL PUBLIC TRANSPORTATION

In 1972, the Commission funded its first rural public transit demonstration project. While the primary purpose was to serve employment trips, ARC emphasized the importance of moving toward coordination of social-service agency transportation. Funds for front-end planning and operational subsidies were provided, but management responsibilities were emphasized to decrease the percentage share of operational subsidies. Other demonstrations were made along with feasibility/management studies in eight states, which gave these projects an advantage in qualifying for Section 147 funds. The U.S. Senate Public Works Committee staff met with Commission staff to determine our rationale for subsidy of operational costs. Several of the most important elements of this program have been the Commission's flexibility and ability to provide front-end planning/administrative costs.

RAIL REORGANIZATION ACT OF 1973

At the request of a number of Appalachian states, the ARC participated in meetings of the Conference of States on Regional Rail Reorganization. The Commission met with the Federal Railroad Administration and the U.S. Railway Association (USRA) on a number of occasions to represent concerns regarding rail abandonment. It conducted a study in which a methodology was developed for measuring community impacts as the result of rail line abandonment and urged USRA to take this into account in the development of its final system plan. Formal presentations were made before the U.S. Interstate Commerce Commission (ICC) regarding the failure of USRA to properly consider economic development issues, including the movement of coal. The USRA and Conrail became more sensitive to these issues, partly because of this effort, and some concessions were achieved.

As part of its rail efforts, which began in 1974, the Commission prepared one report on the impact of the Railroad Reorganization Act on economic development in Appalachia. Another report determined the present operational and condition characteristics of all rail branch lines in Appalachia and examined current assistance programs and the railroads' own efforts to improve their capital investments and quality of service. The abandonment of light-density branch rail lines is a matter of serious concern for Appalachia because of its many rail-intensive but scattered industrial sites. Also, the Appalachian states with their coal resources find the rail system critical to their economies as energy demands grow.

AIRLINE DEREGULATION ACT OF 1978

The Commission joined with its member states in responding to proposals developed by the U.S. Civil Aeronautics Board (CAB) and its staff regarding implementation of the small community air service program and the determination of "essential air service." Several formal presentations were made to the CAB in addition to ARC's technical assistance to a number of small communities in Appalachia. Air transportation was shown to be important in achieving the developmental goals of the region, and a quick review of air service in Appalachia pointed to a severe deterioration since deregulation occurred. In view of this, the Commission has a study under way on the effects of airline deregulation on air service in Appalachia. The purpose of this study is to (a) establish the facts about changes in air service to Appalachian communities since deregulation, (b) identify problems and issues in the transition from a regulated to a deregulated environment, and (c) develop proposed policies and programs for assuring adequate air transportation services in the future. Currently, the ARC is preparing a response to the Federal Aviation Administration's (FAA) rulemaking on the functioning of slot allocation committees at National Airport.

COAL HAUL ROADS

Due to the importance of Appalachian coal, the Commission undertook an assessment of the effects of coal movement on the highways in the Appalachian Region. Some 14 300 miles of roads within the eight coal-providing states were identified as coal haul roads. A conservative estimate set the cost to reconstruct existing roads and bridges to adequate structural standards for coal haulage at about $4.5 billion. This work was completed in November 1977 and was used, in part, by the Federal Highway Administration (FHWA) in its Coal Haul Road Study completed in April 1980. The Commission worked with FHWA to devise a methodology for a more detailed state-by-state assessment of coal road needs throughout the United States. The Commission worked with the U.S. Departments of Transportation and of Energy along with the Office of Management and Budget (OMB) in an effort to define possible funding sources for a coal haul road improvement program. The Commission also joined with the National Governors' Association (NGA) to address how nationwide needs regarding coal haul roads and coal train impacts at grade crossings might be funded. The NGA and ARC transportation representatives agreed that a $10 billion program over a 10-year period, funded from the windfall profit tax, was the proper approach. This was not endorsed by the Carter Administration.

TENNESSEE TOMBIGBEE WATERWAY

Over the period 1975 to 1977, the Commission made a comprehensive assessment and evaluation of the impacts and development opportunities that would result from construction of the Tennessee Tombigbee Waterway. It looked into what public policies and programs would be needed to accommodate future changes to capture the development opportunities of the waterway. This effort was not made to justify the construction of the waterway--it assumed its completion and then focused on the development opportunities that would accrue to the impacted area and how to take advantage of them. As a followup to this study, the Commission provided $12 million for special access roads in Mississippi and made other funds available for port studies and related development planning.

COAL SLURRY PIPELINES

In 1978 the Commission completed a study of the coal flow network in Appalachia. This study identified those coal flows with a potential for slurry pipeline application. These were analyzed and the cost of coal transportation by a slurry pipeline determined and compared with that of the competing mode. The socioeconomic and environmental implications of coal slurry pipelines were analyzed within the context of the more likely applications. Recommendations were made regarding the development of regional and state policies regarding coal slurry pipeline applications.

OTHER ENERGY TRANSPORTATION

The Commission undertook a broad-ranging study on the major movements into, through, and out of the region by various transportation modes (primarily rail, water, and pipeline) of all energy commodities produced or consumed. The purpose of this study, completed in 1978, was (a) to identify potential mainline capacity problems, (b) to develop recommendations on energy and transportation policies, (c) to develop information on energy and energy flow in Appalachia, and (d) to develop an analytical methodology usable for continuing policy analysis.

In 1980, the ARC decided to review its previous transportation efforts relative to the production, use, and transport of Appalachian coal and other energy resources. Based on this review, a review of other agency studies, and an assessment of those issue areas most critical to the region in both the short and long term, the Commission will undertake a series of energy transportation efforts in 1981--concentrating on items it believes can be positively impacted by the Commission effort.

In late 1980, a truck and rail deregulation study was initiated. Its purpose is to assess the impact of deregulation on the quality and quantity of goods transportation service provided to the Appalachian Region.

The results of these efforts may be used to seek legislative changes and administrative/regulatory rulemaking.

APPALACHIAN DEVELOPMENT HIGHWAY SYSTEM

The ARC may be known best for its effort to construct a 3025-mile Appalachian Development Highway System. More than $2.4 billion in federal non-highway trust funds have been obligated on this system since it was designated in 1965 and more than 60 percent of the system is either constructed or under construction at this time. The purpose of this system is to open up areas of Appalachia with a potential for economic development. While much of it is designed close to Interstate standards, it becomes part of the Federal Aid Primary System (FAPS) when completed. This is an economic development highway program, and it was the basic concept used for FHWA's Economic Growth Center Development Highway Program. The Commission is conducting an in-depth review of the Appalachian Development Highway System in 1981 to develop a realistic strategy regarding its completion.

SUMMARY AND CONCLUSIONS

The ARC has conducted studies on highways, rail, air, mass transit, waterways, and pipelines over the years. In addition it has funded construction of the Appalachian Development Highway System to advance economic development in the region. Except for the Appalachian Development Highway System, none of the Commission efforts cited in this paper occurred in the Commission's first seven years (1965-1971). Exactly why more transportation issues were not addressed in the early years is not clear. In 1974, there was a Commission assessment of previous transportation projects and programs, which resulted in recommendations concerning transportation efforts to be undertaken in future years. Prior to this effort, there had been a lack of focus on multimodal issues. A staff reorganization in 1975 to remedy this problem resulted in a Transportation Division within the Commission and consolidation of all transportation responsibilities. There may have been a heightened perception of the Commission's transportation abilities by its member states. There was recognition of a need for concerted multistate action on pressing issues such as the Rail Reorganization Act of 1973, the Airline Deregulation Act of 1978, and the high cost involved in repairing coal haul roads.

What can be concluded from a review of the ARC's transportation efforts and its impact on the development of national transportation policy? What does it imply for multistate efforts in the future? Almost all of the efforts undertaken by the ARC and identified in this paper (except for the Appalachian Development Highway System) were initiated at the request of a number of Appalachian states. Sometimes the request came directly from the governor, since the Commission works directly with his office, and sometimes from the Transportation and Highway Departments, through the governor's office.

Federal agencies have approached the Commission on a more infrequent basis with requests to help on transportation issues. This may be due to a belief that they need less assistance, or because they believe it should be up to the states to decide whether the Commission should have a role to play, or perhaps because the issue involves controversy over their own programs.

Work done by multistate organizations is generally well received by member states, but less success is achieved with federal agencies. A number of reasons for this may be (a) agencies such as the ICC, USRA, and CAB may not be as responsive as others due to their more independent nature; (b) the multistate recommendations may call for significant investments, such as the coal haul road program; (c) the organization may not be perceived as having enough political clout; (d) proposals may be viewed as beneficial to only a small group of states; and

(e) proposals are usually directed more to changes in federal rules, regulations, and legislation.

Multistate proposals to federal agencies are more likely to be implemented when (a) the issue is addressed enough in advance of key decision points that there is time to coordinate efforts by elected representatives, (b) working relationships are developed with professional staffs of congressional committees, transportation interest groups, and the federal agencies; (c) sufficient funds and adequate staff are available at the multistate organization to produce a thorough report; (d) the substance and quality of previous and ongoing work are considered appropriate by the transportation professionals (and acceptable to elected and appointed officials) representing at least the multistate area served by the organization; and (e) state and federal transportation staffs are involved with the effort from its inception.

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

Improving Usefulness of Section 15 Data for Public Transit

JAMES M. HOLEC, JR., DIANNE S. SCHWAGER, AND MARTA J. GALLAGHER

The purpose of this paper is to accelerate the creative and insightful use of a new and powerful data base. The paper focuses on the use of Section 15 data as a surveillance and monitoring tool for statewide transportation planning and management. Use of Section 15 data for this purpose is receiving widespread attention and is advancing from initial consideration to development and implementation in many areas. This activity is likely to increase with the release of Section 15 data by the Urban Mass Transportation Administration. Two principal methods for improving the usefulness of Section 15 data are discussed in this paper. The first method involves improving the potential user's familiarity with the nature and quality of the data. This familiarity will foster informed analysis and limit misrepresentation of a transit system's financial and operating performance. The second method involves enhancing the data base itself through editing and correcting the initial submissions of transit operators, clarifying reporting instructions (and thereby improving the quality of data submitted), modifying reporting forms, refining data-collection techniques, adding or deleting data elements, and/or augmenting the Section 15 data base with other available data. These methods are introduced by first providing a brief perspective on the type of information contained in the Section 15 data base, discussing specific shortcomings with the current data, and concluding with a summary of methods for improving the usefulness of the data base.

In November 1974, the Urban Mass Transportation Act was amended to introduce federal participation in the financing of transit system operating expenses. Provision of funds for this purpose through Section 5 of the Act was accompanied by a directive to the U.S. Secretary of Transportation to develop, test, and prescribe a uniform system of accounts and records "to accumulate public mass transportation financial and operating information." The directive further specified that, after July 1, 1978, no grantee could receive federal operating assistance through the Section 5 program without complying with this reporting requirement. The portion of the Act that established this new requirement was Section 15.

The first full year of the Section 15 reporting system encompasses the reports of transit systems with fiscal years ending between July 1, 1978, and June 30, 1979. The Urban Mass Transportation Administration (UMTA) has received nearly two full years of data under this reporting system and is planning its initial release of industry summaries for the first full year, reflecting data for more than 300 transit systems.

WHAT IS CONTAINED IN THE SECTION 15 DATA BASE

To obtain a complete understanding of the Section 15 reporting system and the information contained in the Section 15 data base, it is essential that potential users of these data review the report, Urban Mass Transportation Industry Uniform System of Accounts and Records and Reporting System, and its complementary reporting manuals: Required Reporting Manual and Sample Forms, Level C Reporting Manual and Sample Forms, Level B Reporting Manual and Sample Forms, and Level A Reporting Manual and Sample Forms. This documentation provides detailed instructions and sample forms for filing Section 15 reports in compliance with federal requirements. The report is available through the National Technical Information Service (NTIS); the manuals can be acquired through UMTA's Office of Transportation Management.

The Section 15 system consists of multiple levels of reporting detail reflecting differences in the size of the transit agency submitting data (measured by the number of vehicles it operates in revenue service). For each level of reporting, data are submitted on the sources and uses of funds for capital and operations, and on the physical, service, and utilization characteristics of the operating system. Financial information is provided on an accrual basis of accounting and the reporting schedule is designed to allow for the reporting of audited financial data as required by the Section 15 system. Physical, level-of-service, and utilization characteristics are based on counts at a point in time (e.g., revenue vehicles are reported at the end of the year level), accumulation of data throughout the year (e.g., annual vehicle miles operated or annual accidents by category), or estimates of annual totals based on sample observations collected randomly throughout the year (e.g., annual passenger trips or annual passenger miles).

Figure 1 presents the type of information contained in the Section 15 data base and Figure 2 illustrates a typical format for summarizing this information. These exhibits begin to suggest the type of analyses that can be conducted by using Section 15 data and showing the compilation of information in selected categories.

The use of Section 15 data for the surveillance and monitoring activities of state agencies is currently in the formative stages. In this developmental period, it is important for these state agencies to be familiar with the quality of information submitted in first-year Section 15 reports and efforts under way to improve the accuracy and completeness of the Section 15 data base over time. The nature, extent, and outcome of a review of first-year Section 15 reports are described below to provide potential users with a preliminary assessment of year 1 data quality.

REVIEW OF FIRST-YEAR SECTION 15 REPORTS

Each of the first-year Section 15 reports submitted to UMTA was reviewed at least four times prior to the preparation of computerized output reports that are currently being prepared by UMTA. Three of the review procedures were manual and one included a limited automated validation as the data base was formed. To the extent possible, corrections were made to the data following each review. The intent of reviewing the data was (a) to ensure that the data were in a form suitable for keypunching; (b) to check that the keypunched data reflected the data submitted by the operator; (c) to check that the data did not contain arithmetic errors; and (d) to ensure that the data appeared reasonable, based on a series of checks performed on selected information in each report.

Figure 1. Summary of Section 15 data by type of information reported.

TRANSIT REVENUE DATA
- Sources of Transit Operating Revenue
- Sources of Public Operating Assistance
- State and Local Taxes Dedicated at Their Source for Transit Operations
- Sources of Public Assistance for Transit Capital
- State and Local Taxes Dedicated at Their Source for Transit Capital

TRANSIT EXPENSE DATA
- Transit Operating Expenses by Mode and Function
- Transit Operating Expenses by Mode and Object Class
- Employer Paid Fringe Benefits
- Employee Paid Fringe Benefits
- Operators Hours and Wages –– Operating Time
- Operators Hours and Wages –– Non-operating Time
- Transit System Pension Plans

TRANSIT NON-FINANCIAL DATA
- Service Supplied and Service Consumed by Mode
- Age Distribution of Revenue Vehicles by Mode
- Operating Labor Employee Count
- Capital Labor Employee Count
- Safety and Energy Consumption Data
- Transit Way Mileage

First Manual Review: Prior to Keypunching the Data

Each Section 15 report was manually reviewed before the data were keypunched. The reports were checked for completeness, rounding, and gross errors. The review for completeness included three steps. First, the forms were reviewed to determine that the appropriate forms were included in each report based on the reporting level and number of modes operated. Extraneous forms were discarded and missing forms noted.

Next, the identification data (identification number, fiscal year, reporting level, and mode code) on each form were checked to ensure that the data were consistent within a report. This check was important since the entry of each line of data into the computerized data base includes the identification information. The final review before keypunching involved checking that all subtotals and totals were filled in. When data were missing, it was entered on the report.

The review for rounding involved checking all financial data to ensure that the data were reported to the nearest dollar. In addition, some of the operating statistics were checked to ensure that data were reported to the level of significance specified in the Section 15 reporting instructions. When errors were noted, corrections were made.

During this first manual review effort, some reports appeared to have substantial errors or missing data. These reports were generally not keypunched. UMTA contacted these transit systems and either a more complete report was submitted or the report was not included in the first-year Section 15 data base. For example, transit systems that changed fiscal years or were new starts in FY 79 and had

Figure 2. Sources of public operating assistance by transit system size.

SOURCES OF PUBLIC OPERATING ASSISTANCE	PERCENT OF TOTAL OPERATING ASSISTANCE							
	SIZE OF TRANSIT SYSTEM: NUMBER OF REVENUE VEHICLES							ALL SYSTEMS
	UNDER 26	26 – 49	50 – 99	100 – 249	250 – 499	500 – 999	1000 +	
Federal Government								
Section 5								
Other Federal								
Total Federal Assistance								
State Government								
General Revenues								
Dedicated Sources								
Total State Assistance								
Local Government								
General Revenues								
Dedicated Sources								
Total Local Assistance								
Total Public Assistance	100%	100%	100%	100%	100%	100%	100%	100%
Number Systems Reporting								
Number Systems UMTA								

reported data for less than a 12-month period were not included.

Automated Review

Each Section 15 report was keypunched and a computerized output report replicating the report submitted by the operator was produced (referred to as an ECHO report). During the production of ECHO reports the data were checked for errors in arithmetic, errors in carrying forward subtotals on multiple page forms, and for missing data or incorrect entry of codes specified in the Section 15 instructions (i.e., fuel type and vehicle ownership codes) on selected forms. When these types of errors were identified, an asterisk was printed next to the appropriate line on the ECHO report.

Second Manual Review: Keypunch Verification

The second manual review was conducted after the ECHO reports were produced. The primary objective of this review was to verify keypunching and ensure that the ECHO report replicated the data submitted by the transit operator. When keypunch errors were noted, corrections were made on the ECHO report. These corrections were later rekeypunched to produce a "clean" ECHO report. As part of this review, additional errors noted in the automated review were also corrected. Missing data and incorrect codes identified in the automated review were noted. Efforts to correct these errors were incorporated in the next phase of the review process.

Third Manual Review: Reasonableness Check and Telephone Contact

The final manual review effort was the most detailed. The objectives of this effort were to assess the reasonableness of selected data and to correct errors when possible. This effort involved reviewing selected data against a series of reasonability checks, noting data that appeared questionable, and contacting the transit operator by telephone to verify the data or make corrections if the data were incorrect.

The checks developed for this review effort reflected an understanding of the transit industry and the performance of transit systems and an understanding of the Section 15 reporting requirements. These checks were not exhaustive and will be refined and expanded over time for verification of future Section 15 reports.

Some of the checks included in the review of the first-year Section 15 reports defined reasonable bounds within which the data were expected to fall. Some examples of this are noted below.

Federal funds for both capital assistance and operating assistance were checked against the state and local funds reported to ensure that the federal funds (Sections 3 and 5) did not exceed the expected match allowed by federal law, i.e., 80:20 for capital assistance and 50:50 for operating assistance.

The average cost per gallon of fuel was calculated as a check on the data reported for fuel consumption and fuel expense. The price per gallon was expected to fall between \$0.30 and \$0.70 (reflecting diesel fuel price before June 1979).

Fare-box revenues represent up to 70 percent of total transit system revenues.

Average mileage per vehicle per day was calculated as a check for total and revenue miles of service supplied and number of vehicles in operation. In general, motor buses travel between 50 and 175 miles per day; therefore, the data were expected to fall within this range.

Average transit speed was calculated for an average weekday, Saturday, Sunday, and each period of the weekday. This was to check total and revenue miles and hours of service supplied. Average speed was expected to fall between 4 and 20 miles/h. (Rapid rail vehicle speed was at a higher range.)

Other checks compared data between Section 15 forms in instances where data should have been identical on more than one form. For example, local state and federal assistance reported on Form 201 (or 202), Revenue Summary Schedule, should match local, state, and federal assistance reported on Form 203, Revenue Subsidiary Schedule Sources of Public Assistance; and total fringe benefits reported on Form 331, Fringe Benefit Subsidiary Schedule, should match total expenses for fringe benefits reported on the appropriate expense form (determined by reporting level and number of modes operated).

A third type of check identified data that should generally be included in all of the Section 15 reports. These included matching local or state funds for federal assistance, expenses for insurance or accidents and claims, employee count data, service period schedule (i.e., time of day service is provided), service supplied by time of day, and service consumed by time of day. A total of up to 57 checks for data reasonableness were made to each Section 15 report.

If a data entry did not pass a check, a note was made regarding why the data were considered questionable. After the review of the report was completed, the transit operator was contacted by telephone and each questionable data entry was discussed. In general, the transit operators were cooperative in verifying the data and most problems were resolved.

Not all data that were questioned were incorrect. In some cases, the operator was able to provide a reasonable explanation for the data. For example, several small bus systems that largely provide commuter service into New York City receive more than 90 percent of their revenues from the fare box and several bus systems that largely provide airport limousine service maintain average miles per day and average speeds higher than those expected for a transit system.

If problems were acknowledged and the operator could adjust the data, the ECHO report was corrected and the correction was rekeypunched. If the transit operators indicated that it would be impossible to correct the data in the first-year Section 15 report, they were advised how to avoid similar errors in their future reports. These errors were not corrected and the data were generally not deleted.

QUALITY OF YEAR-1 SECTION 15 DATA BASE

Through this four-step review process the first-year Section 15 data were considerably improved. Many errors were identified and corrected. Other errors, however, were identified but could not be corrected. Therefore, users of the Section 15 data from the first-year reports should consider the potential shortcomings outlined below.

Corrected Errors

Errors in rounding data, arithmetic, identification, and missing subtotals and totals were corrected on all reports. In most cases, errors that resulted from a misunderstanding of reporting instructions were corrected during the telephone conversation with minimal effort by the transit operator. For example, these errors included

1. Many transit operators apparently did not

understand the instructions on Form 401, Transit Service Period Schedule. The 24-h clock was not used as instructed and there was confusion on how to calculate service periods.

2. Many transit operators interpreted the meaning for "average weekday" on Form 406/407, Transit System Service Supplied, Service Consumed and Personnel Schedule, to mean the average during the day instead of the total for the day.

3. Several transit systems did not understand the difference between dedicated taxes, i.e., funds dedicated to transit at their source by a local or state government, and taxes levied directly by a transit authority.

4. Because UMTA had advised transit systems to report data on Form 406/407 in whole numbers rather than in thousands, as is printed on the form, some inconsistencies existed in these data (UMTA decided to change the instructions when smaller operators complained that their data could not be meaningfully reported in thousands.

5. Some operators did not understand the instructions for vehicle type, ownership, and fuel type code on Form 408, Revenue Vehicle Inventory Schedule.

Some of the errors discussed with the transit operator during the telephone conversation required some effort on the part of the operator to make needed corrections. In these instances the transit operator was allowed the necessary time to make these adjustments. These types of errors occurred most frequently when either capital or operating grants were reported on a cash rather than an accrual basis. Some operators reported multiple-year grants as FY 79 revenue or reported no assistance because monies had been applied for and approved but not yet received. These types of errors resulted in an under or overreporting of financial assistance on Form 103, Capital Subsidiary Schedule, Sources of Capital Assistance, and Form 203, Revenue Summary Schedule Sources of Public Assistance. In the majority of cases, corrections were made and the operator indicated that year-2 data would be correctly reported.

Unresolved Errors

Finally, there were some errors that could generally not be resolved. The large majority of these errors represented data that had not been gathered by the transit systems for their first-year Section 15 report. This occurred most frequently with the service consumed (passenger statistics) reported on Form 406/407. To a more limited extent service supplied data and occasionally other operating data had not been gathered. In almost all of these instances the transit operators indicated that they would be unable to provide the missing data in their first-year Section 15 report but would provide that data in future reports.

Unknown Errors

A final type of error may exist in the first-year Section 15 reports. These are errors for which no review was made in the first-year reports. The checks included to review the data in the first-year reports were performed on selected data and were not exhaustive. During the process of checking the reports the need for additional checks was observed. For example, it was noted that transit operators were overreporting the average minutes per unlinked trip on Form 406/407. Some operators had reported the average trip length as hundreds or even thousands of minutes.

The extent of errors of this type is, of course, unknown at this time. UMTA intends to conduct additional research in the future to further refine and correct the future Section 15 reports. As information is obtained to correct and improve reporting practices, Section 15 report forms will be improved and the transit operators will be advised to ensure the refinement of the Section 15 data base over time.

Summary of Review Process

During the first-year review of Section 15 data, 57 individual checks for data reasonability were conducted. These checks involved the assessment, at varying levels of detail, of 432 out of a total of more than 800 fields of data contained in each Section 15 report submitted. The most intense review was concentrated on 147 of these fields.

A measure of the reasonability of the year-1 Section 15 data base can be obtained by examining a sample of the property submittals as illustrated below:

Factor	No. of Data Checks
Possible checks (57 per property)	1995
Questionable data identified	195 (9.8 percent of possible checks)
Questions resolved	145 (74.4 percent of questionable data)
Questions unresolved	50 (25.6 percent of questionable data)

For this sample, a total of 2.5 percent (50 ÷ 1995) of the data checks made in the year-1 Section 15 review remain unresolved and represent questionable, if not incorrect, data within the data base.

IMPROVING USEFULNESS OF SECTION 15 DATA FOR PUBLIC TRANSIT

Anticipating the use of Section 15 data beyond year 1, UMTA is continuing to implement a program to upgrade the quality of the Section 15 data base in order to improve its usefulness for public transit. This effort is being concentrated in four areas: (a) editing of data submitted by transit operators, (b) clarifying data definitions and instructions, (c) improving data-collection techniques, and (d) organizing for the creative dissemination of Section 15 reports. These efforts are evolving as familiarity with the Section 15 data base improves and specific reporting problems become known.

Three improvements are currently being considered with respect to the editing of data submitted by transit operators. The first involves the development of "tighter" edits by using data reported in year 1 to review subsequent submittals. The second involves the addition of edit checks not included in the year-1 review. These additional checks could take two forms: checks of specific data items not previously reviewed and checks of selected performance indicators not previously reviewed. The third improvement involves the automation of the data-checking process; this improvement is designed to reduce the possible errors introduced in the tedious manual checking process.

In addition to the review and improvement of submitted data, efforts are under way to upgrade the data at its source. These efforts are focused on the clarification of data definitions and instructions and the development of improved data-collection techniques. The principal mechanism being used to clarify data definitions and instructions (aside from direct communication with transit agencies) is the Section 15 Accounting and Reporting Release series initiated by UMTA Circular C2710.5 on February 11, 1980. This series, which is planned as an intermittent publication, elaborates on Section 15 definitions and instructions in a question-and-answer format with extensive use of examples.

Efforts to develop improved data-collection techniques are focused on those items of data for which annual estimates are prepared based on randomly collected samples throughout the year (e.g., the annual passenger trips and annual passenger miles). These efforts are concerned with methods to reduce the required sample size for all transit agencies in general and for smaller transit systems in particular.

In addition to activities designed to improve the quality of the data, efforts are under way to improve the usefulness of the Section 15 data base by organizing for the creative dissemination of reports and analytical access by potential users. These efforts are focusing on the development of core reports providing published access to the bulk of the data and information contained in the Section 15 system, methods for preparing ad hoc, customized reports on request, and methods for allowing potential users direct access to the full data set. These efforts are currently in the developmental stage within UMTA.

APPLYING SECTION 15 DATA BASE FOR STATEWIDE TRANSIT ANALYSIS

The keys to the use of Section 15 data as a surveillance and monitoring tool are (a) confidence in the data base, (b) familiarity with the data base, and (c) understanding the meaning of reported data (recognizing the nature and extent of the information contained in the reported values submitted by individual transit agencies).

In the preceding section, some of the shortcomings with the first-year Section 15 reports were outlined to provide initial guidance to potential users in considering analyses based on these data. This is a sound foundation on which to build an appreciation for the strengths and limitations of the first-year Section 15 data base; however, it provides only broad guidelines to assist with the analysis and interpretation of individual transit agency reports. States considering the use of Section 15 data as a surveillance and monitoring tool must first determine the relevance of these shortcomings for the systems it intends to routinely review. This will require a careful screening of the hard-copy submittals of individual transit agencies to ensure complete, valid, and reasonable reports. Once this prescreening is complete, the more difficult task of taking advantage of the information content contained in these reports can begin. For this purpose, states considering the use of Section 15 data in surveillance and monitoring activities are looking to this data base as a foundation for a performance measurement system using selected performance indicators.

Figure 3 illustrates the type of performance indicators that can be developed by using the most rudimentary reporting level (the required level) within the Section 15 system. This figure represents a subset of the performance indicators being considered for routine performance monitoring by the State of Michigan.

Figure 4 provides an example of the type of instructions necessary for the development of performance indicators from the Section 15 data base. The instructions identify (a) the performance indicator, (b) the data elements included in each indicator (generally by report form and line number), and (c) the arithmetic operation required to develop each indicator. As surveillance and monitoring activities become more routine, these instructions can be automated to eliminate the manual computation of indicator values.

To begin to appreciate the information content of the data it is suggested that the state agency define the measures of performance as shown in Figure 5. This approach gives meaning and content to the measures evaluated by the state and provides a uniform basis for interpreting the results of performance measurement.

With this foundation, three types of indicator analyses can be suggested: (a) comparison of single-year indicator values for each system to a statewide average value (or to an average value for peer systems selected regionally or nationally), (b) comparison of the value of an indicator and its rate of change to an average value for a selected peer group, or (c) comparison of the performance trend of each indicator over time.

Figure 6 shows the type of analysis that can be conducted, by using the first approach. In this exhibit, data from a sample of six Section 15 reports are used to determine the dispersion of individual property data in relation to the reported average indicator values for a peer group. The information obtained from this analysis can be used to focus the attention of the state on specific transit agencies with apparent performance well above or well below average, and direct further analysis efforts to determine improvement opportunities within the state.

Figure 7 shows the type of analysis that can be conducted by using the second approach. The value of each indicator can be tracked over time and compared with the average value and average rate of change for a peer group. As shown in this figure, six scenarios might be observed in comparing each indicator with a peer group average, with each scenario calling attention to different patterns of performance over time. The scenarios represent:

1. A--The indicator is below the statewide average value in year 1, increasing at a slower rate than average, and is below the statewide average value in year 2;
2. B--The indicator is above the statewide average value in year 1, increasing at a slower rate than average, and is below the average value in year 2;
3. C--The indicator is below the statewide average value in year 1, increasing at a faster rate than average, and is below the average value in year 2;
4. D--The indicator is above the statewide average value in year 1, increasing at a slower rate than average, and is above the statewide average in year 2;
5. E--The indicator is below the statewide average value in year 1, increasing at a faster rate than average, and is above the statewide average in year 2; and
6. F--The indicator is above the statewide average value in year 1, increasing at a faster rate than average, and is above the statewide average in year 2.

Scenarios A and F represent the most extreme situations in terms of an indicator differing from the statewide average in both value and rate of change. While it would be informative to assess the values of all of the indicators in relation to each of these scenarios, this could be prohibitively time consuming; the analysis requirements can be conveni-

Figure 3. Illustrative indicator structure: vehicle operations labor expense.

$$\frac{\text{Total Operator Straight Time + Scheduled Overtime Hours}}{\text{Total Platform Hours}} + \frac{\text{Total Unscheduled Operator Overtime Premium Hours}}{\text{Total Platform Hours}} + \frac{\text{Total Nonoperations [Operator] Paid Worktime Hours}}{\text{Total Platform Hours}}$$

$$\frac{\text{Total Vehicle Operations Labor Expense}}{\text{Total Vehicle Miles}} = \frac{\text{Total Platform Hours}}{\text{Total Vehicle Miles}} \times \frac{\text{Total Operator Pay Hours}}{\text{Total Platform Hours}} \times \frac{\text{Total Operator Salaries and Wages}}{\text{Total Operator Pay Hours}} \times \frac{\text{Total Vehicle Operations Salaries and Wages}}{\text{Total Operator Salaries and Wages}} \times \frac{\text{Total Vehicle Operations Labor Expense}}{\text{Total Vehicle Operations Salaries and Wages}}$$

Figure 4. Indicators of total system efficiency: administrative labor expense detail.

Data Source: Section 15 Report Forms

Special Instructions: * These data elements represent the sum of the data reported for all modes operated.

INDICATOR	DATA ELEMENTS	INDICATOR ESTIMATE
Total Administrative Labor Expense / Total Vehicle Miles	(Form 310 (pg. 1) Ln 1e + 2e + 3e) / (Form 408, Sum Col. K)	
Total Vehicle Miles / Total Number of Administrative Employees	(Form 408, Sum Col. K) / (Form 404, Ln 9b + 9c + 10b + 10c)*	
Total Administrative Salaries and Wages / Total Number of Administrative Employees	(Form 310 (pg. 1) Ln 1e + 2e) / (Form 404, Ln 9b + 9c + 10b + 10c)*	
Total Administrative Labor Expense / Total Administrative Salaries and Wages	(Form 310 (pg. 1) Ln 1e + 2e + 3e) / (Form 310 (pg. 1) Ln 1e + 2e)	

Figure 5. Components of major efficiency indicators: administrative labor expense per vehicle mile.

INDICATOR	ANNUAL STATISTIC	RELEVANCE
Administrative Employees Per Vehicle Mile (1)	Total Vehicle Miles (1) / Total Administrative Employees	Measure of administrative labor productivity. Reflects the number of employees relative to service provided.
Administrative Salaries & Wages Per Administrative Employee	Total Admin. Salaries & Wages / Total Administrative Employees	Measures average wage not including benefits. This is a major factor in explaining administrative expense.
Total Admin. Labor Expense Per Admin. Salaries & Wages	Total Admin. Labor Expenses / Total Admin. Salaries & Wages	Measures the fringe benefit multiplier which reflects the amount of labor expense that is made up of benefits. Important in explaining administrative expense.

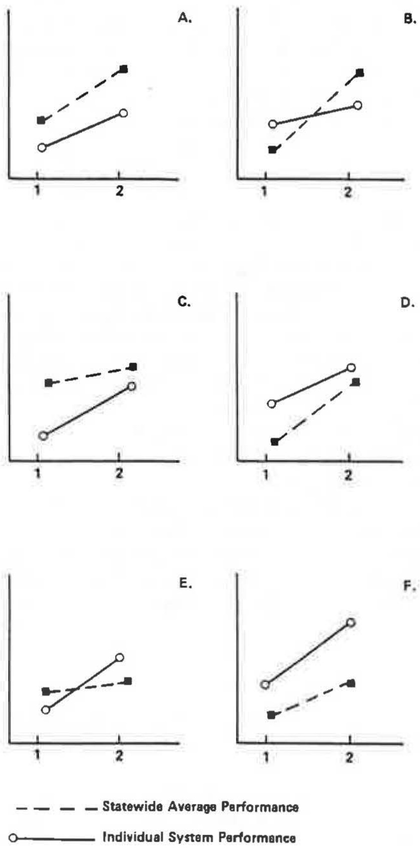
(1) Total vehicle miles includes both revenue and non-revenue vehicle miles.

Figure 6. Indicators with values at least one standard deviation above or below statewide average.

INDICATOR	RANGE	MEAN	ONE STAND. DEVIATION	SYSTEM IDENTIFICATION (1)	(2)	(3)	(4)	(5)	(6)
Total Operating Expense / Total Vehicle Miles	.634–1.484	1.150	± .254	+[1]	–	––		+	
Total Vehicle Operations Labor Expense / Total Vehicle Miles	.284–.831	.558	± .156	+		–		+	
Total Maintenance Labor Expenses / Total Vehicle Miles	.033–.255	.157	± .064	+	–	––	––		
Total Administrative Labor Expenses / Total Vehicle Miles	.022–.112	.062	± .032					+	–
Total Material and Supplies Expense / Total Vehicle Miles	.109–.229	.177	± .047			–	+	+	+
Total Other Expense / Total Vehicle Miles	.040–.347	.166	± .101	–	–		+		

(1) Each plus and each minus represents one standard deviation above or below the median respectively.

Figure 7. Comparisons of transit system performance indicators to statewide average values over time.



ently reduced by focusing on those indicators that fit scenarios A and F. Classifying indicators by using these scenarios does not imply either exemplary or poor performance. Rather, it focuses attention on the indicator for more detailed investigation.

In selecting indicators to review the use of this decision rule, it is particularly important to consider the relationships among indicators. An increase in expense for one indicator may be accompanied by an increase in productivity or a decrease in expense in another indicator. Alternatively, a change in productivity or the rate of wage increases may be accompanied by a negotiated change in fringe benefits. These interrelated factors must be carefully assessed.

In addition to comparing the performance of a transit system with that of other systems, a system can be compared with itself over time. In this type of analysis, the focus is on the magnitude and direction of change for each indicator value to identify indicators that appear to merit detailed evaluation by, for example, (a) comparing the change in expense-related indicators with the consumer price index, (b) comparing the change in fuel price indicators with a nationwide fuel price index, or (c) comparing selected indicators performance with an acceptable limit such as changes in value of more than 20 percent.

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

Planning for Options and Commitments: An Approach to Transport Planning in Uncertainty

HANS L. WESTERMAN

Much transport planning is concerned with proposals of wide-ranging implications that are to be implemented over an extended period. During this time the context and the decisionmakers may change, and the original objectives may no longer be valid. The paper outlines an approach to planning and decisionmaking in such a situation of uncertainty. The approach requires inventing alternative futures for the system as a whole, developing scenarios for proposed intervention and, after evaluation, formulating time-limited commitments and credible options that are worth retaining. The process is incremental and open ended and involves collective learning and selective decisionmaking in which the only firm plans are those that are actually committed. The approach, in a greatly simplified form, is illustrated by a proposal to construct a major road in an inner area of Sydney, Australia. Four alternative futures are invented and examined to determine what strategic options seem worth retaining for the system as a whole. The results may not only show which aspects of the new road require consideration, but also what options are available for its introduction and the kind of commitment that can be made.

Transport planning has become like a game of chess in which it is difficult to plan more than one or two moves ahead. The opponents are many and the game requires great skill because some of the rules are no longer observed. To make matters worse, the board itself is changing.

Uncertainty has always been a critical variable in planning, but it has become more obvious in recent years. Greater public awareness and concern for the environment and the impact of development proposals, energy constraints, technology and its impact on employment and leisure, curtailment of public capital expenditure, changes in population growth and structure, and many other influences have undermined the confidence in longer-term planning.

It has been customary to define objectives and develop proposals that meet these objectives, but uncertainty about the future creates problems in such a closed-system approach. The context giving rise to the objectives may change during the implementation of the proposals. A similar problem exists with forecasting and evaluation models, calibrated on the basis of existing data. Transport planning is particularly vulnerable because it is often concerned with the formulation and implementation of systems within a long time frame, during which both ends and means can change in unpredictable ways.

There is a natural inclination to abandon long-term planning and deal with each case on its merits. This may seem like practical politics but will sell the environment short in the long term. Without longer-term perspectives, options may be foreclosed that should have been kept open and commitments made that need not have been made.

This paper outlines one approach to the problem and illustrates it by applying it to a particular problem: the planning of a major new road in inner Syndey, Australia. In view of the very broad nature of the subject and the difficulty of dealing with it in a short paper, the description of the process and its application in particular can only be sketchy.

PROBLEM DEFINITION

Planning and decisionmaking are part of the same process and the value of planning can be measured by its relevance and usefulness for decisionmakers in a given institutional and political context. In most situations this means a rather short-term planning horizon with greater emphasis on meeting the needs of today than on possible needs at some future time. The value of planning is not limited to today's decisionmaking, but it is also determined by the opportunities it creates for future decisionmakers acting in a different context. This longer-term horizon is particularly important when there are strong and divergent community or political views on matters with a long-term impact such as the introduction of new technology or energy constraints.

Urban planning and transport planning have always operated in both the short- and long-term range. They involve making policy and program decisions, not all at the one time but sequentially and with cumulative effects. The conventional process is to prepare an optimal strategic or structure plan, expose it to the public and decisionmakers, and proceed with the preparation of short-term operational plans and programs after its adoption. The operational plan is subject to cost/benefit analysis, environmental impact assessment, sensitivity tests, and other routines; implementation follows after these hurdles have been successfully overcome.

In the 1950s the transportation planning process was seen as a simple linear sequence. Proposals were developed on the basis of a study of the physical context and future needs, submitted to the decisionmakers who rarely questioned the professional advice, and, once approved, a commitment to both long- and short-term implementation could be assumed.

In the 1960s it was realized that land use and transport interacted and required iterative procedures, but the approach and implementation processes were essentially unchanged.

In the 1970s community concern and the emergence of action groups led to a much closer interest by decisionmakers and others in the formulation of proposals, but there was still the notion that, once the hurdles had been cleared, there was a straight road ahead.

The weakness in this notion lies in the commitment to a single strategic plan. Although the plan has been simplified progressively to that of a diagram or a statement of policies or principles, it nevertheless has tended to become codified in many countries including Australia (1) by legislative or administrative requirements after formal public exhibition procedures. There has been much discussion on the degrees of commitment to such plans (2,3) but the issue is far from resolved. There is usually the qualification that they are to be reviewed regularly, but this rarely happens in practice.

The uncertain nature of the long-term future has two consequences. A commitment to a single strategy and a whole bundle of policies and programs derived from such a commitment will almost certainly create problems. There will be a resistance to change it because of legal and administrative as well as professional commitments to the plan, yet the rapidly changing context will undermine its basic assumptions. When the tension becomes too great, the plan is discarded and ad hoc decisionmaking takes over.

Second, the implementation of a particular proposal over an extended period (as so many transport proposals require) is exposed to high risk as the context, the people, and organizations that make decisions and those that influence them will probably change during this period. Thus uncertainty about the future of the system as a whole and the manner in which it may be controlled create uncertainty about the progressive introduction of a new component of that system, such as a new road, technology, or energy policy.

The problem can then be stated as that of how to formulate and implement a proposal to introduce a new system component over an extended period during which the system as a whole may undergo change not only in its nature but also in the manner in which it is controlled.

REQUIREMENTS

There are a number of requirements for a planning process designed to deal with such a problem (4).

1. There is a need to make forecasts not only of the possible evolution of the new component, but also of the system as a whole.
2. There must be a systematic study of the component and the way it interacts with other parts of the system.
3. The frame of reference must be wide enough to encompass the broad spectrum and long-term implications of the component's introduction.
4. The impact on individuals and groups must be understood and public participation should be built into the process.
5. Values differ among groups and there will be conflict over choice; the process must present options, with their implications, for political decision.
6. Values change over time and the introduction of a new component changes values and behavior; hence, there can be no unalterable choices during its introduction and the process should be ongoing and adaptive.
7. The process must provide the basis for commitments of a strategic and operational kind and must, therefore, be integrated with the decisionmaking process.
8. The demand on resources in using the process should not be excessive.

The process must therefore be anticipatory, systematic, long term, broadly based, participatory, ongoing; present options; allow decisions to be made; and be manageable. The key lies in the relationship between learning and decisionmaking: what needs to be known for what decision, who needs to know and to what extent, how much can be committed with confidence, what options must be kept open.

This is not a once-for-all activity but an ongoing process of exploration, enquiry, reflection, synthesis, consultation, and decision. There is a substantial body of literature on parts of this process: forecasting (5), sensitivity and impact analysis (6), operational research and decisionmaking (7), and many other aspects, but few encompass the entire range of criteria enumerated below. Jantsch (8) attempts to link thinking about the future with action in the present and presents a general framework for long-range exploration and its translation into terms of corporate planning. Etzioni (9) proposes long-term mixed scanning as a means of reducing uncertainty in short-term decisionmaking. Friend and Jessop (10) put forward the concept of strategic choice as a means of making decisions in uncertainty. These contributions are valuable, but the problem remains of how to develop an operational process meeting all the criteria.

OPTIONS-COMMITMENTS APPROACH

The options-commitments approach originated from a study that examined the progressive introduction of a line-haul transit system in Canberra, Australia (11). It was part of an international project on the social assessment of new transport technology. The approach has been under further study since (12) and, although much more requires to be done, the basic structure is simple.

There are four phases in the process: long-term, short-term, and intermediate-term assessments and a repetition of these assessments.

The long-term assessment is essentially a learning phase in which options are generated, awareness of possible implications is created, and a broad indication of preferred direction and bundles of options emerges.

The short-term assessment looks at the preferred options in more detail and in a shorter time frame, analyzes the implications more precisely, indicates what decisions can be made now, establishes the degree of support for them, and identifies how long the decision is likely to remain valid before the next decision has to be made. The short-term assessment is conventional and generally follows well-established procedures and is not elaborated here.

The intermediate assessment is concerned with preparing the ground for the next decision. It assumes that there is a desirable course to pursue but that it does not occur by chance. At the time when the next decision has to be made, the decisionmaker will be influenced not only by the performance of the previous action but also by the attitude of the community toward it. It involves the monitoring of performance and attitudes, the acquisition of new data, recalibration of models, seeking legislative changes, improving institutional arrangements, creating more effective interaction with certain groups (i.e., clients and unions). Although the intermediate assessment is often overlooked, it is no innovation and does not require further comment. The three phases are repeated when the next decision must be made.

The principal difference with currently used processes lies in the long-term assessment and the integration of long-term, short-term, and intermediate-term assessments. A guidance process is established that searches out directions without firm destinations but with an identification of the first likely port of call (Figure 1). Repetition of the process at future points of decision produces a course adjusted to the perceptions at the time (Figure 2).

The long-term framework consists of systemwide forecasting, preparing scenarios for the progressive introduction of a new component of the system, evaluating the scenarios and forecasts together, and delineating possible strategic options and commitments that form the basis for formulating operational decisions during the short-term assessment phase (Figure 3).

LONG-TERM ASSESSMENT

Systemwide Forecasting

The first activity is that of systemwide forecasting. A horizon year must be assumed that is far enough into the future to encompass likely impacts of the new component. In urban planning, a period of 25 years may be appropriate. This is followed by an important step in the process: the inventing of alternative futures. It does not start with alternative proposals (which has often been the practice to date), but with alternative contexts and controls at the horizon year. Assumptions are made about the socioeconomic system, the values, and institutional influences acting on the physical environment (Figure 4).

The conceptualization of quite different futures is an exercise in lateral thinking and a creative act of "imagineering". It requires developing a holistic view of a number of quite different futures, each with a characteristic dominant theme. For instance, one could postulate, as Robertson does (13), two contrasting views of post-industrial society: the hyperexpansive (HE) view with high technology, computing and telecommunications setting the pace, or the sane, humane, and ecological (SHE) view where personal and humane development is the dominant consideration. The Gamma Report (14) considers five futures: doing more with more, doing more with less, doing the same with less, doing less with less, or doing less with more. There are many ways in which alternative futures can be conceived, but the principal criteria are that there is diversity, the assumptions are made explicit, and the futures are comprehensive.

The futures are transformed into physical forms or structures and policies to enable some degree of quantification to be carried out. This conversion requires a good knowledge of the strengths and weaknesses, and the opportunities and constraints of the system, and an understanding of the processes of urban change. For instance, the scope for major changes in a city's form and structure, even in the longer term, may not be as great as is often assumed, but significant changes within an existing morphology can occur, especially in population and employment structure. It is possible to evaluate alternative structures in terms of equity, accessibility, economy, funding, pliability, and probability.

It is also possible, and indeed desirable, to involve groups with a particular view about the future at this stage in order to understand the willingness to trade-off conflicting objectives and outcomes. None of this activity is intended to lead to decisions; it is an attempt at discovering boundary conditions of the future and understanding the impact of possible fundamental rather than incremental changes of the system as a whole.

Developing Scenarios

The second activity of the long-term assessment concerns the preparation of scenarios for the progressive introduction of the new component into the system. There may be a wide choice, ranging from introducing it in one operation to doing nothing, with many forms of incremental development in between.

The development of alternative scenarios should, at this stage of the process, not be a matter of developing a single decision tree and selecting a few paths to it for closer analysis. While it is

Figure 1. Options and commitments.

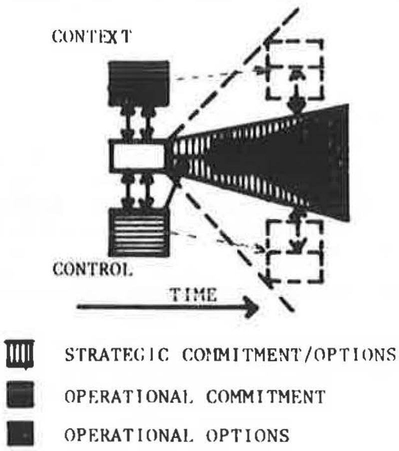
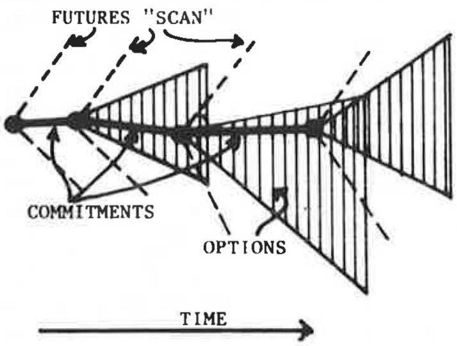


Figure 2. Ongoing nature of the process.



fortunately not a forest, there may be a number of trees because there are quite different criteria and understanding is increased by looking at them individually, at least initially. For instance, the introduction of a new line-haul transport system may involve choices of routes and their progressive development, changing levels of service and rates of technological change in addition to matters such as programming and budgeting (Figure 5). Another significant consideration is the likely policy and community response as means and ends in planning are often difficult to separate.

It is possible to develop scenarios for each of these criteria and, in turn, holding the others constant. For instance, one can develop a systems concept specifying the principal elements of the operation requirements without specifying the precise technology or policy (15). Another approach is to specify policy levels of service and to develop a floating corridor concept (16). A third way is to vary the levels of service and make assumptions about degree of exclusivity of right-of-way, technology class, and operational strategy (17). However, if there is limited time or resources for such a procedure, it is possible to develop the scenarios as different combinations of such variables.

There is scope for systematic approaches as well as for creative short cuts but in all approaches a thorough knowledge of the component is required, a range of alternatives should be explored, and a preliminary appreciation of possible impacts should be obtained. Again, there is no need for any decisions; the purpose is to test incremental changes under different conditions. Selective public participation in some cases is feasible (18).

