Toward Optimal Planning of a Two-Mode Urban Transportation System: A Linear Programming Formulation

GEORGE A. HAY, Graduate Student, Economics Department;
EDWARD K. MORLOK[1], Lecturer and Graduate Student, Civil Engineering Department; and
ABRAHAM CHARNES, W. P. Murphey Professor of Applied Mathematics, Engineering Sciences Department, Northwestern University

The purpose of this study was to develop an analytical methodology or model for finding the optimal combination of two modes in providing transportation service. The specific case treated was that of providing automobile transport facilities and possibly some rapid transit facilities in a radial, downtown-oriented corridor. The objective was to find that combination of facilities which minimized transport costs—including both capital and operating costs of transit and auto transport—during the design or horizon year.

Since transportation is a service to its environment, the services provided were required to have certain attributes. The capacity of the two modes had to be capable of accommodating the peak period flows. Furthermore, the system had to be designed so that the peak period and non-peak period interzonal travel times did not exceed their respective maximum acceptable values. Because a two-mode system was dealt with, the modal choice behavior of travelers had to be incorporated into the model.

In order to insure the usefulness of the model, it was developed with reference to a specific real world situation in the Chicago area. The nature of the cost functions for the two modes and the constraints related to capacity, travel times, and modal choice was such that the prolem could be characterized within the framework of linear programming. This very efficient optimization technique was used to find the solution, which appeared to be quite reasonable.

•THE URBAN transportation planning process has been advanced to a high level of sophistication. It is now possible to predict future demand for transportation and to evaluate any transportation plan against that demand in terms of many different measures of performance. From any given set of alternative plans it is usually not difficult to select that one plan which is best according to some specified criterion, such as minimization of total annual cost subject to service constraints.

Despite the tremendous strides made in planning methodology, at least one serious weakness remains: the time and expense involved in developing and then evaluating an alternative plan precludes the consideration of the large number of plans which are quite different from one another yet are all reasonable alternatives and merit serious

Paper sponsored by Committee on Transportation System Evaluation and presented at the 45th Annual Meeting.

[1]Now with the Technical Analysis Division of the National Bureau of Standards.

consideration. In only a few studies, such as the Chicago Area Transportation Study (4), have a large number of highway alternatives been considered and evaluated. Even when this is done, highway planning and evaluation often proceed independently of that for public transportation. Under these circumstances it is very difficult to believe that near optimal network configurations and combinations of expressways, arterials, and mass transit facilities are found.

In this paper a mathematical model of urban transportation facilities in a radial corridor is presented. The purpose of the model is to circumvent some of the objections to conventional techniques. The essential attributes of various combinations of arterial streets, expressways and rapid transit, and the predicted demand for transportation are characterized within the framework of linear programming. This efficient computational technique is then used to find the optimum plan.

THE MODEL

This model considers the travel along a single corridor in an urban area. The objective function to be minimized is the total annual cost of transportation, during one design year:

$$\min \left\{\begin{matrix}\text{annual road}\\ \text{capital cost}\end{matrix}\right\} + \left\{\begin{matrix}\text{vehicle operating}\\ \text{cost}\end{matrix}\right\} + \left\{\begin{matrix}\text{annual transit}\\ \text{capital cost}\end{matrix}\right\} + \left\{\begin{matrix}\text{annual transit}\\ \text{operating cost}\end{matrix}\right\} + \left\{\begin{matrix}\text{annual parking}\\ \text{facilities cost}\end{matrix}\right\}$$

The specific form of these cost functions is discussed in a later section, since these are based on empirical data. Assumed costs of the time of travelers, while used in several other transportation studies, is not included here. The authors feel that the value of time is so dependent on the amount under consideration and the time of day, as well as the individual, that the concept of an average value is probably not particularly useful. The constraint set includes various types of travel time constraints which we feel are more meaningful than some hypothetical value of time.

With the above formulation, transportation costs would be minimized by producing no transportation. However, since transportation is a service to its environment, this service must have certain attributes in order to meet the needs of the environment satisfactorily. These requirements are reflected in the constraints of the problem.