Figure 3. Long-term assessment.

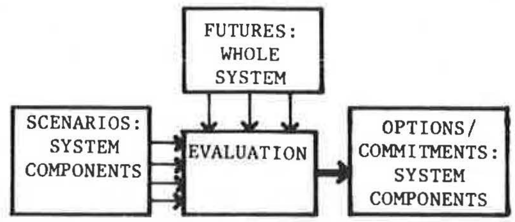
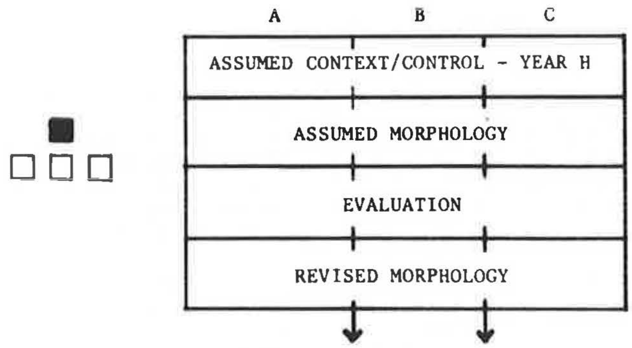


Figure 4. Alternative futures.



Evaluation

It is now necessary to relate the alternative scenarios for a new component of the system to the alternative futures of the system as a whole. This is the core of the long-term assessment. The aim is to reduce uncertainty and complexity and to obtain a picture of what strategic direction may be worth pursuing and which options should be kept open.

Before considering evaluation procedures, it is necessary to comment on the methodological problem of relating time series (scenarios) to fixed states (alternative futures at horizon year). It is possible to take the end of the scenarios and thus have a common horizon year for comparison. It is also possible and preferable to regard the fixed states as the outcome of dynamic action and relate the scenarios to this dynamic context. In essence, one is relating a number of different things to each other: the incremental introduction of a new component against the possible fundamental change of the system as a whole and a view from the present "up" to the future (i.e., the scenario) against a view from possible futures "down" to the present.

The systematic evaluation of possible outcomes and impacts can be exceedingly complex in both conceptual and computational terms. Computer interactive approaches may hold promise in the longer term, but as the evaluation is carried out for the purpose of learning, intuitive procedures based on an assessment of probability and credibility may be sufficient. Obvious inconsistencies will appear, undesirable or improbable associations can be detected, conflicting values can be revealed, and information needs for future decisions can be identified (Figure 6).

The result can be summarized in a simple status report and exposed to the community and decision-makers. It identifies the assumptions made; the issues, possible impacts, options that would appear to be worthwhile to keep open; and the general direction to pursue. It is beyond the scope of this paper to expand on the normative or functional as-

Figure 5. Alternative scenarios.

INDICATIVE SCENARIO ASSUMPTIONS		SHORT TERM	→	LONG TERM
Level of Service	1	D	C	B
	2	C	B	A
Level of Funding	1		Constant rate	
	2	L	H	M
Technology	1	L	L	M
	2	M	H	H
Corridor	1	Ia	IIa	IIIa
	2	Ib		IIb

Figure 6. Evaluation.

SCENARIOS		FUTURES A	FUTURES B	FUTURES C	
1	1	O	O	X	→ STATUS REPORT
	2	O	•	•	
2	1	X	•	O	→
	2	•	O	•	
3	1	O	O	O	→
	2	•	X	O	
4	1	X	•	O	→
	2	O	X	•	

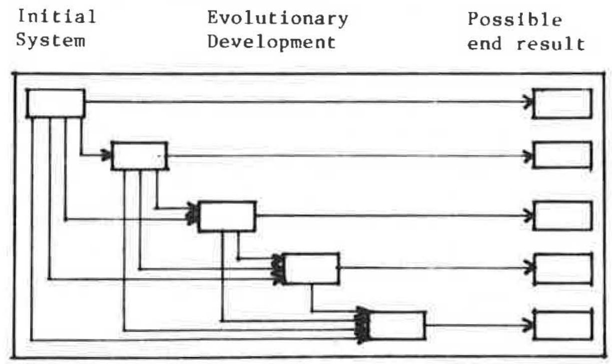
X NOT PROBABLE
O PROBABLE
• CREDIBLE

pects of this procedure (19). However, it may suffice to point out that the credible direction should be determined by the likely policy response of decisionmakers and influential groups, and responses to this status report should therefore be sought.

Options and Commitments

The final activity of the long-term assessment is the structuring of those options that are worth retaining in some form of evolutionary framework (Figure 7). It involves making a strategic commitment on the direction to pursue or, as Jantsch puts it, "determining the future boundary conditions" (20), as it is unlikely that all worthwhile options can be retained.

Figure 7. Framework of options.



For instance, in the case of a new line-haul transit service there may have to be a commitment to a corridor, but there can be options in implementation starting from an express bus in mixed traffic to an automated light rail vehicle on its own right-of-way.

The structuring of the options can be based on single or composite criteria and following analytical (21) or intuitive procedures. However, as the time-limited, operational decision is made during the short-term assessment, great accuracy is not called for and a simple procedure may well suffice.

To conclude, the long-term assessment provides an understanding of the implications of making a strategic commitment to the component to be introduced and particularly of its robustness in an uncertain situation. However, there is no operational commitment. Such a commitment involving a decision on the first stage of implemenation depends on the outcome of the short-term assessment.

SCOPE AND LIMITATIONS

The options-commitments approach can be described as a process that allows long-term possibilities to be taken into account in short-term decisionmaking despite uncertainty about the future. It meets the requirements set out earlier in that

1. There is an open-ended and integrated approach to planning and decisionmaking;
2. There is an emphasis on thinking holistically about the city, not within a single view, but within a range of possible futures;
3. Collective learning in which professionals, politicians, and the community can participate is an integral part of the process;
4. There are no unrealistic and unnecessary commitments but a careful combination of commitments that are achievable and options that are worth retaining; and
5. There is a differentiation between strategic and operational commitments and options.

One of the central questions is the relationship between information collection and analysis and decisionmaking. In view of the very broad nature of the approach, there is a risk of losing the trees for the forest of possible options. The process requires a careful assessment of how much is needed to know for what level of decisionmaking, and it may be more productive to make quick assumptions (so long as they are made explicit) than to undertake time-consuming studies that do not remove the basic uncertainty underlying the future development of the system or its component. A quick assessment may be more relevant for a decisionmaker than a thorough assessment that comes too late to be useful.

There are several procedures to simplify the process without diminishing its essential character (Figure 8). Diagram 1 sets out the process as described. In diagram 2, the alternative futures are evaluated in terms of commonalities so that a bundle of options and commitments for the system as a whole is selected against which alternative scenarios for a specific proposal or new component can be tested. This simplification is used in the illustration that follows and can also be useful where the management of change is centralized (e.g., development corporations). In diagram 3, the alternative scenarios are contracted to one or two basic variants. It was used in the Canberra study of a proposed line-haul system where the route was predetermined but technology and funding levels were the uncertain elements (11). Diagram 4 represents a simplification in both alternative futures and scenarios.

APPLICATION: SYDNEY'S TRANSPORT SYSTEM--A CASE STUDY

My interest in applying the options-commitments approach to Sydney's transport system was aroused by the decision of the state government in 1979 to conduct an inquiry into the location and construction of a major arterial road through the inner suburbs of Sydney. The road was seen to fill a need following the development of a new port and container terminal. The options presented were not all compatible, the case made in support of the road was based on traffic growth, using models not calibrated for changes in energy costs, and the evaluation relied on cost-benefit analysis with its attendant problems of quantifying benefits.

A wide range of differing objectives was put to the inquiry in public submissions. The inquiry typified the complexities of decisionmaking in today's climate: uncertainty about the future, absence of long-term policy guidelines, conflict between long-term and short-term interests, and between metropolitan and local interests, distrust of professional attitudes, rationality clouded by emotion, electoral prospects, union attitudes, differences between public authorities, limitations in public funds, and unequal impact in different groups.

Sydney's growth since 1945 has been rapid but has slowed down in recent years. The city has a population of about 3 million (1980), spread out over an area as large as greater London. It has a reasonably good radial railway system, an extensive system of buses and ferries, but an arterial road system that has not been upgraded sufficiently to keep up with the city's expansion. There has been some freeway construction, confined mostly to the fringes, and there have been longstanding proposals to improve traffic conditions in the intermediate and inner suburbs by the construction of new arterial roads and freeways. These plans were thrown into disarray in 1977 when the government decided to abandon large sections of the proposed inner suburban freeways (Figure 9).

The decision was not unexpected as the freeways would seriously affect the environment and funds were simply not available nor likely to become available in the near future. The context had changed to a point where the concept could no longer be supported. However, the removal of a significant part of a proposed system may invalidate that which remains and makes it all the more difficult to judge whether an ad hoc proposal such as the new road to the port made sense in the long run.

Methodology

The process outlined in the first part of the paper was adapted and simplified because of the need to make a quick assessment, use readily available information, and relate it to the inquiry in

Figure 8. Alternative procedures.

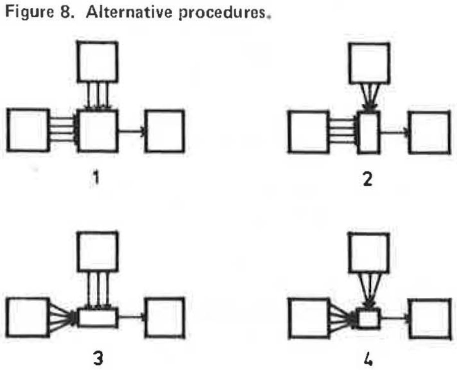
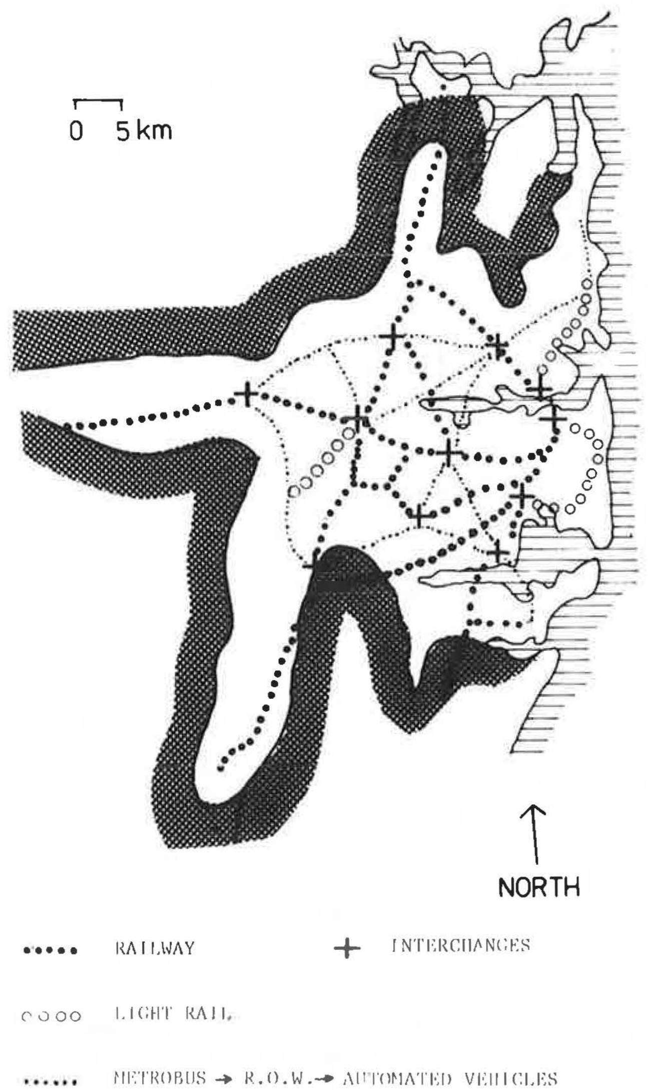


Figure 9. Strategic public transit options in Sydney, 1980.



progress. Procedure 2 (Figure 8) was followed for the long-term assessment; the short- and intermediate-term assessments were not carried out because they required a policy decision first.

A study was made of the historical relationship between context and strategic planning concepts, and this was followed by an examination of current problems (e.g., imbalances in employment and work force, accessibility, equity, and incidence of congestion), existing commitments, opportunities, and constraints.

The long-term assessment commenced with the formulation of alternative futures and a search for commonality that produced a framework of strategic options and possible commitments for the city as a whole. The options for the location and performance of the road as presented to the inquiry were accepted as alternative scenarios. These were then related to the strategic options and commitments for the system as a whole. Some conclusions were drawn from this evaluation used for constructing an options-commitments diagram for the introduction of the new road.

Application

The four different futures were a public-investment-sensitive future, an energy-sensitive future, a pollution-sensitive future, and a future based on accessibility at a price.

In the public-investment-sensitive future, the overriding assumptions are those of doing more with what exists, making small rather than large-scale commitments, and maximizing the utility of any future transport extensions or improvements by supporting land use or other policies. In such a future there would be less emphasis on freeways and expressways and more on selective elimination of congested areas, further extension of clearways, priority lanes for buses, area traffic management, and downgrading of traffic-generating land uses fronting arterial roads.

In the energy-sensitive future the overriding concern can be much higher cost of fuel and/or lack of supply. Lack of supply will mean rationing that would affect every motorist; higher costs will mean that those able to afford it will continue to use their car and those who cannot must have an alternative. Accessibility to public transport and trip length become critical variables. Relocation of jobs and homes, with employers seeking employees closer to where they live and employees moving closer to work, would be some of the consequences.

In the pollution-sensitive future the overriding concern is that of improving the environment by reducing air pollution by cars and industry, eliminating noise and waste, increasing safety, enhancing the environment through higher standards of building, landscaping and urban design, and preserving historic buildings and environments. Stringent controls on vehicle emissions, reducing stop-start driving and traffic densities in critical areas, affecting a shift to public transport, duo-mode or electric buses, electric commercial vehicles, and applying clean air standards to industry would be some of the implications.

In the accessibility-at-a-price future it is assumed that the user pays directly for the price of improved accessibility by private vehicle. It would involve a simple network of freeways, accessibility only by payment with the charge dependent on the distance traveled and the weight of the vehicle. Credit cards could be used, with the level of charges set to recover capital and operating costs. The public transport system would be expected to meet its operating costs, but historic costs would be written off and there would be grants for capital improvements. Special funds would also be available for the development of an intrasuburban metrobus system, suburban paratransport services, and public transport interchanges (with cost recovery through the sale of development rights).

The alternative futures are expressed in diagrams showing their morphology and related policies and are described in more detail elsewhere (12). A comparison of the alternatives shows that there are commonalities. Land use policies directed toward subcentralization and transport policies to improve conditions in congested areas are common to all alternative futures. They are robust, current policies, and continuing commitment can be justified with little risk.

There are others that emerge as recurring themes: the relocation of industrial and related storage functions from inner areas and their replacement by higher-density housing, the concentration of employment into nodes that can be served by public transport, the development of an express intermodal public transport system with proper interchanges, the rationalization of goods movement with greater use of the railway network, and the improvement of public transport with further land use intensification in the areas served to maximize the benefits of such improvements.

All futures provide for high-standard regional connections, at least one major intermediate link from north to south and one intermediate-ring route. These options would appear important to retain. In the accessibility-at-a-price future, additional corridors are envisaged and these may be desirable to retain as longer-term options, which can be reviewed at the next round. There are also a number of arterial roads that may be considered as candidates for progressive upgrading to expressways. Long-term land use policies designed to reduce the impact of frontage access would make their upgrading a more realistic option in the long term (Figure 10).

All futures also envisage, inter alia, significant improvements in the intrasuburban public transport system including the development of interchanges. The evolution of such a system can take different forms and at this stage it would seem that the longer-term option of moving toward a separate right-of-way with a capability of automated vehicles should be kept open (Figure 11).

In terms of commitments, there are some links that seem to be robust, irrespective of the kind of future that may occur. These include the new road as part of a circumferential route, roughly in the position proposed to the inquiry. Its long-term status is that of a high-capacity expressway serving a significant metropolitan function.

THE NEW ROAD--SCENARIOS AND EVALUATION

Four options of the new road were put to the inquiry. The first and second option envisaged a new arterial road with small differences between them in location. The first minimized the disruption of existing residences while the second minimized the effect on open space. The third option involved a partial upgrading of existing roads. The fourth option proposed the construction of a freeway running in a different direction.

The options were not strictly comparable and internally consistent. None considered incremental implementation in the form of scenarios. There was an assessment of environmental impact and costs and benefits of the alternative routes, but there was no consideration of costs and benefits of alternative levels of service.

Evaluation of the options for the road against

Figure 10. Strategic road options.

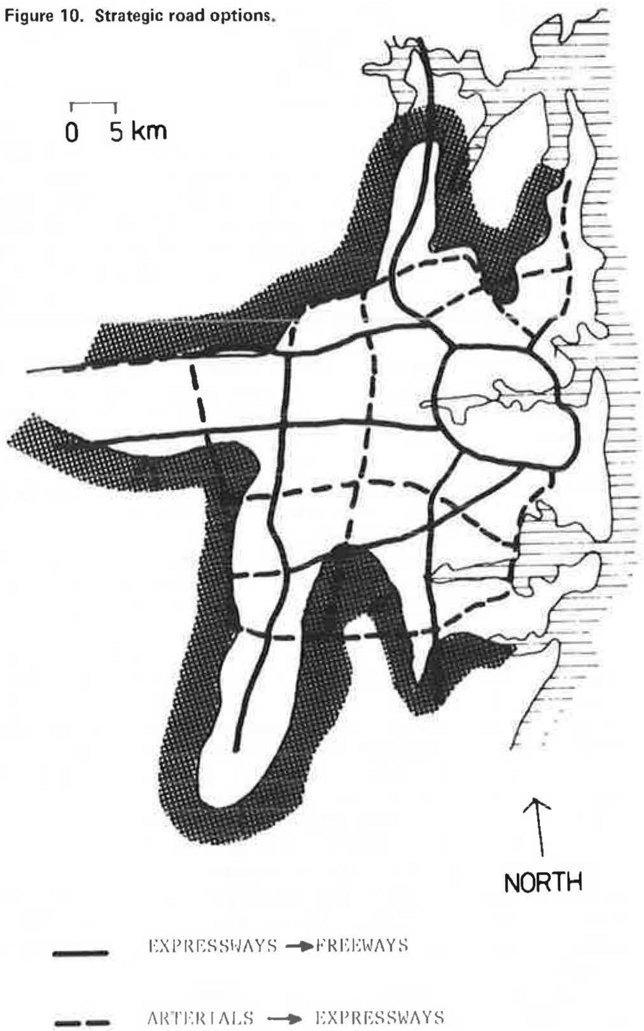
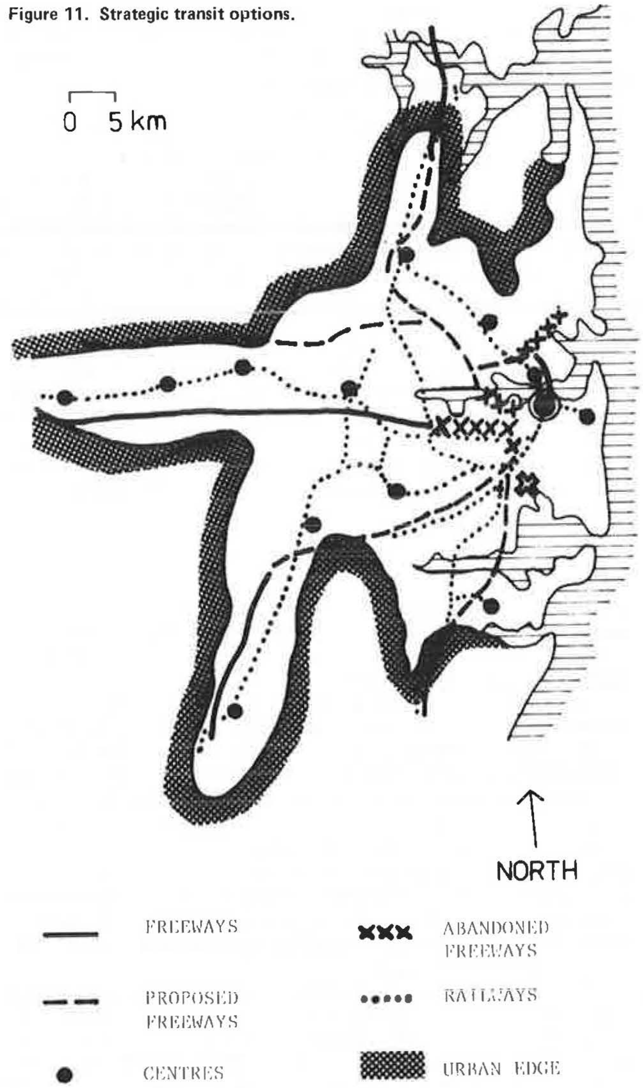


Figure 11. Strategic transit options.



the strategic options and possible commitments for the system as a whole showed that the assumed level of service of the road--an arterial road with many intersections at grade--was ill-considered and that the long-term option should have been that of an expressway with potential grade separation.

At this point, the proper course would have been to prepare new scenarios for the road, assume different levels of service, and reassess its performance. When the credibility and robustness of the road and its general performance characteristics in a metropolitan long-term context had been established and political support had been received in principle, studies of operational options and commitments could have commenced. They would have taken account of the level of resources likely to be available, direct and indirect costs, and the need for parallel programs. Such programs could ensure that loss of open space is compensated for, people affected are rehoused in the locality, traffic management schemes are introduced to reduce the incidence of through traffic, and the new road acts as a catalyst to a general upgrading of the area affected by it. An options-commitments diagram can then be prepared that shows different ways in which the road (and associated programs) can be phased in (Figure 12).

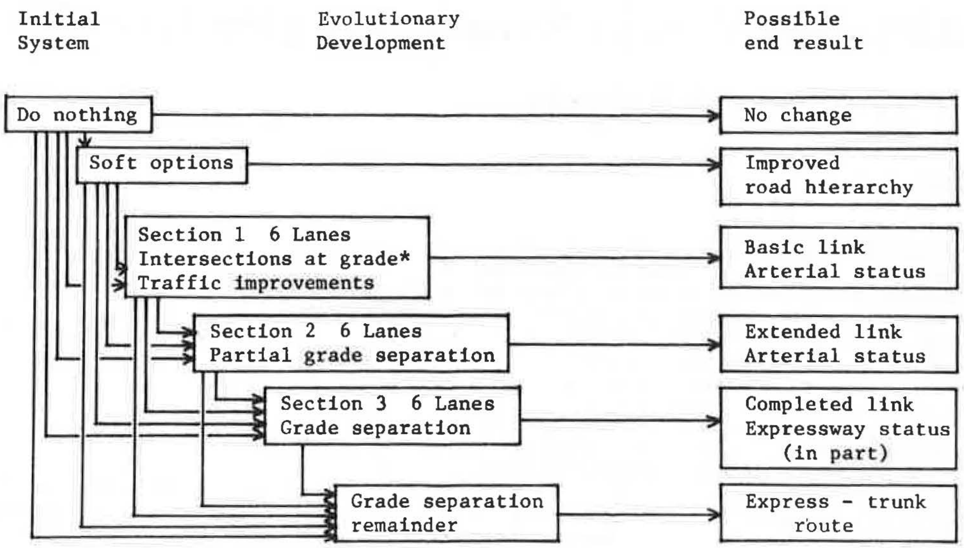
The terms of reference of the inquiry, however, did not foresee the need to distinguish strategic from operational decisions. Had this been the case, abortive work could have been avoided and metropolitan issues could have been resolved before local issues were dealt with. A two-stage approach to the inquiry would have provided the feedback necessary for making strategic commitments.

Since this paper was written, the report by the Commission of Inquiry has been tabled in the New South Wales Parliament. It recommends that the road not be built, that any land already acquired for the road be declared as open space, and that the containers from the new port be moved substantially by rail. The recommendations are understandable in the current context of limited public funds and/or improving the local environment and will be welcomed by local residents and politicians. However, they appear to ignore the longer-term metropolitan needs. The declaration of land acquired as open space removes, for all time, the option of progressively introducing the new road. There is obviously a need for more emphasis on collective learning so that the need to keep options open for future generations is understood and widely accepted.

CONCLUSION

There has been a rapid expansion of specialized knowledge and techniques in recent years. This is valuable as planning and decisionmaking must proceed on a solid base, but there is a danger of fragmenta-

Figure 12. Options for development.



tion if such expansion is not matched by growth in understanding of the totality of the environment and by the development of holistic approaches to the management of physical change.

This will be all the more important and also the more difficult in the 1980s, which will be characterized by uncertainty. The natural reaction in situations of uncertainty is to deal with problems and issues on an ad hoc basis, but this does not constitute a holistic response to the management of change. There is a need, therefore, to develop processes that allow decisions to be made in a broad context and within both a short-term and long-term perspective.

The approach outlined in this paper is one such process. It does not produce immutable plans (with the inevitable psychological commitment by those who prepared them with care and conviction), but options and limited, time-based commitments. The process may be more accurately described as a general approach than a rigorous procedure as it need not be elaborate and can be adapted to suit individual situations. It is an incremental process of collective learning and selective decisionmaking in which plans, policies, and programs can respond to changes in context and control.

ACKNOWLEDGMENT

I wish to acknowledge the contribution of my former colleagues, C.A. O'Flaherty and G.J. Campbell of the National Capital Development Commission and, particularly the consultant to the team, P. Fisher, to the development of the original approach.

REFERENCES

1. Environmental Planning and Assessment Act 1979. N.S.W., Australia.
2. P.H. Levin. Government and the Planning Process. George Allen and Unwin, London, 1976, pp. 297-305.
3. Drake, McLoughlin, and Thompson. Aspects of Structure Planning. Centre for Environmental Studies Research Report 20, Sept. 1975.
4. Social Assessment of Technology in the Field of New Urban Transportation Systems. O.E.C.D., Paris, 1978, p. 2.
5. W. Kennet. The Futures of Europe. Cambridge University Press, Cambridge, England, 1976, pp. 19-44.
6. B.H. Massam. Spatial Search--Applications to Planning Problems in the Public Sector. Pergamon Press, Oxford, England, 1980, p. 278.
7. P.G. Moore and H. Thomas. The Anatomy of Decisions. Penguin Books, London, 1978.
8. E. Jantsch. Technological Planning and Social Futures. Cassell, London, 1972.
9. A. Etzioni. Mixed Scanning: A "Third" Approach to Decisionmaking. In A Reader in Planning Theory (Faludi, ed.), Pergamon Press, 1973.
10. J. Friend and N. Jessop. Local Government and Strategic Choice. 2nd ed., Pergamon Press, 1977.
11. H.L. Westerman. Technology Assessment, A Case Study by the National Capital Development Commission. In Proc., Workshop in Technology Assessment, Australian Government Publishing Service, Canberra, 1978.
12. H.L. Westerman. Planning and Decisionmaking in Situations of Uncertainty: An Approach to Sydney's Transport System. Transactions, Institution of Engineers, Canberra, Australia, Vol. CE 23/1, 1981.
13. J. Robertson. The Sane Alternative. Villiers Publications, London, 1978.
14. Gamma Report on the Conserver Society. Univ. of Montreal and McGill Univ., Montreal, Quebec, Canada, 1977.
15. R.M. Soberman and G. Clark. Operational Requirements for New Transit Technology: Transport Decisions in an Age of Uncertainty. Proc., World Conference on Transport Research, Martinis Nijhoff, The Hague, 1978.
16. M. Scheibe and G.W. Schultz. Design-Synthesis Approach to Transit Planning. TRB, Transportation Research Record 639, 1978, pp. 1-7.
17. S. Mitric. Comparing Modes in Urban Transportation. TRB, Transportation Research Record 639, 1978, pp. 19-24.
18. G. Krampe and U. Weiler. Cooperation with Citizens in Operational Planning of New Transportation Systems. Proc., International Symposium of Transport and Transportation Technologies, Bundesministerium fur Forschung und Technologie, Bonn, Vol. A. 11, 1979, pp. 144-162.
19. A. Faludi. Planning Theory. Pergamon Press, Oxford, 1976, pp. 171-186.
20. E. Jantsch. Technological Planning and Social Futures. Cassell, 1972, p. 54.
21. P.G. Moore and H. Thomas. The Anatomy of Decisions. Penguin Books, London, 1978, pp. 39-73.

Publication of this paper sponsored by Committee on Transportation Systems Design.

Estimating Transit Supply Requirements for Alternatives Analysis

ROBERT E. SKINNER

The estimation of transit supply parameters has received relatively little attention in the technical literature. Yet these supply parameters, such as vehicle miles, vehicle hours, and employees by category, are the major determinants of operating cost estimates that, in turn, are a principal factor in evaluating major transit alternatives and selecting a single alternative for implementation. Furthermore, it is during the estimation of transit supply parameters that basic questions of operating feasibility are addressed, implicitly or explicitly. Through the use of a simple example for a single rail line, the complexity of supply-parameter estimation is illustrated in terms of the number of assumptions and inputs required. The example also illustrates the sensitivity of the supply-parameter estimates to variations in input assumptions. A framework is presented for the process of supply-parameter estimation in context with the overall transit planning process, particularly with respect to federally required alternatives analysis studies. Some general considerations in the use of this process that are discussed include the need for supply-demand equilibrium, the iterative nature of the process, the differences between phase 1 and phase 2 alternatives analysis studies, operating feasiblity, and the relation between predictive techniques used in supply-parameter estimation and "reasonability" checks. Finally, selected examples of specific inputs to the process are discussed and recommendations are made with respect to how these inputs should be developed. The examples include vehicle capacities and loading standards, service hours and service days, background bus networks, special services, and demand estimates.

Supply-parameter estimation has received relatively little attention in the technical literature. Certainly, in comparison with demand estimation, there is a paucity of formalized procedures and technical guidelines for the estimation of supply parameters. In transit planning studies, supply-related assumptions and procedures are sometimes made or selected in ways that are inconsistent between alternatives (e.g., passenger capacity of buses versus rail cars) or that do not fully take into account their impact on aggregated supply parameters and operating costs. Distortions resulting from such practices are particularly serious if they occur in federally required alternatives analysis studies whose intent is to determine the worthiness of major capital investments in new transit facilities.

In the context of this paper, transit supply parameters are those measures that describe and quantify either (a) the amount of transit service a given alternative will provide or (b) the nonmonetary resources needed to maintain this level of transit service, such as vehicles or employees. The following key supply parameters are considered in this paper:

1. Peak-period daily and annual vehicle (train) hours by vehicle type,
2. Peak-period daily and annual vehicle (train) miles by vehicle type,
3. Fleet size by vehicle type,
4. Employees by category,
5. Peak-period daily and annual place hours by transit mode,
6. Peak-period daily and annual place miles by transit mode, and
7. Average system operating speed by transit mode.

Supply-parameter estimation is an important component of alternatives analysis studies because supply parameters are key inputs to operating-cost estimation and to the determination of operating feasibility. The topic of transit supply-parameter estimation is far too broad to be fully addressed in this paper. Instead, the paper presents selected topics drawn from alternatives analysis research undertaken for the U.S. Department of Transportation (1) that provide some overview discussion of supply-parameter estimation as well as specific guidance for planners on some key issues.

EXAMPLE

To illustrate the sensitivity of supply-parameter estimates to input assumptions and procedures, an example is helpful. The assumptions and procedures used to estimate three key supply parameters for a hypothetical rail line--annual train hours, annual car hours, and annual car miles--can be summarized as follows. The alternatives used are defined as

1. A 10-mile rail line with seven stations;
2. Average station spacing of 1.67 miles;
3. The following vehicle characteristics: (a) 630 gross ft^2/car, (b) maximum train consist of three cars, (c) loading standard of 5.4 ft^2/passenger, (d) cruise speed of 60 mph, and (e) service acceleration and deceleration of 3.0 mph/s;
4. Policy headways of 5 min for the 4 peak hours and 10 min for the 12 off-peak hours;
5. Weekend and holiday service resulting in an annualization factor of 310;
6. Station dwell time of 40 s; and
7. Minimum headway of 2.5 min.

The following demand levels are used:

1. Maximum-load-point volume--32 000 riders (two ways),
2. Peak-hour, maximum-load-point volume in the peak direction--4800 riders (15 percent of total ridership), and
3. Off-peak, maximum-peak-load-point volume in the peak direction--1200 riders (3.75 percent of total daily ridership).

The procedures used to estimate supply parameters are as follows:

1. Average speed estimated by using a formula applicable for sufficient station spacing to reach cruise speed (2):

$$S = D/[T + D/C + C(1/2a + 1/2d)] \tag{1}$$

where

S = average transit vehicle speed,
D = average distance between stations,
T = stop time at stations or stops,
C = cruising speed,
a = rate of acceleration, and
d = rate of deceleration;

2. Train hours minimized (headways maximized) subject to policy headway and demand constraints;
3. Adequate capacity provided in 12 off-peak hours to meet peak-load-point, peak-direction demand (1200 riders/h); and

4. Adequate capacity provided in 4 peak hours to meet peak-load-point, peak-direction demand (4800 riders/h).

The baseline estimates for these parameters and their sensitivity to various changes in assumptions or procedures are given in Table 1.

Because additional passenger capacity can only be added in discrete units, excess capacity is usually provided in order to meet a prespecified demand level. The amount of excess capacity provided influences the degree of change in supply parameters that may occur in response to a change in an input assumption. Consequently, Table 1 also indicates the amount of peak and off-peak excess capacity provided by each sensitivity test.

Based on this example, several observations can be made about the estimation of supply parameters:

1. The sheer number of input assumptions alone illustrates the complexity of supply-parameter estimation. The example is really not so "simple" after all. Furthermore, the need for decision rules, such as minimizing train hours, suggests that the procedures are not as straightforward as many have supposed.
2. Because of the "lumpiness" in the supply curve, a minor change in a particular input parameter (such as cruise speed) may sometimes be just enough to reduce equipment requirements. At other times, however, a more substantial change in an input parameter may have a negligible impact. The net change depends in part on site-specific circumstances (thus, it is not possible to generalize based on the results of this example).
3. None of the sensitivity tests revealed dramatic changes in supply parameters resulting from a single change in an input assumption over the ranges tested. They do illustrate, however, that seemingly insignificant assumptions can constrain or partly predetermine results--for example, peak-hour peaking factors, station dwell times, annualization factors, and off-peak policy headways.
4. The impact on supply parameters of multiple changes in input assumptions can be additive so that, instead of 5-10 percent changes, 30-40 percent changes in supply parameters might result. For example, the combination of tests 6 and 10 decreases train hours by 35 percent. Two different sets or packages of assumptions could yield very significant differences in supply-parameter estimates.
5. Train hours, probably the most significant parameter from the standpoint of operating cost, tended to be more sensitive to changes in input assumptions than either car miles or car hours.
6. Changing the objective function from minimizing train hours to minimizing car miles can result in significant changes in supply parameters.

GENERAL CONSIDERATIONS

The following section of this paper first describes, in conceptual terms, the process of supply-parameter estimation and how it relates to the overall planning process for alternatives analysis studies. Next, it discusses critical general considerations that relate to the estimation process as a whole.

Process of Supply-Parameter Estimation

Figure 1 shows schematically the process of supply-parameter estimation and how it relates to other steps in the overall alternatives analysis process. Key inputs to supply-parameter estimation are the definition of alternatives (step 1) and ridership forecasts (step 2). For the most part, supply-parameter estimation takes place in steps 3, 4, and 5. The resulting supply-parameter estimates serve as input to the estimation of transit operating cost (step 6) and, to a lesser extent, the estimation of capital costs.

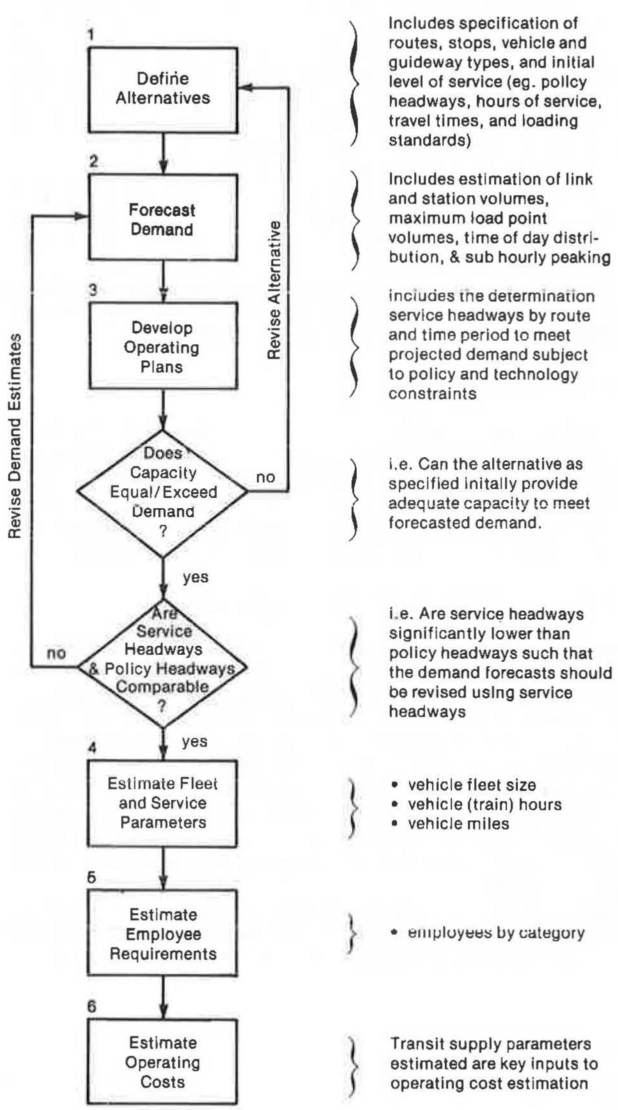
Definition of alternatives includes an initial specification of the level of service (LOS) to be provided by each alternative. Based in part on this

Table 1. Sensitivity of supply parameters to changes in analysis assumptions and procedures.

		Cars Required[a]		Annual Supply Parameter[b]: Train Hours		Car Hours		Car Miles		Excess Capacity (no. of passenger places)	
Test No.	Sensitivity Test	No.	Change[c] (%)	No.	Change[c] (%)	No.	Change[c] (%)	No.	Change[c] (%)	Peak	Off-Peak
1	Baseline estimates	24		24 784		59 483		2 232 000		465	555
2	Decrease loading standard by 10 percent to 4.9 ft^2/passenger	21	−12.5	23 537	−5	55 741	−6	2 090 640	−6	309	555
3	Increase cruise speed by 10 percent to 66 mph (increases average speed by 4 percent)	21	−12.5	23 498	−5	55 645	−6	2 184 260	−2	9	632
4	Increase maximum train consist to four cars	24	0	24 784	0	59 483	0	2 232 000	0	465	555
5	Decrease annualization factor by 10 percent to 279	24	0	22 306	−10	53 535	−10	2 008 800	−10	465	555
6	Remove one station and increase average station spacing to 2 miles, or +20 percent (increases average speed by 18 percent)	21	−12.5	19 809	−20	48 267	−18	2 145 510	−4	641	361
7	Decrease station dwell time to 30 s, or −20 percent (increases average speed by 7 percent)	21	−12.5	19 809	−20	48 298	−18	1 935 950	−13	114	206
8	Decrease maximum train consist to two cars	22	−8	28 463	+15	56 926	−4	2 137 760	−4	25	555
9	Decrease peak-hour, peak-load-point, one-way volume from 15 to 14 percent of daily, maximum-load-point, two-way volume (from 4800 to 4400 passengers/h)	21	−12.5	23 537	−5	55 741	−6	2 090 640	−6	118	555
10	Increase off-peak policy headways from 10 to 15 min	21	−12.5	21 043	−15	52 032	−13	1 952 380	−13	465	116
11	Minimize car miles rather than train hours subject to demand and policy constraints	22	−8	35 899	+45	49 491	−17	1 858 760	−17	25	116
12	Tests 6 and 10	21	−12.5	16 089	−35	48 267	−19	2 144 270	−4	641	358
13	Tests 6, 9, and 10	18	−25	14 849	−40	44 547	−25	1 980 590	−11	200	360

[a]Excludes spares. [b]Revenue service only. [c]Compared with baseline estimate.

Figure 1. Estimation of key transit supply requirements.



LOS, ridership estimates are prepared. Once the ridership estimates are available, the first step in supply-parameter estimation occurs: the development of operating plans. Operating plans are developed for each individual transit service and route included in the definition of a given alternative, and they specify the amount and temporal distribution of transit capacity to be provided. The operating plans should attempt to match supply with demand in ways that tend to minimize operating costs, subject to policy constraints (e.g., maximum headways) and technological constraints (e.g., minimum headways).

Operating plans must be checked to determine whether the maximum capacity that can be provided equals or exceeds the estimated demand. If not, the definition of the alternatives must be revised. For example, a rail alternative might require longer station platforms so that longer trains can be accommodated. In addition, the operating-plan LOS must be compared with the initial LOS used in the demand estimates. If there are substantial variances, revised demand estimates may be necessary. The need for such revisions is discussed below in further detail. Once an operating plan satisfies both of these checks, vehicle requirements and service measures (i.e., vehicle hours and vehicle miles) can be calculated. Next, the estimation of employee requirements by category completes the estimation of supply parameters.

A continuous activity in the process of supply-parameter estimation not shown in Figure 1 is the use of "reasonability" checks. These checks should be made of the supply-parameter estimates or underlying operating plans to determine their reasonability and consistency with respect to transit experience elsewhere that involves similar technology and operating conditions.

Supply-Demand Equilibrium

An imbalance between supply and demand in a transportation analysis can lead to unrealistic forecasts and erroneous conclusions. Consequently, there is a need in the alternatives analysis process to explicitly address supply-demand equilibrium. In the estimation of supply parameters, this need is reflected in the two key checks of the operating plan cited earlier.

The first check determines whether or not the operating plan provides sufficient capacity to meet or exceed demand. In essence, this check is asking the following: Given the physical, technological, and policy constraints of this alternative, is an operating plan feasible that provides adequate capacity to meet demand? If not, these constraints must be modified by revising the definition of the alternative. Of course, certain technological constraints cannot be modified, so that in an extreme case it may mean abandoning a particular technology.

Whereas the first check may lead to a direct change in supply, the second check may lead to a direct change in demand. Demand forecasts are necessary inputs for the development of operating plans and the estimation of key supply parameters. However, demand forecasts are predicated on certain supply-related parameters specified in the initial description of alternatives, particularly (a) headways and (b) line-haul speeds and travel times.

During the development of operating plans, changes may be made in both headways and speeds. Headways will change whenever demand headways (i.e., headways needed to provide adequate capacity to meet projected demand levels) govern rather than policy headways (i.e., maximum headways in a given time period specified as policy). Speed estimates may change as more accurate estimation procedures are used and as more detailed information is available about station volumes, vehicle performance characteristics, and alignment and station location.

Significant discrepancies (or inconsistencies) between headways and speeds used in demand estimation versus supply-parameter estimation imply disequilibrium between demand and supply. If, for example, operating-plan development results in lower headways and faster speeds than those used in demand estimation, greater ridership than initially predicted could be anticipated.

Often such differences in headways and speeds will not be so great that completely new travel demand forecasts will be required. Nevertheless, consistency checks are necessary, and feedback modifications of demand estimates may be warranted to ensure supply-demand equilibrium.

Iterative Nature of Process

The estimation of supply parameters is an exercise that is not restricted to planning. It continues beyond alternatives analysis into preliminary engineering, final engineering, and finally detailed operational planning. As the estimation of supply parameters proceeds from the early stages of alternatives analysis toward final design and detailed operational planning, the input assumptions must be examined more rigorously, the alternatives must be specified in greater detail, and the estimation procedures must be increasingly precise and accurate. The process itself, however, remains essentially the same.

Initially, supply-parameter estimation is directed at providing information in sufficient detail to distinguish between broadly defined alternatives; later, it provides information needed to assess the worthiness of a limited number of alternatives; finally, it is concerned with the design and optimal operation of a single alternative. Although there is considerable overlap, the discussion in this paper is aimed at requirements and procedures needed to conduct planning during the alternatives analysis process.

Phase 1 Versus Phase 2 Alternatives Analysis

As specified in the Urban Mass Transportation Administration (UMTA) policy on major urban mass transportation investments (3), the overall alternatives analysis process is divided into two sequential phases or stages. The first phase, "systems planning", is usually regional in scope and is aimed at establishing relative priorities among individual corridors and identifying a limited set of alternatives for each corridor that merits more detailed analysis. The second phase is a detailed alternatives analysis of a limited set of alternatives in a single corridor performed in conjunction with the preparation of an environmental impact statement.

A major concern in providing suggested guidance for the estimation of supply parameters is the distinction between the two phases of alternatives analysis. Unfortunately, this distinction cannot be drawn clearly and unequivocally because of the variability in site-specific circumstances and the range of alternatives considered. Clearly, studies in which the alternatives examined include additions to extensive rail systems will involve greater operational complexity than studies that consider a single rail line in an area where there are no existing rail services. Greater operational complexity generally requires more sophisticated procedures of supply-parameter estimation.

Although firm guidelines cannot be prescribed, it is possible to characterize, in an approximate sense, requirements and procedures for regional studies versus those of subsequent corridor-level studies. With respect to phase 1 systems planning studies, supply-parameter estimation should generally have the following characteristics:

1. It is geared for simplified operating-cost estimation.
2. Estimated supply parameters include vehicle miles, vehicle hours, and fleet size but not necessarily employees by category.
3. Operating plans focus on peak-period weekday services.
4. Relatively simple manual estimation procedures are used.
5. Only limited sensitivity analyses and interactions are performed.
6. Estimation procedures and reasonability checks may not be independent.

For subsequent phase 2 corridor-level studies, supply-parameter estimation should have the following characteristics:

1. It is geared for comprehensive operating-cost estimation procedures.
2. A complete set of supply parameters is estimated (such as that given at the beginning of this paper).
3. Operating plans specify weekday peak and off-peak service and possibly weekend service as well.
4. More complex estimation procedures are used, and there is likelihood of the use of computer-assisted procedures.
5. The iterative approach with extensive sensitivity analyses is used.
6. Reasonability checks are independent of estimation procedures.

Because supply-parameter requirements and appropriate procedures may vary by stage between different locations, it is important that federal and local sponsors make agreements in advance with respect to supply-parameter requirements, inputs, and estimation procedures. Although these agreements may require modification during the course of the study, they will serve as a blueprint for study conduct.

Operating Feasibility

As the analysis of transit alternatives proceeds, the definition of alternatives is continually refined and increasingly detailed. At the conclusion of the process, it is of obvious importance that the remaining alternatives be feasible from engineering and operational standpoints. Operational feasibility, as noted earlier, is of considerable relevance to the process for estimating supply parameters.

It should be emphasized that operational feasibility, as defined here, is concerned with whether or not a system or technology can actually perform as specified in a planning study. The fact that a system may be operationally feasible does not necessarily mean that it is operationally efficient. Relative operational efficiency must be determined by exploring different operating strategies through the use of the iterative planning approaches discussed earlier.

There are two major prerequisites for operational feasibility. The first is that relations between supply parameters, operating plans, vehicle performance specifications, demand levels, and various descriptors of alternatives be internally consistent. For example, there should be consistency (a) between average speed and vehicle performance, route characteristics, station (stop) spacing, station dwell time; (b) between vehicle passenger capacity and loading standards, vehicle dimensions; and (c) between vehicle requirements and route characteristics, demand levels, vehicle characteristics, minimum-maximum headways, vehicle reliability.

The second prerequisite is that the specification of various performance characteristics be within the capability of available technology and that this technology be specified in the capital cost estimates. Performance characteristics of particular concern to operational feasibility include vehicle performance and signal control (for rail technology).

SPECIFIC CONSIDERATIONS IN DEFINING ALTERNATIVES

A number of inputs and assumptions to the process of supply-parameter estimation are embodied in the definition of alternatives. Several of the most critical of these are discussed below.

Vehicle Capacities and Loading Standards

Some of the most significant assumptions in the determination of supply parameters involve vehicle capacity and associated loading standards. Despite the significance of these assumptions, there is considerable variability in the manner in which they are developed and in the resulting loading standards and vehicle capacities. Of particular concern is the tendency to use different loading standards for different transit modes in the same analysis, which tends to bias the analysis in favor of one mode.

There are really two basic issues to be addressed in determining vehicle capacity for use in supply-parameter estimation:

1. What are appropriate loading standards and seating plans for the service under analysis? Loading standard refers to how much space is made available per passenger in peak demand conditions. Seating plan refers to how this space is used for seated passengers and standees.
2. How can these standards and plans be used to determine vehicle capacities for different transit modes and vehicle types?

The first issue is largely a policy question that must take into consideration the nature of the transit service being analyzed and the cost implications of different loading standards. For services that have long average trip lengths, such as commuter rail operations, the ratio of total to seated passengers for design volumes is usually low, approaching 1.0 where a seat is planned for every passenger. Furthermore, a relatively large amount of space per seat may be used to increase passenger comfort.

As passenger trip lengths decrease and passengers are boarding and alighting more frequently, the total/seated passenger ratio typically increases and smaller seats are used. Table 2 (3-5) gives typical space requirements for seated and standing passengers. There is an obvious trade-off between total capacity and the amount of seating provided, since standing passengers consume less space than seated passengers. To increase available space for standees, seats can be either eliminated or reduced in size and different seating patterns (e.g., longitudinal instead of transverse seating) can be used. Reducing seating, of course, increases the frequency of conditions in which all seats are occupied and some passengers must stand.

In addressing this issue, policy decisions must be made with respect to (a) the percentage of passengers to be seated in the peak design period (ratio of total to seated passengers), (b) the amount of space to be allocated to seated passengers, and (c) the amount of space to be allocated to standing passengers. In using these decisions to determine the capacities of specific transit vehicles, it is likely that the seating configurations used elsewhere will be inconsistent in terms of space per seated passenger and the total/seated passenger ratio. Thus, for consistency, it is necessary to assume modified interior layouts for planning purposes.

One method of doing this is to use a constant loading standard expressed in gross square feet per passenger, where gross square feet is measured by exterior dimensions. Gross vehicle area is convenient to use since it is generally easier to determine than the amount of usable interior space. The Regional Plan Association (6) proposes a standard of 5.4 ft^2 of gross vehicle area per passenger.

One advantage of using a loading standard expressed in square feet per passenger is that the seating configuration does not need to be treated explicitly. By doing so, it is assumed that seating patterns could be developed for different vehicles under analysis that would have the same total/seated passenger ratio and would be equivalent in terms of space per seated passenger and space per standee.

It is important to recognize, on the other hand, that the loading standard expressed in square feet per passenger is not independent of space standards for seated and standing passengers and the total/seated passenger ratio. Thus, the loading standard should be selected based on the type of service being planned and its cost implications. This implies that an appropriate universal standard for all transit loading does not exist and that, in selecting a standard for a specific set of circumstances, consideration should be given to its implications for space for seated passengers and standees and the total/seated passenger ratio.

Another consideration in using loading standards based on gross area is that different transit vehicles may have different percentages of usable interior area compared with gross floor area. For rail transit cars, the Regional Plan Association found considerable similarity in this respect for rail cars. For transit buses and light rail vehicles, however, the percentage appears lower. Hence, using the same loading standard based on exterior dimensions may not be appropriate in comparisons between bus and rail.

In many instances, it may be simpler to compute vehicle capacities directly based on prototypical seating patterns and standee standards rather than to adjust loading standards based on gross floor area. The vehicle capacity calculations are straightforward once the following data are available:

Data Item	Source
Usable floor area	Data and specifications for production vehicles
Square feet per seat	Based on trip characteristics and industry experience (Table 2)
Square feet per standee (minimum)	Based on trip characteristics and industry experience (Table 2)
Total/seated ratio in peak conditions	Based on trip characteristics and policy decision regarding comfort

In computing capacity from these inputs, it must be remembered that the exact seating pattern implied may not be strictly possible. Door locations, passageway and wheelchair requirements, and other considerations are constraints on the interior layout of a transit vehicle. However, the loading plan developed should be feasible, at least in an approximate sense, and will ensure consistency in the estimation of vehicle capacities.

Table 2. Space requirements for seated and standing passengers.

Category	Net Space per Passenger[a] (ft^2)
Seated passengers	
Typical commuter rail	4-6
Typical urban rail transit	3-5
Typical urban bus transit	3-4
Standing passengers	
In unconstrained condition	4-9
Minimum requirement to avoid contact	2.4-2.8
DuWag standard	2.7
NYCTA standard for maximum "practical" capacity	1.8
Moscow Metro minimum standard	1.3

Note: NYCTA = New York City Transit Authority.
[a]Excludes nonusable space. For seated passengers, the data include the space consumed by the seat plus leg space between seats. For standing passengers, the data are based on clear floor area per standee.

Service Hours and Service Days

Transit planning analyses are oriented toward peak-period weekday travel. Off-peak, weekend, and holiday services are not always given a great deal of consideration, though they can significantly affect annual supply parameters and operating cost estimates. Assumptions about such services are often treated implicitly in the factors used to expand supply parameters from peak-period to weekday totals and then from weekday to annual totals.

In the first-phase, regional-level alternatives analysis studies, the use of approximate expansion factors, without a detailed consideration of service levels in off-peak and weekend time periods, can be appropriate provided the factors are reasonable in light of local and national experience. In subsequent corridor-level studies, however, supply parameters should be developed based on explicit assumptions regarding off-peak, weekend, and holiday service levels.

In developing weekday supply-parameter estimates during corridor studies, service levels can be specified in three to five time periods during which service levels are assumed to be reasonably constant: peak (morning and evening), base, night, and "owl". It must be recognized that dividing a weekday into a limited number of constant service time periods is a simplification, albeit a necessary one. Transitions between time periods, special "tripper" runs, and other scheduling considerations typically produce hourly variations so that, in larger transit systems, no two hours are exactly the same.

To develop annualization factors--factors that expand weekday supply parameters to annual supply-parameter estimates--explicit assumptions regarding weekend and holiday service are necessary in corridor-level alternatives analysis studies. The sensitivity of annualization factors to weekend and holiday service assumptions is given in Table 3.

Another important consideration in developing annualization factors is seasonality. Bus transit systems often provide special revenue school runs when school is in session. Fares are charged and anyone may use the service, but it is oriented toward schoolchildren. Such service may represent a sizable proportion of background bus services and should be taken into account in developing annualization factors. In addition to school-related service, some transit systems have seasonal service variations related to recreational travel and climatic conditions, though they tend to be minor.

It should be noted that annualization factors are not necessarily the same for different supply parameters. For instance, in the case of rail lines, trains are likely to be shorter on weekends so that the reduction in vehicle miles, compared with weekday service, is greater than the reduction in train hours. Strictly speaking, such differences imply that different annualization factors should be used for vehicle miles and train hours. For regional-level studies, however, the use of a constant annualization factor for supply-parameter estimation is a reasonable approximation. In corridor-level studies, annualization factors should be developed with great care and, where appropriate, different annualization factors should be used for different supply parameters.

Table 3. Annualization factors: sensitivity to weekend and holiday service.

Service (%)				Annualization Factor
Weekday	Saturday	Sunday	Holiday	
100	-	-	-	251
100	50	-	-	277
100	70	-	-	287
100	70	40	40	311
100	50	50	50	308

Note: Assumes nine holidays per year and no seasonal variation in transit services.

Background Bus Networks

In analyzing capital-intensive transit facilities, facility alternatives are superimposed over existing or proposed bus systems in the corridor. The extent and cost-effectiveness of the background bus system are important because the evaluation of alternatives will consider the ridership, revenues, and costs of all transit services in the corridor.

As a minimum, the background bus system must be modified to interface with the capital-intensive line-haul system under consideration and reduce or eliminate service duplication. In addition, the background bus system may be modified to provide additional feeder-distribution service for the line-haul service and also may be expanded to include new services that are warranted by anticipated growth and are unrelated to the line-haul system. In the latter case, the background bus system is usually based on an improved "all-bus" alternative for the corridor.

In developing background bus systems for alternatives analysis, a number of questions must be addressed. Some of the most critical questions include the following:

1. Should all capital-intensive facility alternatives be superimposed over the same background bus network with only interface changes and route truncations and eliminations to reduce service duplication?
2. Should the background bus network be based on an "existing" or improved bus network?
3. Should there be an attempt to "optimize" the background bus network for individual capital-intensive facility alternatives?
4. Can individual bus lines be aggregated to simplify the analysis?
5. How should the treatment of background bus systems vary between regional-level and corridor-level studies?

In completed alternatives analysis studies, it appears most often that capital-intensive corridor facility alternatives have been superimposed over essentially the same background network and that the basis for this network is usually an "improved" all-bus alternative. Although the background network was often modified to provide feeder and distribution services, formalized attempts to determine the "optimal" bus network were not made. Bus networks were analyzed at the level of detail usually required for an Urban Transportation Planning System transit network (i.e., all background bus transit services were generally coded into the network). Note that completed studies have mostly been corridor-level rather than regional-level studies.

In making recommendations for the treatment of background bus networks in future alternatives analysis studies, no radical departure from past practice appears warranted, but some guidelines for consistency are needed. Recommendations in this regard are summarized below:

1. Essentially the same background bus network should be used for all capital-intensive facility or service alternatives.

2. The basis for this network should be an improved bus network that reflects service increases warranted by projected growth and applicable transportation system management improvements.

3. The background network must be modified for each line-haul facility or service alternative to provide appropriate interfaces and eliminate service duplications. Service duplications should only be eliminated to the extent that they could actually be eliminated. Experience with the Bay Area Rapid Transit system and the Washington, D.C., Metro system indicates that there may be substantial community opposition to the elimination of bus routes that, although they duplicate rail service, may offer a better level of service than rail or may provide local service that rail cannot provide. Bus routes that duplicate a proposed line-haul transit facility and that would offer service clearly inferior to that provided on the other facility can usually be eliminated or truncated without question.

4. No formalized attempt should be made in alternatives analysis to develop an optimal background bus network for capital-intensive facility alternatives. However, obvious modifications to the bus network should be made to provide adequate feeder and distribution services. At the conclusion of the analysis, supply-demand checks should be made to determine whether the services in the background network, especially feeder service, are in equilibrium. Furthermore, reasonability checks of riders per bus mile and riders per bus hour should be made to determine the relative productivity of the background bus network for different capital-intensive line-haul alternatives. These checks may indicate the need for refining the specification of the background network for a particular line-haul alternative. Usually, such refinements can be made manually without changing computer networks and repeating the travel demand forecasts.

5. Aggregation, or the schematic treatment of individual background bus routes, is appropriate for regional-level studies. In corridor-level studies, supply-parameter estimation and demand estimation will require more detail, and therefore the specifications for background bus networks should be at the level of individual routes or lines within the corridor under study.

CONCLUSIONS

Through an example, a discussion of general considerations, and detailed discussions of selected topics, this paper has attempted to reveal the complexity and significance of the estimation of transit supply requirements in planning studies. Although the paper has been written in the context of federally required alternatives analysis studies, the principles and concerns raised are applicable to transit planning generally, particularly any planning efforts in which fixed-guideway transit modes are considered and alternative transit modes evaluated.

The paper is not comprehensive and has focused its most detailed discussions on selected inputs to the supply-parameter estimation process that require special attention. Many other important inputs to the process were not covered, and the entire subject of how these inputs are used to calculate supply-parameter estimates was not addressed at all. It is hoped that this paper will stimulate interest and further discussion of supply-parameter estimation among those planners actively involved in transit alternatives analysis studies.

ACKNOWLEDGMENT

This paper reports the results of research sponsored by the Transportation Systems Center and UMTA of the U.S. Department of Transportation. The views and recommendations presented, however, are mine and do not necessarily reflect the position of the research sponsors.

REFERENCES

1. Estimation of Key Transit Supply Requirements. Alan M. Voorhees and Associates, McLean, VA, draft, spring 1980.
2. Urban Mass Transportation Administration. Major Urban Mass Transportation Investments: Statement of Policy. Federal Register, Vol. 41, No. 185, Sept. 22, 1976, pp. 41512-41514.
3. B. Pushkarev and J.M. Zupan. Urban Space for Pedestrians. M.I.T. Press, Cambridge, MA, 1975.
4. H.D. Quinby. Mass Transportation Characteristics. In Transportation and Traffic Engineering Handbook, Institute of Traffic Engineers, Arlington, VA, 1976, pp. 207-257.
5. E.S. Diamant and others. Light Rail Transit: State of the Art Review. U.S. Department of Transportation, 1976. NTIS: PB 256 821.
6. Where Rail Works: An Exploration of Travel Volume Criteria for Fixed Guideway Transit. Regional Plan Assn., New York, draft, 1978.

Publication of this paper sponsored by Committee on Transportation Systems Design.

Development of Regional Multimodal Transportation Performance Measures for the Twin Cities

WILLIAM R. LOUDON, WENDY P. STERN, AND JOHN F. HOFFMEISTER

Results of a study performed to develop measures for assessing the effectiveness of the transportation policies of the Metropolitan Council of the Minneapolis-St. Paul area are presented. The purpose of the study was to develop a set of performance measures for assessing the degree to which the Metropolitan Council's regional transportation policies were being adhered to and the extent to which the policies have been effective in attaining the basic objectives of the region. Two important innovations in the study were the emphasis on planning versus management performance measures and the development of three sets of measures—one to provide an overview of transportation in the region, one to assess objective attainment, and the other to determine policy

adherence. The planning measures address the degree of movement toward satisfying the planning objectives of the region. This differs from more traditional management-oriented measures, which have focused on transit supply and productivity and are used primarily for allocating funds within transit agencies. The importance of using different measures for different purposes is stressed. Eight purposes are discussed, and example measures are suggested for each. The three primary uses discussed are (a) determining the degree of objective attainment of the regional transportation system, (b) determining the degree of policy adherence within the system, and (c) evaluating the regional transportation policy plan. The framework within which the performance measures were developed is discussed, and a list of the summary measures, including examples of detailed measures, is presented. A plan for implementing these measures is currently being developed.

The importance of urban transportation as a regional issue has long been recognized, and its implications for regional land use, economic vitality, energy consumption, and environmental quality have become more widely accepted. But in spite of the recognition of these relations, few regional planning agencies have been effective in influencing the development of a transportation system at the regional level guided by a clear set of transportation policies.

For many years, the Metropolitan Council of the Minneapolis-St. Paul area has recognized these interrelations and has been a leader among regional planning agencies in consciously addressing the issues. In 1975, the Metropolitan Council adopted a development chapter of its comprehensive Metropolitan Development Guide (1). As a means to help curb the uncontrolled spread of new urban development, the Development Framework defines a metropolitan urban service area (MUSA). It is now the policy of the Metropolitan Council that the bulk of new development and redevelopment is to occur within the MUSA through 1990. Similarly, investments in new urban services and facilities (including transportation) are to be concentrated within the MUSA line, to the extent feasible.

In 1976, the Metropolitan Council adopted a revised transportation chapter of its comprehensive Metropolitan Development Guide. The central focus of the Metropolitan Development Guide Transportation Policy Plan is on more efficient use of existing and committed transportation investments in the region. Among other things, it suggests that future highway and transit improvements should be scaled to serve off-peak rather than peak-period demand levels, that investments in new capacity should be concentrated within the MUSA delineated by the Metropolitan Development Guide, that transit services should focus on providing for travel within subregions and to the "metropolitan centers" rather than providing service between subregions, and that a wide variety of means should be considered for encouraging people to travel as passengers (whether in transit vehicles, taxis, or private automobiles) rather than as drivers.

When the Transportation Policy Plan was adopted, it was intended that the policies would be periodically reviewed and revised as necessary to maximize their effectiveness in meeting the basic objectives of the region. In August 1979, the Metropolitan Council initiated a program to develop performance measures that could be used in the evaluation of the 1976 Transportation Policy Plan. The result of this effort has been a framework for evaluating policy adherence, policy effectiveness, and objective attainment by using three basic types of performance measures.

This paper describes the framework within which performance measures were developed for the Minneapolis-St. Paul area and presents a discussion of the need for and the purpose of the measures. In addition, the measures that were developed by the Metropolitan Council are presented, and the issues related to the application of the measures in a performance review are discussed.

At the time this paper was written, the first phase of the project was complete. In that phase, the range of uses for performance measures was explored, a framework for developing performance measures was specified, and measures were developed for each of the three categories mentioned above. In the second phase of the project, which is to be completed in early 1981, details necessary for making the measures and the performance review operational will be explored. Those issues will include a screening of the measures with respect to eight criteria, specification of a plan for the collection of the data necessary to apply the measures, and specification of a plan or schedule for performance review.

PURPOSES AND USES OF PERFORMANCE MEASURES

The primary purpose of the performance measure project was to enable the Metropolitan Council to evaluate its Transportation Policy Plan by determining (a) the degree to which the transportation policies were being followed and (b) the extent to which these policies were bringing about the attainment of the basic objectives of the region. More formally stated, the performance measures were developed to

1. Provide an indication of policy adherence,
2. Provide an assessment of the attainment of regional objectives,
3. Provide input into a periodic reevaluation of the transportation policies that would result in revisions of the existing policies as necessary, and
4. Aid in the evaluation of the regional transportation system.

In addition to the four primary purposes listed above, five other significant purposes were identified for which the performance measures could be used. They are

1. To monitor and evaluate existing transportation projects and services;
2. To review and evaluate proposed transportation projects;
3. To estimate the incidence of impacts of the transportation system on different socioeconomic groups and geographic locations;
4. To provide information for responding to inquiries from citizens, elected officials, and agencies about the performance of the transportation system; and
5. To provide a set of indicators of the state of transportation in the region.

It is important to point out that the measures that have been developed to satisfy the purposes mentioned above expand considerably on the types of indicators traditionally used to assess transportation performance. Measures such as cost per vehicle hour of service, cost per seat mile of service, or cost per person mile served represent just one component of a more comprehensive set of measures that takes into account regional planning objectives in addition to the efficiency with which the transportation service is being provided. This is required since the policies developed by the Metropolitan Council often specify priorities of market segments to be served. The cost-effectiveness of providing the transportation services required to effectively serve a specific market segment may not be as favorable as for other types of transportation services that can be provided to the general population. In particular, demand-responsive services and other paratransit services provided for the elderly and the handicapped are an example of where cost-effectiveness measures may not be appropriate as an exclusive evaluation tool but should be considered in conjunction with more human-service-oriented measures.

FRAMEWORK OF PROJECT

The framework within which the performance measures were developed consists of three elements structured in a hierarchical relation: the regional objectives (derived from the transportation policies), the transportation policies, and the performance measures. This framework is shown in Figure 1. Each of these elements and the relation of the performance measures to each element are briefly discussed below and elaborated on in subsequent sections of this paper. The regional transportation policies developed by the Metropolitan Council suggest the role that the transportation system can have in the attainment of the regional planning objectives. Careful examination of the 36 policies (of the 50 policies in the transportation policy plan, performance measures were developed for the 36 that were regional in scope) and discussions with the Metropolitan Council staff resulted in the identification of seven basic objectives:

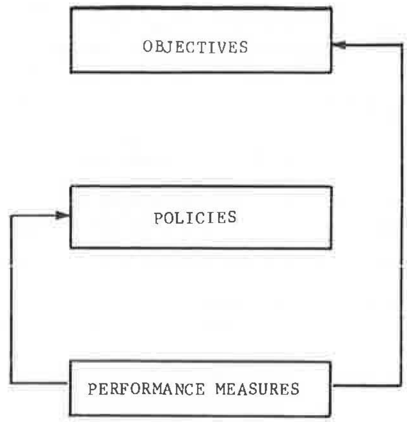
1. Maintain and selectively increase accessibility and mobility,
2. Promote the effective use of existing investments,
3. Improve travel safety,
4. Promote positive social impacts,
5. Promote environmental quality,
6. Conserve energy, and
7. Promote positive economic impacts.

These objectives constitute the highest level of the performance measures development framework shown in Figure 1.

At the next level in the hierarchy of Figure 1 are the 36 regional transportation policies. These policies represent Metropolitan Council statements of the role that the regional transportation system should play in attaining the seven objectives. The policies that share common themes have been grouped for simplicity of presentation into the following 14 policy clusters:

1. Concentrate transportation investments within the MUSA.
2. Place equal emphasis on transit service and investment within subregional areas and to and within the metropolitan centers.
3. Plan transportation services so as to encourage the self-sufficiency of subregional centers with less emphasis on intersubregional travel.
4. Reinforce the attractiveness and vitality of the metropolitan centers by providing a high level of transportation service to and within the centers.
5. Transportation services should be given highest priority in areas that have the greatest demand among persons dependent on public transportation (elderly, handicapped, youth, and low income).
6. Better use of the transportation system should be encouraged, emphasis should be placed on low-capital [transportation system management (TSM)] actions, and both the public and private sectors should be viewed as eligible service providers and as participants.
7. Transportation actions should promote the attainment of state and national environmental standards, giving priority to geographic areas that experience the most severe violations.
8. Safety considerations must be a major consideration in the planning, implementation, and maintenance of any transportation facility or service.
9. Transit actions that demonstrate the highest cost-effectiveness should receive priority consideration.
10. The concept of equity should be used to establish fare structures that reflect the operating cost of the service and the public purpose or need for the service.
11. Transportation facilities and services should promote and be consistent with the land use and development patterns specified in the Development Framework chapter of the Metropolitan Development Guide.
12. Transportation actions outside the urban service area should be limited to the specific needs of the free-standing growth centers and the rural town centers, consistent with the land use and development plans and policies of the region.
13. Innovative transportation demonstration projects should be tested and accompanied by a well-designed evaluation plan in order to assess the merit of continuing and/or expanding the project.
14. Transportation investments should be made on the basis of need and the ability of the metropolitan area to support these investments in relation to other metropolitan system needs and investments over time.

Figure 1. Framework of performance measures project.



Each of the clusters is related to at least one of the seven basic objectives and is often related to several of the objectives. This relation is illustrated in Table 1 for a sample of the policy clusters.

As Table 1 indicates, more than one policy cluster may be related to the same objective. It is for this reason that the complete set of policies and policy clusters must be evaluated in assessing how well the regional transportation system is promoting the attainment of the basic objectives of the region.

Below the policies on the hierarchy of Figure 1 are the performance measures. These measures represent analytic tools for evaluating the extent to which the regional transportation system is consistent with the policies and, in turn, the objectives.

It is important to note that it is possible to achieve policy adherence without necessarily attaining any of the basic objectives of the region. For example, consider the policy that states that the major share of transportation investments should be

Table 1. Example of relation between policy clusters and basic objectives.

Policy Cluster	Objective: Maintain Accessibility and Mobility	Promote Effective Use of Existing Investments	Promote Environmental Quality	Conserve Energy	Improve Travel Safety	Promote Positive Economic Impacts	Promote Positive Social Impacts
1	●	●	●	●		●	
2	●	O	O	O		●	●
3	●	●	●	●			
4	●	●				●	●
5	●						●
6	●	●	●	●		●	
7		O	●	●			

Note: ● = primary relation between objective and policy cluster; O = relation of lower priority than primary relation.

concentrated within the urban service area. Although the rationale for this policy relates to one of the basic objectives of the region--namely, to promote the effective use of existing investments--simply measuring the pattern of investments does not provide an indicator of how well the regional land use and development plan is being achieved. Therefore, it is possible to concentrate investments in the urban service area and satisfy this policy without adhering to the land use and development plan, so that the objective implied by this policy is not attained.

For this reason, it was necessary to develop two types of performance measures for each policy: objective-related performance measures, which assess the objective attainment that results from adherence to a particular policy, and policy-related performance measures, which determine the degree to which the policy as specified is being achieved. Depending on the nature of the policy, the objective-related measures and the policy-related measures may be the same for a given policy. When the two are not identical, however, the policy-related measures should be used in combination with the objective-related measures to evaluate the policy under consideration. The arrows in Figure 1 illustrate that some performance measures would allow the Metropolitan Council to evaluate policy adherence and others would enable the determination of both policy adherence and objective attainment when the two are related.

In addition to objective-related and policy-related performance measures, a set of measures has also been developed that is designed to provide a limited perspective of the state of the region with respect to its transportation system. These measures, called overview performance measures, do not provide any direct indication of policy adherence or objective. They have a much more specialized use and are discussed in a separate section of this paper.

DEVELOPMENT OF PERFORMANCE MEASURES

Objective-Related Measures

A set of summary performance measures was defined for each of the seven basic objectives. Each summary measure represents a general construct from which more specific performance measures can be developed. The summary measures make it possible to show in a simplified way the relations between them and the regional policies and to demonstrate how a performance measure in its general form may relate to several of the policies. This is important because it emphasizes the interrelations between the measures and the policies.

After it was determined which of the summary measures were applicable to each of the policies, based on the relations between the objectives and the policies, the summary measures were defined in greater detail to reflect more specifically the policy under consideration. In this way, a more meaningful measurement of that policy with respect to the related objectives could be achieved.

When an objective contained more than one category, performance measures were developed for each category. For example, the objective to "promote the effective use of existing investments" encompasses two categories--land use and the transportation system--both of which constitute existing investments. Performance measures were defined that address (a) the connectivity of the transportation system to specified land uses and (b) the basic effectiveness of the transportation system as an independent investment unit. The summary performance measures for this objective and three of the other objectives are given in Table 2 for example purposes. The complete set of summary measures for all seven of the basic objectives can be obtained either from the Metropolitan Council or Cambridge Systematics.

Several summary measures were developed for each category in order to provide a range of perspectives for evaluation purposes. Once again, if we use the objective relating to existing investments as an example, the connectivity between the transportation system and land use types can be measured by determining the percentage of the population within a specified travel time of the designated land use type. It can also be measured by determining the extent to which designated land uses are within a specified travel time of a certain percentage of the population. Both of these measures address the connectivity between the transportation system and the specified land use, but each portrays it in a slightly different manner.

The progression from the summary performance measure to a more detailed version of the summary measure in order to be more policy specific is presented for policy cluster 1 ("concentrate transportation investments within the urban service area") as an example. The two basic objectives that relate directly to this policy cluster are to "promote the effective use of existing investments" and "promote positive economic impacts". One summary performance measure that applies to this policy for the objective to "promote the effective use of existing investment" is the "population within a specified travel time of designated existing investment areas". To provide a more meaningful assessment of the policy with respect to this measure, a more detailed specification of the measure is developed. This may be the "change over time of the urban service area population within a specified travel time of designated activity generators or geographic locations compared with the same change for residents outside the urban service area". This

Table 2. Example of objective-related summary performance measures.

Objective	Category	Summary Performance Measure
Maintain accessibility and mobility		Population (activities within specified distance of transit)
		Population within specified travel time of activities (geographic locations)
		Travel times for specified origin-destination pairs
Promote effective use of existing investments	Land use	Population within specified travel time of designated existing investment areas
		Percentage of trips with destination at designated existing investment areas
		Percentage of designated existing investment areas within specified travel time of certain percentage of the population
		Percentage of designated investment areas within specified distance of a transit stop
		Change in assessed valuation (retail sales and employment) in designated existing investment areas due to improvements in level of transportation services attributed to facility and/or service improvements
		Number of displacements (disruptions) to designated existing investment areas due to construction (operation) of a transportation facility (service)
		Visual compatibility of transportation facility with adjacent land uses
Promote environmental quality	Air quality	Assessment of current (predicted) conditions relative to previous (current) conditions
		Contribution of specific travel patterns to ambient air quality
		Air-quality improvements due to selected changes in travel patterns as result of improved level of transportation services (exogenous factors, improved efficiency of transportation vehicles)
	Noise	Activities (persons) affected by excessive noise levels
		Noise level reductions due to selected actions
	Water	Impact on surface water (groundwater) due to facility construction and/or operation
	Land	Acres (areas) of sensitive land use affected
	Biological life	Impact on animal and/or plant life due to facility construction and/or transportation operations
	Other natural resources	Impact on mineral resources due to facility construction
Conserve energy		Energy consumed by various travel patterns (modes)
		Energy saved by changes in travel patterns

procedure is to be repeated for each of the summary measures in relating them to the specific policies.

Policy-Related Measures

The policy-related measures provide the means by which to assess directly the degree of adherence to regional policies without regard to whether or not any of the seven basic objectives are being attained. In some cases, the policy-related measures are identical to the objective-related measures. In other cases, two separate sets of performance measures are required. Which situation applies depends on whether or not the policy can be completely represented by the basic objectives.

As an example of the two kinds of situations, consider Metropolitan Council policies 1, 8, 10, and 23 (in order to better demonstrate the methodology used, this section focuses on some of the specific policies and not on the aggregate policy clusters):

1. Policy 1--The major transit and highway investments should be concentrated within the urban service area (as defined in the Development Framework chapter of the Metropolitan Development Guide).
2. Policy 8--Safety standards must be a major consideration in the planning, design, and maintenance of transportation facilities and services.
3. Policy 10--Discourage the use of automobiles in those areas where air quality is unacceptable if automobile emissions are a major contribution to the degradation of the air.
4. Policy 23--Equal emphasis should be placed on transit service and investment within subregional areas and to and within the metropolitan centers.

Policies 8 and 10 make explicit reference to two of the basic objectives. Policy 8 refers to the objective, "improve travel safety", and policy 10 refers to "promote environmental quality". Using the performance measures developed for these two objectives will provide an indication of how well these two policies are being achieved. Therefore, for these two policies the policy-related measures and the objective-related measures are identical.

For policies 1 and 23, however, the situation is different. While they can also be measured with respect to the basic objectives--namely, the objectives to "promote the effective use of existing investment" and to "promote positive economic impacts" (with the added objective to maintain accessibility and mobility for policy 23), a direct measurement of these policies requires performance measures that do not necessarily measure the attainment of the basic objectives. An example of a performance measure that can be used in the evaluation of each of these policies is the following performance measure for policy 1: local dollars appropriated (spent) on transportation investments within the urban service area in relation to the local dollars appropriated (spent) on transportation investments outside the urban service area.

The extent to which transportation investments are concentrated within the urban service area can be determined by the above measure. However, it is possible to evaluate adherence to this policy--namely, to concentrate transportation investments within the urban service area--without necessarily determining whether the policy is also promoting the effective use of existing investment or promoting positive economic impacts.

A similar situation is represented by the following performance measure for policy 23: dollars spent on intrasubregional transit in relation to dollars spent on transit service from that subregion to the metropolitan center. Once again, it can be seen that satisfying this policy does not necessarily guarantee that the basic objectives will simultaneously be satisfied. It is possible to provide an equal level of service and investment within subregions and to and within the metropolitan centers but not necessarily to satisfy the basic objective of maintaining accessibility and mobility.

A set of proposed policy-related performance measures has been developed for each of the 36 regional policies. In all, 161 policy-related performance measures were developed, of which 75 were more detailed versions of the objective-related performance measures. The 86 that were not related to the objectives can be summarized as follows: 17 are indicators of ridesharing activity and effectiveness, 16 are transit supply measures, 12 are measures of transit use, 10 are measures of vehicle occupancy, 7 are highway supply measures, 7 are measures of public funding priorities, 5 are measures of public transit cost, 5 are measures of effectiveness for demonstration projects, 4 are indications of intermodal terminal location and effectiveness, and 4 are indicators of the locations of transit-dependent populations.

As an example, the policy-related measures for the Metropolitan Council's regional policy 8 are presented below. Regional policy 8 states that "safety standards must be a major consideration in the planning, design, and maintenance of transportation facilities and services." The performance measures related to this policy are as follows:

1. Number of transit accidents per 1 000 000 revenue vehicle miles (route, etc.),
2. Number of automobile accidents per mile of facility (1 000 000 vehicle miles, intersections),
3. Number of transit passenger injuries (fatalities) per mile of facility (vehicle miles, intersections),
4. Number of automobile passenger injuries (fatalities) per mile of facility (1 000 000 vehicle miles, intersections), and
5. Number of accidents significantly attributable to dangerous operating conditions (structural problems of the facility, vehicle design, or inadequate facility or transit vehicle maintenance).

A full set of the policy-related performance measures can be obtained from either the Metropolitan Council or Cambridge Systematics.

OVERVIEW OF PERFORMANCE MEASURES

A set of 15 overview performance measures was developed to satisfy the last of the nine purposes stated earlier in this paper, namely "to provide a set of indicators of the state of transportation in the region". These 15 measures provide no direct indication of policy adherence, policy effectiveness, or objective attainment but do allow for comparison of the state of the region over time or comparison of the Twin Cities area with other regions. The overview measures, which are presented below, are for the most part restatements of either objective-related measures or specific policy-related measures (measure 12, which is not an objectiveor policy-related measure, is designed to reflect travel patterns that could influence the effectiveness of the transportation system):

1. Transportation expenditure (public) per capita;
2. Percentage of transportation expenditures made inside the MUSA;
3. Work-trip modal split;
4. Average trip length (by trip purpose);
5. Average vehicle occupancy (by trip purpose);
6. Accidents per million person miles of travel;
7. Percentage of population in the MUSA with access to the regional transit system;
8. Percentage of employment in the MUSA that is accessible by means of the regional transit system;
9. Percentage of households within a 30-min travel time from any part of a designated subregion to any other part of that subregion by transit and highway during off-peak periods for residents of that subregion;
10. Percentage of households within a 30-min travel time from any part of the MUSA to either metropolitan center by highway during off-peak periods for residents of the MUSA;
11. Percentage of households within a 45-min travel time from any part of the MUSA to either metropolitan center by public transit for residents of the MUSA;
12.

		Place of Residence: Inside MUSA	Place of Residence: Outside MUSA
Place of Work	Inside MUSA		
	Outside MUSA		

13. Annual transportation-related energy consumption (by fuel type) per capita;
14. Comparison of the estimated cost of a single-occupant automobile trip with the cost of a transit trip, when each is of average trip length; and
15. Number of locations in which federal air-quality standards are not met.

These overview measures were developed for informational purposes and not for the purpose of policy analysis as were the other measures.

APPLICATION OF PERFORMANCE MEASURES IN EVALUATION OF REGIONAL POLICIES

An important step in the development of performance measures is the operationalizing of the measures. By this, we mean the process by which a performance measure is estimated and used in evaluation. Operationalization includes the detailed specification of a performance measure to relate it to the situation being evaluated, determination of the data necessary for evaluation, and the appropriate data-collection techniques.

In operationalizing the performance measures, each is analyzed to determine the ease with which it can be measured. Depending on this assessment, the measure may be respecified to improve its measurability or to take advantage of existing data if this will not significantly compromise the intent of the measure. For those measures for which data do not exist and a new data-collection effort is not considered to be feasible, the measure is either eliminated or assigned a low priority.

Once the measurability criterion has been considered, each performance measure is analyzed with respect to eight other criteria that must be satisfied in developing a final set of performance measures. These criteria are cost, pertinence, clarity, sensitivity and responsiveness, appropriate level of detail and aggregation, nonsensitivity to exogenous factors, comprehensiveness, and discrimination between influences.

It should be noted that the set of preliminary measures developed to date has not yet been reviewed with respect to the criteria listed above. This screening will occur during the phase of this study that is currently in progress.

A final step in the operationalizing of performance measures is the specification of data storage and retrieval systems that facilitate performance measure estimation as well as reporting systems that enable the Metropolitan Council to monitor the performance of the regional transportation system over time. The refinement of the Development Framework concepts, the specification of detailed performance measures, and the operationalizing of the measures are currently under way as part of phase 2 of this project.

SUMMARY

The performance measure study for the Metropolitan Council of the Twin Cities area has illustrated the usefulness of multimodal performance measures with a regional planning orientation. The measures developed for the Metropolitan Council include many of the more traditional highwayand transit-supply-oriented measures that have been applied in performance reviews throughout the country. However, the highlight of this study is that it has also produced performance measures that reach beyond the supply characteristics to relate the supply of transportation and the attainment of planning objectives.

Three types of measures have been developed for the Metropolitan Council to meet their many needs: The first type was designed to assess attainment of regional planning objectives, the second was designed to evaluate performance with respect to specific policies, and the third was designed to provide an overall picture of the state of transportation in the region. In all, roughly 200 measures were necessary to satisfy the needs implied by the many types of applications that the Metropolitan Council can make of the measures in performance review; each type of use requires a different type of measure.

Although a large number of performance measures were presented to the Metropolitan Council, a methodology for their use was also developed that results in a practical, straightforward, and comprehensible program for performance review. The methodology ensures that there is a performance measure appropriate to each need.

ACKNOWLEDGMENT

We wish to recognize the support and assistance provided by the Metropolitan Council of the Twin Cities area and the partial funding provided by the Urban Mass Transportation Administration, U.S. Department of Transportation. In particular, we wish to thank the Metropolitan Council Performance Measures Task Force and the transportation staff. We are especially grateful to Larry Dallam of the Metropolitan Council for his enthusiastic support of this effort and his technical guidance and diligent review of all materials prepared for the project and this paper.

Further helpful insights were provided by Marvin L. Manheim of Cambridge Systematics, Inc., G.J. (Pete) Fielding of the Institute for Transportation Studies at the University of California at Irvine, and Earl Ruiter of Cambridge Systematics, Inc. We are grateful for those insights.

Our special thanks go to Carol Walb, Susan Billings, Patti Kinnear, and Sarah Sly for their contributions in the preparation of this paper and to James Wojno for his role in the preparation of the graphical material.

Although we acknowledge the contributions of those mentioned above, the views expressed in this paper are ours and do not necessarily reflect the views of the above-mentioned persons, the Metropolitan Council, or the U.S. Department of Transportation.

REFERENCE

1. Metropolitan Development Guide. Metropolitan Council of the Minneapolis-St. Paul Area, St. Paul, MN, 1975-1977.

Publication of this paper sponsored by Committee on Transportation Systems Design.

Disaggregate Model of Mode Choice in Intercity Travel

ALAN GRAYSON

The development of a policy-sensitive model of mode choice in intercity travel is discussed. The disaggregate logit model is based on the National Travel Survey of 1977, supplemented by service information from industry guides. Automobile, air, bus, and rail market shares are estimated from information on cost, travel time, frequency, terminal access, automobile availability, and trip purpose. By all measures, the model performs well. It is applicable on a national, regional, or route-by-route level. Forecasts are performed for a variety of national and regional scenarios.

This report summarizes the results of a recent project to develop a policy-sensitive behavioral model of mode choice in intercity travel. The approach is based on the economic theory of consumer behavior embodied in the logit statistical model. The model is calibrated with data from the National Travel Survey of 1977, supplemented by common-carrier service information for selected routes.

The purpose of this project is to develop a research tool to forecast the impact of transportation policies and controls on national patterns of intercity travel. These policies include gasoline pricing, modernization of the Interstate Highway System, air traffic control, interstate common-carrier pricing and service, etc., as well as general concern for competition and efficiency, energy conservation, regional equality, and other broad goals that intercity travel may impinge on. The government needs reliable forecasting techniques to weigh various policy alternatives.

The disaggregate, cross-sectional logit model of mode choice in intercity travel employed is based on a sample of intercity trips from a given period. The sample includes information for each trip on the attributes on each mode, characteristics of the travel party, and the mode chosen. The logit model is used to estimate the relative importance of different explanatory variables determining mode choice. In the base situation, the observed distribution of modal attributes is matched with the observed mode split. In alternate scenarios, changing the values of these variables yields a different mode split. The causal relationship between mode choice and the predictor variables specified in the model is based on a specific conception of the choice grounded in consumer behavior theory and experience with mode choice in urban contexts. The disaggregate logit model views transportation demand as demand for the "best" mode from each traveler's point of view, rather than demand for a specific mode per se (cf., time series multiple regression models of intercity travel). The "derived-demand" assumption in the model emphasizes competition among modes; the model is not concerned with changes in total demand.

What are the elements of an ordinary traveler's logit utility function when he or she evaluates an intercity travel mode? Four measurable factors that the traveler considers are cost, travel time, frequency of service, and terminal access. For each of these variables, the data only approximate what is believed to be the "actual" element of the subjective utility function, but each is a reasonable and useful approximation.

Modal attributes become observable as the circumstances of the trip are defined. Obviously, for instance, there is no one universal cost for air travel; the cost depends on the circumstances. Some circumstances are determined for all trips prior to mode choice. These circumstances may involve decisions too, but it is assumed that these decisions are made prior to mode choice in order to simplify the problem.

It is assumed that automobile ownership, destination, and travel party membership are determined prior to mode choice. It is also assumed that the decisionmaker considers all available alternatives, has perfect information about their attributes, and bears all costs. Some exceptions to these rules, e.g., business travelers, are isolated through stratification described below.

By making these assumptions about the circumstances of the trip, the values of the attributes of each mode become less and less ambiguous. At some point, a single value has been defined for, say, the cost of air travel. The attributes to be considered (cost, travel time, frequency of service, and terminal access) are admittedly biased toward the economic and the measurable. This choice is conditioned by the available data. Variables are specified according to how they might matter to the decisionmaker. To illustrate, a traveler has no inherent interest in the frequency of common-carrier service between Boston and Washington, but he or she may know that arrival at Washington National Airport must be by 9:00 and the closest earlier arrival is 8:13. So from the traveler's point of view, frequency itself is meaningless; but the reciprocal of frequency, the average waiting time at the destination between actual embarkment and the "ideal" preferred time of embarkment, is important.

So far only modal attributes and the assumptions that define these attributes have been considered. But many characteristics of the travel party itself affect mode choice. There are several ways in which characteristics of the travel party or its members might be modeled: The characteristics might be used to scale a modal attribute, foreclose an alternative, enter the utility function as a dummy variable, or stratify the data. Examples of scaling in this model include multiplication of common-carrier fare by travel-party size to yield common-carrier cost and multiplication of travel time by family income, following the hypothesis that the value of time varies linearly with income. The lack of automobile availability forecloses automobile as an alternative. Stratification was preferred to intercept dummy variables to avoid the assumption of equal slope coefficients and allow the testing of behavioral hypotheses. Stratification was employed for trip purpose and annual automobile use, but not for routes or regions, since the basic principles of mode choice decisionmaking were not thought to vary by route or by region.

This introduction is intended as an outline of the general considerations behind the structure of this model of mode choice in intercity travel. The utility function remains simple to reflect a simple decision-making process. Most characteristics of the travel party are manifested through stratification or elimination of alternatives to leave the behavioral assumptions of the model clear and challengeable.

MODEL DEVELOPMENT

The core of the model of mode choice in intercity travel is information on several attributes of each mode: cost, travel time, frequency, and access. The goal of this analysis has been to gather this information for a sample of trips within the 1977 National Travel Survey (NTS) and create a working forecasting model with broad geographic applicability.

The NTS collected information on all trips of 100 miles or more made by members of 20 000 households during 1977. This represents a sample of about 1 in every 4000 intercity trips.

The NTS contains no information on the frequency of common-carrier service for the trip. The information on travel time (number of days en route) is too imprecise to be useful. The information on cost (transportation cost for common-carrier trips) tells nothing about automobile trips and modes not chosen and is also apparently unreliable. As for access, there are data on the distance from place of residence to common-carrier terminals.

There are two methods of estimating intercity cost, travel time, and frequency of service: as a function of distance and region or of origin and destination. The first method is somewhat imprecise, obscuring differences among different routes and different stages of the trip (e.g., access and line haul). The second method is more precise, but it involves secondary data collection (e.g., looking up the number of flights between New York and Washington). It is applicable only to trips originating and terminating in standard metropolitan statistical areas (SMSAs) because only geographic coding by SMSA is available.

Following Stopher (1,2), a per-mile estimate of automobile cost and travel time was used together with estimates of common-carrier service attributes collected for specific routes from industry guides. As will be seen later, there is evidence that per-mile estimates of common-carrier attributes may be sufficiently accurate. The origin-destination approach required selection of specific routes within the national sample. Models based on selected routes are accurate only to the extent that the decision process of travelers along these routes (the subjective utility function, which is estimated in the logit model) is the same as the decision process elsewhere. Clearly, the mean value and distribution of modal attributes will be different for the sample as a whole. But there is no reason to believe a priori that the basic decision process will be different.

The model was calibrated for samples based on two different sets of routes: a sample of 1658 trips along 46 routes that were the most heavily sampled in the NTS (generally corresponding to the routes with the most person trips) and a sample of 1062 trips along 41 routes representing the greatest number of passenger miles. The first sample comprises New York to Boston, Hartford, Albany, Syracuse, Philadelphia, Washington, Chicago, and Miami; Los Angeles to San Diego, Bakersfield, Santa Barbara, Phoenix, Las Vegas, San Francisco, and Sacramento; San Francisco to Fresno, Salinas, Sacramento, Reno, and San Diego; and 26 other routes. The second sample includes New York to Boston, Philadelphia, Washington, Miami, Chicago, Dallas, Las Vegas, San Francisco, and Los Angeles; Los Angeles to San Diego, Phoenix, Las Vegas, Santa Barbara, San Francisco, Sacramento, Seattle, Honolulu, Dallas, Chicago, Washington, and Boston; Chicago to Honolulu, San Francisco, Las Vegas, Phoenix, Miami, Tampa, and Washington; San Francisco to Honolulu, San Diego, Sacramento, Reno, Washington, and Boston; Miami to Washington, Philadelphia, and Boston; Washington to Philadelphia, Orlando, and San Diego; and Dallas to Houston.

For both samples, information was collected on coach fare, fastest line-haul time, and trips per week for each common-carrier mode along each route as of June 1977. Common-carrier cost was defined as coach fare multiplied by travel party size. "Waiting time" was defined as one-half the average number of hours between common-carrier departures, i.e., the reciprocal of frequency. Travel time was defined as fastest line-haul time. Access was defined as distance to common-carrier terminals. A unit of access based on cost or time would have been preferable, being a better estimate of subjective utility, but the data would only allow an arbitrary transformation based on some assumed access speed or cost per mile.

Estimates of automobile cost and travel time were based on a U.S. Bureau of the Census-generated distance estimate called Place Identification, Characteristics and Area, Direction and Distance (PICADAD). Although only origin and destination SMSA information was released by the Census Bureau, the original survey included origin and destination addresses. Knowing the geographic location of every significant "place" in the United States, the Census Bureau calculated exact straight-line distance. The final PICADAD estimate was the straight-line distance scaled by an elaborate system of estimated "circuity factors" documented in the Census report, Travel During 1977. This measure of distance was found to be more reliable than traveler estimates or atlas listings. Minor adjustments were made in PICADAD distances by different modes to make them comparable. Automobile travel time was defined as PICADAD distance multiplied by an assumed average speed of 50 mph. This is in-car time. After consulting information on 1977 gasoline cost and fleet fuel economy, automobile cost was defined as distance times \$0.05/mile (gasoline plus maintenance costs) plus \$25/350 miles (overnight costs) plus toll costs. This formula may seem arbitrary, but experience with the model has shown that reasonable modifications of each component have little effect on coefficient estimates or forecasts. Intercity mode choice is not very sensitive to minor changes in automobile costs.

The basic utility function is defined in simple terms. It is a linear combination of cost, line-haul time, waiting time, and access for each mode.

In certain obvious cases some modes were eliminated as alternatives (i.e., assigned a probability of zero). In 1977 there was no rail service to Las Vegas, so the rail alternative was eliminated for trips to and from Las Vegas. On several other routes, the only rail service was connecting service between routes served once daily. A train trip from Cleveland to Columbus, for instance, would take 40 h 25 min (via Chicago), while a bus trip would last 2 h 50 min. Rail trips along these routes were also eliminated as alternatives. Automobile, bus, and rail travel was precluded for trips to Honolulu. Automobile was eliminated as an alternative if the travel party did not own a car, or if no member held a driver's license. Finally, air, bus, and rail were eliminated as alternatives for outdoor recreational trips. Because these destinations were not served by common carriers and automobile storage space was necessary for equipment and luggage, virtually every sampled outdoor recreational trip was an automobile trip. Almost all of the few exceptions were children traveling by bus in large groups. Captive Honolulu and outdoor recreation trips were retained in the sample so that it would not be necessary to adjust forecasts for these two groups. All the categories of eliminated alternatives together represent 9 percent of total alternatives.

The factors of automobile availability (automobile ownership and possession of a driver's license) and outdoor recreation (one category of trip purpose) represent the first allowance for characteristics of the travel party. Other relevant characteristics that do not preclude alternatives but still affect the choice will be discussed later.

The data below summarize the simple model of mode choice outlined so far:

$$U_m = aC_m + bYT_m + cY/2F_m + dYA_m + e_m \quad (1)$$

where

U_m = "utility" of mode M,
a, b, c, d, e_m = coefficients to be estimated,
C_m = cost of mode M,
T_m = travel time of mode M,
F_m = frequency of mode M (average departures per hour),
A_m = access of mode M (miles to common carrier terminal),
Y = family income/2000,
N = number of observations,
L = log-likelihood value,
L_0 = pre-calibration log-likelihood value,
ρ^2 = 1 - L/L_0, a measure of goodness of fit,

and

P_{AUTO} = 0 for (1) (2),
P_{AIR} = 0 for (4),
P_{BUS} = 0 for (2) (4), and
P_{RAIL} = 0 for (2) (3) (4) [(1) = the household owned no car, or no member of the travel party held a driver's license; (2) = Honolulu-San Francisco, Honolulu-Los Angeles, Honolulu-Chicago; (3) = Las Vegas-Los Angeles, Las Vegas-Chicago, Las Vegas-New York, Los Angeles-Bakersfield, Knoxville-Chattanooga (no service); Cleveland-Pittsburgh, Cleveland-Columbus, Cincinnati-Columbus, Charlotte-Columbia, Los Angeles-Bakersfield (connection); and (4) = outdoor recreation trips].

There are four modal attributes and four classes of trips where alternatives have been eliminated.

Figure 1 shows the results of calibration of the logit model for the two samples derived from the NTS. The coefficients represent the traveler's subjective weighting of each unit of a service attribute. The ratio of these coefficients repre-

Figure 1. Model I.

$$U_m = aC_m + bYT_m + cY/2F_m + dYA_m + e_m$$

Group I			Group II		
N = 1658			N = 1062		
a = -0.0161	(- 5.65)	(COST)	a = - 0.0036	(- 4.22)	(COST)
b = -0.0240	(-12.66)	(TIME)	b = - 0.0050	(-11.63)	(TIME)
c = -0.0055	(- 1.81)	(WAIT)	c = - 0.0050	(- 1.37)	(WAIT)
d = -0.0007	(- 1.66)	(ACCESS)	d = - 0.0005	(- 1.05)	(ACCESS)
e_{AIR} = -2.700	(-14.60)		e_{AIR} = -0.582	(- 2.03)	
e_{BUS} = -2.552	(-14.32)		e_{BUS} = -2.728	(- 8.75)	
e_{RAIL} = -3.027	(-16.86)		e_{RAIL} = -2.1850	(- 7.55)	
L = -932.6, L_0 = -1338.0, ρ^2 = .303			L = -585.5, L_0 = -861.9, ρ^2 = .321		
82.7% correctly classified			76.4% correctly classified		

(pseudo t-ratios in parentheses)

sents the relative importance of the units. The quotient of the time (income hours) and access (income miles) coefficients, for example, represents an imputed time expenditure equivalent to each mile of access. The t-ratio determines whether each coefficient is statistically significant. The ρ^2 statistic is a measure of goodness of fit, how well the model explains observed choice. As a rule of thumb, ρ^2 values above 0.2 represent an excellent fit.