All transportation systems are characterized by a capacity limitation, and in this case, it is required that design capacity meet or exceed the predicted demand. It should be borne in mind that it is a choice of transportation decision-makers whether to provide sufficient capacity to meet demands—often at considerable expense—or to limit capacity and thereby force a displacement of some trips in time and space.

Another set of constraints refers to the maximum travel time which will be permitted for trips between the various possible origin-destination combinations. These constraints can be precisely the interzonal travel times assumed in trip distribution; thus, this concept of travel time constraints is useful in fitting this model into the existing urban transportation planning methodology.

These constraints also point out a public policy question regarding the level of service and accessibility which are to be given to each region. Within the technological constraints on speed there is nevertheless a wide range of choice as to level of service and accessibility, and these must be dealt with directly. Of course, public expectation as to reasonable travel times and the willingness to pay the price of speed and capacity significantly influence decisions here. In this program lower bounds on speed are specified along with the upper bounds imposed by technology, permitting the program to choose any speed within this range.

Since the program deals with both road transport and rapid transit, the behavior of people regarding modal choice must be taken into account. This is also done in a constraint set, which attempts to duplicate one of the more sophisticated modal choice models currently used in planning studies. The choice is based on such factors as door-to-door travel time, transit waiting time, out-of-pocket costs, trip purpose, and socio-economic status of the traveler.

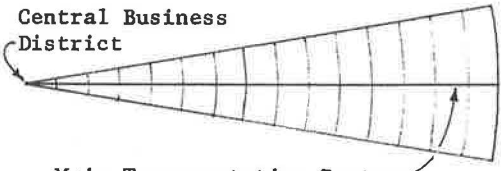

Figure 1. The region.

In order to demonstrate the usefulness of this model, it was implemented using cost and demand data from the Chicago area (1, 4, 6, 11, 14) and a modal choice model developed for the Washington area (8). The sole reason for these choices was the availability of data. It should be remembered that we are not solving for the actual optimum solution to Chicago's problem, because we are treating the problem assuming no rapid transit or expressways exist, and we are treating a hypothetical average Chicago area corridor, not a real one. These conditions were imposed by the difficulty of obtaining more complete and detailed data in the short time available. It is also emphasized that in any real world application it is necessary to obtain detailed cost and demand predictions for the region in question before this model can be expected to yield valid results.

Only the problem of transportation improvements in a radial corridor is considered. Because generally neither corridor travel demand nor costs follow any simple mathematical relationship with distance from the central business district, the 30-mi long corridor was divided into 2-mi long zones (Fig. 1). This permits approximations of any demand and cost distributions and greater accuracy, if desired, can be achieved by reducing zone length. Demands and travel times are treated on an interzonal basis, while cost parameters are uniform in each zone. In this particular application only travel to and from the central business district is considered, in order to simplify the computations, but the model can be used for all interzonal travel in the corridor.

Before discussing the model in detail, mention should be made of the relationship of this paper to the existing literature in the area. In this paper the concern is with the addition of capacity and improvement in the level of service in an existing network, where these additions can be in the form of incremental changes in existing streets or in the form of entirely new expressways or freeway-type facilities with the associated high threshold costs. The studies by Garrison and Marble (9), Carter and Stowers (3), and Quandt (12), however, are solely concerned with essentially continuous additions of capacity to existing facilities, while the work of Roberts and Funk (13) is concerned only with new investments of a very lumpy sort. Also, in our study the level of service to be provided is treated explicitly as a choice variable, subject to explicit constraints, whereas none of the other studies deal with this directly.

Beckmann (2) presents a very general and sophisticated model for freight flows, but this model is implemented by use of the calculus of variations. Unfortunately, algorithms for solving such problems have yet to be developed, so that his model is effectively not operational.

Creighton et al (7) treat investment in a two-mode system, but make some very questionable assumptions regarding the transport network configuration and the nature of cost characteristics. Moreover, it does not appear that systemic effects even on a link—much less a network—can be taken into account. We have attempted to develop the model so that it follows known cost functions and demand interrelationships closely.