In both samples, every coefficient has the expected sign. This is uncommon in logit models of mode choice. The cost and time coefficients in both samples are statistically significant. The wait and access coefficients are significant at the 0.05 level (one-tailed test) in the first sample, statistically insignificant in the second sample. The goodness-of-fit measures are unusually large. By prevailing standards of logit mode choice modeling, this model appears effective.

The marginal significance of access is perhaps attributable to the fact that many trips, especially business trips, do not originate at home. Also, a specific value of access distance may encompass wide variations in access time and cost, depending in part on the access mode. Similarly, a specific "wait" value represents quite different scheduling delays. Even if there is only one train per day from San Francisco to Sacramento that arrives at 2:40 p.m., this represents no inconvenience to the traveler who wants to arrive at 2:40 p.m.

All of the common carriers show constant terms significantly different from the automobile base. Even when the effects of cost, line-haul time, and waiting time are allowed for, automobile is still favored. Factors omitted from the model (e.g., use of a car at destination, instant free egress, convenience in carrying luggage, etc.) in aggregate strongly favor automobile use. In the first sample differences among the common carriers are small, but in the second sample, where longer trips predominate, air travel is much less disadvantageous (relative to automobile travel) than bus or rail travel. The values of the constant terms relative to the coefficients of the modal attributes seem substantial.

The values of the attribute coefficients are smaller for the second sample than the first. The second sample included more long trips, and each trip was weighted by distance. If either of these factors is eliminated, the coefficients assume intermediate values. Longer trips are associated with more expense, more time, and less frequency. One interpretation of this phenomenon is that the difference between $10 and $11 is subjectively greater than the difference between $100 and $101. The model as specified assumes the contrary, constant marginal utility.

Much of the research through this section has been directed toward reducing the complex phenomenon of intercity mode choice to a tractible one embodied in a simple model based on explicit assumptions. The next section examines the consequences of varying some of these assumptions.

MODEL VARIATIONS

The fundamental assumption in the basic model outlined above is that all travel parties "think the same way," i.e., they all perceive and weight each modal attribute according to the same utility function. This is no more than a fiction, but the success of the basic model suggests that it is not absolutely necessary to include a large number of traveler characteristics in the model to produce a workable result.

Some of the traveler characteristics included in the basic model deserve closer examination. The value of time clearly varies with family income; the explanatory power of the model improves markedly when line-haul time, frequency, and access are scaled accordingly. Yet it seems unlikely that different family subgroups (e.g., husband, wife, husband-wife, parent-children, children, and complete family) manifest the same relationship between family income and the value of time. However, stratifying by subgroup to incorporate travel party would strain data resources and weaken the behavioral framework. It was thought that much of the residual variation in the value of time was captured by another variable: trip purpose. For instance, business trips tend to be associated with lone travelers and a characteristic relationship between family income and the value of time. Trip purpose is also important in its own right.

The NTS enumerates nine different trip purposes but these resolve themselves into four basic categories: business, social, entertainment, and outdoor recreation. Virtually all outdoor recreation trips were by automobile, so this trip purpose was incorporated into earlier models by eliminating alternatives. The components of the remaining groups are (a) business and convention (business), (b) visiting relatives or friends and personal and family affairs (social), and (c) entertainment and sightseeing (entertainment).

Model II stratifies the first sample by trip purpose and calibrates each subsample separately. The most prominent characteristic (Figure 2) is the submodels' similarity to each other and to Model I. The attribute coefficients remain the same order of magnitude. Each common carrier displays a large, statistically significant negative coefficient compared with automobile. The goodness-of-fit is uniform.

Differences seem to conform to common-sense perceptions of trip purpose. The cost coefficient is much higher for business travel than other travel. One might expect that cost would be a less important factor in business travel because often the business, not the traveler, pays for the trip. But it is also true that the business, not the traveler, chooses the mode. If the business makes the mode choice, it is no more likely to downgrade travel costs (for which it is paying) than it is to downgrade the value of the traveler's time (for which it is probably paying). (This assumes that the business is a decision-making entity distinct from the traveler.) Moreover, business travel may entail greater actual costs than coach fare model estimates indicate. Also, per-mile automobile reimbursement far exceeds out-of-pocket costs. The model probably underestimates business travel costs relative to other trip purposes, causing a larger cost coefficient.

The travel time coefficient is much smaller, although still negative, for entertainment trips. This is hardly surprising, since sightseeing travelers may actually value time spent en route, unlike destination-oriented business and social travelers. For some travelers, getting there is half the fun. The extreme case of travel for the sake of travel are trips by railroad afficionados, who may not even disembark at the destination before returning.

The waiting time (frequency) coefficient is not statistically significant for entertainment and social travelers. Only business travelers are likely to have to arrive at a destination at a specific time of day. For business travelers, waiting time is comparable with travel time. Others traveling on weekends or on vacation can be more flexible.

Access (i.e., distance from home to common-carrier terminal) appears to be less important to business travelers, probably because many business trips originate outside the home.

The air constant is smaller for business trips than others. Some of the factors making air travel less attractive than automobile travel (e.g., use of the car at the destination and convenience in carrying luggage) are not important to business travelers. The bus coefficient is smaller for entertainment trips. This may be attributable to nonscheduled bus package tours or charters for sightseeing trips that reduce the cost and inconvenience of bus travel. Charter trips account for about one-sixth of all intercity bus travel. The cost of these trips may be significantly overestimated in the model, which is based on scheduled bus tariffs.

Although they bear the proper signs, the modal attribute coefficients in the entertainment model are not statistically significant. This is perhaps attributable to the unusually small sample size, less than 200 trips. Only seven of these trips were by air and only eight were by rail.

Coefficient differences for stratified samples cannot always be explained in behavioral terms. The ability to do this provides additional confidence in the model.

The next travel party characteristic to be considered is a function of the number of miles driven by each travel party member annually. The hypothesis is that, if the travel party contains no one who drives a great deal, driving will be perceived as a burden and the group would be more likely to travel by common carrier. Some people (e.g., the elderly or students) hold driver's licenses and have access to automobiles but are generally reluctant to drive. There is some circularity in this definition, but, for most people, intercity mileage is a small percentage of the total.

The basic variable is the annual mileage of the most experienced driver in the travel party. There are many different ways in which the variable might be incorporated into the model. It was decided to stratify the model, isolating those trips made by travel parties without any heavy drivers. The median mileage for licensed drivers is about 20 miles/day. This was chosen as the threshold. About 15 percent of the sampled trips included only light drivers, or only light drivers and non-drivers, owning cars.

Model III (Figure 3) is a calibration of the base model for the subsample of light drivers. The results are somewhat confusing, since the supposed aversion to cars can interact with the wait and access variables whose automobile values are zero and common-carrier values positive. The model clearly has less explanatory power when calibrated for this group. But the common-carrier modal constants are only about half as large here. Driving experience does seem to have some effect on the set of intangibles embodied in the modal coefficients, including driving attitudes. But the relationship is difficult to model accurately.

Some other readily available variables whose

Figure 2. Model II.

$$U_m = aC_m + bYT_m + cY/2F_m + dYA_m + e_m$$

BUSINESS		SOCIAL		ENTERTAINMENT	
N = 550		N = 717		N = 191	
a = -0.0328	(- 4.06)	a = -0.0111	(- 2.62)	a = -0.0111	(- 1.03)
b = -0.0200	(- 5.55)	b = -0.0238	(- 8.01)	b = -0.0079	(- 1.22)
c = -0.0244	(- 2.54)	c = -0.0090	(- 1.51)	c = -0.0030	(0.56)
d = -0.0006	(- 1.04)	d = -0.0016	(- 1.55)	d = -0.0015	(- 0.89)
e_{AIR} = -1.313	(- 3.66)	e_{AIR} = -3.166	(-11.49)	e_{AIR} = -3.435	(- 5.98)
e_{BUS} = -3.159	(-8.57)	e_{BUS} = -3.016	(-11.47)	e_{BUS} = -1.797	(- 3.69)
e_{RAIL} = -2.461	(-7.15)	e_{RAIL} = -2.828	(-10.97)	e_{RAIL} = -3.288	(- 6.35)
ρ^2 = .299		ρ^2 = .306		ρ^2 = .301	

Figure 3. Model III. $L_m = aC_m + bYT_m + cY/2F_m + YA_m + e_m$

LIGHT DRIVERS			TOTAL		
N	= 233		N	= 1658	
a	= -0.0129	(- 1.96)	a	= -0.0161	(- 5.65)
b	= -0.0167	(- 3.49)	b	= -0.0240	(-12.66)
c	= -0.0246	(- 2.15)	c	= -0.0055	(- 1.81)
d	= -0.0003	(- 0.39)	d	= -0.0007	(- 1.66)
e_{AIR}	= -1.995	(- 5.19)	e_{AIR}	= -2.700	(-14.60)
e_{BUS}	= -1.137	(- 3.30)	e_{BUS}	= -2.552	(-14.32)
e_{RAIL}	= -1.253	(- 3.62)	e_{RAIL}	= -3.027	(-16.86)
ρ^2	= .140		ρ^2	= .303	

relationship to mode choice in intercity travel might be significant, aside from those considered above, include the automobile needs (e.g., journey to work) of household members not in the travel party, the existence of secondary destinations along the travel route, the gasoline mileage and condition of available family automobiles, the availability of non-household drivers and vehicles, general discounts based on travel party composition, etc.

As was mentioned above, much of the data for the model was collected from outside the NTS. This limited the data base to trips along a relatively small set of routes. To extend the model to the national sample and make it applicable to intercity travel in general, it is necessary to generate service information from the NTS itself.

A first effort toward this goal was made by regressing cost and time for air and bus service against distance, and replacing the actual values with the fitted values. R^2 was 0.97 or above in each case and the regression coefficients were in accord with prior expectations. By using the fitted values in the model, the logit coefficients were largely unchanged and goodness of fit actually improved slightly. These results make a national model of mode choice in intercity travel appear feasible.

This section was intended to explore some of the ways in which the basic model can be modified and strengthened by relaxing some of the assumptions inherent in the model. Stratification by trip purpose seems to be a simple and instructive adjunct to the basic model. The next section reverts to the basic model to demonstrate its applications to forecasting and policymaking.

APPLICATIONS

The model of mode choice in intercity travel developed here can be applied to a wide range of forecasting problems. Basically, a forecast of mode shift can be derived for any posited change in circumstances by changing the values of affected variables in the model and reestimating what choices would be made under the new circumstances. A change in highway speed limits, for example, would manifest itself through different automobile travel times and bus travel times. Forecasts will be accurate if the model explains observed behavior well and if the posited changes have a clear, direct effect on elements of the model (and an insignificant effect on factors omitted from the model).

The means of estimating changes in modal demand is the "success matrix." The logit model yields the probability of choosing each mode for each trip in the sample. The success matrix is the product of the choice matrix, with ones for the mode chosen and zeros for all other modes, and the matrix of estimated probabilities. If the model is applied to a different set of values, e.g., higher automobile and bus travel times, the sum of all the estimated probabilities will change. The ratio of the new sum to the old sum is the forecast.

To illustrate possible applications, various scenarios were applied to the model by using the first sample that included 46 routes chosen on the basis of sampled trips. The model is used here to estimate the change in the number of trips (not passengers or passenger miles) on these 46 routes (which are by no means representative of all intercity travel in the United States) for different modes, assuming that total demand remains constant. These caveats should be kept in mind as the forecasts are examined.

Table 1 shows the impact of six different scenarios on mode choice. The figures generally conform to common-sense expectations about what the effects of the posited scenarios might be. Perhaps the most important finding is the minimal mode shift arising from a significant gasoline price rise. Of course, the real price of gasoline has risen by more than 50 percent since the sample was taken in 1977.

Forecasting can also be applied to particular corridors (Table 2). The latter group of scenarios is applied to the corridor between Sacramento and San Diego, including San Francisco and Los Angeles. The third column suggests that the low present speed limit, partly justified on fuel economy grounds, might be a spur to air travel and therefore have perverse effects on fuel savings. In general, travel time appears to be a very important factor in mode choice.

The model can also be applied to specific routes. For 10 routes, the sample size is more than 200 trips. Samples of this size could support reasonably precise forecasting. Because most of this large-sample data became available only recently, route-by-route results cannot be reported here.

Elasticities were derived for the complete sample as well as for the Northeast and Southwest Corridors (Table 3). Unlike reported elasticities for many other intercity travel models, these are consistent with expectations. The cost elasticities, for instance, are all less than one. As expected, automobile travel is more sensitive to travel time than cost changes. Rail sensitivity to frequency changes seems fairly high, but apparently this is

Table 1. Impact analysis: group 1.

	Scenario					
Mode	1[a]	2[b]	3[c]	4[d]	5[e]	6[f]
Automobile	-1%	-2%	-0.3%	-2%	-0.1%	+0.1%
Air	+7%	+5%	-0.3%	-5%	-0.4%	+0.1%
Bus	-2%	+4%	-0.4%	+27%	+2.3%	+0.2%
Rail	-2%	+4%	+7.4%	-4%	-0.5%	-2.4%

[a]Family income rises 10 percent.
[b]Gasoline costs rise 50 percent.
[c]Frequency of rail service outside the northeast doubles.
[d]80-mph speed limit for buses.
[e]Bus fares decline by 10 percent.
[f]Rail fares rise by 10 percent.

Table 2. Impact analysis: Southwest Corridor.

	Scenario			
Mode	1[a]	2[b]	3[c]	4[d]
Automobile	-1%	-7%	+4%	-0.4%
Air	+4%	+19%	-19%	-0.7%
Bus	+3%	+20%	+7%	-1.1%
Rail	+3%	+17%	-9%	+18.2%

[a]Two cents per mile toll for automobiles.
[b]Two-h increase in automobile travel time.
[c]70-mph speed limit for automobiles and buses.
[d]Rail frequency triples.

Table 3. Elasticities for group 1 and corridors.

Mode	Cost	Time	Wait	Access
Group 1				
Automobile	-0.076	-0.303	0.0	0.0
Air	-0.618	-0.159	-0.048	-0.107
Bus	-0.321	-1.101	-0.054	-0.061
Rail	-0.373	-0.251	-0.463	-0.100
Northeast Corridor				
Automobile	-0.112	-0.555	0.0	0.0
Air	-0.538	-0.163	-0.031	-0.108
Bus	-0.267	-0.954	-0.022	-0.062
Rail	-0.315	-0.825	-0.050	-0.059
Southwest Corridor				
Automobile	-0.072	-0.292	0.0	0.0
Air	-0.359	-0.164	-0.023	-0.141
Bus	-0.311	-0.209	-0.038	-0.070
Rail	-0.361	-0.335	-0.419	-0.168

less true in the Northeast Corridor where service is already competitive. On the other hand, travel time is a more important factor in rail demand in the Northeast.

As noted above, these sensitivity analyses can be performed for any scenario affecting model variables. The model is sensitive to changes in various costs, travel time, frequency, income, access, service availability, and automobile ownership. With little additional effort, the changes in demand can be expressed in policy-relevant terms such as energy saved, tax revenues earned, change in vehicle miles of travel, and change in transportation expenditures.

Various technical improvements, such as sample expansion and sensitivity to total demand, have been undertaken to expand the applicability and usefulness of the model. Even in current form, it can be a useful forecasting and policy tool.

ACKNOWLEDGMENT

This research was performed under the sponsorship of the U.S. Department of Transportation. I gratefully acknowledge the guidance of Carl N. Swerdloff of the Office of the Secretary of Transportation.

REFERENCES

1. P. Stopher and J. Prashker. Intercity Passenger Forecasting: The Use of Current Travel Forecasting Procedures. Proc., Transportation Research Forum, 1976, pp. 67-75.
2. P. Stopher, J. Prashker, B. Smith, and E. Pas. Final Report to Amtrak on Intercity Passenger Forecasting from the National Travel Survey Data of 1972. Northwestern Univ., Evanston, IL, 1976.

Publication of this paper sponsored by Committee on Computer Graphics and Interactive Computing.

Interactive UTPS: Implementation Under a Timesharing Environment

JEROME M. LUTIN AND MATTHEW LIOTINE

This paper reports on the development of interactive computer programs for the Urban Mass Transportation Administration's Urban Transportation Planning Systems (UTPS). The programs, originally designed to run under an IBM 360 or 370 OS environment, were executed under a conversational monitor system (CMS) timesharing environment. The aim was to reduce turnaround time and explore future interactive capabilities of the programs. Interactive versions of programs INET, UPATH, UPSUM, ULOAD, UROAD, NAG, UMATRIX, UFIT, and ULOGIT were developed. The paper describes the process involved in creating CMS exec programs to control the program compilation and data set manipulation without any job control steps. Each UTPS program exec is described along with other supporting software that was developed. Finally, a summary of the problems encountered in transforming the software and data files from CS to CMS is presented.

This paper summarizes the development of an interactive version of several Urban Transportation Planning System (UTPS) computer programs. UTPS is a collection of computerized and manual techniques to aid planners in the assessment of urban transportation systems. It was developed and maintained by the Urban Mass Transportation Administration (UMTA) and the Federal Highway Administration (FHWA). The computerized element of UTPS is a battery of computer programs designed to operate on an IBM 360/370 computer system. These programs were designed to operate under batch processing (1). Part of this research focused on the application of these programs in an interactive timesharing environment. The aim was to facilitate an advanced UTPS training course, where the actual use of the programs would be provided.

The course was held at Princeton University in March 1980 and was taught as an advanced use course for transportation professionals with previous exposure to UTPS programs as a prerequisite. The course emphasized a more systematic analysis of transportation system management (TSM) alternatives through the use of existing and newer UTPS programs. The programs used were ULOGIT, UFIT, INET, UPSUM, UMATRIX, UPATH, ULOAD, NAG, and URCAD (2).

The course described a typical corridor analysis for a large metropolitan area and involved intensive use of the previously mentioned UTPS programs not covered in the existing one-week course. The case study was broken down into three phases:

1. Phase I--Build and update networks and build paths and skims,
2. Phase II--Calibrate and apply mode split model and assign transit trips, and
3. Phase III--Prepare highway subarea and assign all trips.

A brief description of the function of each UTPS program set up on the interactive system is described below:

1. INET--Transit network builder. Takes as input a description of the highway network, transit routes, and links. Produces as output a transit data base, network files, and reports on transit supply parameters.
2. UPATH--Transit path finders. Takes network files from INET as input. Produces weighted impedance paths via transit between zones. Produces file of minimum path trees, and optional interzonal fare and distance matrices.
3. UPSUM--Impedance summarizer. Takes path file produced by UPATH and "skims" impedances. Produces output files containing impedance matrices of transit wait time, transit in-vehicle time, numbers of transfers, and total transit travel time.
4. UMATRIX--Matrix manipulator. Takes as input a variety of data set formats and performs user-specified arithmetic and logical operations. Can be used to factor trip tables and apply demand model formulas to produce transit trip tables.
5. ULOAD--Transit assignment program. Takes as input transit trip tables and interzonal transit paths. Assigns transit trips to system to obtain transit line volumes. Assignment rules can be varied by user.
6. NAG--Network aggregation program. Takes network files and trip tables as input. Allows user to analyze a portion or "window" of the network in detail while reducing detail of network outside the window. Helps simplify output and reduce cost of analysis. Produces aggregated files as output.
7. UROAD--Highway network model. Takes as input an historical record file of the highway network produced by program HR or HNET. Calculates highway travel times between zones. UROAD also assigns automobile trip tables to the highway network to produce link volume reports. Capacity restrained and stochastic traffic assignments can be performed. Output includes reports and updated network files with speeds and traffic volumes.
8. UFIT--Demand model calibration. Takes as input card image files or binary calibration files and performs linear least-squares regressions to fit models. User can establish conditional expressions to screen observations and constrain parameters.
9. ULOGIT--Logit model calibration. Takes as input a binary calibration file of observations of individual user (e.g., disaggregate) mode-choice data. Estimates parameters to fit observed mode choice to an S-shaped (logit) curve. Used to calibrate disaggregate multinomial logit mode split models.

A network for the Shirley Highway Corridor in Virginia, developed at UMTA, was used for the case-study problem. The network data were shipped to Princeton on tape, and the Princeton APL graphics system was used to prepare master maps of the corridor on mylar film. Zone centroids, load nodes, access links, and highway links were plotted with distinguishing graphical conventions. Centroid and node numbers were superimposed. Two maps were produced, one showing the entire Shirley Highway Corridor case-study area, and a second showing downtown southwest Washington at an enlarged scale. Each team received several black-line copies of the maps to lay out transit alternatives.

The course attendees were required to make changes to a transit system network in order to achieve better operational characteristics than its original ones. Within the larger transportation network, a detailed section of the highway network was focused on to evaluate the impacts of the changes previously made. The course emphasized the instruction of newer UTPS programs as opposed to the older programs, specifically the use of INET and NAG in network analysis, the use of ULOGIT and UFIT in the calibration of demand models, and the use of a new version of UMATRIX. The actual interactive use of these programs was featured that had never before been incorporated in any previous UTPS training session. This facilitated actual program use during the session, and quick turnaround time for output.

OBJECTIVES

There were several objectives sought in the development of software to execute the UTPS programs under a time-sharing environment. First, the software would have to allow the user to interact with each program at a CRT terminal and would enable the user to operate each program with a minimum amount of knowledge of job control language and data set manipulation. The turnaround time and output retrieval would be quick. The user would be able to examine the results of a program run at the terminal as well as in the form of a paper printout.

The end result of attaining these objectives was a highly intelligent CMS exec program that would function with a minimum amount of computer knowledge on the part of the user. This would allow more concentration on the use of UTPS for a particular analysis without coding job control language (JCL) steps and little use of the conversational monitor system (CMS) language. The following sections describe in greater detail the execs that were developed.

CMS EXECS

The IBM Virtual Machine Facility 1370 (VM/370) is a system control program that controls "virtual machines". A virtual machine is the functional equivalent of a real computer whereby the user can control its operation from a terminal using a command language. In effect, the computer simulates for each user an entire computer system, which appears dedicated to the user. The language used to operate UTPS under this environment is called the conversational monitor system, or CMS (3).

CMS operates under yet a higher command language, the Control Program (CP). CP controls the resources of the physical computer machine and also manages the communications among several virtual machines and between a virtual machine and the physical or "real" system. CMS is the conversational operating system designed specifically to run under CP. It can simulate many of the functions of the IBM Operating System (OS).

The file is an essential unit of data in the CMS system. CMS disk files are unique to the CMS system and cannot be read or written using other operating systems. CMS files are named according to a file identifier consisting of three fields: a filename, filetype, and filemode. CMS files are written on disk in 800-byte physical blocks, regardless of whether they have fixed or variable length records.

CMS is a language of statements consisting of active verbs and nouns. An exec is a CMS file that contains many of these executable statements instead of data items. The statements may be CMS or CP commands or exec control statements. The execution can be conditionally controlled, have variables, and may expect arguments to be passed to it. In its most complex form, an exec can contain thousands of records and may resemble a program written in a high-level programming language. It was in this form that the CMS exec was used to operate UTPS.

Under CMS, it is possible to execute many OS language processors: Assembler, VS Basic, CS FORTRAN IV, OS COBOL, and OS PL/1. This enabled the execution of UTPS, whose programs are mostly written in FORTRAN IV, but which does have some subroutines written in Assembler and COBOL. By using CMS, one can assemble and invoke compilers by using special commands. Thus, a typical UTPS program such as UPATH could be implemented in the following fashion. First, a previously compiled object module (a machine language version of the program) is put through a linkage editor loader (software that assigns the program to certain memory addresses in the system) using CMS commands. File definitions must have been previously made for all input and output files, similar to JCL file definitions. This step produces a load module, or load program, which is yet another form of the original program. Once a load program is present in a virtual machine, it could be executed using a CMS command. A CMS exec was written for each UTPS program that was highly robust in that it made this process virtually invisible to the user. The exec performed this operation, overseeing the data file manipulation and program execution.

SPECIFICATIONS AND PROGRAM DESCRIPTIONS

All of the UTPS programs have some features in common. A control card file is needed as input along with other input files. As output, a printout file is produced along with other files. The execs had to allow the user the option of naming all input and output files or use default names supplied by the exec. A default-naming convention was developed and is described in the following section. The execs also had to allow the user to make changes to the control card file interactively, and to view the printout at the terminal screen before deciding whether a hard paper copy is to be printed.

The execs also had to deal with user errors. Typing mistakes at a terminal keyboard are inevitable and thus the execs had to alert the user of the error. In addition, a feature was needed to allow the user to exit the exec at any point and to provide the option of deciding whether to proceed with the program running or to revert back in the exec to make additional changes.

The execs written for each UTPS program were somewhat similar in structure, but were individually designed to accommodate the unique features of each program. Some standard subexecs were written to perform functions required by most of the programs. The following are brief descriptions of each UTPS program exec and the specifications required of it.

1. INET. This program operates in two modes: update and build. In the build mode, the program reads in a historical record (HR) file and, with "NET=F" on the control card, produces a transit data base (TDB) file. In the update mode, the program expects only the TDB file and produces a new or updated TDB file in subsequent runs. In either mode, if NET=F is specified, five new files are produced for use by other programs. Thus, the exec for INET queries the user as to whether he or she wishes to completely build or update a TDB. It also has the capability of examining the NET parameter to see if NET files are to be produced, so that the appropriate file definitions could be called. In addition, the exec accommodates a file of ROUTE cards and allows the user to make changes to these cards if necessary.

2. UPATH. This exec was not very complex in comparison with INET and was designed to resemble a canned process. The exec expects four NET files from a previous INET run, with the default naming of these files unchanged from INET. The exec allows the user to name the output path file and the two non-transit link files. It also examines the control card options "DIST", "FARE", and "IMPED", to see if the user wished to output a distance and/or fare impedance matrix.

3. UPSUM. This exec receives the path file from UPATH with the same default name and produces a skim file.

4. UFIT. This exec, like INET, operates in two modes. Program UFIT reads in a binary calibration file compatible with program ULOGIT and from this file creates a new calibration file as output, which is conditional on the specification of the "FILE: keyword in the control card file. Thus, the exec determines whether this keyword is present. Another mode of operation is the case in which the user wishes to create a new calibration file from card images. These images are usually attached to the control card file and are used only if the option "BUILD=T" is specified. The program then ignores any input calibration file and processes only the card images. This feature was compensated for in the exec by again examining the control card file for the BUILD=T specification.

5. ULOGIT. This exec is structurally the simplest since only one file is input in addition to the control card file and no files are output, except printout files. However, a special FORTRAN program was written to produce a prediction success table as an optional output.

6. UMATRIX. Because of the flexibility involved in the names and numbers of input and output files, this exec must search the control card file for certain characteristics. First, the number of input J-files is determined by examining the specifications of files J1 through J8. A J9 specification alerts the exec of an output file. The exec then queries the user for the exact name of each J-file specified in the control cards. Z-files are handled in the same manner. The use of look-up tables can also be accommodated.

7. ULOAD. This exec was designed as an almost canned process, with an extra provision for an input-loaded legs file from a previous ULOAD run and the file of selected volumes conditional on the specification of the "ALINE/CLINE" criteria in the control card file. In addition, the exec can determine if the user specified the "GENT" parameter, eliminating the need for an input J1 trip matrix file.

8. NAG. The main feature of this exec is the detection of "NET" and "GENT" parameters in the control card file that eliminates the need for the user to input an HR file and trip matrix.

9. UROAD. It was decided that this exec would not implement any of the plotting features of program URCAD and would only address the traffic assignment capabilities of the program. For the purpose of this research, the exec only accommodates the insertion of multiple trip tables. Other features will be incorporated later.

10. ULOG. This exec facilitates the printing of the user's log report after a specified number of UTPS program runs. This exec allows the user to specify this number and maintains the log file on disk for subsequent updating.

GENERAL PURPOSE EXECS AND PROGRAMS

As mentioned earlier, several subexecs were written to perform general purpose tasks used by most of the main execs. These execs are listed below and are briefly described.

1. DEEM. This exec resides on the user's disk and obtains the default filemode for files used in the program. As mentioned earlier, CMS file designations consist of three fields: a filename, filetype, and filemode. A default-naming convention was developed whereby the user's disk mode serves as the default filemode for all files. In most cases, files would maintain a filetype of the name of the UTPS program that produced them. Filenames, in most cases, resembled the original filenames contained on the DD statements in the program's catalogued procedure. Thus, for example, a file designated as TDB INET A is a TDB file produced by program INET and resides on a disk in the user's virtual machine that has been accessed as A, which is the default filemode. Route card files and control card files were handled in a different manner as described in the following section.

2. UTPS. This exec links the user's disk to two software disks. The first is a disk containing the execs and load modules. The second disk contains utility software supported by the Princeton University computer system.

3. GETFID. This exec responds to an input file specification by the user and searches all disks in the user's virtual machine for file and verifies whether it can be used.

4. REPLY. This exec simply checked to see if a user's reply was correct or not, i.e., contained no typing errors for a yes or no reply.

5. CPUNIT. This exec is executed when a program begins running and informs the user that this has occurred.

6. CPUFINL. This exec is invoked at the end of the program run and informs the user of the CPU time of the run and the return code if the program terminates abnormally.

7. DISPLAY. This exec takes the printout file and displays it on the terminal screen. The user can specify the pages and report of the printout he or she wishes to view. It also allows the user the option of printing the file on the system printer.

8. ANTEST. This exec is a subexec of GETFID and, like REPLY, examines the user's entry of a filename to see if it was typed correctly.

9. PRMCLR. This exec clears all the file definitions in the virtual machine after the program run.

10. ODSK. This exec is a subexec of GETFID and determines whether an accessed disk can have information written on.

11. ST. This exec intercepts the print file from the real system printer and spools it to the user's disk. The return code is also read from the print file and passed to CPUFINL.

12. STCLEAR. This exec clears the console stack.

Two general purpose programs were written to supplement the above execs. The first was an Assembler program called FIND. A key element in the overall exec structure, FIND was designed to specifically search the control card file for any character string argument. This enabled the execs to determine the required input/output file based on keywords such as NET in INET or BUILD in UFIT.

The second program was a FORTRAN program called REPS2. This program was used in the exec DSPLY and performed a minor task of determining which UTPS reports are available on the printout file.

FINAL EXEC STRUCTURE

The final versions of the execs conformed to the specifications mentioned earlier. Each was designed to guide the user through the program and demanded little knowledge of CMS from the user. Implicit in the design of each exec was the assumption that the user had at least some knowledge of the UTPS program use. A virtual machine is established for the interactive UTPS environment. A typical user would have read or write access to his or her own personal disk where input and output files could be maintained. All of the necessary software, including the program execs and load modules, reside on another disk from which the user can read only, leaving the software protected and intact. A temporary disk is also formatted on which the user may be able to read and write scratch files and temporary data sets. This disk provides space for scratch files and any extra temporary space if required by the user. Since this is a temporary disk, its contents would be wiped out once the user logs off the system.

The virtual machine configuration is established when the user invokes the exec called UTPS. Once this is done, the user can invoke any exec by simply entering the UTPS program's name.

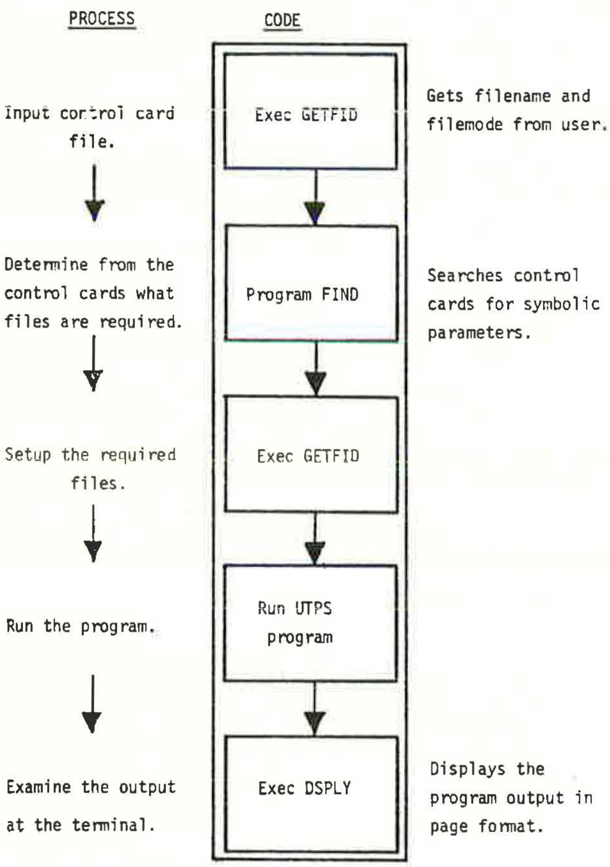
Each exec called several general purpose execs and programs described in the previous section. A general flow-chart for the execs is shown in Figure 1. Each exec follows a process that is summarized as follows:

1. Input Control Card File. Exec GETFID is invoked first, retrieving from the user the name of the control card file specified. The exec then seeks to locate this file in the virtual machine.

2. Determine the Required Files. The control card file is then searched by program FIND for symbolic parameters and keywords that inform the exec what files are to be either read in or written. For example, the presence of the keyword BUILD=T in the INET control card file would inform the exec that five NET files will be written by the program.

3. Set Up the Required Files. Exec GETFID once again is invoked. This time it will ask the user to specify the names of both the input and output files. Input files are then located within the virtual machines. The specified names are inserted in the file definitions so that output files are written with the names given by the user.

Figure 1. General flowchart for UTPS execs.



4. Run the Program. The program's load module is then called and the program is executed.
5. Examine the Output at the Terminal. The output file is retrieved and, through exec DSPLY, is displayed on the terminal screen. The user then has the option of requesting a paper printout of the file.

Mention should be made of the default-naming convention for route card and control card files. Unlike the file convention described earlier, these files assumed a filetype of the UTPS program name by which they were to be used. For control cards, a default file name of CNTRLXX was used. The suffix XX is substituted by a two-digit code from 01 to 99, which is assigned by the user. For example, a control card file might be designated as CNTRL03 UPATH A. The filename of CNTRL03 designates this as a control card file. The filemode A is the default mode of the disk in the user's virtual machine on which this file resides.

The two-digit suffix enables the handling of many sets of control card files on the user's disk and provides the user with a convenient way of identifying them. It also facilitates any easy default way of specifying a particular control card file designation to the exec. By entering "." followed by a two-digit number corresponding to the two-digit numerical suffix in the control card file's filename, the exec will locate this file in the user's virtual machine. For example, if in using program ULOAD, the user responds to the exec's prompting for the name of the control card file by entering ".05", the exec will search for a file designated as CNTRl05 ULOAD A, assuming A is the default filemode. If this file is residing on a disk accessed under a different filemode, say B, then the user needs only to enter ".05 B". Route card files for program INET were handled in a similar manner, with ROUTEXX as the default filename prefix.

UTILITY EXECS

Several utility execs were written during the software development to perform general tasks. One is an exec called SCOPY, which enabled the transfer of files formatted as variable block spanned (VBS) from one disk to another. The normal CMS "COPY" command tends to disrupt the format of these files and thus SCOPY was used in lieu of this command. The original version of SCOPY was written at UMTA.

SUMMARY OF PROBLEMS ENCOUNTERED IN SOFTWARE DEVELOPMENT

Implementing the UTPS programs, which were originally designed to operate in an OS batch environment, in a timesharing environment was not a direct process. Certain inherent OS features of the programs that are not handled by CMS had to be overcome.

A problem arose with respect to data files. It was unclear in the case of files that were originally formatted as VBS as to what the proper data control block (DCB) parameter was required in the CMS file definition. By using the original parameters from the DD card in the catalogued procedure failed in most cases. The problem was resolved by literally guessing at the record format, using a trial-and-error procedure that iterated different permutations of the RECFM parameter until a successful run was obtained. VBS files also presented a problem mentioned earlier, whereby they could not be transferred by using a normal CMS copy command. A special exec was required to copy files.

Problems arose in INET with respect to sorting. Within its internal structure, INET makes several calls to the IBM OS Sort/Merge routine. Under VM/370, this utility is unavailable. A counterpart to this is the CMS SORT utility that is called in lieu of the OS SORT. However, INET uses, in addition to the OS SORT, a UTPS sorting routing called SORT. Problems arose in calls to this routine. It was later discovered that the calls to this routine were identical to the CMS SORT and thus the calls had to be changed.

A problem occurred in ULOAD with respect to the sorting of the loaded legs file. This file is passed through two subroutines, E15 and E35, which are user entries in the OS SORT routine. E15 processes the input loaded legs file for the OS SORT while E35 processes the output file. These routines were designed to process a blocked file. The CMS SORT failed to sort this file due to this characteristic. Thus, E15 and E35 were replaced by two routines, UNBLK and BLK, that, respectively, unblocked and blocked the file.

Program NAG produces an output trip table in a compressed format that deletes rows for which all cells are zero. UROAD did not accept this file. A modification to the file was done by using UMATRIX. UMATRIX will insert a row of zeros where rows are deleted. It was also used to change the file format by using the keyword "OUTPBT".

CONCLUSIONS AND RECOMMENDATIONS

The interactive implementation of UTPS was not a direct process. Before proceeding on this endeavor, a knowledge of the use of the UTPS programs in transportation system planning as well as computer science must be obtained. The transportation planning and UTPS experience enabled the specification of the exec structures. Implementation of the execs involved a substantial degree of knowledge of the IBM VM/370 system, CMS, several OS compilers, and UTPS. Although this process, in general, did not warrant the entering of the UTPS internal source code, in some instances it was unavoidable. Because the execs were developed with the intention of being used as an instructional tool for a one-week course, many of the accessory features of the programs, such as the plotting capabilities, were not incorporated into the execs. The end result was an exec that required little knowledge of CMS and VM/370 on the part of the user but some knowledge of the use of the UTPS programs in transportation planning.

It is recommended that this process be further crystalized and documented so that typical users may find it easier to implement UTPS in a CMS or any other timesharing environment. Interactive computing is now gaining widespread interest, and any interactive capabilities of software would thus make it more attractive.

The interactive software described in this paper has been turned over to the UMTA Office of Planning Methods and Support and is undergoing further development. It is expected that interactive versions of many UTPS programs will be used in further advanced UTPS training sessions.

ACKNOWLEDGMENT

This research was sponsored by UMTA's Office of University Research and Training. We wish to sincerely thank Philip Hughes, Nathaniel Jasper, and Judy Z. Meade of the Office of Policy Research for their support of this work. A large measure of credit for the short-course development goes to Larry Quillian of UMTA's Office of Planning Methods and Support. Ed DeLong, Chief of the Software Support Division, played a major role in helping debug the system. Most of the programming was done by William Collins, a 1981 graduate of Princeton University.

REFERENCES

1. M. Liotine and J.M. Lutin. Report on the Development of the Workshop on Interactive Applications of UMTA/FHWA Planning Tools. Urban Mass Transportation Administration, U.S. Department of Transportation, Sept. 1980.
2. Urban Transportation Planning System Introduction. Urban Mass Transit Administration, U.S. Department of Transportation, Feb. 1980.
3. IBM Virtual Machine Facility 1370: CMS User's Guide, 3rd ed. IBM Corp., White Plains, NY, March 1979.

Publication of this paper sponsored by Committee on Computer Graphics and Interactive Computing.

Interactive Model for Estimating Effects of Housing Policies on Transit Ridership

JEROME M. LUTIN AND BERNARD P. MARKOWICZ

This paper reports on computer graphics developed as part of an interactive computer model designed to assess the impact of housing policies on transit ridership in urban transit corridors. A set of programs was written in APL to implement the model in an interactive computer environment, with computer graphics used for both input and model output. A mode-split model that uses U.S. Bureau of the Census data predicts ridership for the transit line, based on discrete combinations of mode and access mode including walk-and-ride, park-and-ride, kiss-and-ride, and feeder bus. The program permits the analyst to input alternative residential patterns, with respect to location and density, in the transit corridor and to evaluate the effects on transit ridership by comparing various alternative housing policies. Computer graphics are used at two levels. First, as an input mode, graphics allow the planner to create new transit route alignments and station locations by using a screen cursor. The program then models station choice from the zones, on the basis of a number of variables, including driving or walking times to stations, transit fares, line-haul travel times, etc. As an output mode, graphics are used to display socioeconomic data, mode-split results, or any algebraic combinations of input or output data. Different types of graphic displays are used for data presentation at the zone level or station level. Throughout the development of the graphics, special attention was given to the readability of the output. The paper reflects the general effort to produce more visually attractive and commonly understandable outputs. Included in the paper are a description of the program design and organization, examples of graphic output, and a discussion of the ability of the model to provide useful output to policymakers.

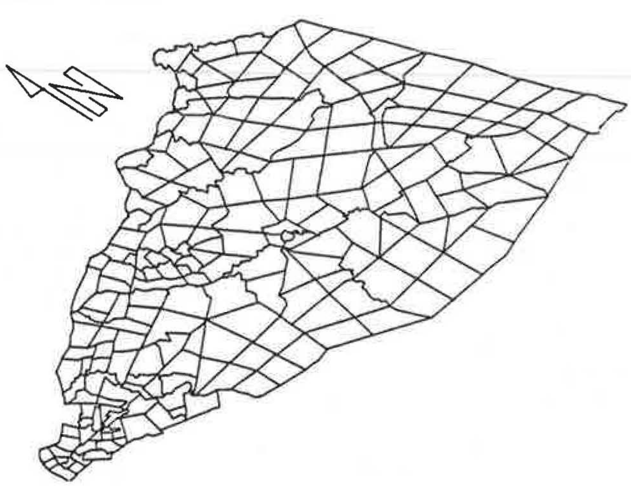
Planners and urban policymakers have long recognized that a strong relationship exists between urban development forms and the existence of rapid transit systems in cities. In recent years, new transit systems have not led to significant positive changes in urban development. It is believed that the existing high level of automobile accessibility tends to obscure the increases in mobility achieved by transit. Many planners and policymakers believe that transit systems can be more effective in meeting the travel needs of the public, more energy efficient, and require less subsidy if land use planning in transit corridors can be coordinated with the planning of the transit system itself.

To achieve better coordination between transit planning and land use planning, the Urban Mass Transportation Administration (UMTA) has been making grants to cities to encourage urban development in transit corridors. However, there are major questions that need to be answered about the kinds of policies to be implemented. Planners need to know, for example, what kinds of housing should be encouraged in transit corridors. Should land close to transit stations be reserved for high-density apartments or be kept open to provide large lots for park-and-ride patrons? Given that land use regulations are difficult to enact and enforce, how does noncompliance with the plan affect the desired result? Because of the many unanswered questions, this research was directed toward the development of some quantitative tools that would provide planners with the ability to determine the likely effects of alternative land use plans on transit ridership.

The objective of the research was to develop a model that would take as input various housing policies and translate these results into transit ridership figures. The model was designed to estimate the proportion of commuters traveling by transit, the mode split, given that population could be clustered at various distances from the transit stops. By changing the location of population clusters, one alters the relative travel times and costs encountered in traveling to both transit stations and the central business district.

To test the model, a case study area was chosen in southern New Jersey. A proposed branch line extension to an existing rail rapid transit system is currently under study, and the corridor it is projected to serve was chosen as a test area for the model. The triangular transit corridor, 30 miles long and 15 miles wide at the maximum, covers parts of Camden and Burlington Counties in New Jersey and includes a population of about 450 000. The initial data set used by the model comprises 60 variables recorded in the 1970 census, population projections for the year 2000, and developable land areas for each of the 116 census tracts that comprise the corridor. For the purposes of the research, some of these tracts were further subdivided into subzones, which increased the total to 212 subareas or zones for analysis. A basic map of the zone boundaries was digitized and stored on disk along with a map of the zone centroids (see Figure 1).

Figure 1. Base map.



INTERACTIVE MODEL STRUCTURE

The nature of the research suggested that a number of alternative policies would be tested. This, coupled with the magnitude of the data base, led the investigators to use an interactive computer approach that would permit quick evaluation of many policy scenarios in a short period of time, with minimum data manipulation.

The interactive system is run under a virtual machine VM370/CMS operating system on an IBM 3033 computer at the Princeton University Computer Center. The model is programmed in VSAPL, enhanced by graphical processors, and uses auxiliary processors interfacing with CP, FORTRAN, and ASSEMBLER. The user accesses the timesharing system via a Tektronix graphics terminal.

The program is composed of a transit line input routine, a housing allocation model, a mode split model, and a routine to produce graphic output. These four routines are managed by a conversational program, called "POLIS", which controls the sequence of model execution and accesses the various routines and subroutines. The POLIS program contains user options to review or send output to a regular line printer, to modify interactively the numerous parameters used in the mode split model, to indicate which parameters bear non-standard values, and to control for valid parameters or command input names.

The transit line input routine allows the user to input a new transit route alignment, to reset the program to a previous alignment, and to add or modify the number and location of stations. The functions of the input routine are to calculate the distance between each zone centroid and each station, to select the station nearest each zone based on the least "weighted" distance to all stations, to create around each station a new 0.5-mile² (0.8-km²) "special development district" zone to be superimposed on the original zones, and to assign to each zone a classification code based on the zone's location relative to both the destination--in this case the Philadelphia central business district (CBD)--and the nearest station. The 0.5-mile² zone is created in order to enable the user to apply special housing allocation policies to those areas within walking distance of the stations. The line input program is functionally divided into three subroutines. The first is designed to gather the input data, the second to perform the computations, and the third to create a new working data set. This data set includes characteristics for the new zones as well as adjusted data for the original zones.

Once the line input routine has been executed, the POLIS program takes control and allows the user to review parameters, produce base data maps, and run the housing allocation model or the mode split model. The housing allocation model simulates a 1980 housing distribution by allocating specific increments of dwelling units to the 1970 base year. A distinction is made between free-market and policy-allocated dwelling units. The policy-allocated number of dwelling units strictly conforms to location and density patterns set by the user to simulate policies to increase development at locations within the corridor. The free-market dwelling units replicate population gains and losses projected by the regional planning commission if no transit-related development were to be induced. Developable land is calculated as a percentage of total vacant land. Housing units are allocated to zones until target densities have been reached, based either on existing density levels, or on the basis of user specified growth policies. Since there is ample vacant land to accommodate growth, the allocation process fills zones according to a priority index set in the transit line routine.

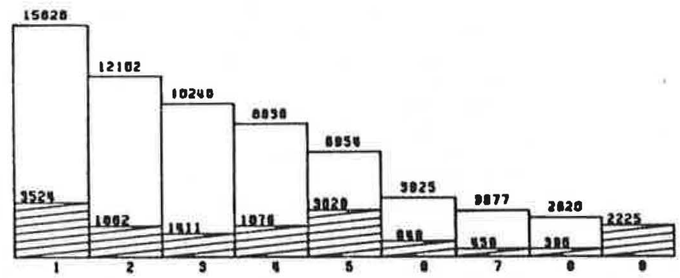
Each zone is assigned an index from 1 to 12. In turn each index number is associated with one of 12 classes of specific zones that share similar locational properties. The allocation model uses this index as a mean for specifying both the density and the priority class of each locational group of zones. The model takes the group of zones with the highest priority index and allocates to those zones a number of dwelling units compatible with the remaining vacant land and the specified target density, but not greater than the pool of dwelling units available for allocation. If the allocation of dwellings to this class of zones exhausts the vacant land, the program goes to the next priority class, and so forth, until the pool of dwelling units is allocated. If the pool to be allocated is greater than the capacity of the developable land, given the user-specified densities, the user is informed and allowed to get back to the top of the routine. Because of the large amount of vacant land in the corridor, this restriction applies only to the very low densities or large increments of dwelling units.

The housing allocation model first asks the user to input the total growth projected for the corridor, then to input the "percent effectiveness" that limits the percentage of vacant land available for policy allocation of housing. The allocation priority for each of the 12 classes of zones and the associated target density in dwelling units per acre are then input. Once the first subroutine has collected the information interactively from the user, a second subroutine computes the allocation and passes the output to a third subroutine that prints out a summary of the output to the screen and prepares a detailed report to be sent to the line printer. The POLIS program again takes control and, at user request, passes to the mode split model.

The mode split model is an eight mode and access mode stochastic choice model. The core of the program is a logit function that calculates the probability of choosing a given mode. The modes are automobile, carpool, express bus, rapid rail via park-and-ride access, rapid rail with kiss-and-ride access, rapid rail with feeder bus access, rapid rail with walk access, and rapid rail with bike access.

Figure 2. Screen summary of mode-split model.

STATION AND CUMULATIVE RIDERSHIP

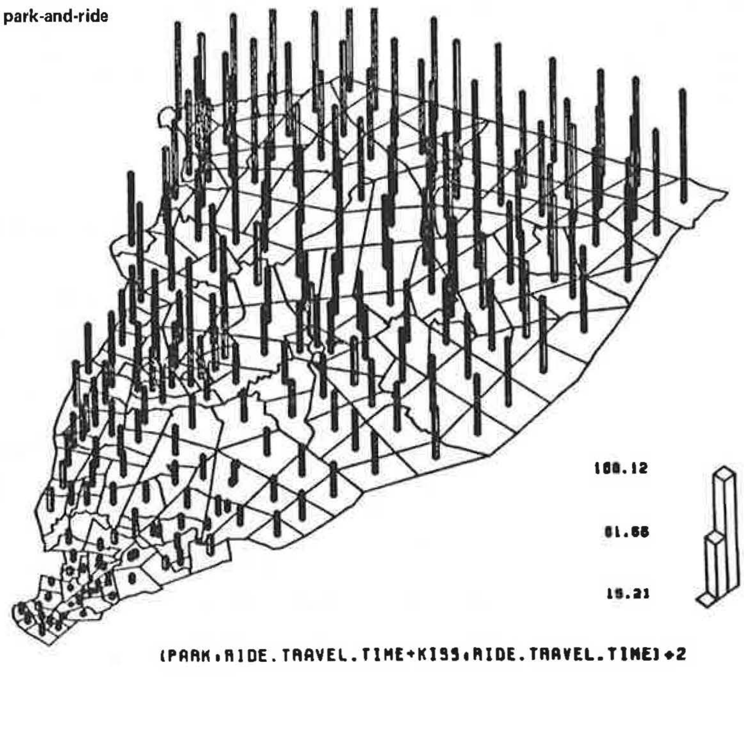


MODE SPLIT | TOTAL

AUTO :	26830	PERCENT :	.44			
CARPOOL :	18953	PERCENT :	.31	COMMUTERS :	61410
EXPR.BUS :	0	PERCENT :	.00		
PARK◦RIDE :	6208	PERCENT :	.10	TRANSIT :	15626
KISS◦RIDE :	5109	PERCENT :	.08		
FEEDER BUS:	2648	PERCENT :	.04	PERCENT :	.25
WALK◦RIDE :	1212	PERCENT :	.02		
BIKE◦RIDE :	450	PERCENT :	.01		

POLIS : INPUT COMMAND FOR :
TRANSIT LINE INPUT (L), THE HOUSING MODEL (H),THE MODE SPLIT MODEL (M)
TO REVIEW PR OUTPUT (R), SEND CASE PAGE (S),CHANGE PARAMETERS (C),
TO DISPLAY MAPS (D) OR TO QUIT ( ).

Figure 3. Screen output: average park-and-ride and kiss-and-ride travel times.



The mode split program calculates the impedance of each commuting trip at the zone level, including (a) travel time spent in vehicle, (b) travel cost (cost in dollars later transformed to income-earning minutes), and (c) excess time--time spent in waiting, transferring to, or accessing a mode. A subroutine makes the translation between real trip characteristics and perceived total impedance. This is done by multiplying travel time, cost, and excess time by weighting coefficients, and summing the terms. This sum, when exponentiated, represents the total trip impedance or disutility. The probability of choosing one mode is the ratio of its utility to the sum of all modal utilities. The zonal mode choice is expressed as the probability that an individual will commute to the CBD, multiplied by the population of the zone, and multiplied by the probability of selecting each mode. The model sends a summary of the corridor mode split to the user's screen, and prepares a detailed report to be sent to the line printer. The screen summary shown in Figure 2 also displays a graphic representation of the cumulative transit line ridership, along with the station loads (the shaded areas).

Following mode split calculation, the user again has the choice to go back to any main routine, to send the output to the printer or to produce computer graphics output. The computer graphics routine allows the user to display a variety of images on the screen, which can be copied via a hard copier or directed to a film plotter or a Calcomp plotter. The basic display includes a three-dimensional bar chart of a variable or combination of variables superimposed on a map of the corridor. The display routine offers a series of catalogued subcommands for plotting zonal mode split, total transit ridership, percentage of mode split, percentage by access modes, ridership versus density, or income. For documentation purposes, the routine also allows the user to plot any zonal census variable. Other subcommands enable the user to graph any variable by station.

BASIC COMPUTER GRAPHICS TOOLS

The program is implemented for use with a Tektronix graphics terminal attached to a Tektronix photo-processing hard copier, and through the main computer to a Calcomp 936 multicolor plotter. The Tektronix screen is composed of a 1023 by 780 matrix of addressable points, which are activated by an auxiliary graphic processor. The two dimensions of the matrix define the X and Y axes in screen coordinates. In addition to addressing screen coordinates, the processor accepts virtual coordinates for which a virtual window has been specified. The virtual window transforms the screen into a new cartesian system of coordinates with user-specified origin and scales along the X and Y axes. The graphics processor enables the programmer to define a straight line segment by two pairs of X and Y coordinates, absolute or virtual, or by two pairs of relative coordinates, each pair specifying the move along each axis from the previous point. Those two properties, the virtual system of coordinates and the relative coordinate system, are the basis for the computer graphics implementation.

The main objective of the graphics was to communicate quickly and efficiently the changes in mode split and transit ridership caused by changes in housing density. Figure 3 shows an example of the type of graphics developed in this case study. The zone boundary map is displayed in a simulated perspective mode. The perspective effect is obtained by rotating the map by an angle selected between 30° and 45°, and by compressing the figure along the Y axis. The first operation (rotation) is done only once, and the map coordinates are stored that way in the computer. The second operation (Y axis compression) is performed each time the map is drawn on the screen or on the plotter. It is achieved by setting a virtual window with dimensions not proportional to the 1023/780 ratio of the screen addressable matrix. If the ratio of the X range to the Y range is 1023 to 1560 for instance, the figure will be compressed along the Y axis and give an impression of perspective. The first advantage of this method over a spatial projection is that the computer is not required to solve complicated equations to locate coordinates. Another very important property is that, although virtually distorted, the computed distance between any two pairs of coordinates remains a scaled distance and not a projected distance.

GRAPHIC INPUT MODE

The graphic input mode is an essential component of the transit line input routine. At the user's request a base map is drawn on the screen along with the planned transit route.

A graphical cross-hair cursor appears on the screen and allows the user to input station locations. After each station input the program responds by drawing the link to the previous station. Each station location is read by the computer as a pair of virtual coordinates. To quit the input

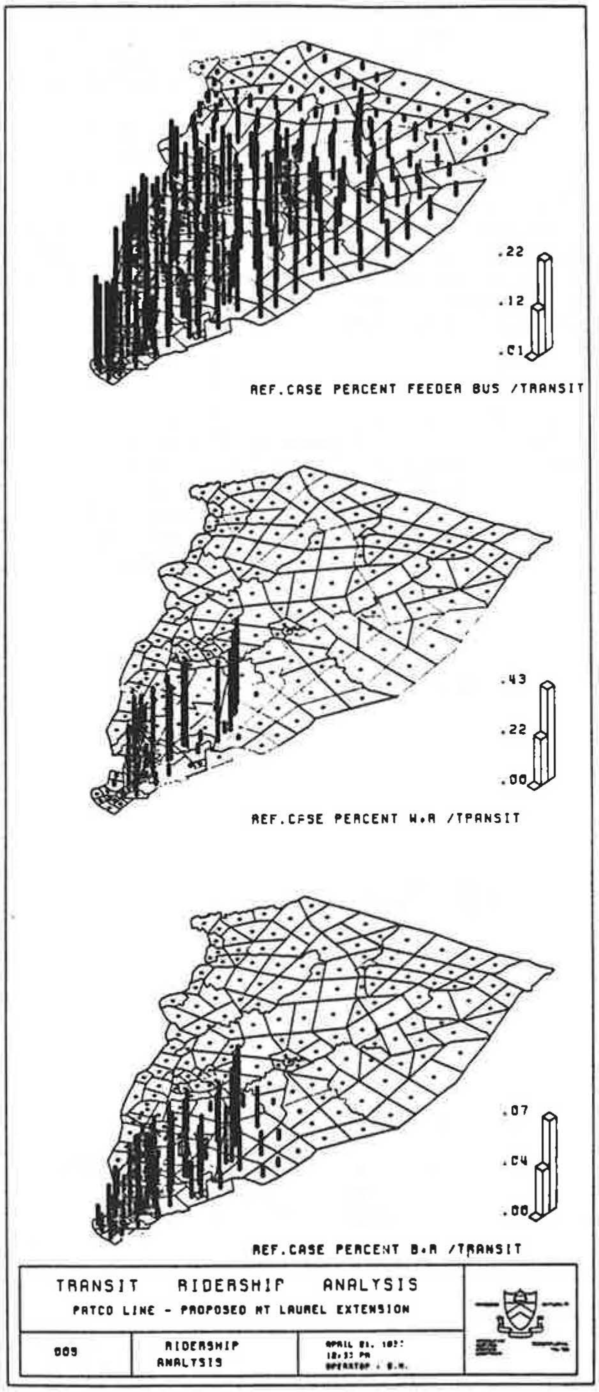
Figure 4. Graphic output mode: list of variables and input commands.

```
D
DISPLAY :INPUT COMMAND FOR :
         LIST AND BRIEF INSTRUCTIONS (L), MAP (M), OR SUB/QUIT ( )
L
CODE AND LIST OF VARIABLES
IN (M), INPUT ANY VALID APL EXPRESSION USING CODE VALUE
FOR NUMBERS OR NUMERICAL EXPRESSIONS, USE THE [   ]

 1  AUTO.DRIVERS          9  TRANSIT.RIDERSHIP  17  PROJECTED.POPULATIO
 2  CARPOOL              10  DWELLING.UNITS     18  AREA.OF.ZONE.(SQ.M.
 3  EXPRESS.BUS          11  POPULATION         19  VACANT.LAND.(ACRES)
 4  PARK∘RIDE            12  FREE.MARKET.ALLOC  20  WEIGHTED.AVG.TIME
 5  KISS∘RIDE            13  POLICY.ALLOCATION  30  WEIGHTED.AVG.COST
 6  FEEDER.BUS           14  TOTAL.COMMUTERS    40  WEIGHTED.AVG.EXC.TI
 7  WALK∘RIDE            15  DESTINATION.PROBAB 50  WEIGHTED.AVG.MILEAG
 8  BIKE∘RIDE            16  MEDIAN.INCOME
       21 - 28 : TRAVEL TIMES BY MODE / 31 - 38 : TRAVEL COST BY MODE
       41 - 48 : EXCESS TIME BY MODE  / 51 - 58 : MILEAGE BY MODE

DISPLAY :INPUT COMMAND FOR :
         LIST AND BRIEF INSTRUCTIONS (L), MAP (M), OR SUB/QUIT ( )
M
INPUT ALGEBRAIC EXPRESSION :
(24+25)÷[2]
PROPOSED TITLE :(PARK∘RIDE.TRAVEL.TIME+KISS∘RIDE.TRAVEL.TIME)÷2
ENTER NEW TITLE OR HIT RETURN IF O.K.
```

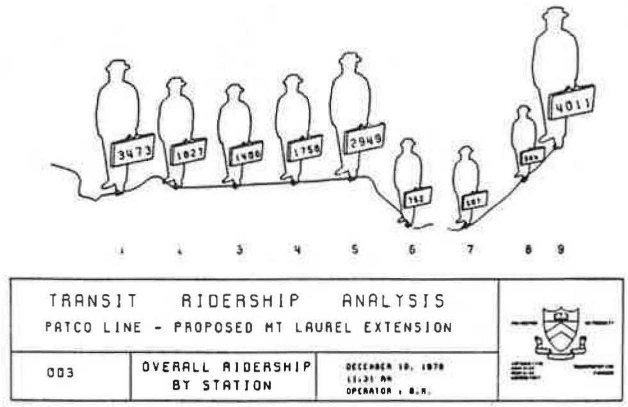
Figure 5. Sample of printed output.



mode, the user presses the Q key to signal that all stations have been input.

In order to create the 0.5-mile² (0.8-km²) special development districts mentioned earlier, the computer has to relate those districts to the original zones they intersect. The problem of boundary recognition is computationally difficult, and through interactive graphics, is left to the user's visual capabilities. A non-distorted portion of the map surrounding the first station is drawn on the screen along with the special development district. The user indicates which zones are included in the development district by pointing at their centroids with the cursor. Pressing a specific key directs

Figure 6. Height of character (commuter) is a function of projected loadings at each station.



the program to go on to the next station. Once all stations have been reviewed, the routine passes to the computing phase.

GRAPHIC OUTPUT MODE

The basic format used in the graphic output mode is a perspective-like map on which an output variable is represented by a vertical rectangular prism for each zone. The prism is located at the zone centroid and its height indicates the value of the variable. A subroutine automatically scales the output according to the absolute maximum and minimum values of the variable. To allow visual comparison between several outputs, the user has the ability to specify virtual maximum and minimum values. As shown in Figure 3, a scale automatically appears on the right of the display.

The plotting of variable size prisms is easily achieved in the relative coordinate system. Each pair of point coordinates (x and y) is expressed relative to the previous point in the array by the distance in units on the x and y axes. Therefore, by modifying the relative y coordinates of only three points in this case one can specify any height for the prism.

The display routine has been designed to allow the user a maximum of freedom. Entering the display program the user can review the list of variables that can be currently accessed as seen in Figure 4. By using the numerical code, the user then inputs any algebraic expression of the variables. Numbers of numerical expressions are expressed in brackets to differentiate them from numerical codes referring to variables. By using its own algebra, the program checks for a number of mistakes in the input expression and, if valid, proposes a title based on the algebraic expression to be shown on the map. The user is allowed to input his or her own title or to keep the proposed one.

Taking advantage of the high resolution of the Calcomp 936 plotter, as well as its increased useful surface, the programmer has the option to group plots by three. Figure 5 shows an example of this format. The size of the plot has been set to be reducible to an 8.50x11-in sheet on a Kodak 150 copier. A scale is automatically calculated and displayed with each map.

Finally, the display routine allows the user to access special graphic devices for station-by-station or housing class representation. Figure 6 shows a map of the proposed rail transit route with stations indicated by diamonds. At each station, a figure of "Mr. Commuter" is drawn. The size of the figure is proportional to the number of workers using the line to commute to work and selecting each station. The number of commuters is shown on the figure's briefcase. This display, while rather whimsical, may prove useful in communicating results to the public. This graphic is processed by using another property of the relative coordinate system. By multiplying all elements of a two-dimensional array of relative coordinates by a single factor, one can enlarge or reduce the object to be represented.

CONCLUSIONS

The software developed for this research takes advantage of the capabilities of interactive computer graphics in several ways. First, the quick turnaround of the system permits the analyst to explore a larger number of alternatives in a given time than is possible for a batch-mode computer model. The ease with which a model run can be performed encourages the user to explore a wide range of solutions and gives the user the opportunity to follow analytical paths that might not otherwise have been pursued. Subsequent runs can be made quickly, so that an idea can be tested while it is still fresh.

Second, the graphic form of the output shows at a glance the results of the model run. By using hard copies made from images on the terminal screen, one can immediately acquire an understanding of how results change by comparing visual outputs from one run to the next. One can pick out major shifts much more quickly from examining visual images than from examining numerical output. In addition, the pictorial quality of the output allows policymakers and non-technical individuals to grasp the implications of the analysis with greater clarity than can be achieved through examining reams of computer printout.

In an academic environment, the use of conversational computer graphics programs allowed undergraduates with little computer training to participate in the research. Undergraduates received 2 h of classroom instruction and then were assigned various policies to test. The easy manipulation of model parameters and sophisticated output gave students the sense of using a powerful lever and increased their motivation for working on the project.

ACKNOWLEDGMENT

This research was partially funded under a contract from the Program of University Research, U.S. Department of Transportation. We wish to thank Robert J. Ravera, Director, Office of University Research, and Richard I. Cohen, Office of Policy and Program Development, Urban Mass Transportation Administration, for their support and encouragement of this project.

Publication of this paper sponsored by Committee on Computer Graphics and Interactive Computing.

Network Simulation Interactive Computer Graphics Program

SHIH-MIAO CHIN AND AMIR EIGER

An overview of the deficiencies of the network simulation (NETSIM) program with regard to data input, data debugging, and analysis of the output is presented. Interactive computer graphic (ICG) enhancements are suggested as measures to eliminate many of the difficulties. The NETSIM/ICG program, which provides ICG capability in input-data preparation, input-data display, and both real-time and passive displays of link-specific measures of effectiveness, is described. The interactive data input, both graphical and keyboard in free format, follows a systematic procedure for obtaining the necessary information needed by NETSIM without reference to the user's manual. By using input-data display and input-data modification capabilities provided by the pre-NETSIM (PRENET) enhancement program, the user can easily comprehend and debug the NETSIM input data. As a consequence, significant reductions in the costs associated with data preprocessing are anticipated. The capabilities of providing both real-time and passive displays enable the user to more easily assimilate the information generated by the NETSIM simulation model and to comprehend the overall operation of the network. Consequently, these programs provide a heuristic approach to determining high-performance solutions at a minimal cost of both personnel and computer time.

As traffic flows through street networks, it experiences periods of congestion that may result from inadequate geometric design, signalization, or simply excessive demand. Traffic-simulation techniques are important tools for the traffic engineer in investigating the impacts of various traffic-control strategies. These simulation experiments can yield an enormous amount of data that could not be obtained in real life for economic or other reasons.

Among the network traffic simulation models, the network simulation (NETSIM) model produced for the Federal Highway Administration has been the most popular. NETSIM is an extension of the UTCS-1/SCOT simulation model, developed originally as an analytical tool for studying computer control of urban traffic networks. It has been extensively validated and is generally considered to yield reasonable results. The program is a microscopic simulation that deals with the movement of individual vehicles in an urban street network according to car-following, queue-discharge, and lane-changing theories. NETSIM defines the traffic network in terms of streets (links) and intersections (nodes). Each vehicle that travels through the network has associated with it data for, among other things, current speed, acceleration, and position. Detailed information concerning the operational characteristics of NETSIM may be found elsewhere (1,2).

THE PROBLEM

Although the NETSIM model is useful for accurately simulating traffic flows within an urban street network, certain deficiencies quickly become evident. These can be classified into three groups--data input, data debugging, and output analysis.

NETSIM is quite expensive to operate. Short (15-min) simulation runs of even a relatively small network generally require large amounts of computer time. Furthermore, input-data preparation requires excessive personnel time. One study (1) shows that 85 percent of the total cost of an initial NETSIM run consists of information-coding costs. For succeeding runs, about 65 percent of the total costs is in input-data modifications. There are several reasons for such high costs associated with input-data preparation. First, intuitive physical meaning of the network geometry and signal information is often lost, since all that information must be digitized. The coder is consequently faced with the problem of constantly referring to the network diagram and user's manual. This is very time consuming and confusing. Second, network information has to be divided due to the data-input limitation and, as a result, some input information is duplicated. This also requires the coder to recall prior input data, a situation that in many cases leads to inconsistencies. Finally, option spaces have to be provided within the input-data field in order to accommodate a variety of situations. Such option spaces are scattered throughout the input-data field and may not follow any apparent logic from the user's point of view.

Because of the above-stated conditions, many errors may occur in the input-data file. The NETSIM preprocessor has the capability to check the consistency of the input data and inform the user of such inconsistencies via error messages. However, in some cases error messages are given in terms of the link numbers assigned by NETSIM according to the order in which the links were input on the link-geometry cards. It is likely that the coder is unaware of the link-number assignment until faced with an error message. As a result, decoding time is spent in relating the error messages to the links on the link-node diagram. Problems such as the one described above can be classified as data-debugging problems.

NETSIM has the capability of generating printed measures of effectiveness (MOEs) either at predetermined times or at the end of the simulation subinterval and performing statistical analysis on multiple subinterval runs. The printouts may be voluminous and, although they are presented in an appealing format, they are sometimes difficult to interpret. Although the outputs are useful in defining the existence of potential problems, it may be difficult for the user to understand how such problems have evolved during the simulation time interval. Following through all the intermediate outputs is an impossible task.

With regard to the three problem elements associated with the use of the NETSIM program, interactive computer graphics (ICG) can aid in reducing or even eliminating many of the difficulties. The use of graphic output displays from computer programs is becoming common. Many computer programs are now being written in such a manner that the user can follow elements of the program execution (dynamic displays) and change the program during execution (interactive programs). The utility of computer graphics in the field of transportation was initially brought into focus at a conference held in Seattle in 1973 (3). Since then, applications in this area have been numerous. An article by Schneider (4) reviewed various applications of computer graphics in the transportation field. It is believed that alone or when combined with existing packages, computer graphic routines are cost-effective tools that can be used in analyzing data, generating and evaluating alternative solutions, and presenting the results.

Previous application (5) of computer graphics in conjunction with the NETSIM program has produced animated displays of vehicle movements and signal indications on an urban street network. The display of the results of the simulation model can be used in searching for high-performance traffic management strategies. The NETSIM/ICG computer graphic program described in this paper is an attempt to enhance the capabilities of the existing NETSIM program in addition to the animation of vehicular flow. The modifications and further developments of the NETSIM computer graphics program are described. The overall framework of the NETSIM/ICG program is presented in Figure 1. The program consists of three interactive computer graphics programs: pre-NETSIM (PRENET), NETSIM display (NETDIS), and postdisplay (POSDIS). The PRENET program provides the capabilities for interactive data input and modification, data display, and preparation of the input-data files for both NETSIM and NETDIS. The NETDIS program runs in conjunction with NETSIM and provides real-time displays of link-specific MOEs generated by NETSIM and, in addition, prepares a display file for POSDIS. The POSDIS program provides passive displays of user-selected link-specific MOEs.

SYSTEM DESCRIPTION

NETSIM/ICG is a FORTRAN-based program. It was designed for use on Rensselaer Polytechnic Institute's IBM 3033 computer system that has an IBM 3277 graphics attachment and a graphics attachment support program PRPQ. The IBM 3277 graphics attachment uses a dual-screen work station concept. All alphanumeric data that are related to, but not part of, the contents of the graphic display are managed by an IBM 3277GA terminal. Graphic displays and graphical data manipulations are done on a Tektronix cathode-ray tube (CRT) that has "write-through" ability (limited refresh vectors) together with supporting interactive devices. The PRPQ is made up of a collection of FORTRAN and ASSEMBLER language subroutines that are available from the user's application program for a variety of graphical and control functions.

OBJECTIVES

The objective of the NETSIM/ICG project was to develop interactive computer graphics programs that have minimum alterations to the existing NETSIM program and the following capabilities:

1. Allow user to interactively create and/or modify link-mode diagram that represents street network;
2. Allow user to interactively input and/or modify information needed for NETSIM (efforts were made to categorize NETSIM input information into major groups and to incorporate logic that eliminates all unnecessary information requests);
3. Allow user to input all required information without reference to user's manual and in free format;
4. Provide graphical displays to as great a degree as possible of network-related NETSIM input data;
5. Create input data files necessary for both NETSIM and NETDIS from information previously input and read in these data files for modification;
6. Provide real-time displays of selected link-specified MOEs;
7. Allow user to interrupt execution, change display variables, and then resume execution of sim-

Figure 1. Overall framework of NETSIM/ICG computer program.

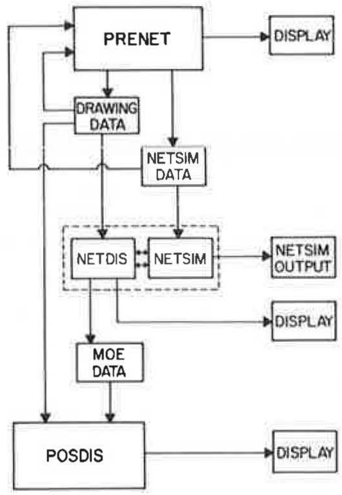
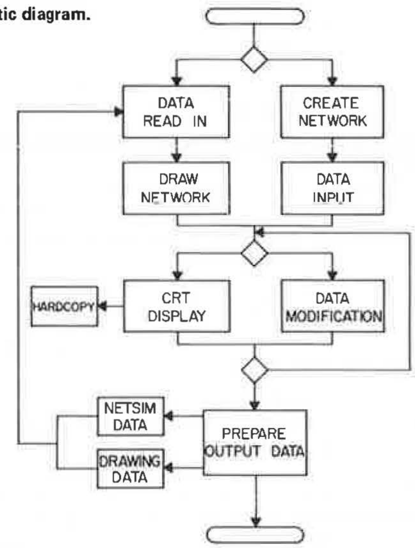


Figure 2. PRENET schematic diagram.



ulation and to prematurely terminate execution without losing standard cumulative statistical outputs generated by NETSIM;

8. Generate a graphical link-specific MOE data file for future passive displays; and

9. Provide passive displays of link-specific MOEs.

THE PROGRAM

Due to the size and the complexity of the NETSIM/ICG project, only portions of the features will be presented here in order to demonstrate the capability of the computer graphic aid. The description of the program is divided into three major sections. Each section addresses one of the problems identified in the problem statement, namely, input-data preparation, data display (debugging aid), and output-data interpretation.

Input-Data Preparation

This capability is provided by the PRENET enhancement program. The conceptual structure of the PRENET program is presented in Figure 2.

Initially, the user is asked to type in the information concerning the title of the run, the name of the network, the city, the state, and the date. This information is displayed at the top of the CRT screen and used as an identification code for this run.

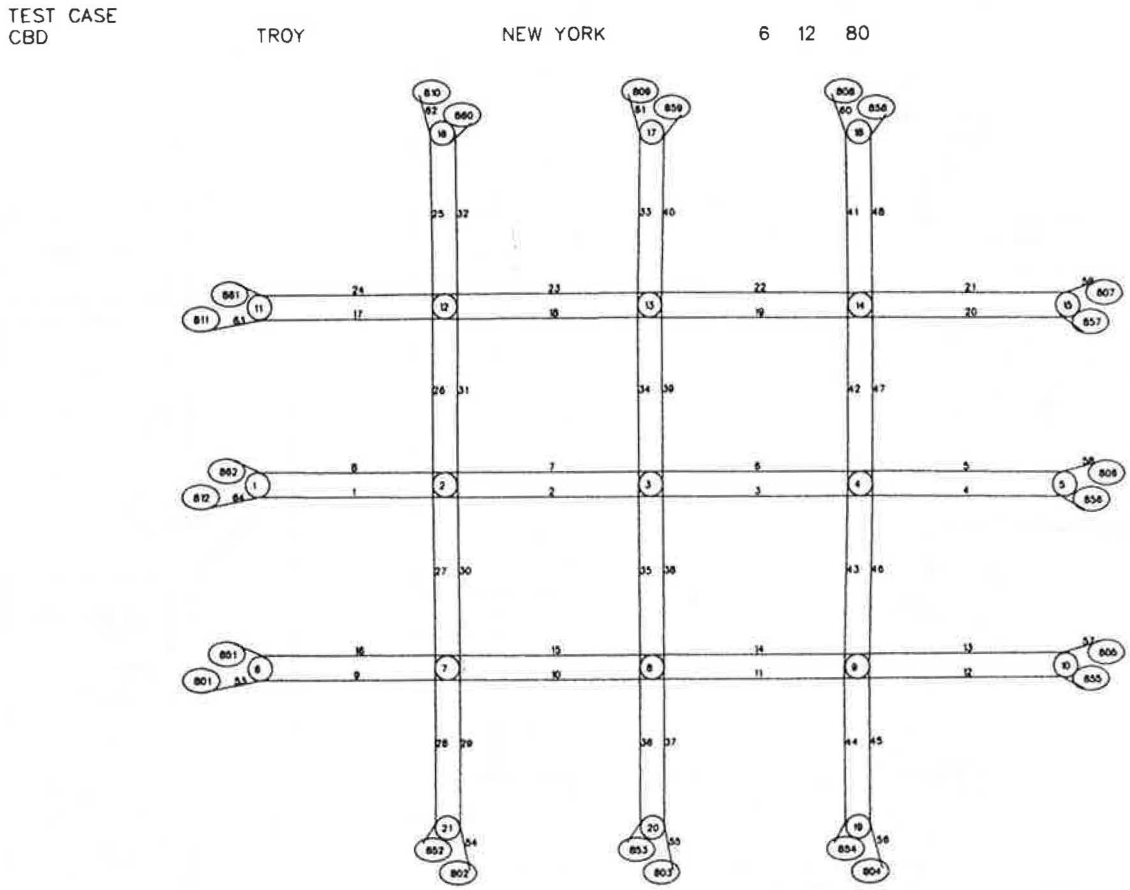
The NETSIM model uses a link-node network description of the actual system, and in order to accommodate a variety of urban street networks, the program has to have the capability of allowing the user to interactively create such a link-node diagram. Initially, the user is asked to use the cross-hair cursor to identify the location of nodes on the CRT screen. Each time a node location is identified, a circle that has an identifying node number is drawn on the screen (the circle is drawn within a subpicture for future subpicture detection). During the node-input stage the user has the capability of changing or deleting an already defined node. After the user exits from this input stage, the node numbers and their coordinates are fixed and cannot be changed. Second, the user must identify the upstream and downstream nodes of each link with the cross-hair cursor. Each time a link is thus identified, a line is drawn that connects the nodes and a link number appears on the screen adjacent to the link. As in the node-input stage, the user has the option to change or delete any link; thereafter, the links remain fixed. After the user has input all the nodes and links (NETSIM classifies nodes and links as entry nodes, exit nodes, internal nodes, exit links, entry links, and internal links), the complete link-node diagram is drawn and pertinent information is stored in appropriate arrays for later use. Such a link-node diagram is illustrated in Figure 3.

The NETSIM input data can be categorized into six major groups: link information, signal information, input traffic volumes, bus information, short-term and long-term events, and simulation control information. The program prompts the user for input and, if necessary, provides a short description of the appropriate code. Certain displays appear on the link-node diagram at appropriate times to aid the user. For example, the particular link for which input is being prompted is displayed in the flashing mode. Such a display is very helpful in identifying destination nodes that receive left-turning, through, and right-turning vehicles. Precautions have been taken in the program to reject obviously erroneous input data. For example, NETSIM requires that a bus route always start at an entry node. If the first node input for a bus route is not an entry node, the program will print an error message and prompt for corrected input. Other options, such as redoing portions of the input data, have also been provided for so that apparent mistakes can be corrected immediately. These options are provided in various places in the program, including the end of each link-information input, phase-information input, bus-station information input, bus-route information input, etc.

An effort has been made to organize the input-information procedure according to its logical flow. For example, during the link-information input stage, the program will prompt the user for through-traffic information if no left-turn destination node is designated. Questions concerning the left-turn percentage and left-turn pocket capacity will be bypassed. The logical structure of the input procedures for link and bus information is outlined in Figure 4.

In many cases, the information needed by NETSIM is graphically fed directly from the CRT screen. To illustrate, consider the bus-information input, which consists of bus station information and bus route information. For bus station information, the link on which the station is located is identified on the link-node diagram by the cross-hair cursor.

Figure 3. Sample link-node diagram.



Other alphanumeric data such as lanes blocked, type of station, dwell times, etc., are typed in by means of the keyboard. At this point, a rectangle, which represents the station, is drawn. Adjacent to the rectangle are shown the station number and dwell time (Figure 5). For bus route information, the links are erased and only the nodes and bus stations are displayed on the screen (Figure 6). For each bus route, the user must identify a series of succeeding nodes that connect this bus route, beginning with an entry node and terminating at an exit node. Concurrently, a line is drawn between the nodes each time a new node is identified (Figure 7). After having defined a particular bus route, the user is asked to identify the bus stations associated with this route by using the cross-hair cursor. A circle is drawn around each of the bus stations as they are identified (Figure 8). This is very useful in indicating data that have already been entered. Finally, the user types in the bus headway for this route. The above options, provided by PRENET, illustrate how the user, working on the link-node diagram, can input the bus information interactively.

Similar procedures are used in the input of signal information. For each phase, the user works with a signal-phasing diagram, which is displayed on the screen. Variable initial-green and gap-reduction information for actuated signals is displayed on top of the phasing diagrams, if applicable. Detectors are drawn and identified (the letters I and A are used to identify inactive and active detectors, respectively) on the "lane-detailed network plot" on the appropriate link and at the appropriate distance from the stop line. The phase numbers that are serviced by the detectors are also indicated on the screen. The user, in effect, puts the detectors on each approach at an actuated intersection (Figure 9). The data input logic is such that separate surveillance information is not required.

Currently, the interactive data input portion of the PRENET enhancement program cannot process optional input information such as link names, embedded data changes, and subinterval update data.

Data Display

After the necessary information for NETSIM and NETDIS is obtained by using the interactive data input capabilities provided by PRENET, the information is written into data files in a format that conforms to the NETSIM data input requirements. The capability of displaying these data is also provided by PRENET.

Since NETSIM is actually modeled on a lane-detailed microscopic network and keeps a record of each vehicle in each lane on the links of the network, it is desirable for the user to have a graphic display of such a lane-detailed microscopic network. A schematic lane-detailed microscopic network plot serves that purpose. An algorithm has been developed to produce such a plot. The word "schematic" is used because the length of the link and the width of the lane are not of the same scale. An example of such a schematic lane-detailed microscopic network diagram produced by this algorithm is presented in Figure 10. This schematic lane-detailed network plot is useful in displaying lane-specific information such as turning pockets, lane channelization, signal detectors, etc.

In coding the existing NETSIM signal information,

Figure 4. Link- and bus-information input procedure.

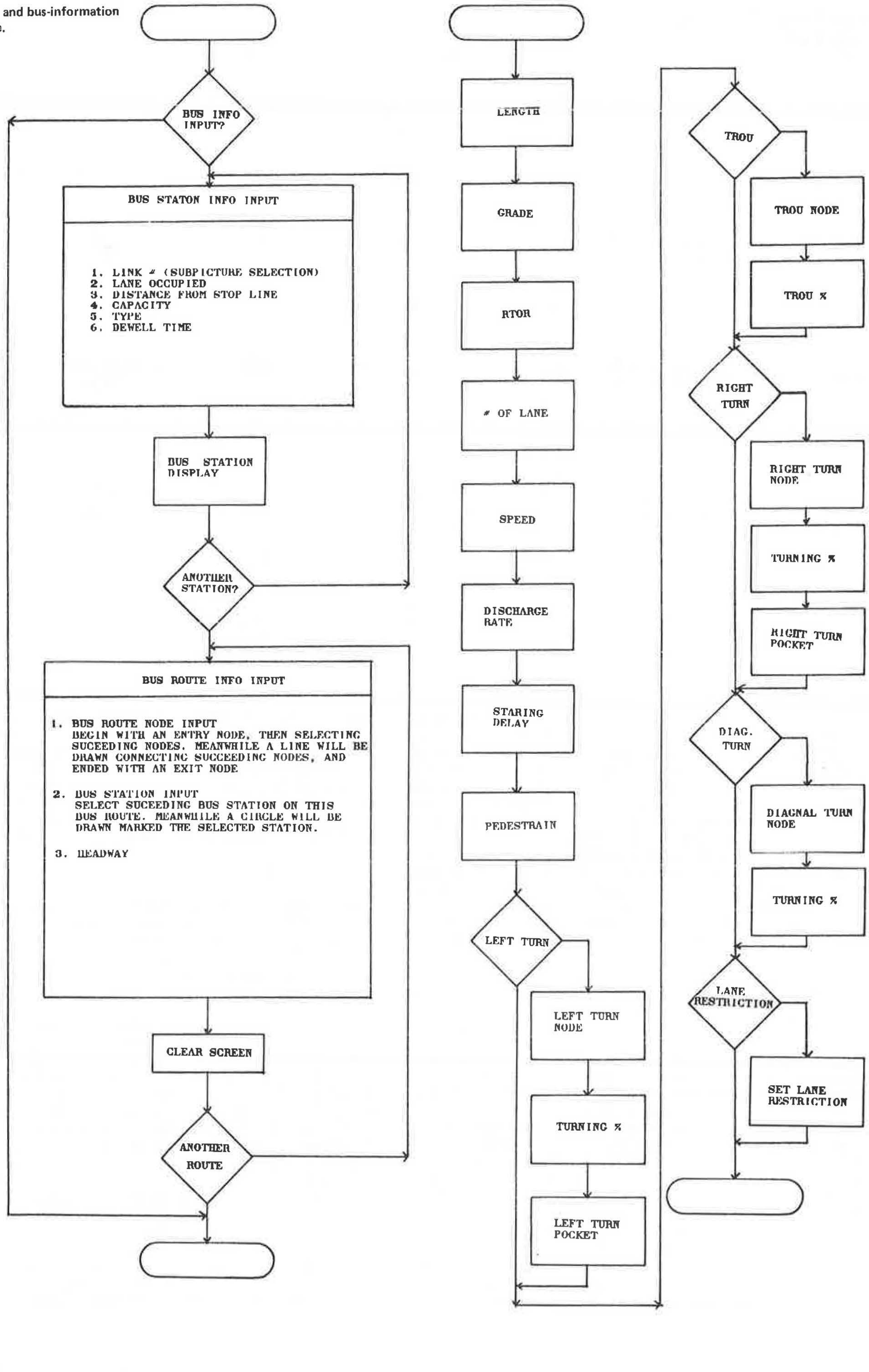


Figure 5. Bus station display.

TEST CASE
CBD TROY NEW YORK 6 12 80

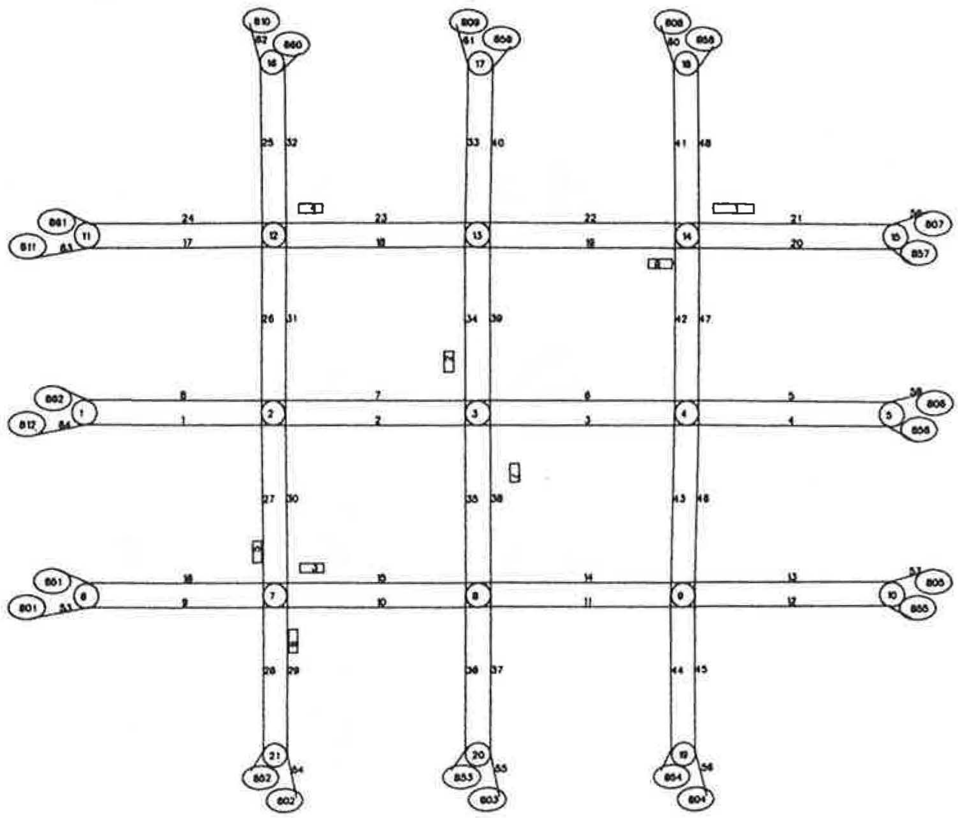


Figure 6. Bus station and node diagram.

TEST CASE
CBD TROY NEW YORK 6 12 80

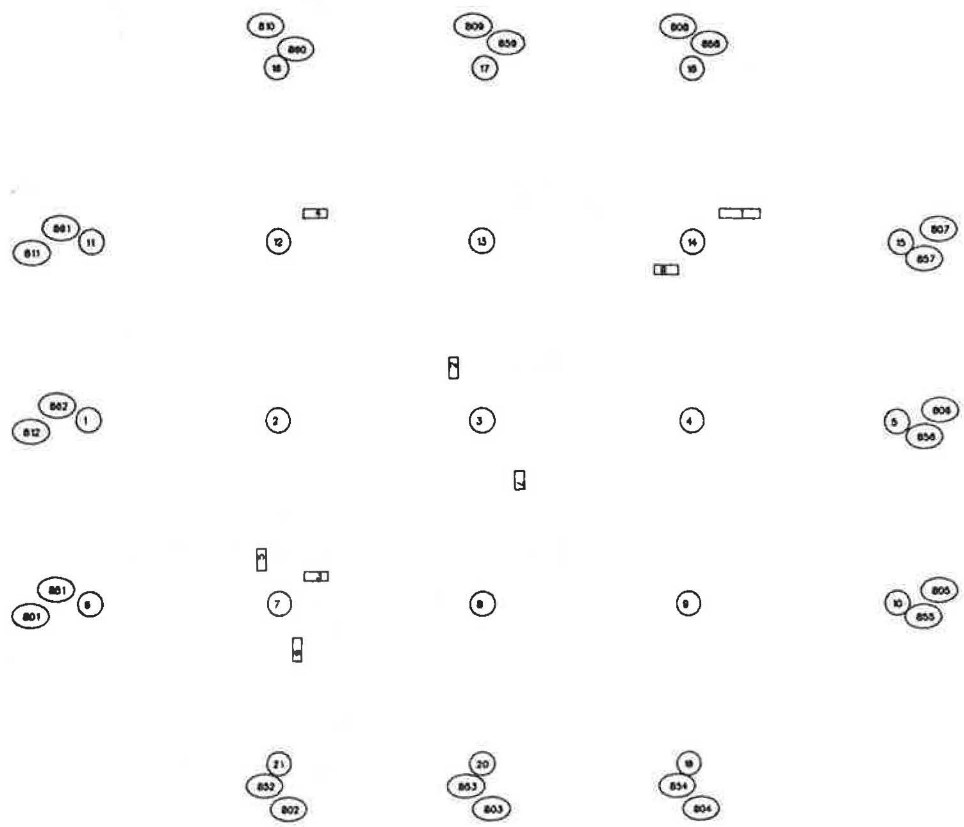


Figure 7. Bus route display.

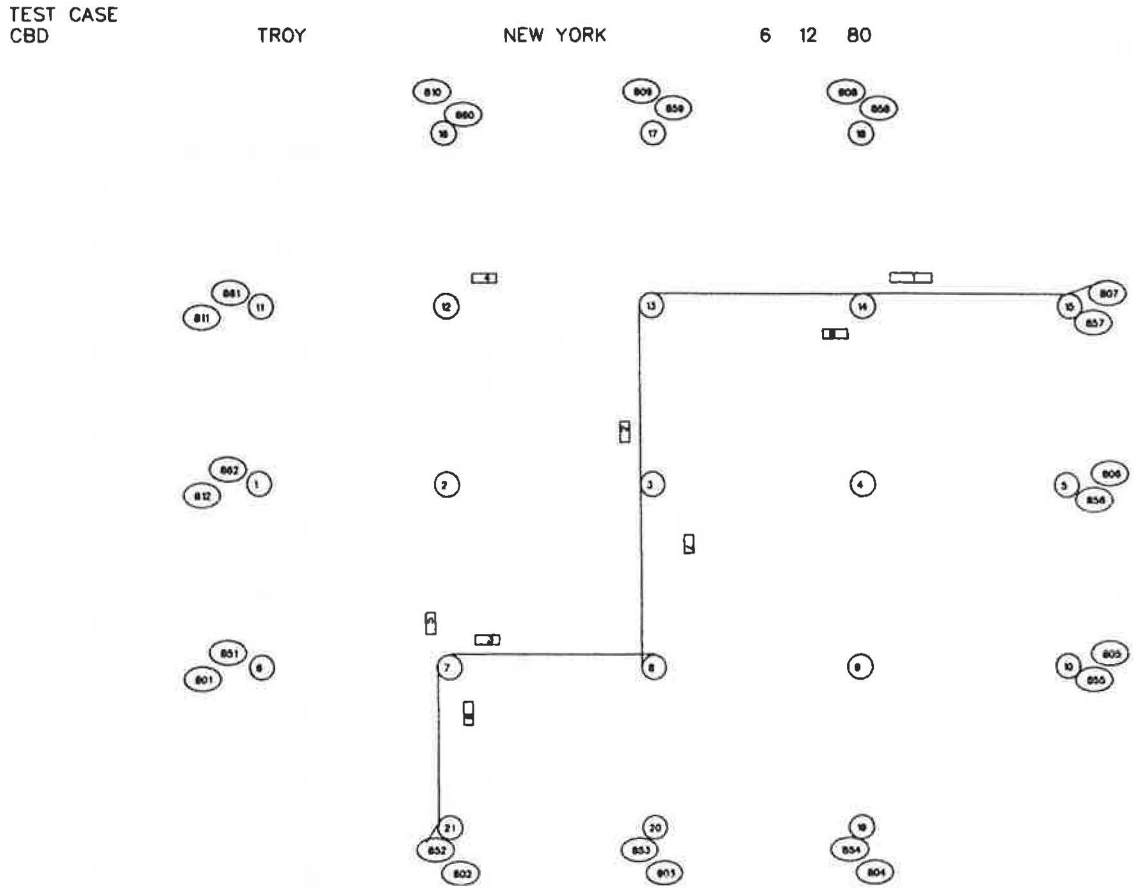


Figure 8. Bus station display.

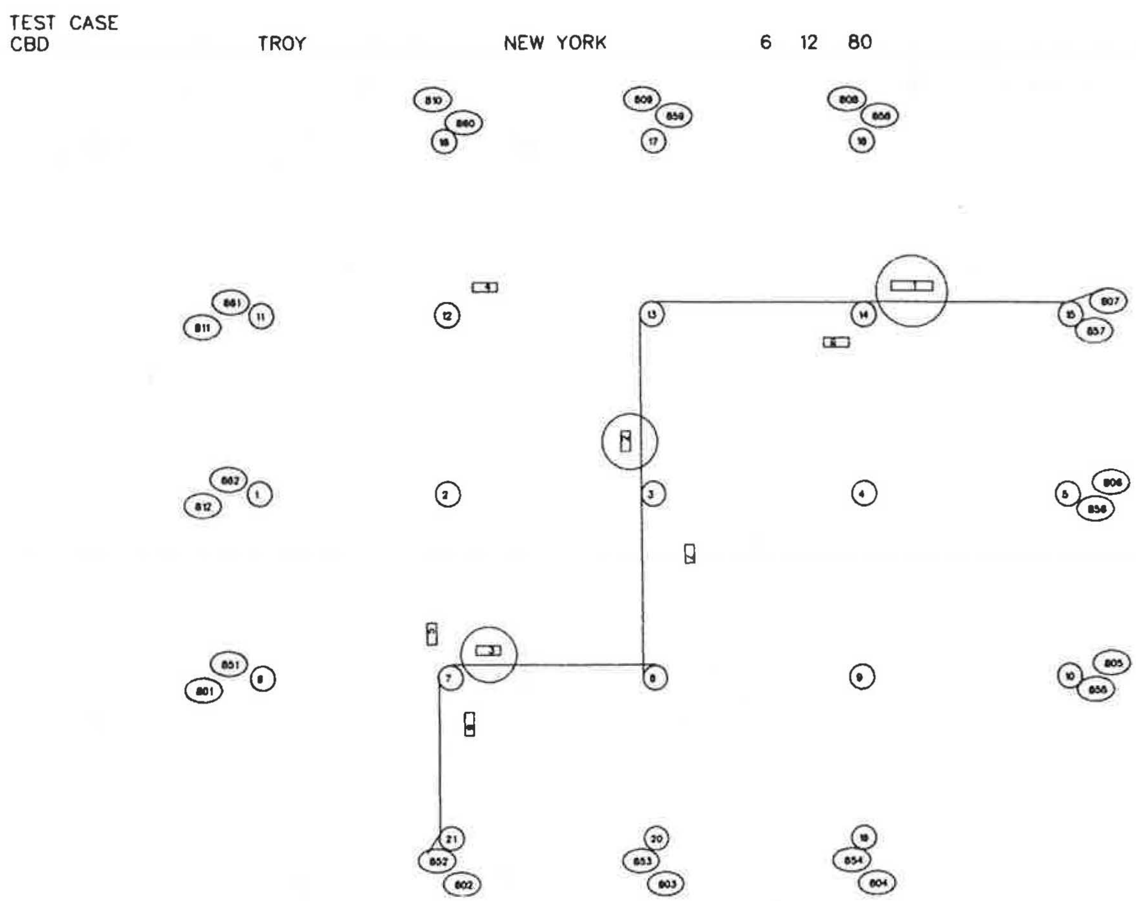
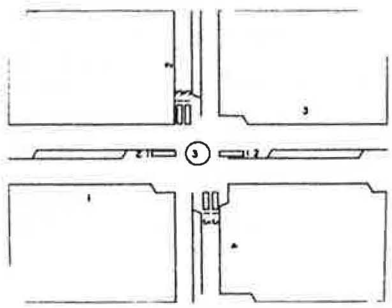
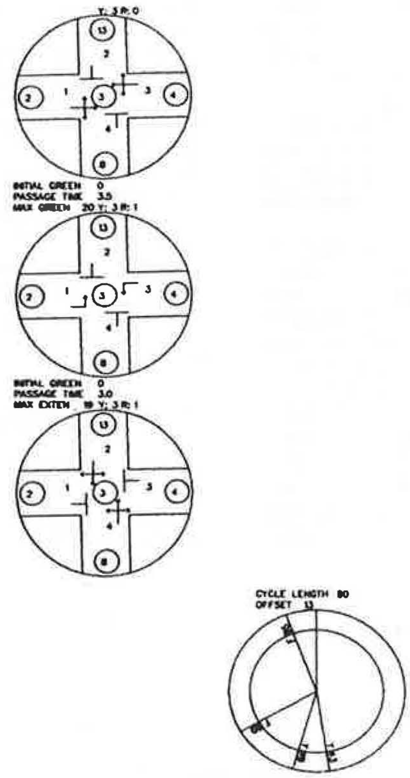


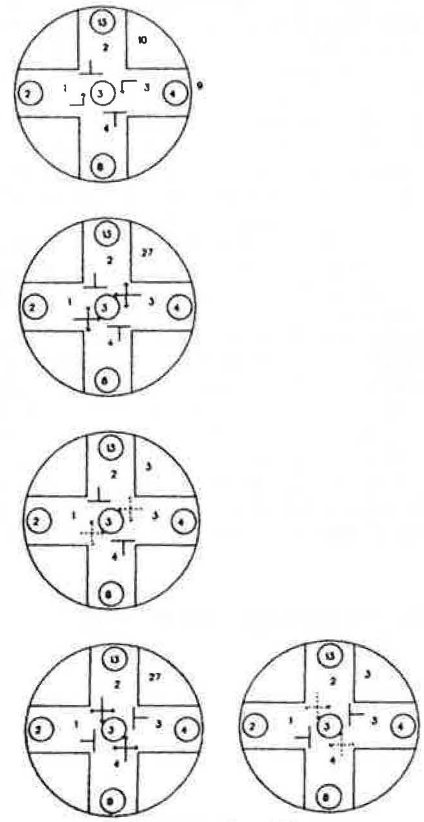
Figure 9. Schematic lane-specific network drawing.