This model also differs from the others mentioned in that it deals with only one corridor, not a complete network. Therefore, certain systemic effects cannot be dealt with. Nevertheless, it is felt that this model is useful, because in certain corridors—particularly radial corridors in larger cities—the flows are much larger than in the intersecting corridors.

THE OBJECTIVE FUNCTION

The cost functions for road transport, rapid transit, and parking are considered separately. The specific purpose of these functions is to relate the cost of producing transportation to measures of the amount and quality of the transportation produced. In each case some theoretical considerations are discussed first, and then the models are developed from data on actual systems.

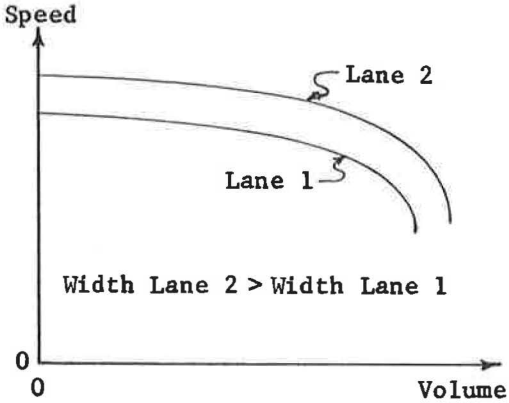

Figure 2. Lane capacity.

Road Capital Costs

The capital cost of highway facilities is related to the capacity of the road (the maximum vehicular flow rate it can accommodate) and the average speed at which this traffic moves. For one lane of any given road without signals or stop signs the speed and flow are related (Fig. 2). Thus an increase in speed with no change in capacity can be obtained by an increase in the number of lanes or, up to a point, in the width of lanes, both at an increase in cost. Similarly, increases in capacity with constant speed are associated with the expense of wider lanes or more lanes.

Because of these characteristics of flow, the capital cost function for a given length of road resembles the surface in Figure 3. This surface is drawn without discontinuities to represent changes in the number of lanes, because it is felt that changes in road width will provide for the specified changes in speed and capacity. In Figure 3 it is assumed that some sort of road already exists, for non-zero speeds and capacities can be obtained at zero cost. This is generally the case in urban areas, but the alternative can also be considered within the framework.

In the case where different road technologies are available the best for any given combination of speed and capacity can be determined rather easily. For graphical simplicity, consider one speed, with varying capacity. The cost curves might resemble those in Figure 4, with the choice of road type being that which yields lowest cost.

The above considerations lead to the linear road capital cost model for a road spanning zone i

$$C_i c_i + M_i \left(1 - \frac{m_i}{\bar{M}_i}\right) + S_i (m_i - s_i)$$

$$m_i \geq \underline{M}_i$$

$$m_i \leq \bar{M}_i,$$

$$s_i \leq m_i, \text{ and}$$

$$s_i \geq \underline{M}_i$$

where

C_i = annual unit capacity cost, \$ per vph;

c_i = capacity, vph;

M_i = annual unit peak period speed cost, \$;

m_i = peak period slowness, min/mi;

$\bar{M}_i$ = maximum (technological) slowness, min/mi;

$\underline{M}_i$ = minimum (technological) slowness, min/mi;

S_i = annual unit cost of additional non-peak period speed, \$ per min/mi; and

s_i = non-peak period slowness, min/mi.

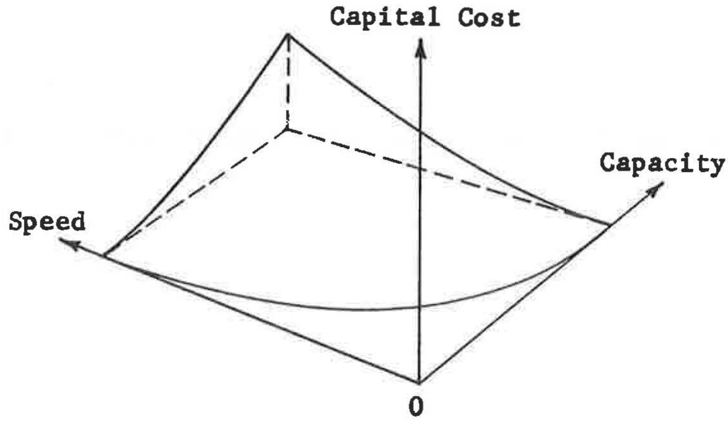

Figure 3. Road capital cost function.