Detectors Display

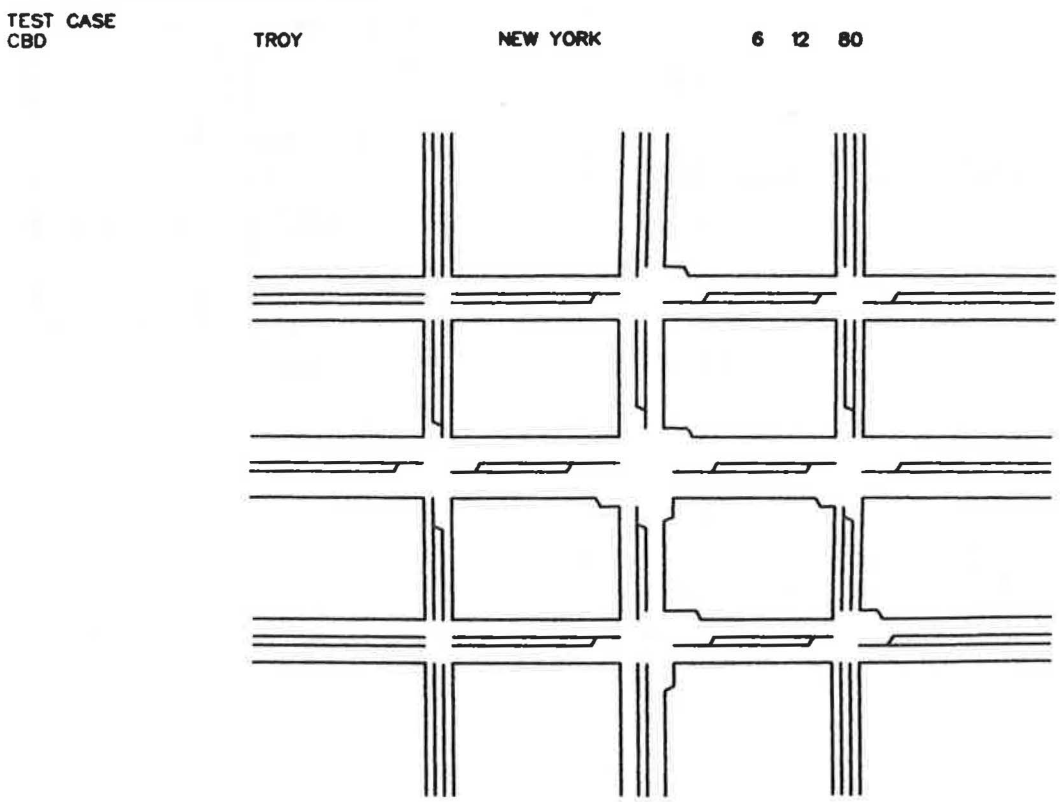


Actuated Signal Phasing with Coordination Dial



Fixed Time Signal Phasing

Figure 10. Link-specific information display.



the user is constantly dealing with approach numbers and signal codes. It is desirable for the user to visualize the signal operation by providing signal-phasing diagrams. PRENET has the capability of providing up to nine different phasing diagrams along the left and bottom edges of the CRT screen. For the actuated signals, controller timing and detectors are also displayed. Furthermore, a plot of the coordination dial is provided if the actuated signal is coordinated.

PRENET has the capability of allowing the user to interactively display the NETSIM input data by means of menu selection. For link information, the user can display (a) the schematic lane-detailed diagram with turn-pocket capabilities, channelization, link length, and free-flow speed (Figure 11); (b) destination nodes that receive the left-turn, through, and right-turn vehicles (Figure 11); (c) turning percentages (Figure 11); and (d) other miscellaneous information such as right-turn-on-red code, pedestrian code, grade code, queue-discharge rate, and starting delay (Figure 11). For signal information, only one node can be displayed at a time (Figure 9). The user can have all detectors within the network displayed. For input traffic volumes, only one display is needed (Figure 11). For bus information, the user can display (a) all bus stations on a schematic lane-detailed network (Figure 11) and/or (b) either all or some bus routes and their associated bus stations (Figure 12). For event information, the user can either display both short-term and long-term events in one plot (Figure 11) or can display them separately.

PRENET has the capability of allowing the user to interactively modify an already defined data base by using menu selection. The geometry of the link-node diagram cannot be modified. The user can delete, change, or add to the already defined NETSIM information.

Output-Data Interpretation

This section addresses the third identified problem element associated with the use of the NETSIM program, namely, the analysis of the outputs. As alluded to previously, the aggregated nature of the

Figure 11. Signal phase and detector display.

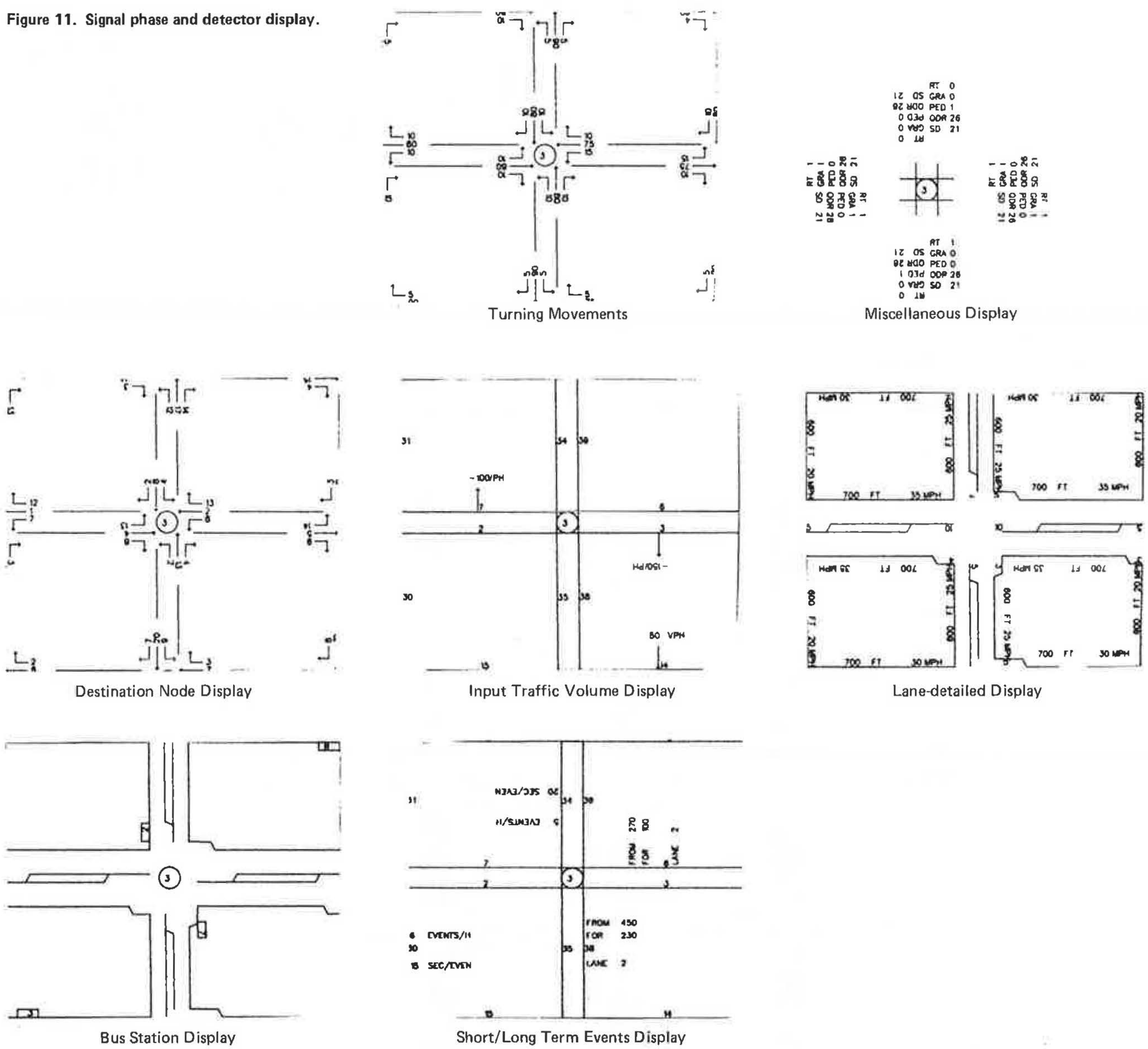
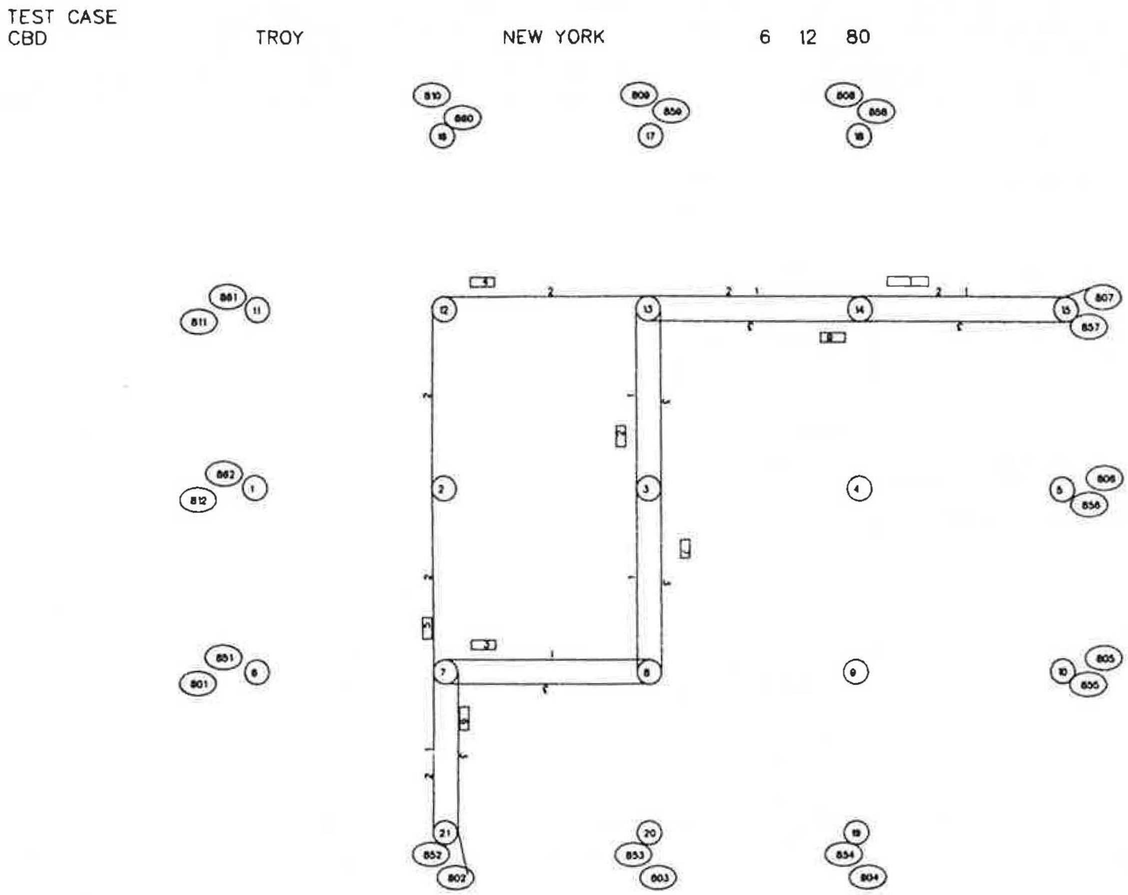


Figure 12. Bus route and bus station display.

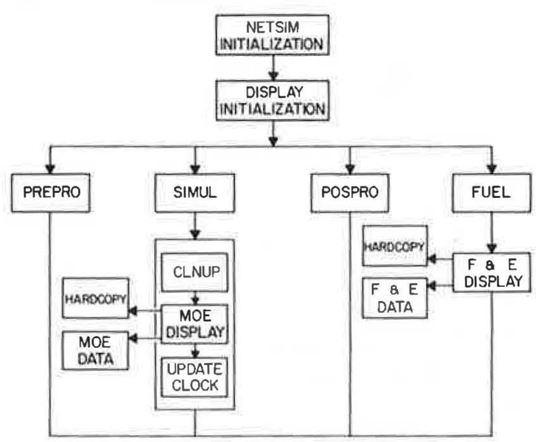


final simulation output tends to obscure much of the information that the program generates. Moreover, the analysis of the voluminous intermediate outputs is essentially infeasible. The following section discusses the concepts and approaches followed in designing the NETDIS and POSDIS enhancement programs that are a part of the developed NETSIM/ICG program. Both NETDIS and POSDIS were developed with the goal of facilitating the assimilation and interpretation of the network-simulation outputs.

Other researchers have previously applied computer graphics techniques to display the results of NETSIM. Joline (6) has produced a film that displays the movement of individual vehicles through the network. It has been useful in showing potential users of NETSIM what the model does. His work has also helped to identify errors in the model and has led to subsequent modifications. Eiger, Chin, and Woodin (5) have also developed a program that produces animated displays of the results of the simulation model that can be used in searching for high-performance traffic management strategies. Schneider, Combs, and Folsom (7-9) have taken a different approach, in which the interest is in displaying data that describe the overall performance of the system over time. By examining these systemwide displays, the user should be able to generate some ideas on modifying the parameters of the system to obtain higher levels of performance. This is a holistic approach to the problem of discovering ways to drive the performance of a multiobjective system in the desired directions.

The operational frameworks of the NETDIS and POSDIS programs are shown in Figures 13 and 14, respectively. The MOEs generated by the NETSIM

Figure 13. NETDIS schematic diagram.



program and displayed by NETDIS and POSDIS are shown in Table 1. Fuel-consumption and emissions data can be obtained only at the end of each simulation subinterval. The remaining MOEs can be obtained each simulation second. Queue length and occupancy are instantaneous displays; the others are cumulative or aggregate.

The cumulative link-specific MOEs are displayed on a three-dimensional perspective plot. The link-node network is drawn on the x-y plane, and the MOEs are represented by perpendicular lines (Figure 15) or rectangles (Figure 16) in the z-direction with the height scaled appropriately. The scale factors for all the MOEs are predetermined along with the parameters involved in generating the perspective plots. The user currently does not have the option of selecting these scale factors or perspective plot parameters. This maintains consistency among plots for different runs so that they can easily be compared. Often, portions of the selected MOE displayed on the perspective plot may be obscured by others. In other instances, the displays can be confusing due to the size of the network. Under these conditions, the user may wish to concentrate on certain areas or links of the network. The

Figure 14. POSDIS schematic diagram.

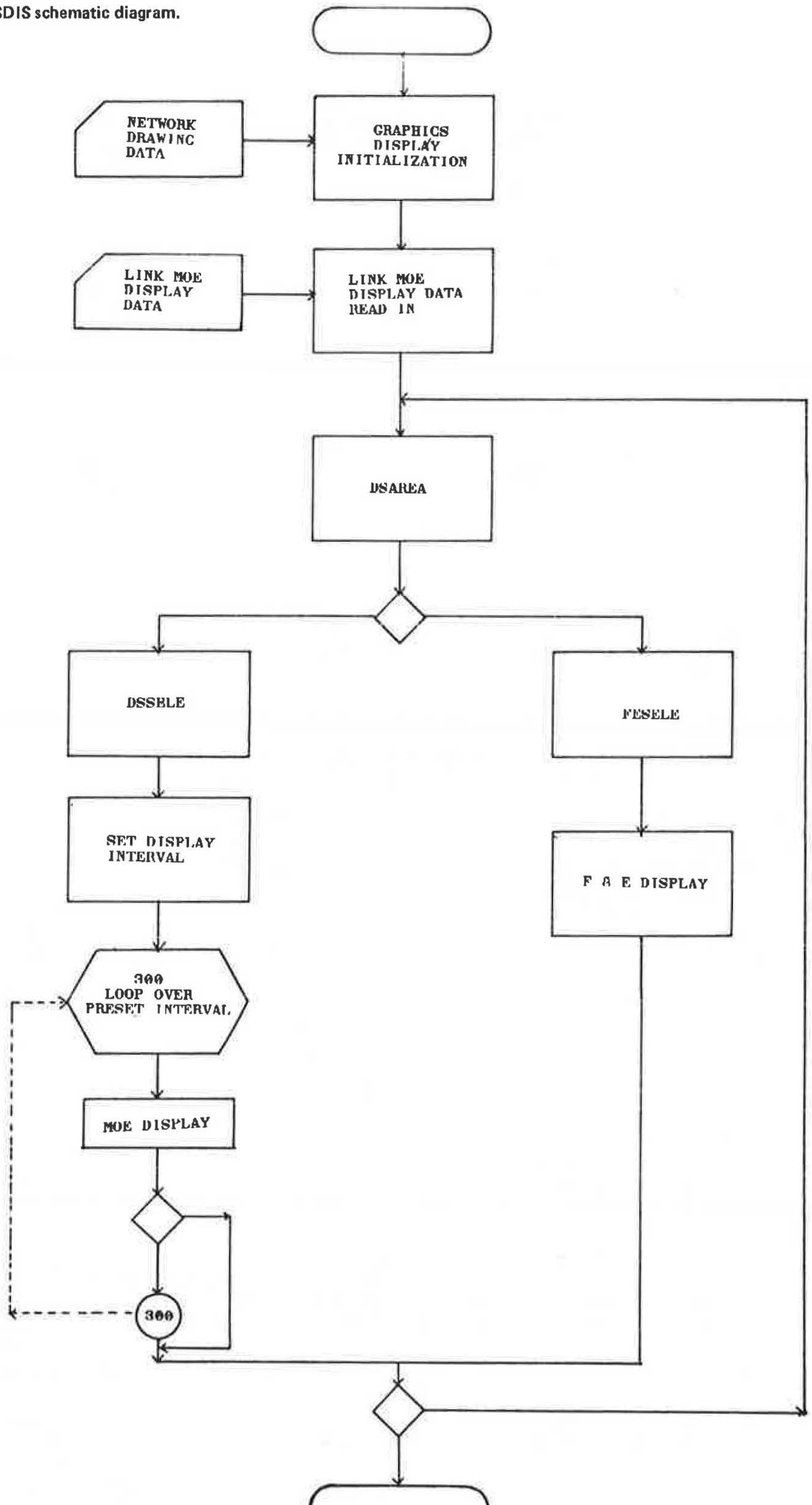


Table 1. Link-specific MOEs displayed by NETDIS and POSDIS.

Link-Specific MOE	Unit	Menu Abbreviation	Availability	Statistic	Display Plot
Queue length	Vehicles	Q-LEN	Every second	Instantaneous	2-D lane-specific
Occupancy	Vehicles	OCCU	Every second	Instantaneous	3-D perspective
Vehicle miles	Miles	V-MI	Every second	Cumulative	3-D perspective
Vehicle discharged	Trips	V-DIS	Every second	Cumulative	3-D perspective
Moving time	Minutes	MOV-T	Every second	Cumulative	3-D perspective
Delay time	Minutes	DLY-T	Every second	Cumulative	3-D perspective
Moving time/total trip time	Percent	M/T	Every second	Cumulative	3-D perspective
Total travel time	Minutes	T-TIME	Every second	Cumulative	3-D perspective
Travel time/vehicle	Seconds	T-TIM/V	Every second	Cumulative	3-D perspective
Travel time/vehicle mile	Seconds	T-TIM/V-M	Every second	Cumulative	3-D perspective
Delay time/vehicle	Seconds	DLY-T/V	Every second	Cumulative	3-D perspective
Delay time/vehicle mile	Seconds	DLY-T/V-M	Every second	Cumulative	3-D perspective
Stop delay/total delay	Percent	PCT ST-DLY	Every second	Cumulative	3-D perspective
Average speed	Mph	AV SP	Every second	Cumulative	3-D perspective
Average occupancy	Vehicles	AV OCC	Every second	Cumulative	3-D perspective
Stops/vehicle	Stops	ST/V	Every second	Cumulative	3-D perspective
Average saturation	Percent	AV SAT PCT	Every second	Cumulative	3-D perspective
Cycle failures	Failures	CYC FAL	Every second	Cumulative	3-D perspective
Fuel consumption (automobile)	Gallons	FUEL-AUTO	At end	Aggregate	3-D perspective
Fuel consumption (truck)	Gallons	FUEL-TRUCK	At end	Aggregate	3-D perspective
Fuel consumption (bus)	Gallons	FUEL-BUS	At end	Aggregate	3-D perspective
Fuel efficiency (automobile)	Mpg	MPG-AUTO	At end	Aggregate	3-D perspective
Fuel efficiency (truck)	Mpg	MPG-TRUCK	At end	Aggregate	3-D perspective
Fuel efficiency (bus)	Mpg	MPG-BUS	At end	Aggregate	3-D perspective
HC (automobile)	Grams	HC-AUTO	At end	Aggregate	3-D perspective
CO (automobile)	Grams	CO-AUTO	At end	Aggregate	3-D perspective
NO X (automobile)	Grams	NO-AUTO	At end	Aggregate	3-D perspective

NETDIS and POSDIS programs have the option of restricting the display area (Figure 17).

Queue lengths, by contrast to the cumulative MOEs, are displayed on a lane-detailed schematic network plot. The queue of a particular lane is drawn on that lane and scaled to the corresponding link length on the lane-detailed plot (Figure 18). By using these and other displays, the user can gain an intuitive understanding of the overall network operation.

Since NETDIS and POSDIS can only display one selected MOE at a time, the capability is provided to interactively select the MOE to be displayed. This capability is through menu-option selection.

As execution proceeds, the user has the option to change the selected MOE or the display area or to terminate the simulation prematurely without losing the link-performance statistics accumulated to that point. These options are provided by using the ATTENTION-TRAP feature inserted at the end of the NETSIM simulation control loop.

In displaying the selected MOEs, the following user-selected options have been provided:

1. The selected MOE can be displayed in either the storage or the refresh vector mode.
2. The user can choose either lines or rectangles to represent the magnitude of the MOE (Figures 15 and 16).
3. The selected MOE can be displayed with or without numerical values of its magnitude (Figures 15, 16, and 17).
4. The user can obtain hard copies of all displays.
5. The user can write all the MOE graphic data into a data file for future passive displays.

After the NETSIM simulation finishes executing through the predetermined simulation time subinterval and the NETSIM subroutine FUEL has been executed, the user can display selected link fuel and emissions data (Figures 15 and 16).

The user may wish to examine the link-specific MOE displays more than once. This is particularly true since, as is recalled, NETDIS can only display one MOE at a time. Instead of resimulating, the user can call a separate program (POSDIS), which can repetitively display the MOE data generated by NETDIS. The options described above can help the user gain a full understanding of the overall operation of the network over the simulation subinterval without resimulating.

IMPLEMENTATION

Further work in this area would address the problem of portability of the NETSIM/ICG computer program. Although efforts have been made to minimize the use of system-dependent subroutines, modifications are still needed in implementing the NETSIM/ICG enhancement on other graphics systems. In general, the following functional capabilities on other graphics systems are needed to accommodate the NETSIM/ICG program without major modifications:

1. MOVE to initial point;
2. DRAW line from initial point;
3. Ability to create subpicture (entities, cells);
4. Device to interactively pick subpictures on CRT screen;
5. Device to interactively pick x-y location input on CRT screen;
6. ATTENTION-TRAP feature; and
7. Graphics processor that has capability to handle large network.

Other software subroutines such as circle drawing, ellipse drawing, perspective drawing, intersection of two linear angles of line, etc., are supported by the NETSIM/ICG program.

SUMMARY AND CONCLUSIONS

The NETSIM program is a useful tool in evaluating alternative traffic-control strategies. However, the numerous and highly complex initial data preparation tasks significantly increase the costs associated with using the program. The interactive data input elements of the PRENET enhancement program demonstrate the potential usefulness of interactive data input for both NETSIM and its other enhancements, NETDIS and POSDIS. The program provides a systematic method for obtaining the information needed by NETSIM without reference to the user's manual or particular attention to correct input-data

Figure 15. Perspective plot with line display.

FUEL CONSUMPTION IN GALLONS FOR COMPOSITE AUTOMOBILE

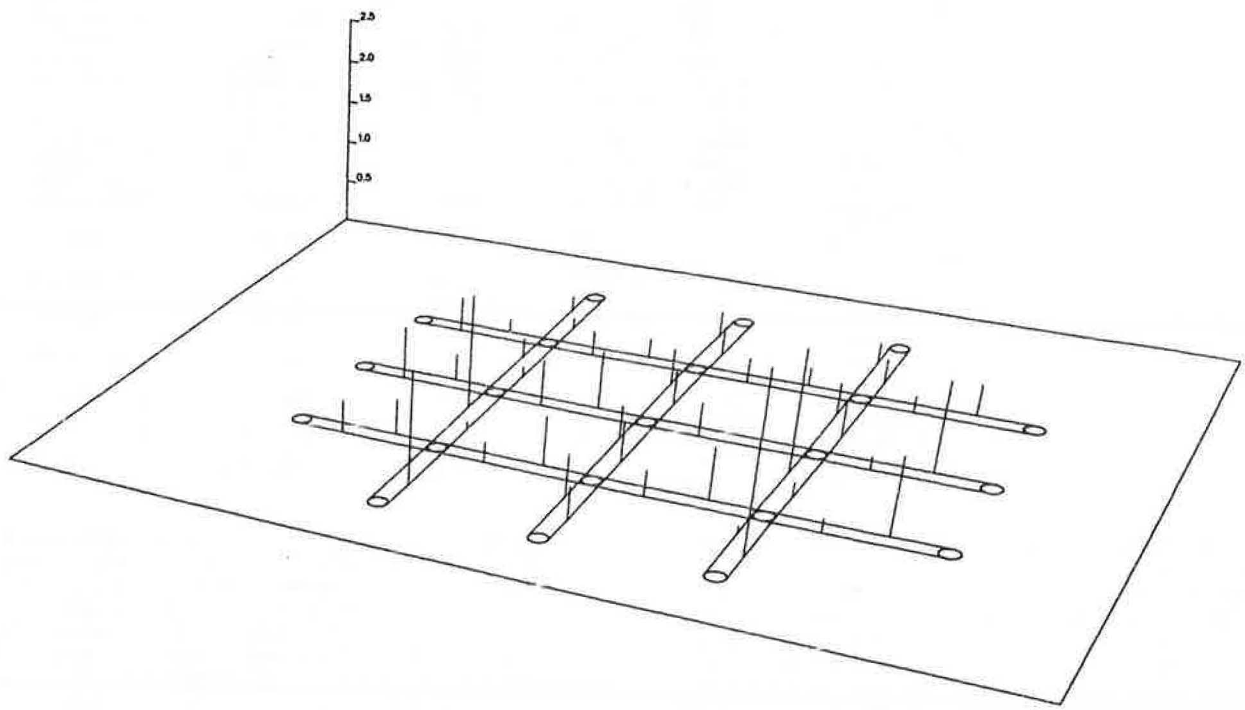
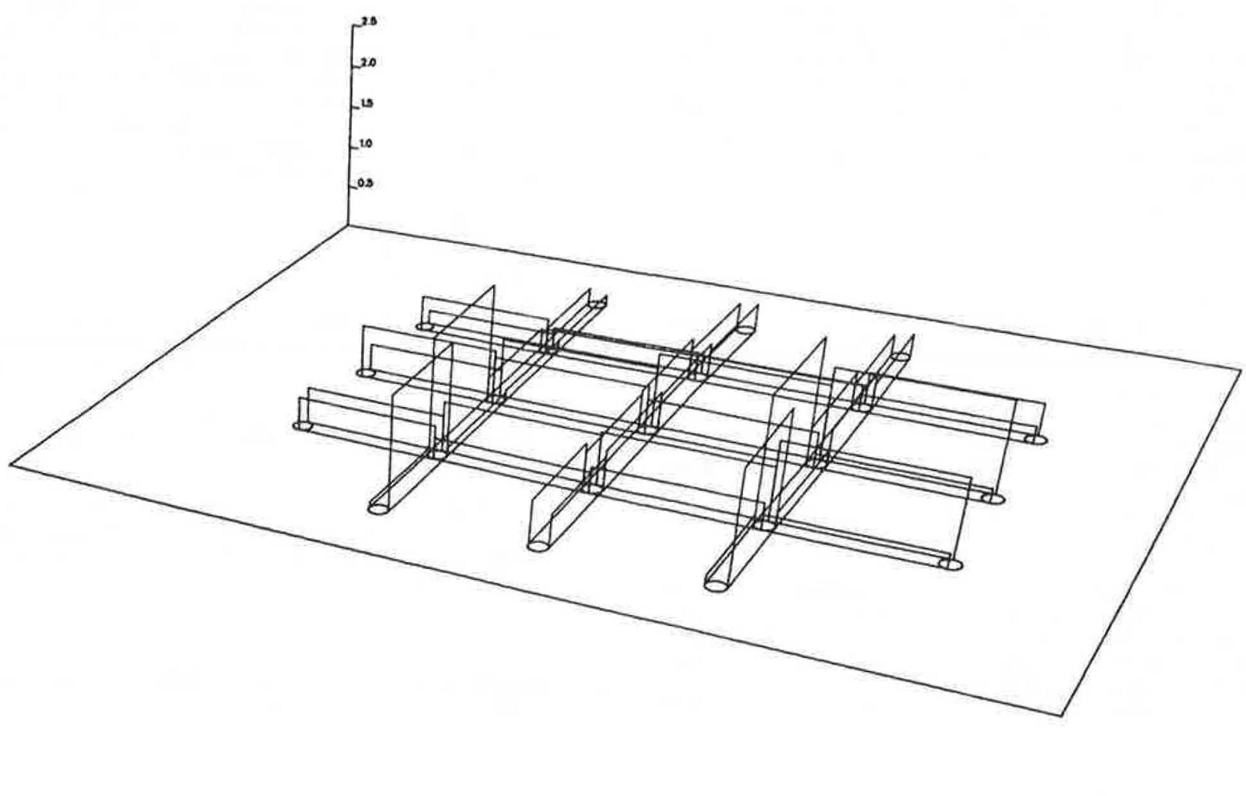


Figure 16. Perspective plot with rectangle display.

FUEL CONSUMPTION IN GALLONS FOR COMPOSITE AUTOMOBILE



fields and codes. In addition, certain information is input graphically and directly on the CRT screen. The capabilities provided by PRENET permit the processing of complex situations with minimal data input problems. In addition, the program has been designed so that terminal sessions can be interrupted without loss of the previously input data. Providing for interrupted terminal session capability is extremely important in an interactive data input process.

The microscopic nature of the NETSIM simulation model requires the user to prepare numerous input data, a process that can be tedious. Very often, inconsistencies, contradictions, or other errors result in the input-data file. The interactive computer graphics program PRENET, by displaying the NETSIM input data, provides the user with an intuitive understanding of the data. Errors within the input data file will produce either an erroneous network display or obvious inconsistencies on a net-

Figure 17. Partial perspective plot with rectangle display.

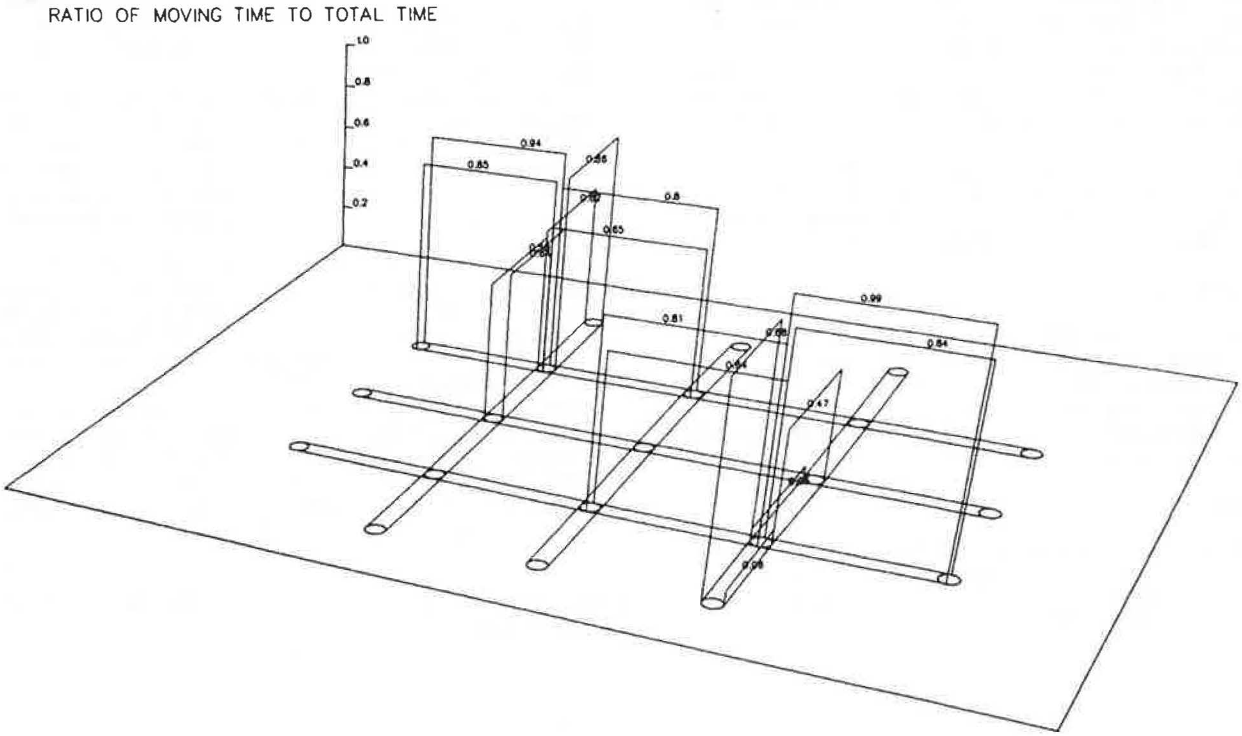
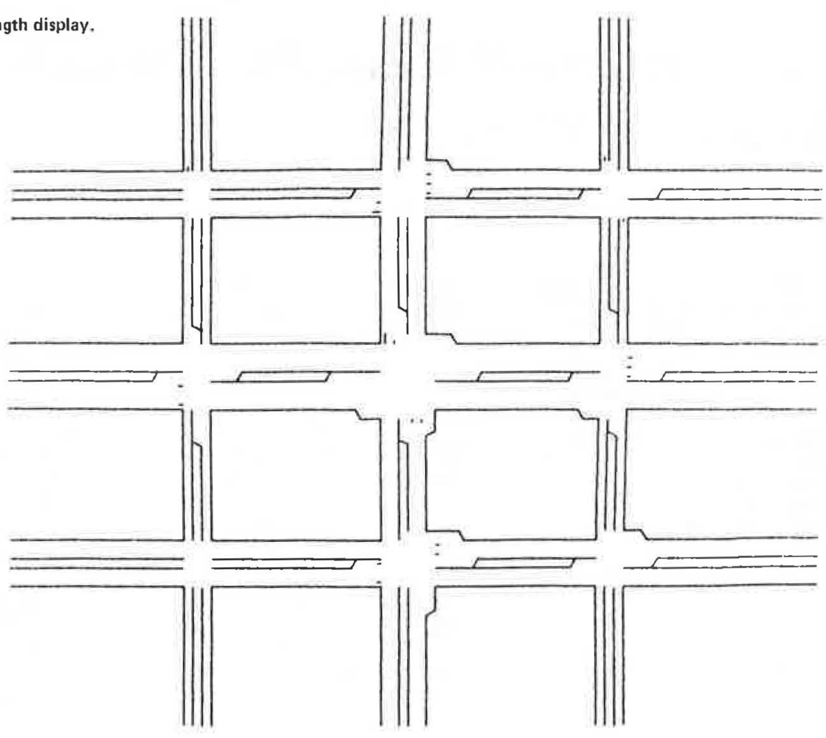


Figure 18. Queue-length display.



work diagram. Subsequently, these errors can be corrected by using the interactive data modification capability provided. This portion of the PRENET enhancement provides the user with an intuitive, direct, and efficient method to debug the NETSIM input data.

Having used NETSIM to define potential problems that relate to traffic control strategies in an urban street network, traffic engineers must gain a full understanding of how and why such problems have evolved. The ICG enhancements NETDIS and POSDIS demonstrate the feasibility and usefulness of graphically displaying link-specific MOEs as generated by the NETSIM simulation model. Such displays provide the user with more easily assimilated information with which operation of the network can be comprehended. More precisely, the lane-specific queue build-ups can be visualized, and the overall network relative performance measures can be obtained at a glance.

With respect to the development of additional graphics capabilities, work is required to generate time-space diagrams and displays of signal indications on the lane-detailed network plots.

ACKNOWLEDGMENT

This project was partly funded by the Research and Special Programs Administration of the U.S. Department of Transportation. The results and views expressed herein are ours and do not necessarily reflect the policies or views of the U.S. Department of Transportation.

REFERENCES

1. L. Kubel, G. Bloodgood, F. Workmon, and D. Gibson. What Network Simulation (NETSIM) Can Do for the Traffic Engineer. Public Roads, Vol. 41, No. 4, March 1978.
2. W.D. Labrum, R.M. Farr, W.J. Kennedy, and D. Gibson. Analyzing Intersection Performance with NETSIM. Public Roads, Vol. 42, No. 1, June 1978.
3. J.B. Schneider. Interactive Graphics in Transportation System Planning and Design. Final Rept. DOT-TST-74-10, Jan. 1974.
4. J.B. Schneider. Applications of Computer Graphics in the Transportation Field. Computer Graphics, Vol. 1, No. 13, 1978.
5. A. Eiger, S.M. Chin, and D. Woodin. Computer Graphic Animation of UTCS-1/NETSIM Traffic Flow. TRB, Transportation Research Record 729, 1979, pp. 7-10.
6. E.S. Joline. Computer-Generated Films to Document and Demonstrate Transportation Simulation Models. TRB, Transportation Research Record 619, 1976, pp. 34-37.
7. J.B. Schneider, D.M. Combs, and T.C. Folsom. NETGRAF: A Computer Graphics Aid to the Operation and Interpretation of NETSIM, a Traffic Simulation Model: Part 1--Overview and Experimental Results. U.S. Department of Transportation, Univ. Res. Rept. DOT-RSPA-DPB-50/80/10, 1980.
8. D.M. Combs, R.C. Folsom, and J.B. Schneider. NETGRAF: A Computer Graphics Aid to the Operation and Interpretation of NETSIM, a Traffic Simulation Model: Part 2--User's Guide. U.S. Department of Transportation, Univ. Res. Rept. DOT-RSPA-DPB-50/80/11, 1980.
9. D.M. Combs and J.B. Schneider. NETGRAF: A Computer Graphics Aid to the Operation and Interpretation of NETSIM, a Traffic Simulation Model: Part 3--Programmer's Guide. U.S. Department of Transportation, Univ. Res. Rept., DOT-RSPA-DPB-50/80/12, 1980.

Publication of this paper sponsored by Committee on Computer Graphics and Interactive Computing.

Transferability and Analysis of Prediction Errors in Mode-Choice Models for Work Trips

YOUSSEF DEHGHANI, BRENDA KOUSHESHI, ROBERT SIEVERT, AND THOMAS McKEARNEY

Some analyses of predictive accuracy and transferability of disaggregate work-trip mode-choice models are reported. The prediction error is separated into three components: model error, aggregation error, and transfer error. The results show that the weighted root mean square of total error is between 25 and 60 percent of the predicted shares and the distribution of total error between error sources seems to depend on how well the model is transferred. The main results of the research are that (a) total forecasting errors may be large, especially if the model transfers poorly; (b) transferability between cities in which the transit shares are very different is poor; (c) market segmentation improves forecasting accuracy only marginally, if at all, and; (d) the type of level-of-service data, i.e., manually coded versus network based, used in model estimation and prediction has some bearing on forecasting accuracy, and the use of zonally averaged socioeconomic attributes appeared to be somewhat detrimental to prediction. These and other results are to be held tentative for reasons discussed in detail.

The sources of the total error for work-trip mode-choice models are identified and their contributions are analyzed separately. In addition, citywide predictions of travel demands are also investigated. Four data sets were used in this study. These were the data from the Minneapolis-St. Paul area (collected in 1970); the two urban travel demand forecasting surveys from the San Francisco Bay area conducted before and after the introduction of Bay Area Rapid Transit (BART) service (collected in 1972 and 1975, respectively); and the Baltimore travel demand data set, a comprehensive set of information that describes travel behavior of 967 households in Baltimore, Maryland (collected in 1977).

The effect of market segmentation on forecasting accuracy is studied by using the same model specification as that for the unsegmented market. Three types of market segments were used: households that had one car versus those that had two or more; commuters bound for the central business district (CBD) versus others; and low-income versus high-income households (annual household incomes of $12 000,

$13 000, and $15 000 were used as dividers between high and low income for travelers in Minneapolis, San Francisco, and Baltimore metropolitan areas, respectively).

It had previously been found (1) that the model coefficients for the two income groups were different for the models that used post-BART data; except for the alternative-specific constants, the coefficients were equal for the other two market segments. In the models that used Baltimore data, only the travel-time coefficient was statistically different for all the market segments, and for the pre-BART models the alternative-specific constants were different for all the aforementioned market segments (2). An interesting question, therefore, is whether statistical inequality of coefficients means substantially different forecasts.

The results are presented in seven sections. First, the method of analysis is described. Then a discussion of model errors follows. Third, the error in prediction due to zonal averaging of the explanatory variables is analyzed. Fourth, errors due to transfer of models over time and space are discussed. In the fifth section, total prediction error is analyzed. Then citywide modal predictions are examined. Seventh, conclusions are presented.

METHOD

Average absolute error (AAE) and the root-mean-square error (RMSE) are used as error measures. Both are expressed as a fraction of the predicted share and calculated as follows. Actual and predicted modal shares are first calculated from 50 draws (observations) selected at random from the appropriate set of data. This process is repeated 50 times and AAE and RMSE are calculated to obtain error in the predicted share of each alternative.

AAE and RMSE are defined as follows:

$$AAE_j = (1/NT)\left[\sum_{i=1}^{NT} |P_j - A_j|/P_j\right] \tag{1}$$

$$RMSE_j = (AE_j^2 + SE_j^2)^{1/2} \tag{2}$$

where

$$AE_j = (1/NT)\left[\sum_{i=1}^{NT} (P_j - A_j)/P_j\right] \tag{2a}$$

$$SE_j = \left(\left\{\sum_{i=1}^{NT}\left[(P_j - A_j)/P_j\right]^2 - NT * AE_j^2\right\}/(NT-1)\right)^{1/2} \tag{2b}$$

and

JT = total number of alternatives in choice set,
NT = number of times 50 random draws are repeated (i.e., 50 times),
AE_j = average error as percentage of predicted share for alternative j,
SE_j = standard error for alternative j,
P_j = average predicted share of alternative j calculated from 50 random draws, and
A_j = average observed share of alternative j calculated from 50 random draws.

The overall error measures are the weighted average absolute error (WAAE) and weighted root-mean-square error (WRMSE) (3). These are defined by the following equations:

$$WAAE = \sum_{j=1}^{JT}\left[AAE_j * (1/NT)\left(\sum_{i=1}^{NT} P_j\right)\right] \tag{3}$$

$$WRMSE = \sum_{j=1}^{JT}\left[RMSE_j^2 * (1/NT)\left(\sum_{i=1}^{NT} P_j\right)\right]^{1/2} \tag{4}$$

WRMSE can also be disaggregated into weighted average error (WAE) and weighted standard deviation of the error (WSDE) as follows:

$$WAE = \sum_{j=1}^{JT}\left[AE_j * (1/NT)\left(\sum_{i=1}^{NT} P_j\right)\right] \tag{5}$$

$$WSDE = (WRMSE^2 - WAE^2)^{1/2} \tag{6}$$

Disaggregation of Errors in Prediction

Total error in prediction may be attributed to the source of the error in the following way. Total WAAE is the sum of WAAEs contributed by each component, and total WRMSE is the sum of WRMSEs contributed by each component. They are defined as follows:

$$WAAE_T = WAAE_M + WAAE_A + WAAE_F \tag{7}$$

$$(WRMSE)_T^2 = (WRMSE)_M^2 + (WRMSE)_A^2 + (WRMSE)_F^2 \tag{8}$$

where

T = total error,
M = error in choice model,
A = error in aggregation, and
F = error in transfer.

Mode-Choice Models Used in Analysis

The work-trip mode-choice models used in the analysis are a five-mode-choice model (drive alone, local bus, express bus, rail, shared ride) developed by using post-BART data (1); a three-mode-choice model (drive alone, bus, shared ride) developed by using Baltimore data (2); and a four-mode-choice model (drive alone, bus with walk access, bus with car access, and shared ride) that used pre-BART data. The model specification and coefficients are given in Tables 1 and 2. The same specification was also used for travelers in different market segments. [A

Table 1. Model specification and coefficients: Baltimore data.

Variable	Alternative Entered[a]	With WORKRS Coefficient	With WORKRS t-Value	Without WORKRS Coefficient	Without WORKRS t-Value
TTIME	1-3	-0.008 56	3.10	-0.008 65	3.30
COST/INC	1-3	-29.091	1.97	-24.881	1.70
CARS	1,3	0.421	3.10	0.365	2.80
WACCESS	2	0.350	0.84	0.292	0.73
CBD	1,3	-1.114	2.24	-0.892	1.87
WORKRS	2	0.403	5.61	-	-
INC	1,3	0.000 031 2	2.35	0.000 019 9	1.70
CONST	1	0.779	1.44	-0.102	0.20
CONST	3	-0.495	0.94	-1.359	2.80

[a]Alternatives: 1 = drive alone, 2 = bus, 3 = shared ride (≥2 occupants). Model used is multinomial logit fitted by maximum-likelihood method.

note on the nomenclature is in order. Two types of coefficients or post-BART models are used in the analyses. The first set of coefficients, called "true coefficients," were estimated by using the individual socioeconomic attributes and observed (manually coded) level-of-service attributes (TRUE-LOS). The second set of coefficients, called "network coefficients," were estimated by using the individual socioeconomic attributes and zone-to-zone network level-of-service attributes (NET-LOS). In addition, two specifications (coefficients) for Baltimore and pre-BART models are used. In the first, the number of workers in the household (WORKRS) is included; in the second, it is excluded from the model.] For detailed information regarding model specification and estimated coefficients of market-specific models, see the papers by Dehghani and Talvitie (1,2) and by Dehghani (4). The explanatory variables are defined below:

Variable	Definition
INVT	In-vehicle time or time spent inside a vehicle when traveling from origin to destination, door to door (round-trip time) (min)
WKT	Walk time to and from bus stop or rail station, in transfer, or to and from car's parking place (min)
WT	Sum of wait times of all transit vehicles, normally one half of first headway plus transfer wait (min)
TTIME	Sum of in-vehicle time, walk time, and wait time (min)
COST	Out-of-pocket travel cost (cents)
INC	Household income (dollars per year)
WORKRS	Number of workers in household
CARS	Number of cars owned
EMPD	Employment density in neighborhood (employees per acre)
CBD	Dummy variable constructed to differentiate trips destined to CBD from those destined to other locations (takes value of 1.0 if trips are destined to CBD, zero otherwise)
WACCESS	Walk access to transit facility (takes value of 1.0)
CONST	Constant (takes value of 1.0 for specified alternative, zero otherwise)

The supporting statistics are given below (the success index is the weighted average of differences in correct prediction between the full model and the model that has only alternative-specific constants). For Table 1:

Statistic	With WORKRS	Without WORKRS
L*(0)	-564.802	-564.802
L*(β)	-445.632	-463.02
Percent right (maximum utility classification)	64.89	62.5
Sample size	544	544
Success index	0.124	0.104
Proportion successfully predicted (expected value)	0.513	0.493
Proportion of prediction success due to variables other than alternative-specific constants	0.124/0.513 = 0.24	0.104/0.493 = 0.21

and for Table 2:

Statistic	With WORKRS	Without WORKRS
L*(0)	-1197.66	-1198.35
L*(β)	-776.98	-791.90
Percent right	66.33	65.67
Sample size	906	906
Success index	0.115	0.107
Proportion successfully predicted	0.523	0.514
Proportion of prediction success due to variables other than alternative-specific constants	0.22	0.21

MODEL ERROR

Definition

The model error captures errors caused by several factors. These are the specification error due to omitted variables and the model form (i.e., logit); the sampling errors in the model parameters; and the sampling errors in the estimated shares. Except for the sampling errors in the estimated shares, the other components of the model error are present in the forecasting situation also; they are an inherent part of the model. The possible sampling errors in estimated shares can be easily quantified and subtracted out of the model error. For example, SD (sampling error) for an alternative that has a sample size of 50 and an estimated share of 10 percent is $\delta_p = [0.10 \quad (1 - 0.10)/50.0]^{1/2} = 0.0424$, and the sampling error as a percentage of the estimated share is $(\delta_p/p) = (0.0424/0.10) = 0.424$. But in reality the error in estimated shares always exists, now as well as in the future. For this reason, the logic of subtracting it out of the model error is not self-evident.

The model error is calculated by using Equations 1 through 4 and the appropriate data set for this calculation in the estimation sample itself. The results are presented in Table 3.

It appears that the variable WORKRS, in spite of its statistical significance, does not reduce the

Table 2. Model specification and coefficients: pre-BART data.