Two factors require additional explanation: the use of minutes per mile or slowness rather than miles per hour or speed, and the inclusion of non-peak period speed. "Min/mi" was used so that the travel time between zone pairs would be a linear combination of the choice variables. Non-peak period slowness was included because it might be desirable to let a road operate at 30 mph, for example, during the peak period to keep capital cost low, but to design it so that non-peak drivers could safely drive at 50 mph during uncongested periods. Clearly there is an additional expense due to such features as longer acceleration and deceleration lanes on expressways and more adequate signing and signaling on arterial streets.

The parameters in this cost expression were estimated using data on some existing and proposed facilities in the Chicago area. Because of data limitations, no general validity is claimed for these estimates. As mentioned earlier, the purpose here is to demonstrate that this model is operational, not necessarily to solve a specific real world problem.

Costs for two different types of urban roads located near the central business district are plotted in Figure 5. The lower curve is for an arterial street with through-lane overpasses at major intersections, on which traffic can flow at about 2 min/mi. The upper curve is for a freeway type facility, designed for flow at about 1.2 min/mi. The other curves are based upon extrapolation with the linear model.

Costs for the arterials are taken directly from Haikalis (10), with an adjustment for the location. It was assumed that the ratio of downtown arterial to Haikalis' outlying arterial costs is the same as that ratio for freeways, \$15,500,000/\$12,000,000 = 1.29. This yielded an arterial cost of \$3,400,000 for a road with a 2,000 vph capacity at 2 min/mi. This total capital cost is converted to an annual cost with the assumption of a 30 yr life and an interest rate of 6 percent per annum, for an annual cost of about \$250,000 per mile of road.

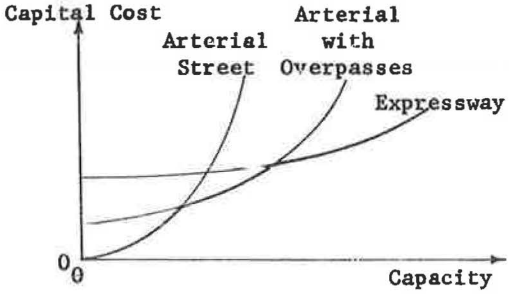

Figure 4. Choice of technology.

The freeway costs are based on the work of Aitken (1) and Satterly (14). Aitken reports that 6.4 mi of an 8-lane freeway entering the downtown area cost in average of \$15,500,000 per mi, or \$1,130,000 annually. According to Satterly the construction (but not right-of-way) cost of a

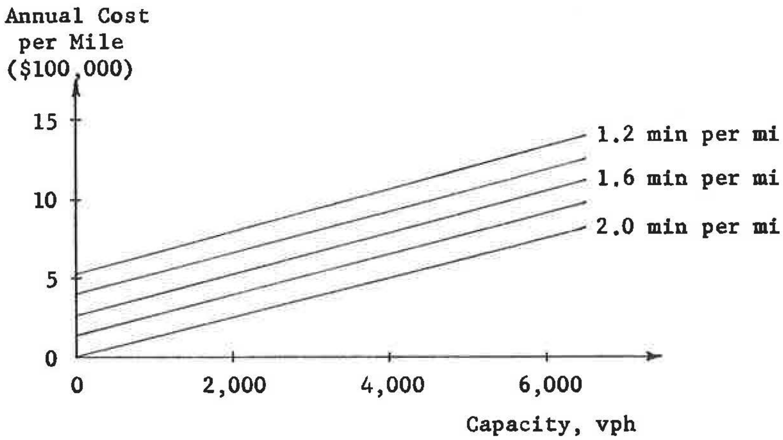

Figure 5. Cost surface for road near the CBD.

10-lane freeway with grade separations every one-half mile and interchanges every mile is $690,000 per lane-mi. Taking the marginal cost of a lane-mile to be Satterly's average construction cost plus one-half of Aitken's freeway right-of-way cost for one lane, the annual marginal lane-mile cost becomes $75,000. Each lane of a freeway can accommodate about 1100 vph at 1.2 min/mi. The resulting cost curve is as shown in Figure 5.

If the cost-slowness relationship is linear in the range of speeds under consideration (Fig. 6) the parameters can be evaluated readily. Unfortunately no data were available on high capacity urban roads built for travel times within the range of 1.2-2.0 min/mi, probably because none have been constructed recently, so that this assumption could not be tested.

The resulting parameter values for a two-mile roadway are as follows:

C = $250 per vph,

$\bar{M}$ = 2.0 mi/mi,

$\underline{M}$ = 1.0 min/mi,

M = $3,000,000, and

S = $600,000 per min/mi.

Since no information was available on the costs associated with S, this value was established from the educated guess that it would cost about $4,000,000 to improve the roadway design so as to permit an increase in speed, at a very low traffic volume, from 30 mph to 60 mph.

All autos entering the central business direct must be stored, and therefore parking costs must be included. The cost of constructing ramp garages in the Chicago Loop during 1954 and 1955 varied from $2,260 to $2,830 per space (6), so an approximation of $2,500 per space is used here. Assuming that demand patterns dictate that three spaces be provided for each one required during the peak hour, the annual cost coefficient becomes $550 per peak hour space. This can be added directly to the road cost coefficient for zone 1 to yield the capacity cost coefficient for that zone.

The cost coefficients for all other zones were developed in a similar manner, with the omission of parking costs. These gave a reasonably accurate representation of the rather scanty historical costs available. All of the coefficients are given in Table 1.

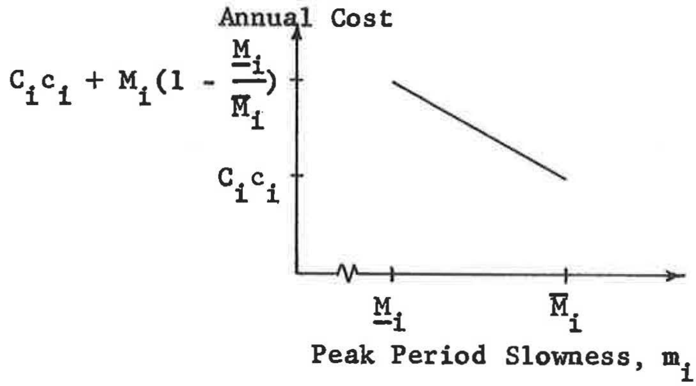

Figure 6. Cost as a function of m_i, with $m_i = s_i$.

TABLE 1

OBJECTIVE FUNCTION COEFFICIENTS

Zone	Coefficients				
i	C_i ($/vph)	M_i ($)	S_i ($/min/mi)	P_i (2/pass./hr)	V_i ($/veh-mi)
1	800	3,000,000	600,000	253	16.80
2	200	2,500,000	500,000	253	16.80
3	170	2,000,000	400,000	253	16.80
4	120	1,600,000	300,000	253	16.80
5	70	1,200,000	200,000	253	16.80
6-15	40	800,000	100,000	313	16.80

Vehicle Operating Costs

Vehicle operating costs are based on information given in Smith (15). We assume that one-half of the drivers using transit would get rid of the car which would otherwise be used for the trip, and therefore one-half of the auto users should be charged the marginal operating costs of driving, whereas the other half should be charged with the full costs.

Total operating costs, exclusive of garaging, parking, and tolls	$0.0751 per veh-mi
Marginal operating costs	0.0368
	$0.1119 per veh-mi

Average operating costs = 11.19/2 = $0.0560 per veh-mi. Converting this figure to an annual basis, using 300 equivalent days per year, V = $16.80 per daily veh-mi.