Variable	Alternative Entered[a]	With WORKRS Coefficient	With WORKRS t-Value	Without WORKRS Coefficient	Without WORKRS t-Value
TTIME	1-4	-0.015 7	5.98	-0.015 3	5.96
COST/INC	1-4	-16.154	2.70	-16.251	2.70
CARS	1,4	1.326	7.50	0.928	6.0
EMPD	2,3	0.002 34	4.0	0.002 16	3.82
EMPD	1	-0.003 73	5.50	-0.003 76	5.50
WORKRS	2,3	0.925	5.31	-	-
CONST	1	0.955	2.99	-0.014 4	0.06
CONST	3	-1.437	-7.30	-1.441	7.35
CONST	4	-0.354	1.14	-1.338	5.48

[a]Alternatives: 1 = drive alone, 2 = bus and walk, 3 = bus and car, 4 = shared ride. Model used is multinomial logit fitted by the maximum-likelihood method.

Table 3. Prediction error due to model specification.

Market Segment	Model Coefficient	Baltimore Data: With WORKRS WAAE	WRMSE	Baltimore Data: Without WORKRS WAAE	WRMSE	Pre-BART Data: With WORKRS WAAE	WRMSE	Pre-BART Data: Without WORKRS WAAE	WRMSE	Post-BART Data: NET-LOS Coefficient WAAE	WRMSE	Post-BART Data: TRUE-LOS Coefficient WAAE	WRMSE
Total	C	12.1	16.3	14.0	18.3	15.1	22.8	15.3	22.6	17.8	25.8	14.0	22.0
High-income	C	15.4	20.8	17.6	23.1	16.0	29.6	16.0	29.6	37.0	27.8	18.0	27.8
	MS	17.1	22.7	14.4	20.4	19.7	52.5	19.7	52.5	13.0	21.2	13.0	20.0
Low-income	C	17.2	20.7	24.4	28.2	14.6	19.7	14.6	19.7	16.0	28.0	18.0	23.7
	MS	11.8	15.1	12.1	15.4	34.6	30.2	24.6	30.2	16.0	25.8	14.0	22.5
CBD	C	NA	NA	NA	NA	13.4	18.7	12.9	18.3	19.0	24.5	16.0	21.7
	MS	NA	NA	NA	NA	13.6	22.5	13.1	21.8	16.0	20.9	15.0	19.8
Non-CBD	C	NA	NA	NA	NA	17.7	38.7	18.5	36.8	17.0	35.9	14.0	27.9
	MS	NA	NA	NA	NA	10.9	17.0	10.9	17.0	15.0	29.6	13.0	26.6
One-car household	C	20.1	29.6	20.4	29.4	16.7	23.3	17.1	23.3	15.5	30.7	13.7	21.9
	MS	12.5	16.7	12.9	17.4	17.1	26.2	24.8	37.9	16.0	25.5	14.0	22.3
Household with two or more cars	C	14.6	19.9	15.0	20.3	16.8	40.6	16.2	41.6	17.8	38.0	14.0	29.1
	MS	10.5	15.8	10.5	15.6	14.2	24.5	24.5	32.8	14.0	27.3	13.0	23.8

Notes: NA = not applicable due to small sample size.
Model coefficients: C = common, MS = market-specific.

model error in either the Baltimore model or the pre-BART model. [Note that the variable WORKRS was selected not only for its statistical significance but for its contribution to improving the model's summary prediction success indices as well. Other specifications were also studied.] The post-BART model with TRUE-LOS (i.e., true coefficients) does have a smaller model error than the NET-LOS-based model does, but only marginally. The ranges of WAAE and WRMSE are 12.1-17.8 and 16.3-25.8, respectively. The median values of WAAE and WRMSE are 15.1 and 22.8. [Note that error measures, i.e., WAAE and WRMSE, are expressed as the percentage of predicted values and include the sampling error.]

The magnitude of WRMSE had been observed previously by Koppelman in his study of error analysis by using disaggregate choice models (3). WRMSE was 15.9 percent when the average sampling error was subtracted and 25.9 percent otherwise. Note that Koppelman performed the analysis by using 50 observations, on the average, per prediction group.

Market Segmentation

Examination of Table 3 by market segment shows that, again, the variable WORKRS does not reduce the model error except in a few isolated cases. The post-BART model that uses TRUE-LOS is slightly better than its network-based counterpart.

Overall, the grouping of travelers into population segments does not always reduce the model error. Market segmentation by car ownership but not by income does substantially reduce the model error, yet the resulting error is no less than the overall error of the common-market model. These results lead to one important inference: Statistical inequality of coefficients does not necessarily mean gross dissimilarity in the overall accuracy of predictions. It may be recalled that the model coefficients for the two income groups in the post-BART and Baltimore models were statistically unequal, yet the market-segment-specific models appear to perform only marginally better than the common-market model does. For the car-ownership groups the Baltimore model's travel-time coefficients and the coefficients as a group were statistically different, and in post-BART models the coefficients were statistically equal. Yet both models show smaller model error for the market-segment-specific model than for the common-market model. These contradictory results cannot be easily explained, and no attempt is made to do so.

Another observation is that the accuracy of the market-segment-specific model is often worse than the overall accuracy of the common-market model. This result and the previous results may be obtained from the reduction in sample size to roughly one-half for estimating the market-segment-specific models and reduces the precision of the coefficients. This increase in sampling error of the parameters offsets the decrease in taste variation presumably gained by market segmetation.

A second set of observations can be made with regard to model specification and type of data. (It may be recalled that the variable WORKRS was one of the variables that did improve the model's summary prediction-success indices. Also, the issue of model complexity is still under study.) It may be seen that the WORKRS variable, in spite of its statistical significance, does not systematically reduce the model error in either the pre-BART or the Baltimore models. The post-BART model that has true coefficients (i.e., observed rather than network LOS variables) has only marginally smaller model error than the network-based model error (it may be recalled that some of the alternative-specific constants were statistically different in these two post-BART models also). Without documentation it is also mentioned that the use of total travel time in place of excess and line-haul travel times with separate coefficients did not increase the model error; the values of WRMSE were 22.8 and 24.8 for the models that used observed and network data, respectively.

In conclusion, then, it may be said that market segmentation by car ownership seems somewhat promising in at least curtailing the model error, if not reducing it, but that minor improvements in model specification or type of data have no appreciable impact on the model error. It is to be noted, however, that the model error is only one component of the error; the other two components--aggregation error and transfer error--cannot be ignored when one attempts to assess the total error.

ERROR IN PREDICTION DUE TO ZONAL AVERAGING OF VARIABLES

The error in prediction due to zonal averaging of variables (aggregation error) was calculated from Equations 7 and 8 by comparing the predicted shares from the model by using observed LOS and/or socioeconomic data with the predictions from the same model by using zonally averaged LOS and/or socioeconomic attributes. The results can be seen in Table 4. Note that the values that appear in Table 4 are the net contributions of predicted errors due to averaging of these attributes. They are calculated as the differences in the WAAEs and WRMSEs due to the model alone (e.g., 15.3 and 22.6, respectively, from Table 3 for the unsegmented market of the pre-BART model) and the model error by using average values of the socioeconomic attributes (WAAE and WRMSE of 18.5 and 26.4 for the unsegmented market, respectively). For example,

$$WAAE_A = WAAE_{M+A} - WAAE_M = 18.5 - 15.3 = 3.2 \quad (9)$$

$$WRMSE_A = (WRMSE_{M+A}^2 - WRMSE_M^2)^{1/2} = (26.4^2 - 22.6^2)^{1/2} = 13.5 \quad (10)$$

WAAE and WRMSE are 3.2 and 13.5 percent of the predicted share, respectively. Note that the aggregation errors presented in Table 4 for Baltimore and pre-BART models are due only to the zonally averaged values of socioeconomic attributes because of the lack of non-network-based values for the LOS attributes. Note also that the zonal averages were calculated from the sample for the BART data but provided externally for the Baltimore sample.

In their analysis of prediction error, Talvitie, Dehghani, and Anderson (5) found that the overall prediction errors due to the use of zonally averaged values of the LOS and socioeconomic attributes were each about the same magnitude and had WRMSFs of 9.30 and 10.30, respectively. However, when the results were examined by market segment, the average LOS attributes sometimes reduced and sometimes increased the aggregation error of the models. One plausible explanation was the existence of a strong correlation between the true values of LOS attributes and socioeconomic attributes, such as wait time and car ownership. The use of zonal averages for socioeconomic data caused no error for many market segments. It was also noted that the calculation of the averages from the sample itself (which often contained only a few data points) and also sampling errors due to lightly used modes might have prevented the detection of the effect of averaging socioeconomic variables in that study.

The error committed by the use of zonally averaged values of independent variables is often referred to as "aggregation error by naive procedure." This aggregation error appears to vary for each data set used in this study. For overall predictions the ranges of WAAE and WRMSE were found (Table 4) to be from 3.2 to 9.6 and from 0.0 to 20.5; the median values were 4.0 and 12.7, respectively. Koppelman (3) obtained an error of 8.0 percent for the "naive" method of aggregation. It is worth noting that, except for the Baltimore data, the average socioeconomic variables are computed from the sample; they are not true zonal averages. Koppelman's study used network (i.e., average) LOS attributes.

The examination of error values given in Table 4 reveals that market segmentation does not necessarily reduce the aggregation error. There are some cases in which market segmentation has substantially increased the aggregation error. Visual examination of Tables 3 and 4 suggests that there is an inverse dependency between model error and aggregation error. If market segmentation reduces model error, it increases aggregation error, and vice versa. It appears that, for some reason, market segmentation by income is the most desirable if the objective is to reduce aggregation error only.

The most interesting result in Table 4 concerns the size of the aggregation error for the "true versus network" coefficients. It is seen that, nearly uniformly, the use of network coefficients results in a lower aggregation error and that furthermore this aggregation is often zero. This result is not totally unanticipated. Talvitie (6) showed that unbiased forecasts are possible even with (biased) network coefficients, provided that the curvature of the logit model is not too large and out-of-range forecasts are not required. Which types of coefficients result in more-accurate forecast can be studied only by examining model-transfer errors. The magnitude of these transfer errors is examined next.

ERRORS DUE TO MODEL TRANSFER

The transfer error is calculated by applying the mode-choice models presented in Tables 1 and 2 to predict modal shares that have values of LOS and socioeconomic attributes by using Twin Cities, pre-BART, post-BART, and Baltimore data from 1970, 1972, 1975, and 1980, respectively. It is noted that the transferability being studied concerns transferability over both space and time (at most for eight years). The results are presented in Tables 5 and 6.

It can be seen from Tables 5 and 6 that the overall magnitude of transfer error in predictions is large when the models' coefficients are applied to Twin Cities data. The values of WAAEs and WRMSEs

Table 4. Prediction error due to aggregation.

		Baltimore Data[a]				Post-BART Data[b]				Pre-BART Data[a,b]	
		With WORKRS		Without WORKRS		NET-LOS Coefficient		TRUE-LOS Coefficient		Without WORKRS	
Market Segment	Model Coefficient	WAAE	WRMSE	WAAE	WRMSE	WAAE	WRMSE	WAAE	WRMSE	WAAE	WRMSE
Total	C	7.0	16.5	9.6	20.5	5.0	0.0	4.0	12.7	3.2	13.5
High-income	C	1.5	0.0	4.9	0.0	25.0	0.0	10.0	8.9	0.0	0.0
	MS	5.0	0.0	4.03	12.8	2.0	7.5	15.0	27.6	0.6	0.0
Low-income	C	34.4	50.6	25.7	46.1	6.0	15.5	0.0	14.3	3.3	14.3
	MS	25.9	36.6	19.9	31.7	0.0	0.0	4.0	14.0	1.6	8.9
CBD	C	NA	NA	NA	NA	1.0	8.7	9.0	25.8	6.9	19.7
	MS	NA	NA	NA	NA	1.0	4.1	6.0	17.0	5.5	19.0
Non-CBD	C	NA	NA	NA	NA	2.0	0.0	5.0	0.0	1.4	6.1
	MS	NA	NA	NA	NA	1.0	16.3	4.0	9.9	1.5	8.5
One-car	C	3.0	0.0	0.0	0.0	0.0	0.0	2.8	9.8	0.7	9.3
household	MS	19.3	36.0	17.8	34.6	2.0	11.6	8.0	21.7	2.0	0.0
Household	C	0.0	0.0	8.9	20.0	0.0	0.0	6.8	6.9	0.8	0.0
with two or more cars	MS	19.6	33.0	17.6	28.9	1.0	0.0	7.0	6.6	6.3	23.5

Note: NA = not applicable due to small sample size.
[a]LOS variables were from the networks and already are zonal averages; aggregation error is due to averaging of socioeconomic attributes only.
[b]Socioeconomic averages from sample.

Table 5. Prediction error due to model transfer: Baltimore model.

Market Segment	Model Coefficient	Pre-BART Data: With WORKRS WAAE	Pre-BART Data: With WORKRS WRMSE	Pre-BART Data: Without WORKRS WAAE	Pre-BART Data: Without WORKRS WRMSE	Post-BART Data: NET-LOS Coefficient WAAE	Post-BART Data: NET-LOS Coefficient WRMSE	Post-BART Data: TRUE-LOS Coefficient WAAE	Post-BART Data: TRUE-LOS Coefficient WRMSE	Twin Cities Data WAAE	Twin Cities Data WRMSE
Total	C	4.2	16.5	8.5	22.2	16.1	32.6	6.0	25.6	44.5	57.9
High-income	C	4.7	19.8	7.5	25.6	16.6	33.6	0.4	18.8	29.9	44.6
	MS	16.7	33.3	7.4	22.3	17.8	34.3	0.7	16.8	30.6	43.8
Low-income	C	2.0	8.3	5.5	0.0	2.2	0.0	8.4	0.0	29.6	49.7
	MS	5.7	15.9	14.8	26.8	19.5	33.8	10.0	22.3	44.8	57.6
One-car household	C	0.0	0.0	0.0	18.4	19.4	28.3	6.1	17.9	43.2	58.42
	MS	6.3	19.3	7.0	17.5	43.7	51.4	16.9	29.4	40.5	52.4
Household with two or more cars	C	1.3	18.1	11.4	31.8	12.0	6.2	3.6	12.6	38.1	51.33
	MS	10.3	22.8	16.8	31.7	42.4	55.0	10.1	27.1	25.6	37.8

Table 6. Prediction error due to model transfer: pre-BART model.

Market Segment	Model Coef-ficient	Baltimore Data with WORKRS: TRUE SE Variable WAAE	WRMSE	Baltimore Data with WORKRS: Avg SE Variable WAAE	WRMSE	Baltimore Data Without WORKRS: TRUE SE Variable WAAE	WRMSE	Baltimore Data Without WORKRS: Avg SE Variable WAAE	WRMSE	Post-BART Data: NET-LOS Coefficient WAAE	WRMSE	Post-BART Data: TRUE-LOS Coefficient WAAE	WRMSE	Twin Cities Data WAAE	WRMSE
Total	C	0.6	0.0	18.3	38.6	5.4	18.4	15.9	32.9	2.1	17.4	2.9	23.6	30.5	43.7
High-income	C	2.0	0.0	0 0	0.0	6.4	17.9	0.0	0.0	4.6	17.5	2.8	19.2	3.9	12.2
	MS	1.6	0.0	0.0	0.0	3.7	0.0	0.0	0.0	4.9	19.8	2.6	4.6	6.8	0.0
Low-income	C	2.5	0.0	27.2	47.6	4.9	11.8	34.9	56.2	1.7	12.4	1.6	10.3	21.8	40.7
	MS	10.8	0.0	19.6	48.9	2.6	18.9	22.2	53.0	12.7	28.9	8.7	28.2	15.9	39.6
CBD	C	NA	NA	NA	NA	NA	NA	NA	NA	18.1	33.8	15.1	30.9	21.2	33.4
	MS	NA	NA	NA	NA	NA	NA	NA	NA	30.2	60.6	28.4	54.6	28.2	38.9
Non-CBD	C	NA	NA	NA	NA	NA	NA	NA	NA	4.2	0.0	5.4	0.0	26.2	31.7
	MS	NA	NA	NA	NA	NA	NA	NA	NA	11.4	38.4	5.0	23.3	42.4	53.3
One-car household	C	9.4	27.2	7.3	16.2	8.7	20.0	5.8	14.0	2.6	13.8	0.0	29.4	36.1	50.8
	MS	7.5	17.6	5.8	10.4	4.7	5.1	0.0	6.1	0.1	0.0	10.0	0.0	42.2	57.0
Household with two or more cars	C	1.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3	21.2	17.1	0.0
	MS	3.3	11.6	13.1	60.3	11.9	45.1	10.3	42.8	10.3	.9.0	10.3	0.0	9.4	20.0

are 44.5 and 57.9 and 30.5 and 43.7 by using Baltimore and pre-BART models, respectively. In general, the median value of error for the overall prediction due to model transfer is about the same magnitude as model misspecification with WRMSE of 23.6 percent, or about the same as the model error.

The inclusion of the variable WORKRS in model specification does not consistently reduce the transfer error. However, it does reduce it often and independent of whether true socioeconomic variables or zonal averages are used. It is noted that it is the latter type of variable that is a standard in transportation studies.

Interesting results are obtained with respect to the use of zonal averages of the socioeconomic and LOS attributes. It can be seen from Table 6 that WAAEs and WRMSEs for model transfer obtained by using the Baltimore data and zonally averaged socioeconomic attributes are about the same and even smaller than the counterpart values obtained by using disaggregate data. For example, the WAAEs and WRMSEs for the common-market pre-BART model to the Baltimore data for the above two cases are 5.4 and 18.4 and 15.9 and 32.9, respectively, when the number of workers is excluded from the model. Thus, the use of zonally averaged socioeconomic attributes seems to result in forecasts that are somewhat worse than those that use disaggregate values. It can also be seen from Tables 5 and 6 that the use of true LOS attributes results in more-accurate forecasts that the use of network-based values does for the Baltimore model but not for the pre-BART model. This section can be concluded with a remark about the market segmentation. Taken together, Tables 5 and 6 suggest that the application of the pre-BART model favors market segmentation by car ownership but the application of the Baltimore model does not support such market segmentation in order to reduce the transfer error.

TOTAL PREDICTION ERROR

As interesting as the examination of different error sources is, the question uppermost in a practitioner's mind is the total prediction error and whether it can be reduced by market segmentation or other means. So far, we have obtained conflicting information. Market segmentation by car ownership or income may reduce the model specification error somewhat. This was suggested by both the statistical tests of the equality of model parameters and the RMSEs in Table 3. Aggregation error, on the other hand, seems to increase as a result of market segmentation by car ownership. Finally, market segmentation sometimes increases and sometimes decreases the transfer error.

The distribution of the total error among the three error sources is also unsystematic. Table 7 shows the total error for the entire travel market and its distribution among the three sources by using the common-market model. If one wants to glean an average from Table 7, one would assign 40 percent of the total error to model and transfer error and assign the remaining 20 percent to aggregation error. With market-segment-specific models the distribution of the total error shifts 10 percent of

Table 7. Total prediction error.

	Baltimore Model					Pre-BART Model				
	Pre-BART Data		Post-BART Data			Baltimore Data Without WORKRS		Post-BART Data		
Error Type	With WORKRS	Without WORKRS	NET-LOS	TRUE-LOS	Twin Cities	TRUE SE Variable	Avg SE Variable	NET-LOS	TRUE-LOS	Twin Cities
Total Market										
WAAE	23.8	32.1	39.7	29.6	68.1	23.9	34.4	20.6	21.4	49.0
WRMSE	28.5	35.3	42.6	37.6	64.0	32.0	42.13	31.6	35.3	51.0
Model Error (%)										
WAAE	51.0	44.0	35.0	47.0	20.0	25.0	45.0	74.0	70.0	31.0
WRMSE	33.0	27.0	18.0	24.0	8.0	50.0	29.0	51.0	41.0	20.0
Aggregation Error (%)										
WAAE	32.0	30.0	24.0	32.0	14.0	13.0	9.0	16.0	15.0	7.0
WRMSE	33.5	34.0	23.0	80.0	10.0	18.0	10.0	18.0	15.0	7.0
Transfer Error (%)										
WAAE	17.0	26.0	41.0	21.0	66.0	22.0	46.0	10.0	15.0	62.0
WRMSE	33.5	39.0	59.0	46.0	82.0	32.0	61.0	31.0	44.0	73.0

the model error to the transfer error.

One has the feeling that the numbers tell two things. First, and most conspicuous, if the model transfers poorly due to the overwhelming influence of the alternative-specific constants (that is, drastically different model shares between the estimation data and the transfer data), then the transfer error is dominant. Second, if there is substantial within-zone variation in the LOS data, then the aggregation error is large. This is often the case for low-income or CBD-bound travel in which the number of transit users, and hence great variance in excess time components, exists.

In general, the message is that the total prediction error is large and little is gained by market segmentation and complex specification. Good judgment in model application and careful preparation of data are keys to forecasting success; even then the forecasts are marked with uncertainty.

CITYWIDE PREDICTIONS BY MODE

It would be inappropriate to conclude this paper without taking a brief look at the predictions (transferability) of the models by mode of travel between different cities. The more complex calculations of RMSE include the variations in individual predictions and provide a convenient one-number measure of forecasting accuracy. On the other hand, such a one-number measure seems to hide information and prevent drawing useful conclusions. The simple share predictions are easy to calculate and comprehend and provide results that seem to be in accordance with statistical tests of coefficient equality or inequality. Results are given in Table 8 (Baltimore and pre-BART models) for the unsegmented market. The results in Table 8 tell us that the WORKRS variable does make a contribution to the model accuracy, especially for the transit share. The two post-BART experiments show that the observed (true) LOS attributes make an important contribution to model accuracy. It may be recalled that the post-BART network coding was found to be quite different from the manual coding and, by inference, faulty. The results here confirm this inference.

The Twin Cities predictions are the worst of all. The reasoning is that, because the actual shares in Twin Cities are so different from the shares in the estimation sample and because the alternative-specific constants account for the bulk of the model power, that alone renders the estimated model nontransferable to cities that have vastly different modal use.

The same type of results can be read from Table 8

Table 8. Citywide predictions by mode.

	Mode			
Data Source	Drive Alone	Transit	Shared Ride	N
Baltimore Model				
Predicted (with WORKRS)	0.53	0.28	0.19	900
Predicted (without WORKRS)	0.51	0.32	0.17	900
Actual	0.55	0.24	0.21	
Post-BART (NET-LOS)				
Predicted	0.48	0.35[a]	0.17	623
Actual	0.57	0.22[a]	0.21	
Post-BART (TRUE-LOS)				
Predicted	0.54	0.28[b]	0.18	565
Actual	0.55	0.25[b]	0.20	
Twin Cities				
Predicted	0.56	0.27	0.17	665
Actual	0.86	0.05	0.09	
Pre-BART Model				
Baltimore				
Predicted (with WORKRS and TRUE SE)	0.50	0.33	0.17	544
Predicted (without WORKRS and TRUE SE)	0.57	0.21	0.22	544
Actual	0.51	0.29	0.20	
Baltimore				
Predicted (with WORKRS and avg SE)	0.64	0.17	0.19	561
Predicted (without WORKRS and avg SE)	0.63	0.19	0.18	561
Actual	0.51	0.29	0.20	
Post-BART (NET-LOS)				
Predicted	0.55	0.24[c]	0.21	623
Actual	0.56	0.23[c]	0.21	
Post-BART (TRUE-LOS)				
Predicted	0.56	0.23[d]	0.21	565
Actual	0.55	0.25[d]	0.20	
Twin Cities				
Predicted	0.64	0.17	0.19	665
Actual	0.87	0.05	0.08	

[a]The predicted and actual BART shares are 0.12 and 0.08, respectively.
[b]The predicted and actual BART shares are 0.09 and 0.10, respectively.
[c]The predicted and actual BART shares are 0.08 and 0.084, respectively.
[d]The predicted and actual BART shares are 0.07 and 0.098, respectively.

for the pre-BART model. Again, the WORKRS variable makes a contribution to the forecasting accuracy, as is evident from the first two rows. The next two rows provide a partial contradiction; if zonal averages are used for socioeconomic variables, the advantage of the better model specification is lost. The behavior of the model is exactly according to the theory; small shares (<0.50) are predicted as being even smaller and large shares (>0.50) are predicted as being even larger than they really are.

The predictions for the post-BART situation are excellent and, again, contradictory to the results obtained by using the Baltimore model. An assumption can be made that the extra mode-specific constant available in the pre-BART model is very helpful. The prediction of the Twin Cities modal shares is done as poorly as the case with the Baltimore model. The comment made then applies now, too. Constants were assigned as follows:

1. Alternatives in estimation samples:
   a. Drive alone, shared ride, bus
   b. Drive alone, shared ride, local bus, express bus
2. Model whose modal constant was assigned to rail or express bus, or both:
   a. Shared ride
   b. Express bus

An interesting complement to these predictions is provided by the market-segment-specific models. The results of the application of the Baltimore and the pre-BART models are shown in Table 9 for the case in which the modal shares for the one-car households and the households that have two or more cars are aggregated. Note that no-car households are not included in this table. Examination of the data in Table 9 shows that market-segment-specific (Baltimore) models are better predictors only occasionally; generally the common-market models are better or at least consistent. Two other things also stand out: The more-accurate true LOS variables yield much better predictions, and the Twin Cities' predictions remain very poor.

The comments made above apply here for both types of pre-BART models as well. To repeat, the market-segment-specific models are not better, the observed LOS variables are better, and the Twin Cities predictions are poor.

The fact that the use of zonal averages has such a drastic detrimental impact on forecasting accuracy, especially on the drive-alone and the bus predictions, would merit serious study. However, to do so would require the assessment of the errors and differences in the extraneous zonal averages versus those calculated from the sample itself. This was felt to be outside the scope of the present study. At any rate, zonal averages and market segmentation do not mix.

CONCLUSIONS

The results presented here are complicated, but the following conclusions can be drawn on the basis of the predictions. Data accuracy is clearly important; this is shown by the clear superiority of the true LOS data over the often glaringly erroneous coding of the post-BART network.

The use of zonal averages in predictions seems to be a source of serious concern. Unfortunately, it is not known to what extent data error rather than the model or aggregation error is responsible for the results. The fact that the Baltimore sample is a stratified random sample should also be factored in, and this was not done here. An interesting thought is to add up the drive-alone and shared-ride percentages and use a car-occupancy model to convert travelers to vehicles. Of course, this could not be done in all applications.

Finally, there is the fact that the common-market model performed very well; the Twin Cities data were the only (occasional) exception. This result argues in favor of aggregated total-market forecasts.

Much remains to be done to ensure accuracy in travel forecasts. Foremost among these is the updating of modal constants to apply in cases when the modal shares in the estimation sample are very different from those likely to be experienced in the city in which the model is to be transferred. The second item that needs constant attention is the accuracy of both the socioeconomic and LOS data. Data must be carefully prepared if forecasting errors are to be avoided. These are the first steps.

ACKNOWLEDGMENT

During the course of this research we have benefited greatly from the assistance and contributions of Antti Talvitie, professor and chairman of the Civil Engineering Department at the State University of New York at Buffalo. This research was supported in part by a Department of Transportation contract to the State University of New York at Buffalo.

Table 9. Predication of modal shares by using common and market-segment-specific coefficients: households with one and with two or more cars.

	Mode								
	Drive Alone			Transit			Shared Ride		
Data Source	C	MS	Actual	C	MS	Actual	C	MS	Actual
Baltimore Model									
Pre-BART (with WORKRS)	0.55	0.51	0.59	0.28	0.27	0.19	0.17	0.22	0.22
Pre-BART (without WORKRS)	0.55	0.51	0.59	0.28	0.27	0.19	0.17	0.22	0.22
Post-BART	-	-	0.59	-	-	0.19[a]	-	-	0.22
NET-LOS	0.51	0.37	-	0.32[b]	0.43[c]	-	0.17	0.20	-
Obs-LOS	0.58	0.49	-	0.23	0.27	-	0.19	0.24	-
Twin Cities	0.57	0.64	0.87	0.24	0.13	0.05	0.19	0.23	0.08
Pre-BART Model									
Baltimore	-	-	0.59	-	-	0.21	-	-	0.20
With WORKRS	0.54	0.62	-	0.30	0.22	-	0.16	0.16	-
Without WORKRS	0.61	0.70	-	0.22	0.13	-	0.17	0.17	-
Post-BART	-	-	0.60	-	-	0.19[d]	-	-	0.21
NET-LOS	0.59	0.51	-	0.18[e]	0.29[f]	-	0.23	0.20	-
Obs-LOS	0.63	0.60	-	0.14	0.19	-	0.23	0.21	-
Twin Cities	0.65	0.61	0.87	0.16	0.21	0.05	0.19	0.18	0.08

[a]The actual BART share is 0.09.
[b]The predicted NET-LOS and Obs-LOS BART shares are 0.11 and 0.08, respectively.
[c]The predicted NET-LOS and Obs-LOS BART shares are 0.24 and 0.14, respectively.
[d]The actual BART share is 0.09.
[e]The predicted NET-LOS and Obs-LOS BART shares are 0.04 and 0.04, respectively.
[f]The predicted NET-LOS and Obs-LOS BART shares are 0.14 and 0.11, respectively.

REFERENCES

1. Y. Dehghani and A. Talvitie. Model Specification, Modal Aggregation, and Market Segmentation in Mode-Choice Models: Some Empirical Evidence. TRB, Transportation Research Record 775, 1980, pp. 28-34.
2. Y. Dehghani and A. Talvitie. Model Specification and Market Segmentation in Modal Choice Models with Baltimore Data. Department of Civil Engineering, State Univ. of New York at Buffalo, Working Paper, 1980.
3. F. Koppelman. Methodology for Analyzing Errors in Prediction with Disaggregate Choice Models. TRB, Transportation Research Record 592, 1976, pp. 17-23.
4. Y. Dehghani. Prediction, Models, and Data: An Analysis of Disaggregate Choice Models. Department of Civil Engineering, State Univ. of New York at Buffalo, Ph.D. dissertation, 1980.
5. A. Talvitie, Y. Dehghani, and M. Anderson. An Investigation of Prediction Errors in Mode Choice Models. Presented at the International Conference on Research and Applications of Disaggregate Travel Demand Models, Univ. of Leeds, England, July 1980.
6. A. Talvitie. Disaggregate Travel Demand Models with Disaggregate Data, Not with Aggregate Data and For What. Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper 7615, 1976.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.

Equilibrium Model for Carpools on an Urban Network

CARLOS F. DAGANZO

Traffic equilibrium methods are presented in which the population of motorists consists of individuals who are minimizers of a linear combination of cost and travel time. The relative importance of travel time versus cost varies across the population, but fairly mild conditions for the existence and uniqueness of the equilibrium can nevertheless be identified. The paradigm is of particular interest for carpooling studies because the occupants of carpools can divide the cost among themselves but they cannot do the same with the travel time. Thus, vehicles that have different occupancy levels will have different relative values of travel time and cost. The model is specially well suited to the analysis of how vehicles that have different occupancies compete for segments of the roads that are crowded or have tolls. It is therefore very useful to predict the impacts of special carpooling lanes, lower tolls for high-occupancy vehicles, and other transportation-system-management strategies on the distribution of traffic over an urban network.

Current traffic-assignment practice takes two principal forms, which are applicable to congested and uncongested networks. Stochastic traffic-assignment models (1-5) ignore congestion but do not allocate all the traffic from an origin-destination (O-D) pair to the shortest route. Instead, they spread it over the network as if travel time was perceived with some random noise by a motorist population of travel-time minimizers.

Deterministic-equilibrium models assume that motorists are accurate minimizers of travel time but that travel time depends on the traffic flow because of congestion. Textbook-level treatments of deterministic equilibrium models can be found (6-9). The equilibrium condition for these models was stated by Wardrop (10). It can be paraphrased as follows: at equilibrium (a) routes that have flow are the shortest routes, or (b) no user can improve route travel time by unilaterally changing routes, or (c) links that have flow for a given destination are on a shortest path to the destination. Since a problem that is more closely related to deterministic-equilibrim models than to stochastic-assignment models will be addressed here, the discussion of the former is expanded below. A question that arises immediately is that of the existence and uniqueness of an equilibrium-flow pattern that satisfies all three equilibrium conditions.

Beckmann, McGuire, and Winsten (6); Netter (11); and Smith (12) have provided progressively more general existence results. It is currently known that if travel time on every link of the network is a continuously differentiable positive function of the link flows, Brouwer's fixed-point theorem guarantees the existence of the equilibrium flows.

Uniqueness was first studied for networks in which the travel time on a link depends only on its own flow (6). In this case and if travel time increases with flow for all links, the equilibrium exists and the resulting link-flow pattern is unique. This is because the equilibrium problem admits a formulation as the minimization of a strictly convex function subject to linear constraints. This formulation can be expressed in terms of link flows as follows:

$$\text{(MP)}\ \min \sum_i \int_0^{x_i} c_i(w)dw$$

subject to

$$\sum_{i\in I(r)} x_i^s - \sum_{i\in E(r)} x_i^s = q^{rs} \quad \forall r \neq s, \forall s$$

$$\sum_s x_i^s = x_i \quad \forall i$$

$$x_i^s \geq 0 \quad \forall i, s$$

In this program, the letters r and s represent nodes, and the letter i represents a link. I(r) represents the set of links that point to node r; E(r), the set of links that point out of node r; and $c_i(\cdot)$, the link-cost function that relates the flow on link x_i to the link travel time c_i. In addition, x_i^s is the total number of trips that have final destinations s and that use link i, and q^{rs} is the total number of trips that go from origin r to destination s.

In order to write equilibrium problems more succinctly, the set of feasible link-flow patterns is denoted by X; thus, program (MP) is written as follows:

$$\text{(MP)}\ \min_{x\in X} \sum_i \int_0^{x_i} c_i(w)dw$$

Link flows that are optimal for (MP) are equilibrium flows (and vice versa) because the Kuhn-Tucker conditions of (MP) are the mathematical expressions for Wardrop's principle as paraphrased under item (c) at the outset of this paper. This happens because the partial derivatives of the objective function are the link-cost functions, as follows:

$$\partial/(\partial x_j) \sum_i \int_0^{x_i} c_i(w)dw = c_j(x_j); \forall j \quad (1)$$

These results can be generalized for models in which link costs depend on the flows of other links. Dafermos (13) seems to have been the first to have studied this class of problems. She showed that the equilibrium problem admits an extremum formulation if the link-cost functions satisfy a condition similar to that in Equation 1. That is, if there is a function C(x) whose partial derivatives are the link-cost functions $\partial C(x)/\partial x_i = c_i(x)$; $\forall i$, the equilibrium problem is as follows:

$$\text{(MP)} \min_{x \epsilon X} C(x)$$

To solve equilibrium problems, one does not have to find the function C(x), since to solve (MP) only the derivatives of C(x) are necessary. Furthermore, the existence of C(x) can be verified from the (continuous) cross-derivatives of $c_i(x)$:

$$C(x) \quad \text{exists if } [\partial c_i(x)/\partial x_j] = [\partial c_j(x)/\partial x_i],\ \forall i,j$$

The uniqueness of the equilibrium link-flow pattern x* can be established from the strict convexity of C(x) or the positive definiteness of the Jacobian $J(x) = [\partial c(x)/\partial x]$. That is, if J(x) is symmetric, there is an extremum formulation and if it is positive definite, the equilibrium solution is guaranteed to be unique. Recent research shows that this uniqueness condition holds even if J(x) is not symmetric (12,14).

Another area of research that is closely connected is multimodal-equilibrium models. In these models, each vehicle type has a different impact on the overall congestion and imposes a different amount of delay on vehicles that share the road with it. In addition, vehicles of different types may exhibit different link travel times under the same link congestion. The most general formulation, short of letting vehicles on a link affect the travel times on another link (11), assumes that one has K vehicle types and that the travel time on link i for the kth vehicle class $c_i^{(k)}$ is as follows:

$$c_i^{(k)} = c_i^{(k)}[x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(K)}] \quad (2)$$

For example, if k = 1 represents automobiles and k = 2 represents trucks, one would expect $c_i^{(1)}$ to be smaller than $c_i^{(2)}$ for any combination of $x_i^{(k)}$'s, and one would also expect $x_i^{(2)}$ to influence $c_i^{(k)}$ more than $x_i^{(1)}$. Hypothetical curves could be as follows:

$$c_i^{(1)} = 100 + x_i^{(1)} + 5x_i^{(2)} \quad (3a)$$

$$c_i^{(2)} = 150 + 1.1[x_i^{(1)} + 5x_i^{(2)}] \quad (3b)$$

in which a truck is depicted as having the same effect on congestion as that of five passenger cars but also requiring more time units to travel the same distance.

Uniqueness results for multimodal networks can also be derived. Since one can visualize each traffic type as moving on its own transportation network, Equation 2 can be interpreted as an interaction among links of a network that consists of K copies of the original network instead of an intermodal interaction. With this mental picture, it is easy to see that multimodal networks are special cases of the single-mode network model that have general link-cost functions. Therefore, they share the same existence and uniqueness results (15). That is, the following equation guarantees existence of (MP) if the derivatives are continuous:

$$[\partial c_i^{(k)}/\partial x_j^{(\ell)}] = [\partial c_j^{(\ell)}/\partial x_i^{(k)}],\ \forall (i,j,k,\ell) \quad (4a)$$

$$J(x) = \partial[\ldots, c_i^{(k)}, \ldots]/\partial[\ldots, x_j^{(\ell)}, \ldots] \quad (4b)$$

Because $c_i^{(k)}$ depends only on the flows of link i, Equations 4 can be simplified as follows:

$$[\partial c_i^{(k)}/\partial x_i^{(\ell)}] = [\partial c_i^{(\ell)}/\partial x_i^{(k)}],\ \forall (i,k,\ell) \quad (5a)$$

$$J(x) = \begin{bmatrix} J_1(x) & 0 & \\ 0 & J_2(x) & \\ & & \ddots \end{bmatrix};\ J_i(x) = \partial[\ldots, c_i^{(k)}, \ldots] \div \partial[\ldots, x_i^{(\ell)}, \ldots] \quad (5b)$$

For uniqueness, it is sufficient thus that for all links, $J_i(x)$ be positive definite. Of course, the symmetry of all $J_i(x)$'s would also guarantee the existence of an extremal formulation. Unfortunately, these conditions are much too restrictive for multimodal networks because the off-diagonal terms of $J_i(x)$ can be large and asymmetric (14). For example, the hypothetical link defined by Equations 3 yields the following:

$$J_i(x) = \begin{bmatrix} 1 & 5 \\ 1.1 & 5.5 \end{bmatrix}$$

which violates the conditions because it is neither symmetric nor positive definite. Typically, vehicles of different sizes will result in asymmetric nondefinite Jacobians as in the example. In recognition of these problems, papers on multimodal public and private traffic-assignment problems tend to focus on computational schemes to finding equilibrium solutions but always recognize that multiple equilibria may exist [see papers by Florian (16) and Abdulaal and LeBlanc (17), for example].

It is shown next that there is a family of link-cost functions that have symmetric semidefinite Jacobians that well describe multimodal networks of similar-size vehicles and have application to carpooling problems.

The generalization of this family to vehicles of different sizes that is mentioned in the conclusion is in agreement with Jeevanantham's conjecture for general networks (18).

CARPOOLING MODEL

Assume that Equation 2 is of the following form:

$$c_i^{(k)} = c_{oi}^{(k)} + c_i(x_i);\ x_i = x_i^{(1)} + \ldots + x_i^{(K)} \quad (6)$$

where $c_i(x_i)$ is continuously differentiable and $c_{oi}^{(k)}$ can represent a constant that is independent of flow but can vary across traffic classes. Note that Equation 6 specifies that all traffic types have the same impact and are affected equally by congestion. Then

$$J_i(x) = [\partial c_i/\partial x_i] \begin{bmatrix} 1 & 1 \ldots 1 \\ 1 & 1 \ldots 1 \\ \cdots & \cdots \\ 1 & 1 \ldots 1 \end{bmatrix}$$

and if $c_i(x_i)$ is increasing, $J_i(x)$ is a positive semidefinite symmetric matrix. That $J_i(x)$ is positive semidefinite is seen by noting that for any vector, $\alpha = (\alpha_1 \ldots \alpha_K)$:

$$\alpha[J_i(x)]\alpha^T = [\partial c_i/\partial x_i] \left[\sum_{k=1}^{K} \alpha_k\right]^2 \geq 0$$

This implies that the equivalent minimization problem is a convex programming problem with a set of equilibrium solutions that is convex. Equilibrium in terms of modal flows $[\ldots, x_i^{(k)}, \ldots]$ is not necessarily unique (as it is not for the route flows for the single-mode problem) because C(x) is not strictly convex. Nevertheless, it is possible to show that all equilibrium-flow patterns must have the same link costs $c_{oi}^{(k)}$ and total link flows x_i (19).

In practical applications, the constants $c_{oi}^{(k)}$ may represent a number of things, including direct costs (expressed in travel-time units) that are independent of flow and may change across the motoring population. For example, the model could be applied to study a futuristic scenario in which a mixture of roadway powered vehicles (RPVs) and internal combustion engine vehicles share a transportation network. An RPV is an electrically powered vehicle that can draw its power from specially equipped links of the network. If we assume that these vehicles do not pay every time they use these special links (presumably they would be taxed differently from gasoline-powered vehicles), their routing incentive will tend to deviate from shortest routes within reason to take advantage of the lower operating costs on those links. The value of $c_{oi}^{(k)}$ on such links will be small for RPVs and relatively larger for gasoline-powered vehicles. On the standard links of the network, the values of $c_{oi}^{(k)}$ would be similar for both vehicle types.

The type of model implied by Equation 6 is particularly useful to study the effects of current transportation-system-management (TSM) strategies to encourage carpooling. In this case, the index k represents the number of people in an automobile and $c_{oi}^{(k)}$ represents the cost to any one of the occupants. If, reasonably, we assume that the k persons in the carpool divide all the costs (tolls and mileage, mainly) proportionately, $c_{oi}^{(k)}$ can be expressed as follows:

$$c_{oi}^{(k)} = \alpha^{(k)}\{[\tau_i^{(k)} + \beta^{(k)} d_i]/k\}, k = 1, 2, \ldots K \qquad (7)$$

where $\tau_i^{(k)}$ represents the tolls (if any) on link i, d_i represents the distance of link i, and $\beta^{(k)}$ and $\alpha^{(k)}$ are factors that convert distance traveled into monetary units and monetary units into travel-time units, respectively.

The following TSM strategies can be studied:

1. Differential tolls,
2. Ramp metering,
3. Special lanes for high-occupancy vehicles, and
4. Parking privileges for carpools.

To model the effect of tolls that depend on vehicle occupancy, one defines $\tau_i^{(k)}$ accordingly. For example, if (βd_i) is negligible and no toll is levied for vehicles that have three or more occupants, we have the following:

$$c_{oi}^{(1)} = \alpha\tau, \quad c_{oi}^{(2)} = 1/2\,\alpha\tau, \quad c_{oi}^{(3)} = c_{oi}^{(4)} = \ldots = 0$$

To model differential treatment of vehicles that have different occupancy levels in a ramp-metering situation (e.g., vehicles that have more than three occupants may bypass the metering queue), one should represent the metered link as two parallel links--one that is not metered and is restricted to cars that have more than two passengers and a metered link for all other vehicles. To forbid the use of the unmetered link to motor vehicles of types 1 and 2, one simply sets a very high toll for these vehicle classes. For example, if the distance component is negligible and the original link is represented by metered link i and unmetered link i', one would have the following:

$$c_i^{(k)} = c_i(x_i), \quad c_{i'}^{(k)} = c_{oi'}^{(k)}$$

where $c_i(x_i)$ represents the delays encountered at the metering ramp when the metered flow is x_i and

$$c_{oi'}^{(k)} = 0 \quad \text{if } K \geq 3$$
$$= M\ (M \to \infty) \quad \text{if } k \leq 3$$

To model lanes for high-occupancy vehicles, one represents the special lane by a separate link and in the same way assigns it a very high differential toll that is applied only to vehicle classes that are forbidden to use it. Special parking privileges for carpools can be modeled similarly by assigning classes that are not allowed to park a very high fixed cost on links that go into the parking lot.

Example

Figure 1A is a graphic representation of the transportation problem from Marin County to San Francisco. It displays the central business district (SF) and a suburb (M) of a metropolitan area that are separated by a toll bridge. The central business district (CBD) can also be reached in 10 time units by using a ferry system. The ferry fleet is supposed to be large enough (or flexible enough) to guarantee this travel time independent of flow. The ferry fare is neglected, but the over-land taxicab fare (assessed jointly to the members of a carpool) is 10 monetary units. We assume that α equals 1. The toll bridge, on the other hand, is so short that its distance and free-flow travel time are negligible. However, congestion sets in very quickly and, for flow different from zero, the travel time is equal to the flow. Figure 1B summarizes this information. It also displays the parameters α and β and the O-D table for the morning rush hour: 10 vehicles per unit time that have one occupant and 10 more that have two occupants; all the traffic goes from M to the CBD.

The cost functions are as follows:

$$c_1^{(1)} = \tau + x_1 \qquad c_2^{(1)} = 20$$

$$c_1^{(2)} = (\tau/2) + x_1 \qquad c_2^{(2)} = 15$$

where τ is the toll on the bridge and $x_1 = x_1^{(1)} + x_1^{(2)}$. We will attempt to study the equilibrium flows on this problem as the toll τ is increased from zero. [For the simple network that is being studied, the equilibrium link costs are unique even though $c_2^{(1)}$ and $c_2^{(2)}$ are constant.]

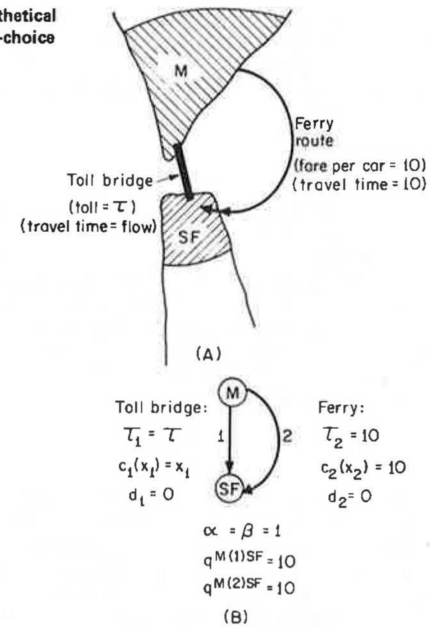
The function C(x) for our problem is (up to an additive constant) as follows:

$$C(x) = (x_1^2/2) + \tau x_1^{(1)} + [(\tau/2)\,x_1^{(2)}] + 20x_2^{(1)} + 15x_2^{(2)}$$

since $\partial C(x)/\partial x_i^{(k)} = c_i^{(k)}$. The flow conservation and nonnegativity constraints are as follows:

$$x_1^{(1)} + x_2^{(1)} = 10$$

Figure 1. Hypothetical two-mode route-choice problem.



$$x_1^{(2)} + x_2^{(2)} = 10$$

$$x_1^{(1)}, x_2^{(1)}, x_1^{(2)}, x_2^{(2)} \geq 0$$

Alternatively, we can solve the following:

$$\min C = [x_1^{(1)} + x_1^{(2)}]^2/[2 + \tau x_1^{(1)}] + [(\tau/2)\,x_1^{(2)}] + 20[10 - x_1^{(1)}] + 15[10 - x_1^{(2)}] \quad 0 \leq x_1^{(1)}, x_1^{(2)} \leq 10$$

Figure 2 plots the negative gradient field of C for different values of τ over the feasible range of $x_1^{(1)}$ and $x_1^{(2)}$:

$$-\nabla C^T = \begin{bmatrix} -x_1^{(1)} - x_1^{(2)} - \tau + 20 \\ -x_1^{(1)} - x_1^{(2)} - (\tau/2) + 15 \end{bmatrix}$$

The equilibrium solution is a point in the feasible region at which the gradient is perpendicular to the boundary.

As seen from Figure 2, there are two cases that result in different equilibrium solutions. If the toll is less than 10 monetary units, all the type-1 traffic takes the toll bridge and the type-2 traffic is split between the two routes so that the total cost to type-2 vehicles will be equal on both routes (type-2 vehicles are diverted from the bridge because they can split the taxicab fare). As the toll is increased more and more, type-2 vehicles are shifted to the long route in order to keep the costs on both routes equal for these vehicles. In the process, however, the relative attractiveness of route 1 for vehicles of type 1 is decreased (the toll affects these vehicles twice as heavily) until eventually route 2 becomes more attractive than route 1 to these vehicles. At that toll value, $\tau = 10$, the equilibrium solution is achieved by using 10 vehicles of any type on each route since they are then equally attractive to both classes. (This illustrates well the possible nonuniqueness of the modal link flows despite the uniqueness of the total flow.) A slight increase of τ beyond 10 shifts all type-2 vehicles to route 1 and all type-1 vehicles to route 2 because the toll is now sufficiently high to make the bridge unattractive to those who are not carpoolers. Increases beyond 10 will result in further decreases in bridge traffic as those who carpool find the bridge increasingly expensive.

Table 1 summarizes the results. Note also that although vehicle traffic on the bridge decreases smoothly with an increasing toll, there is a critical point when the composition of traffic changes drastically with an increase in the total number of bridge users. Figure 3 illustrates this.

The transportation cost to society (tolls are internal transfers) can be decreased by increasing tolls. This is logical because in this way the bridge is used only by cars that have high occupancy. The maximum revenue on the bridge is achieved when $\tau = 15$, but the maximum combined revenue (which also yields the minimum total travel time) is obtained for $\tau = 20$.

Methods

To solve problem (MP), one can use the Frank-Wolfe algorithm. Because the gradient of the objective function C(x) is the set of link costs for all vehicle classes, the linear subproblem is an all-or-nothing traffic-assignment problem. LeBlanc, Morlok, and Pierskalla (20) were the first to propose this algorithm for the one-vehicle traffic-assignment problem. The steps are as follows:

Step 0 (initialization): Set an arbitrary (nonnegative) cost vector $c = (\ldots, c_i^{(k)}, \ldots)$, assign the O-D table of each vehicle type to the corresponding shortest paths, and obtain a feasible link-flow pattern $x = (\ldots, x_i^{(k)}, \ldots)$.

Step 1 (cost updating): Recalculate c by using the new set of flows, c = c(x).

Step 2 (assignment): Calculate the shortest paths and assign the O-D flows to them. Do this for all vehicle types. Label the flow pattern $y = (\ldots, y_i^{(k)}, \ldots)$.

Step 3 (interpolation): Find the value of ω, $\omega^* \in [0,1]$ that minimizes $f(\alpha) = C[x + (y - x)\omega]$ and let $x' = x + (y - x)\omega^*$ be the new flow pattern.

Step 4 (convergence check): If the new pattern is not substantially different from the old pattern, stop. Otherwise, repeat the process from step 1.

The easiest way of performing step 3 is to find the value of ω at which the derivative of $f(\omega)$ vanishes. If $f'(\omega)$ does not have a root in [0, 1], $\omega^* = 1$ because $f(\omega)$ is convex. In this way the objective function is never used, and one does not have to integrate the link-cost function:

$$f'(\omega) = \sum_{i,k} \left([y_i^{(k)} - x_i^{(k)}] \times c_i^{(k)} \left\{ x_i^{(k)} + \omega [y_i^{(k)} - x_i^{(k)}] \right\} \right)$$

Alternatively, one can use the method of successive averages (21-23) or, for sketch planning problems that have few links, some unconstrained methods (19,24).

Example

We do the example in Figure 1 by using $\tau = 0$ and start by using a cost vector that corresponds to an empty network:

$$c = [c_1^{(1)}, c_2^{(1)}, c_1^{(2)}, c_2^{(2)}] = (0, 20, 0, 15)$$

Step 0. The all-or-nothing flow pattern $x = [x_1^{(1)}, x_2^{(1)}, x_1^{(2)}, x_2^{(2)}]$ is x = (10, 0, 10, 0).

Step 1. The revised cost vector is c = (20, 20, 20, 15).

Step 2. The all-or-nothing flow vector is y = (0, 10, 0, 10).

Step 3. The $f'(\omega)$ function is as follows:

$$f'(\omega) = -10 \times [10(1-\omega) + 10(1-\omega)] + 10 \times 20$$
$$-10 \times [10(1-\omega) + 10(1-\omega)] + 10 \times 15$$
$$= -10[5 - 40\omega]$$

and $\omega^* = 0.125$.

The new flow vector is x = (8.75, 1.25, 8.75, 1.25).

Another iteration yields x = (9.3, 0.7, 4.9, 5.1), which is fairly close to the equilibrium solution x* = (10, 0, 5, 5). The Frank-Wolfe algorithm, however, tends to slow down when the equilibrium is approached. This example is no exception, since, as the reader can verify, the next two flow vectors are x = (9.47, 0.53, 6.13, 3.87) and (9.57, 0.43, 4.97, 5.03).

Figure 2. Negative gradient field and equilibrium solutions.

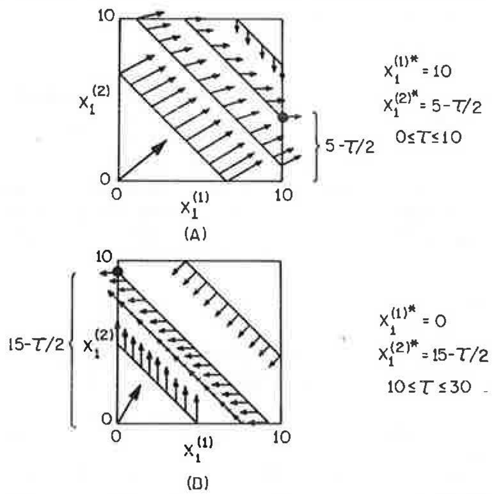
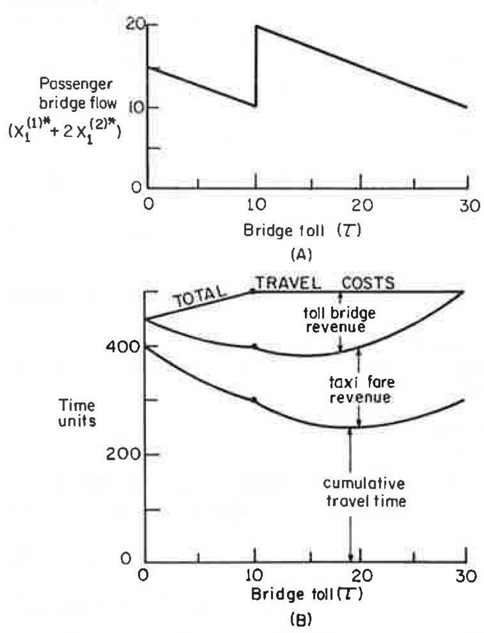


Table 1. Components of the equilibrium solution for different tolls.

τ	$x_1^{(1)*}$	$x_1^{(2)*}$	x_1^*	$c^{(1)*}$	$c^{(2)*}$
$\tau \le 10$	10	$5 - (\tau/2)$	$15 - (\tau/2)$	$15 + (\tau/2)$	15
$10 \le \tau \le 30$	0	$15 - (\tau/2)$		20	

Figure 3. Effect of varying tolls on passenger bridge flow, revenues, and total travel time.



CONCLUSION

This paper has demonstrated that many current TSM strategies related to carpooling can be investigated by using equilibrium theory. It was argued that multiple-vehicle-type network models that have link-cost functions of the following form are good descriptors of carpooling cost functions because the independent constant $c_{oi}^{(k)}$ can capture the difference in the fixed costs of a link to the different vehicle classes:

$$c_i^{(k)} = c_{oi}^{(k)} + c_i[x_i^{(1)} + x_i^{(2)} + \ldots + x_i^{(K)}] \quad (8)$$

It was mentioned that if the $c_i(\cdot)$'s were increasing functions, the total equilibrium flows $x_i = [x_i^{(1)} + \ldots + x_i^{(K)}]$ existed and were unique. Furthermore, because the Jacobian $\{\partial[\ldots, c_i^{(k)}, \ldots]/\partial[\ldots, x_i^{(k)}, \ldots]\}$ is continuous and symmetric, the equilibrium problem admits an extremum formulation that can be solved by using optimization procedures.

The following generalizations are of some merit:

1. Vehicles of different sizes: If $c_i^{(k)}$ can be expressed as follows:

$$c_i^{(k)} = c_{oi}^{(k)} + \delta^{(k)} c_i[\gamma_1 x_i^{(1)} + \gamma_2 x_i^{(2)} + \ldots + \gamma_K x_i^{(K)}]$$

where the γ_k's are nonnegative constants that reduce the flow of k-type vehicles to an equivalent flow of standard-type vehicles, the link costs and the standardized link traffic levels are given as follows:

$$x_i = \sum_{k=1}^{K} \gamma_k x_i^{(k)}$$

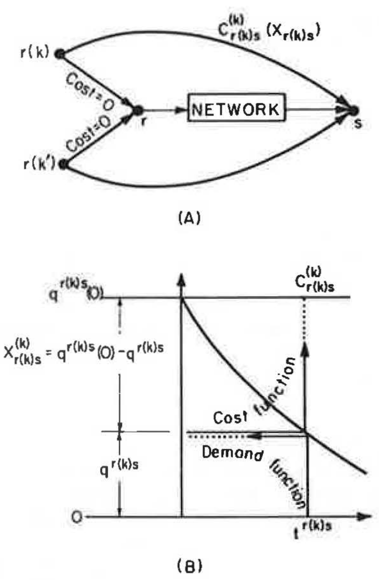
At equilibrium, these link costs and traffic levels exist and are unique, provided the $\delta^{(k)}$'s and γ_k's are positive constants and the $c_i(\cdot)$'s are nonnegative increasing continuously differentiable functions (19). [This is true even though the Jacobian J(x) in this case is no longer positive-definite or symmetric.] This formulation is of interest in cases in which vehicles of different types are of different sizes and is of potential interest to study urban goods movement by trucks in an urban area.

2. Elastic demand: If the k-type O-D flow $q^{r(k)s}$ decreases with an increasing O-D cost for k-type vehicles $t^{r(k)s}$ and is also independent of other O-D costs, the O-D tables, link costs, and link-flow levels are unique at equilibrium. This is not difficult to see because, as with single-vehicle-type elastic-demand problems (24), multiple-vehicle-type problems admit an equivalent formulation that has fixed O-D tables. Figure 4A depicts the equivalent network, which includes additional nodes r(k) and r(k'). These nodes are only connected to r and s by dummy links. The k-type flow from r to s is now assumed to start at r(k) and to be equal to $q^{r(k)s}(0)$. Link r(k)r has zero cost and link r(k)s has a cost function $c_{r(k)s}^{(k)}(\cdot)$, which is defined from the demand function as shown in Figure 4B:

$$c_{r(k)s}^{(k)}(w) = q^{r(k)s^{-1}}[q^{r(k)s}(0) - w]$$

Similar definitions apply to flows of other types.

Figure 4. Fixed-demand network equivalent to elastic-demand network.



Since link r(k)s can carry k-type flow, it is clear that at equilibrium $c^{(k)}_{r(k)s}$ must equal $t^{r(k)s}$, and therefore the following holds:

$$q^{r(k)s^{-1}}\left[q^{r(k)s}(0) - x^{(k)}_{r(k)s}\right] = t^{r(k)s}$$

$$q^{r(k)s^{-1}}\left[q^{r(k)s}\right] = t^{r(k)s}$$

$$q^{r(k)s} = q^{r(k)s}\left[t^{r(k)s}\right]$$

which is the condition for elastic demand.

Elegant as it is, the elastic-demand formulation just explained is not realistic for carpooling problems because it fails to capture the important phenomenon of passenger jockeying among modes. That is, if the travel cost for carpools were to decrease substantially, one would expect to see an increase in carpooling but at the expense of noncarpooling traffic. Modal-split models that assume that the total O-D passenger flows remain unchanged are much better suited for this and other public transportation applications. In these cases, as passengers switch to high-occupancy vehicles, total O-D vehicular flows decrease. Unfortunately, equilibrium results have not been derived for these models.

Further research should concentrate on establishing uniqueness results for multiple-vehicle-type equilibrium flows in which there is passenger jockeying and on further exploration of optimal pricing strategies by using tolls. Some of these issues will be discussed in a forthcoming publication.

ACKNOWLEDGMENT

This research was supported by a grant from the National Science Foundation to the University of California at Berkeley.

REFERENCES

1. H. VonFalkenhausen. Traffic Assignment by a Stochastic Model. Proc., Fourth International Conference on Operational Research, 1966, pp. 415-421.
2. J.E. Burrell. Multiple Route Assignment and Its Application to Capacity Restraint. Proc., Fourth International Symposium on the Theory of Traffic Flow, Karlsruhe, West Germany, 1968, pp. 210-219.
3. H. Beilner and F. Jacobs. Probabilistic Aspects of Traffic Assignment. In Traffic Flow and Transportation (G.F. Newell, ed.), American Elsevier, New York, NY, 1971, pp. 183-194.
4. R.B. Dial. A Probabilistic Multipath Traffic Assignment Model Which Obviates Path Enumeration. Transportation Research, Vol. 5, 1971, pp. 83-111.
5. C.F. Daganzo and Y. Sheffi. On Stochastic Models of Traffic Assignment. Transportation Science, Vol. 11, 1977, pp. 253-274.
6. M.J. Beckmann, C.B. McGuire, and C.B. Winsten. Studies in the Economics of Transportation. Yale Univ. Press, New Haven, CT, 1956.
7. R.B. Potts and R.M. Oliver. Flows in Transportation Networks. Academic Press, New York, NY, 1972.
8. P.A. Steenbrink. Optimization of Transport Networks. Wiley, New York, NY, 1974.
9. G.F. Newell. Traffic Flow in Transportation Networks. MIT Press, Cambridge, MA, 1980.
10. J.G. Wardrop. Some Theoretical Aspects of Road Traffic Research. Proc., Institute of Civil Engineering, Vol. 2, 1952, pp. 325-378.
11. M. Netter. Equilibrium and Marginal Cost Pricing on a Road Network with Several Flow Types. In Traffic Flow and Transportation (G.F. Newell, ed.), American Elsevier, New York, NY, 1971, pp. 155-181.
12. M.J. Smith. The Stability and Uniqueness of Traffic Equilibria. Transportation Research, Vol. 13B, 1979, pp. 295-304.
13. S.C. Dafermos. An Extended Traffic Assignment Model with Applications to Two-Way Traffic. Transportation Science, Vol. 5, 1971, pp. 366-389.
14. S.C. Dafermos. Traffic Equilibrium and Variational Inequalities. Transportation Science, Vol. 14, 1980, pp. 42-54.
15. S.C. Dafermos. The Traffic Assignment Problem for Multiclass-User Transportation Networks. Transportation Science, Vol. 6, 1972, pp. 73-87.
16. M. Florian. A Traffic Equilibrium Model of Travel by Car and Public Transit Modes. Transportation Science, Vol. 11, 1977, pp. 166-179.
17. M. Abdulaal and L.J. LeBlanc. Methods for Combining Modal Split and Equilibrium Assignment Models. Transportation Science, Vol. 13, 1977, pp. 292-314.
18. I. Jeevanantham. A New Look at the Traffic Assignment Problem. In Traffic Flow and Transportation (G.F. Newell, ed.), American Elsevier, New York, NY, 1971.
19. C.F. Daganzo. Unconstrained Extremal Formulation of Some Transportation Equilibrium Problems. Transportation Science (in press).
20. L.J. LeBlanc, E.K. Morlok, and W.P. Pierskalla. An Efficient Approach to Solving the Road Network Equilibrium Traffic Assignment Problem. Transportation Research, Vol. 9, 1975, pp. 309-318.
21. J. Almond. Traffic Assignment with Flow-Dependent Journey Times. In Vehicular Traffic Science (L.C. Edie, R. Herman, and R. Rothery, eds.), American Elsevier, New York, NY, 1967, pp. 222-234.
22. C. Fisk. Some Development in Equilibrium Traffic Assignment Methodology. Transportation Research (in press).
23. W.G. Powell and Y. Sheffi. The Convergency of Equilibrium Algorithms with Predetermined Step Sizes. Transportation Science (in press).
24. C.F. Daganzo. Equilibrium Analysis in Transportation: The State of the Art and Some New Results. Univ. of California, Berkeley, Research Rept. UCB-ITS-RR-81-1, 1980.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting and Committee on Traveler Behavior and Values.

Demand-Supply Modeling for Transportation System Management

ADOLF D. MAY

A review is given of the development and application of a family of operational planning models that are used to predict impacts and traveler responses resulting from traffic management strategies in freeway corridors, arterial networks, dense networks, and rural highways. An overview of the long-term research program and the identification of current research efforts are also included. One of the major goals of the research program is to propose policy guidelines for implementing traffic management strategies. Initial policy guidelines for freeway priority lanes, freeway-entrance control, arterial priority lanes, arterial-signal control, dense networks, and rural highways are included. The following conclusions are drawn: (a) increased attention should be given to controlling the demand side of operational problems by spreading demand over space, time, and mode and by reducing the total demand level; (b) increased attention should be given to assessing energy and environmental impacts of improvement alternatives as well as to continuing the assessment of safety and levels of service; (c) creative techniques need to be devised to generate and screen traffic management strategies prior to analytical evaluation; and (d) greater use of operational planning models by facility operators is essential if our existing transportation system is to be managed effectively.

Operational problems are encountered in the existing transportation system, and inefficient use results when traffic demands exceed traffic capacities. The byproducts of such operational situations are increased travel time, less-reliable service, higher accident rates, greater fuel consumption, and increased vehicle emissions. Historically in the United States, the normal approach to a solution was to increase capacity when such operational problems were encountered. Such actions were generally very expensive, often disrupted the environment, and encouraged further growth in the traffic demand.

Transportation system management (TSM) proposes that greater attention be given to controlling the demand side of the equation by spreading demand over space, time, and mode and/or reducing the total demand level. It is implied that such actions may be accomplished through low-cost improvements, may improve the environment, may conserve energy, and tend to curtail further growth in traffic demand.

Demand-supply computerized models have been developed and applied for impact assessment and traveler responses to traffic management strategies along freeway corridors, arterial networks, dense networks, and rural highways. The traffic management strategies included priority lanes, ramp metering, priority entry, signal optimization, bus priority signals, redesign improvements, parking restrictions, speed-limit control, traffic-resistant measures, and other related actions.

Impact assessment included traveler impacts (travel time, fuel, emissions, noise, and safety), fleet costs, and facility costs. Traveler responses included spatial, temporal, modal, and total demand-level responses as well as public acceptability. The computerized models provide a time series of impact and traveler-response evaluations. This paper will highlight the development and application of these analytical models for freeway corridors, arterial networks, dense networks, and rural highways. It is based on earlier work (1).

OVERALL RESEARCH DIRECTION

The overall research direction for the development of a system of operational planning models for evaluating TSM projects is shown in Figure 1. As shown, the overall research direction consists of four activities that integrate operational and planning techniques: demand-supply modeling, strategy selection, impact assessment, and traveler responses. These four activities and their interactions will now be described.

The research plan calls for the development of a set of demand-supply models for four operating environments--freeway corridors, arterial networks, dense networks, and rural highways. Such models are now operational, but research continues in attempting to improve and extend them. The input to each model consists of supply-related design features, origin and destination (O-D) demand patterns, and the initial control state. The simulation submodel predicts impacts for the existing state, that is, the day before the traffic management strategy will be implemented. The optimization submodel searches for an improved or optimum strategy and then calls the simulation submodel, which predicts impacts for the improved or optimum state, that is, the day after the traffic management strategy has been implemented.

Differential impacts are assessed to evaluate the first day of operation and to provide input data for the forecasting of traveler responses. The impacts assessed include user impacts (travel time, fuel, emissions, noise, and safety) and system impacts (facility and fleet costs).

Travelers may respond to impact changes. For example, individual travelers may change their routing (spatial response), their mode of travel (modal response), and/or their time of travel (temporal response). Strategies implemented further may affect the total demand level over a longer period of time by eliminating nonessential trips, modifying O-D pairs, changing employment and/or residential locations, modifying normal demand growth, etc. Finally, strategy implementation may not be publicly acceptable and such schemes may be terminated.

If traveler response changes the traffic demand, model input demands are modified and the simulation submodel predicts longer-term impacts. These activities are interactive in such a way that a time stream of states can be predicted (impacts and traveler responses) that starts the day before implementation and continues for several points in time for a period of two to five years.

FREEWAY CORRIDORS

Demand-supply modeling efforts for freeway-corridor operating environments were initiated at Berkeley in 1968 when a California Department of Transportation (Caltrans) research project required the evaluation of alternatives for improving 140 miles of the existing San Francisco Bay Area Freeway System. The system was too extensive and the alternative improvements were too numerous for manual analysis to be considered. This first model, called FREQ (2) (now referred to as FREQ1), was developed; it was the forerunner of the system of models displayed in Figure 2. The three freeway-corridor models (FREQ6PL, FREQ6PE, and FREQ6T) that have been most extensively developed and are currently operational will each be described briefly. Before this is

Figure 1. Overall research direction.

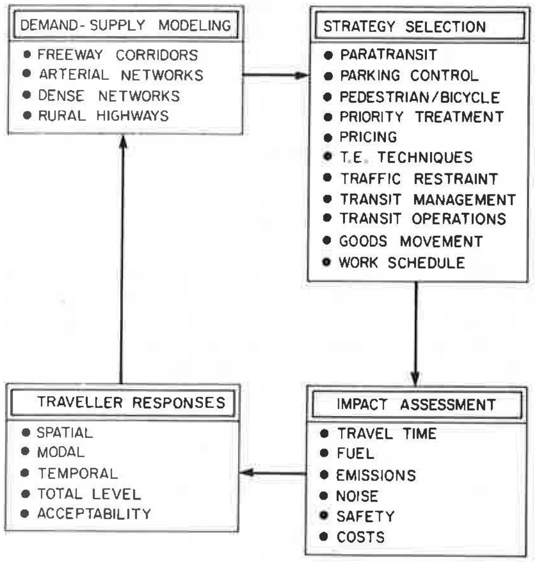
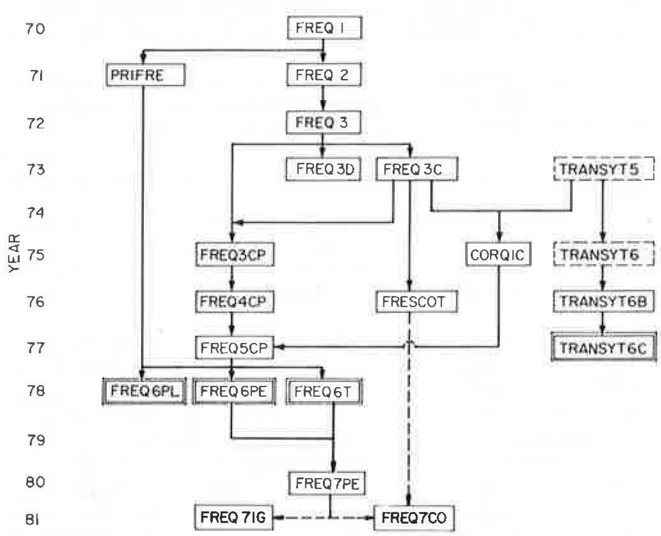


Figure 2. Freeway corridor and arterial network models.



done, some highlights of this systematic model development will be discussed.

The FREQ2 and FREQ3 models were extensions and refinements of the earlier model in which particular attention was directed to shock-wave analysis, computer efficiency, and output format (3,4). The PRIFRE model was developed for the evaluation of priority lanes on freeways (5).

By the early 1970s, the need for decision models (those that incorporated simulation and optimization submodels) was recognized. Three models in this family (FREQ3CP, FREQ3D, and FREQ3C) were developed and incorporated priority-entry control, design improvement, and normal entry-control optimization submodels, respectively (6-8). An on-line version of the FREQ3C model was developed and designated the FRESCOT model (9). One of the significant results of this work was the development of a technique for generating synthetic O-D tables from on-ramp and off-ramp counts. As the modeling effort continued, attention was given to the surrounding street system, (CORQlC) (10), impact assessment (FREQ4CP) (11), and traveler-demand responses (FREQ5CP) (12).

FREQ6PL Model

FREQ6PL (13) is a macroscopic model of a freeway corridor and is used primarily for the evaluation of reserving a lane or lanes on freeways for carpools and/or buses. The model can also be used for evaluating design improvements with or without priority-lane operations. It incorporates many features from the earlier-developed PRIFRE and FREQ5CP. The user selects the priority lane or lanes, design configuration, priority-cutoff level, and time duration of priority operations. Then the model automatically modifies the demand and supply sides and predicts a time stream of impacts and traveler responses.

The impact assessment includes travel time, fuel, emissions, and facility costs, and the model combines these impacts into one performance index. The demand forecasting includes spatial and modal traveler responses in increments during the first year of operation. The model has been applied to the Santa Monica Freeway in Los Angeles and, through sensitivity analysis, initial policy guidelines have been proposed.

Highlights of these initial policy guidelines are listed below [before implementation, the report by Cilliers, May, and Cooper (13) should be consulted for a description of the study and limitation of research].

1. A with-flow median lane used exclusively by priority vehicles on a congested freeway is expected to compare unfavorably with the previous situation in both the short-term and the long-term situations that follow, considering total travel time, fuel consumption, and vehicle emissions;
2. A with-flow median lane used exclusively by priority vehicles on an uncongested freeway is expected to perform equally well or slightly worse than in the previous situation in both the short-term and the long-term situations that follow, considering total travel time, fuel consumption, and vehicle emissions; and
3. There may be some operating environments significantly different from the Santa Monica Freeway environment in terms of occupancy distribution, level of bus service, mode-shift propensity, and parallel arterials (if a with-flow median lane used exclusively by priority vehicles is being considered in such an environment, it is recommended that an in-depth analysis be undertaken before it is decided to implement it).

The FREQ6PL model is written in FORTRAN IV, currently operational on Control Data Corporation (CDC) and International Business Machines (IBM) computer systems, and the program and user's guide are available for distribution from the Institute of Transportation Studies, University of California at Berkeley.

FREQ6PE Model

FREQ6PE (14) is a macroscopic decision model of a freeway corridor and is used primarily for the evaluation of priority-entry and normal-entry control on a directional freeway. The model can also be used for evaluating design improvements with or without freeway-entry control. The user selects the type of entry control combined with any desired design analysis and the objective function and operational constraints, and the model selects the ramp-control plan through a linear programming optimization process. The model predicts a time stream of impacts and traveler responses due to the intersection between ramp-control strategy and traveler responses.

The impact assessment includes travel time, fuel, emissions, and noise, whereas the demand forecasting includes spatial and modal traveler responses in increments during the first year of operation. The model has been applied extensively to the Eastshore Freeway in the San Francisco Bay Area and to the Santa Monica Freeway in Los Angeles, and initial policy guidelines have been proposed as a result of sensitivity analysis.

Highlights of these initial policy guidelines are listed below [before implementation, the report by Jovanis, May, and Yip (14) should be consulted for a description of the study and limitation of research]:

1. Normal-entry freeway control on a previously congested freeway is expected to result in (a) reduced passenger hours of travel, carbon monoxide emissions, and hydrocarbon emissions; (b) little effect on fuel consumption and noise emissions; and (c) increased nitrogen oxide emissions;
2. Priority-entry freeway control on a previously congested freeway is expected to result in slightly improved impacts as compared with normal-entry freeway control; the degree of improvement depends on passenger occupancy distribution and level of service;
3. Extensive freeway congestion and the availability of underutilized alternate routes were positive factors in assessing levels of freeway-corridor improvements; if entry control is being considered in environments where such conditions do not exist, it is recommended that an in-depth analysis be undertaken before a decision is made to implement it.

The FREQ6PE model is written in FORTRAN IV, currently operational on CDC and IBM computer systems, and the program and user's manual are available for distribution from the Institute of Transportation Studies.

FREQ6T Model

FREQ6T (14) is a simulation model of a directional freeway used primarily for the evaluation of temporal traveler-demand responses that result from on-freeway and off-freeway strategies. The model can also be used for evaluating design improvements and longer-term traveler-demand responses.

The research effort for off-freeway strategy investigations was primarily directed toward the evaluation of the influence of flex-time work scheduling on freeway-corridor operations. A series of working papers (15-18) and a final report (19) were prepared. Highlights of these initial policy guidelines are listed below [before implementation, the working papers just mentioned (15-18) should be consulted for the description of the study and limitation of the research]:

1. In situations where intense congestion is encountered for an extended time, congestion improvement or degradation is possible; detailed site-specific studies with the model are recommended;
2. In situations where congestion is encountered for a brief time, there is a strong likelihood of travel-time savings;
3. In situations where little or no congestion is encountered, only a slight traffic impact is expected; congestion may be shifted to earlier time periods.

The research effort for on-freeway strategy investigations was primarily directed toward temporal demand responses due to freeway-entry control. A working paper (20) and a final report (21) have been prepared. The new model, FREQ7PE (19), which incorporated many features from the FREQ6T and FREQ6PE models (14), was developed and applied in this study. A description of current research on the FREQ7PE model is included in the next section.

Current Freeway-Corridor Research

Four freeway-corridor research activities are currently under way. A brief description of each is given below.

The FREQ7PE model (19) is now being used in a research mode, but before being distributed as an operational-planning tool, the model will be thoroughly tested and applied to a wide variety of freeway-corridor environments. FREQ7PE is a revised and extended version of FREQ6PE. The model structure has been reorganized for improved efficiency and flexibility. Many new operations are available to increase the program's power and usefulness in practical traffic-engineering problems. Improvements to simulation include input and output flexibility, fuel and emissions options, and mainline-delay calculations. Improvements to optimization include user-supplied metering plans, queue-length limits, congestion optimization, and overcontrol protection. Improvements to traveler response include temporal response and user-sequenced and/or simultaneous spatial, modal, and temporal response options.

Current research is under way on the subject of on-line traffic-responsive freeway-entry control. The major research objectives are as follows:

1. The modification of a large-scale freeway model,
2. The evaluation of existing and proposed control strategies by using the freeway model, and
3. The evaluation of selected control strategies by their implementation in the field.

Current research is also under way that is concerned with deriving O-D information from routinely collected traffic counts. The first working paper (22) contained a review of current research and applications. Current efforts are being directed along three lines: development of a single-path network model and a multipath network model and data-base acquisition for evaluating models. The freeway-corridor model requires freeway O-D data and it is anticipated that the developed single-path network model will have direct application.

Exploration continues with regard to an interactive-computer-graphics version of the freeway-corridor model. Such a model (FREGRAF) has been developed by the University of Washington (23) and there are plans to expand it. This scheme will permit the user to interact with the intermediate computer results and give direction for further analysis.

ARTERIAL NETWORKS

Demand-supply modeling efforts for arterial corridors and networks were initiated in 1973 when a very extensive freeway-corridor model (CORQ1C) (10) was being developed. The intent was to integrate the previously developed FREQ3C model (8) with a new arterial-network model and to develop the coordination routines for freeway-arterial interactions. Fortunately, a literature review revealed the existence of the TRANSYT model, which was compatible with the concept of the freeway-corridor model and performed the same tasks for arterials that the FREQ series was performing for freeways. The TRANSYT5

(24) and TRANSYT6 (25) models were extended (Figure 2). THe TRANSYT6B (26) model was an extension of the TRANSYT6 model that had impact assessments added. The TRANSYT6C (27) model is an extension of the TRANSYT6B model and not only includes impact assessments but a comprehensive performance-index objective function and traveler-demand responses.

TRANSYT6C Model

TRANSYT6C (27) is a macroscopic decision model of an arterial corridor or arterial network used primarily for the evaluation of traffic-signal settings. The model can also be used for evaluating design improvements combined with the search for improved traffic-signal settings. Further, the model can be used to evaluate priority lanes. A branch-and-bound optimization technique is employed that searches for an improved traffic-signal setting by means of minimizing a performance index. The model predicts a time stream of impacts and traveler responses due to the interaction between signal settings (with or without priority treatment) and traveler responses.

The impact assessment includes travel time, fuel, and emissions, and these impacts are combined into one performance index that includes nine terms; the user can select coefficients or weights for each term. The demand forecasting includes spatial and modal traveler responses. The model has been applied to the Wilshire Boulevard corridor in Los Angeles and the San Pablo Avenue corridor in the San Francisco Bay Area. From these applications and sensitivity analyses, some initial policy guidelines have been proposed (27).

Highlights of these initial policy guidelines are listed below [before implementation, the report by Jovanis and May (27) should be consulted for description of the study and limitations of research]:

1. Signal timing optimized to minimize vehicle delay is expected to result in reduction in delay, fuel consumption, and vehicle emissions; only marginal benefit to buses; and improved arterial productivity.
2. Signal timing optimized to minimize passenger delay is expected to result in moderate reduction in delay, fuel consumption, and vehicle emissions; moderate savings in bus time and bus fuel consumption; and improved arterial productivity.
3. Exclusive bus lanes that have shared stop lines are expected to result in negative short-term impacts; savings in bus time and bus fuel consumption; possible slight mode shift to buses; return to their existing conditions of time spent and environmental impacts in the long term; and little improved arterial productivity.
4. Operating environments significantly different from those studied in terms of level of bus service, degree of saturation, and availability of alternate routes may be encountered; in such cases, it is recommended that an in-depth analysis be undertaken before it is decided to implement such strategies.

The TRANSYT6C model is written in FORTRAN IV, currently operational on CDC and IBM computer systems, and the program and user's guide are available for distribution from the Institute of Transportation Studies.

Current Arterial-Network Research

No sponsored research is currently under way on arterial-network modeling. Efforts continue to extend and refine the initially developed policy guidelines (28). Extensive sensitivity analysis is currently under way to investigate the effect of route characteristics, traffic-flow levels, cycle lengths, and objective functions on various measures of effectiveness.

DENSE NETWORKS

Demand-supply modeling efforts for dense networks were initiated in 1978. Dense networks are distinguishable from arterial networks in three ways: smaller spacing between parallel routes, stronger traffic-resistant actions, and greater concern for environmental impacts. Residential areas, central business districts (CBDs), shopping centers, and sports centers exemplify the dense networks being considered.

The four major tasks were to prepare a state-of-the-art report of existing experience and models, to evaluate and select the most appropriate existing dense-network model or models, to apply them to a residential area and a CBD to assess TSM-type strategies, and to refine and reapply the model in order to work toward initial guidelines for traffic management policy.

A state-of-the-art working paper (29) of existing experience and demand-supply models for traffic management in dense urban networks was prepared. Attention was directed toward identification of problems encountered in dense networks, various types of traffic management strategies implemented, measures of effectiveness considered, foreign and U.S. case studies, and availability of existing demand-supply models.

It was found that existing models fell into five major categories: planning, optimization, equilibrium, simulation, and operational models. More than 30 models were identified within these categories. A general evaluation of model categories was accomplished and this indicated that the category of operational models best met the evaluation criteria established for the purpose of this project. The six operational models (Figure 3) were studied in greater detail, authors were contacted for further information, and copies of computer programs were obtained when possible. Finally, two models [micro-assignment (30) and CONTRAM (31)] were selected for initial application on this project.

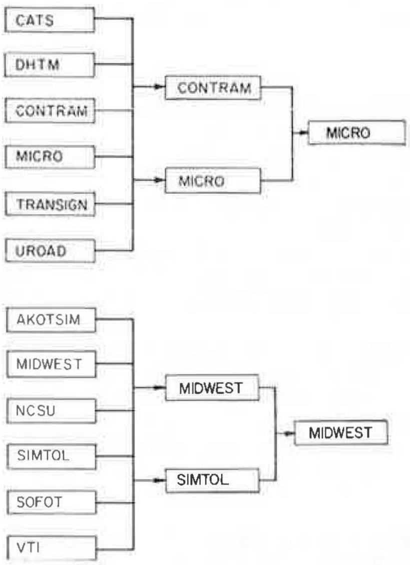
City officials in California were contacted to determine their interest in cooperating, to locate appropriate residential and CBD sites, and to assess available traffic demand and supply data. The major constraint in selecting sites was the availability of adequate traffic demand and supply data, particularly O-D data. Agreements were reached with the cities of Palo Alto and San Jose; the selected residential site was in Palo Alto, and the selected CBD site was in San Jose. The selection of the two models for initial application and the selection of study sites are described in the second working paper (32).

The micro-assignment and CONTRAM models were made operational and applied to the Palo Alto study site. The micro-assignment model was selected as the one to be used in the final phases of this project. This work is described in the third working paper (33).

Micro-Assignment Model

In the final phase of the project, the micro-assignment model was refined to include fuel consumption as a measure of performance and to provide a clearer output and a more-flexible process of input. The refined model was then applied to the residential and CBD study sites and several TSM control plans were evaluated and preliminary guidelines pre-

Figure 3. Dense network and rural highway models.



sented. This phase of the project as well as earlier phases are described in the two-volume final report (34,35).

Highlights of these initial policy guidelines for residential dense networks are listed below [before implementation, the final report (34,35) should be consulted for the description of the study and limitation of research]:

1. Providing a good environment for residents while maintaining reasonable neighborhood accessibility and through-traffic mobility are proposed as specific goals.
2. Reduced traffic on neighborhood streets, residential connectivity, and travel time on adjacent arterials are proposed as measures of effectiveness.
3. TSM strategies to be considered within the residential dense network include pedestrian/bicycle treatment, traffic-restraint measures, and intersection geometries; TSM strategies to be considered on adjacent arterials include parking control, bus routing and stops, truck routing and loading and unloading zones, and traditional traffic-engineering techniques.
4. Involvement of citizens, providers of neighborhood services, and bus and truck management are essential for successful implementation.

Current Dense-Network Research

One of the major obstacles to dense-network analysis is the required O-D data for input. As mentioned in the discussion of freeway corridors, current research is under way that is concerned with deriving O-D information from routinely collected traffic counts. The multipath-network-model development portion of the project will be applicable to dense networks (22).

It is anticipated that work in the dense-network operating environment will continue. Emphasis will most likely be directed to further model refinement, development of an interactive version, screening techniques for selecting strategies for evaluation, application of the model to other dense networks, experimentation with synthetic O-D input, and improving traffic-management guidelines.

RURAL HIGHWAYS

Demand-supply modeling efforts for two-lane, two-way rural highways were initiated in 1975. Primary attention was given to capacity and levels of service under the condition of trucks and grades. This led to the development of the SIMTOL model, which will be described below.

In 1978, Caltrans became interested in sponsoring research on the development of a decision-making framework for the evaluation of climbing lanes. This research activity led to the refinement of the MIDWEST model and the formulation of climbing-lane guidelines. This work will be described later.

SIMTOL Model

SIMTOL (36) is a microscopic simulation for two-lane, two-way rural highways. The model is primarily intended to predict the traffic performance as a function of facility design (design speed and vertical alignment) and traffic loads (quantity and composition). Special attention is given to the performance of trucks on grades and to driver behavior (following cars and passing maneuvers). Only travel-time impacts are assessed, and no demand forecasting is undertaken. The model has been applied to several situations in California. Some sensitivity analysis has been undertaken to predict levels of service and capacities.

MIDWEST Model

In 1978, work began on this development of a decision-making framework for the evaluation of climbing lanes (37-39). The four major tasks were to prepare a state-of-the-art report of existing experience and models, to evaluate and select the most appropriate existing model for this project, to modify the model as required and perform field validations, and through model application and sensitivity analysis to develop policy guidelines for climbing lanes.

A state-of-the-art working paper of existing experience and models was prepared (38). Attention was directed to two-lane rural highway research studies, availability of existing two-lane highway models, and vehicle-driver performance. Six candidate models were identified and included (Figure 3): (a) Akonteh's model from Stanford University (AKOTSIM), (b) St. John's model from Midwest Research Institute (MIDWEST), (c) Heimbach's model from North Carolina State University (NCSU), (d) Stock's model from the University of California (SIMTOL), (e) Kaesehagen's model from Australia (SOFOT), and (f) Gynnerstedt's model from Sweden (VTI). Authors were contacted for further information and computer programs were obtained when possible. The characteristics of the six models were evaluated by using a 19-point list of criteria, and the MIDWEST model was selected for the purposes of this project (39,40).

The MIDWEST model had several deficiencies for use on this project and required some modifications. These included adding a climbing-lane option, adding additional measures of effectiveness such as safety and operating costs, and improving input/output features. The next step was validation of the modified model with particular emphasis on the climbing-lane option. Field studies were conducted in northern California, and the model was refined to represent actual traffic conditions.

In early 1980 the modified MIDWEST model was employed to develop cost-effectiveness curves for a wide variety of traffic (volume and composition) and climbing-lane (percentage of grade, length of grade, length of climbing lane, and position of climbing lane) situations. By using the extensive set of cost-effectiveness curves, initial policy guidelines were formulated with regard to climbing lanes on grades and a final report was prepared (41).

Highlights of these initial policy guidelines are listed below [before implementation, the paper by Botha (41) should be consulted for a description of the study and limitation of research]:

1. The optimum location for construction of a climbing lane is at the midpoint of the upgrade (symmetrical).
2. The optimum length of a climbing lane is 1500 ft; shorter climbing lanes were not considered, since safety implications may restrict the use of shorter climbing lanes.
3. It is more efficient to construct one 1500-ft climbing lane on several upgrades than to have more than one 1500-ft climbing lane on one upgrade.
4. The benefit (travel time) that can be obtained from the construction of a climbing lane is most sensitive to the gradient.

Current Rural Highway Research

A new project, "Further Investigation of Traffic Operations on Two-Lane Two-Way Rural Highways," is currently under way. The two parallel efforts on this project are the further refinement and enhancement of the modified MIDWEST model and the development of a macroscopic-modeling approach for the evaluation of the performance of design elements of a highway.

ACKNOWLEDGMENT

This work resulted from the efforts of many individuals and the support of several organizations. The California Department of Transportation, the Federal Highway Administration, the Office of University Research of the U.S. Department of Transportation, and the National Science Foundation not only provided a large portion of the financial budget but gave excellent guidance and advice on a continuous basis. The Transport and Road Research Laboratory provided copies of TRANSYT5, TRANSYT6, and CONTRAM models and provided advice during periods of model refinement and application. This activity received the continual support of the Institute of Transportation Studies, and many Institute staff members are due a special word of thanks.

Finally, I would like to acknowledge the many individuals who directly participated in the development of these operational planning models. This effort would not have been possible without their enthusiastic involvement. It was a pleasure working with this fine group of researchers.

REFERENCES

1. A.D. May. Impact Assessment and Traveller Responses to Traffic Management Strategies. Institute of Transportation Studies, Univ. of California, Berkeley, Rept. UCB-ITS-RR-79-10, June 1979, 26 pp.
2. Y. Makigami, L. Woodie, and A.D. May. Bay Area Freeway Operations Study--Final Report, Part I: The Freeway Model. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Aug. 1970.
3. R.C. Blankenhorn and A.D. May. Freeway Operations Study--Phase III: FREQ2, A Revision of the "FREEQ" Freeway Model. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Interim Rept. 4-1, June 1972.
4. W.A. Stock, R.C. Blankenhorn, and A.D. May. Freeway Operations Study--Part III: The FREQ3 Freeway Model. Univ. of California, Berkeley, Rept. 73-1, June 1973.
5. R.D. Minister, L.P. Lew, K. Ovaici, and A.D. May. Computer Simulation Model for Evaluating Priority Operations on Freeways. HRB, Highway Research Record 461, 1973, pp. 35-44.
6. K. Ovaici, A.D. May, R.F. Teal, and J.K. Ray. Simulation of Freeway Priorities Strategies (FREQ3CP)--Final Report. Federal Highway Administration, U.S. Department of Transportation, June 1975, 471 pp.
7. M. Eldor and A.D. May. Freeway Operations Study--Phase III: Cost-Effectiveness Evaluation of Freeway Design Alternatives. Univ. of California, Berkeley, Institute of Transportation and Traffic Engineering, Rept. 73-1, June 1973.
8. J.J. Wang and A.D. May. Computer Model for Optimal Freeway On-Ramp Control. HRB, Highway Research Record 469, 1974, pp. 16-25.
9. M. Eldor and A.D. May. Integrated Freeway On-Ramp Control Strategies Under Normal and Incident Conditions. U.S. Department of Transportation, Interim Rept., March 1975, 100 pp.
10. M. Orthlieb and A.D. May. Freeway Operations Study--Phase IV: Freeway Corridor Control Strategies. Institute of Transportation and Traffic Engineering, Univ. of California, Berkeley, Rept. 74-5, Aug. 1975.
11. A.J. Kruger and A.D. May. The Analysis and Evaluation of Selected Impacts of Traffic Management Strategies on Freeways. U.S. Department of Transportation, Oct. 1976, 116 pp.
12. A.J. Kruger and A.D. May. Further Analysis and Evaluation of Selected Impacts of Traffic Management Strategies on Freeways. U.S. Department of Transportation, Oct. 1977, 47 pp.
13. M.P. Cilliers, A.D. May, and R. Cooper. Development and Application of a Freeway Priority-Lane Model. California Department of Transportation, Sacramento, Final Rept., Sept. 1978, 516 pp.
14. P.P. Jovanis, A.D. May, and W. Yip. FREQ6PE--A Freeway Priority Entry Control Simulation Model. California Department of Transportation, Sacramento, Final Rept., Sept. 1978, 516 pp.
15. P.P. Jovanis and A.D. May. Flex-Time Travel Choice: A Literature Review and Framework for Analysis. Institute of Transportation Studies, Univ. of California, Berkeley, Rept. UCB-ITS-WP-79-1, Jan. 1979, 56 pp.
16. P.P. Jovanis and A.D. May. Flex-Time Travel: Research Framework and Preliminary Findings. Institute of Transportation Studies, Univ. of California, Berkeley, Rept. UCB-ITS-WP-79-8, Aug. 1979, 97 pp.
17. F. Harrison, D. Jones, and P.P. Jovanis. Flex-Time and Commuting Behavior in San Francisco: Some Preliminary Findings--Summary Report. Institute of Transportation Studies, Univ. of California, Berkeley, Rept. UCB-ITS-RR-79-12, Aug. 1979, 19 pp.
18. D.W. Jones, Jr., F.D. Harrison, and P.P. Jovanis. Work Rescheduling and Traffic Relief: The Potential of Flex-Time. Institute of Governmental Studies, Univ. of California, Berkeley, Public Affairs Report, Volume 21, Feb. 1980, 10 pp.
19. P.P. Jovanis. Analysis and Prediction of Travel Responses to Flexible Work Hours: A Socioeconomic Workplace and Transportation System Perspective. Univ. of California, Berkeley, Ph.D. dissertation, Dec. 1979, 368 pp.

20. D.B. Roden. The Temporal Demand Aspects of Traveler Response to On-Freeway Entry Control Strategies. Institute of Transportation Studies, Univ. of California, Berkeley, Rept. UCB-ITS-WP-79-17, Nov. 1979, 154 pp.
21. D.B. Roden, W. Okitsu, and A.D. May. FREQ7PE--A Freeway Corridor Simulation Model. Institute of Transportation Studies, Univ. of California, Berkeley, Res. Rept. UCB-ITS-RR-80-4, June 1980, 348 pp.
22. V.C. Chan, R.G. Dowling, and A.E. Willis. Deriving Origin-Destination Information from Routinely Collected Traffic Counts--Phase I: Review of Current Research and Applications. Institute of Transportation Studies, Univ. of California, Berkeley, Rept. UCB-ITS-WP-80-1, Jan. 1980, 92 pp.
23. J.B. Schneider, C.L. Jette, and B.T. Lewis. FREGRAF: A Computer Graphics Aid to the Operation and Interpretation of FREQ6PE, A Freeway Simulation Model. Univ. of Washington, Pullman, Sept. 1979.
24. D.I. Robertson. TRANSYT: Traffic Network Study Tool. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, LR 253, 1964, 37 pp.
25. D.I. Robertson and P. Gower. User Guide to TRANSYT Version 6. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, SR 255, 1977, 107 pp.
26. T.J. Clausen and A.D. May. The Analysis and Evaluation of Selected Impacts of Traffic Management Strategies on Surface Streets. U.S. Department of Transportation, Oct. 1976, 119 pp.
27. P.P. Jovanis and A.D. May. Further Analysis and Evaluation of Selected Impacts of Traffic Management Strategies on Surface Streets. U.S. Department of Transportation, Oct. 1977, 258 pp.
28. P.P. Jovanis and A.D. May. Alternative Objectives in Arterial-Traffic Management. TRB, Transportation Research Record 682, 1978, pp. 1-8.
29. S.M. Easa, A.E. Willis, and A.D. May. Traffic Management of Dense Networks--Existing Experience and Analytical Techniques. Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper UCB-ITS-WP-78-11, Dec. 1978, 111 pp.
30. Micro Assignment--Final Report. Creighton-Hamburg, Inc.; Bureau of Public Roads, 1969.
31. D.R. Leonard, J.B. Tough, and P.C. Bagulez. CONTRAM: A Traffic Assignment Model for Predicting Flows and Queues During Peak Periods. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, TRRL Laboratory Rept. 841, 1978.
32. A.E. Willis, S.M. Easa, and A.D. May. Traffic Management of Dense Networks--Model Evaluation and Site Selection. Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper UCB-ITS-WP-79-5, July 1979, 60 pp.
33. S.M. Easa, A.E. Willis, and A.D. May. Traffic Management of Dense Networks--Application and Evaluation of Two Network Traffic Models. Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper UCB-ITS-WP-79-14, Oct. 1979, 111 pp.
34. S.M. Easa, T-L. A.Yeh, and A.D. May. Traffic Management of Dense Networks--Volume 1: Analysis of Traffic Operations in Residential and Downtown Areas. Institute of Transportation Studies, Univ. of California, Berkeley, June 1980, 223 pp.
35. S.M. Easa, T-L. A. Yeh, and A.D. May. Traffic Management of Dense Networks--Volume 2: User's Guide of the Refined Micro-Assignment Model. Institute of Transportation Studies, Univ. of California, Berkeley, June 1980, 176 pp.
36. W.A. Stock and A.D. May. Capacity Evaluation of Two-Lane, Two-Way Highways by Simulation Modeling. TRB, Transportation Research Record 615, 1976, pp. 20-27.
37. A.D. St. John and D.R. Kobett. Grade Effects on Traffic Flow Stability and Capacity. NCHRP, Rept. 185, 1978, 110 pp.
38. J.L. Botha and A.D. May. A Decision-Making Framework for the Evaluation of Climbing Lanes on Two-Lane, Two-Way Rural Roads--Literature Search. Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper UCB-ITS-WP-78-5, Sept. 1978, 37 pp.
39. J.L. Botha and A.D. May. A Decision-Making Framework for the Evaluation of Climbing Lanes on Two-Lane, Two-Way Rural Roads--Preliminary Evaluation of Candidate Models for Use in the Project. Institute of Transportation Studies, Univ. of California, Berkeley, Working Paper UCB-ITS-WP-78-12, Dec. 1978, 28 pp.
40. J.L. Botha and A.D. May. A Decision-Making Framework for the Evaluation of Climbing Lanes on Two-Lane, Two-Way Rural Roads--Final Model Selection, Simulation Methodology, and Proposed Data Collection. Institute of Transportation Studies, Univ. of California, Berkeley, Rept. UCB-WP-79-12, Oct. 1979, 93 pp.
41. J.L. Botha. A Decision-Making Framework for the Evaluation of Climbing Lanes on Two-Lane, Two-Way Rural Roads. Univ. of California, Berkeley, Ph.D. dissertation, May 1980, 454 pp.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.