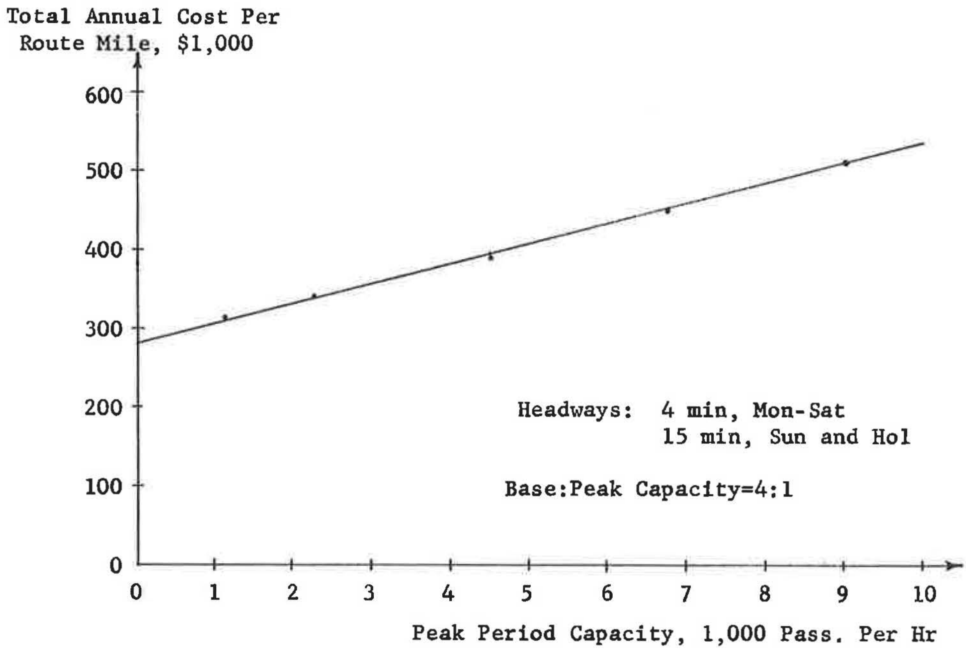

Figure 7. Annual rapid transit cost and capacity.

Rapid Transit Costs

The rapid transit costs used in this study are based entirely on a separate model of rail transit costs developed previously by Morlok. This model relies heavily on the data found in Lang and Soberman (11), and is similar to their model. It does, however, distinguish more completely between fixed and marginal costs than their model.

Although it would be inappropriate to discuss this model in detail here, the essential characteristics are as follows. The model represents a conventional rail transit line operating on a two-track elevated structure of modern design, which costs about $3.5 million per mi. Only high-speed, air-conditioned cars are used, but no automation of train operation or fare collection is assumed. Trains are operated for 16 hours of every day, and all passengers are provided with seats during the entire operating period.

In this model total annual costs can be predicted from capacity and headway during weekday peak periods, weekday non-peak periods, Saturdays and Sundays and holidays. Headways are fixed at 4 min during weekdays and Saturdays, and at 15 min on Sundays and holidays. Saturday, Sunday, and holiday capacity was set at one-eighth of that during weekday peak periods, and weekday non-peak period capacity was set at one-fourth peak capacity. These specifications leave peak period capacity as the only choice variable, yielding a cost function of the form

$$C_{rt} + \sum_{i=1}^{n} P_i p_i$$

where

C_{rt} = annual capital cost of rapid transit, $;

P_i = annual cost per unit of peak capacity, $ per pass./hr;

p_i = peak period transit demand from zone i, pass./hr.

Cost estimates based upon Morlok's model are given in Figure 7. The cost coefficients for a 10-mi long route—the length of the transit line to be considered in this example problem—are fixed cost, $2,870,000, are variable cost, $253 per pass./hr. (The use of an externally specified transit line length will be explained in the final section. Suffice it to say here that for each run of the program, costs will be minimized for a given transit route length, which is arbitrarily chosen at 10 mi for this example. The program is run for each of the possible transit line lengths, of which there are generally only a few reasonable alternatives.)

In addition to train expenses, parking costs vary directly with the capacity of the system. The unit parking cost used was that of constructing the lot at the outer terminal of Chicago's Skokie Swift line in 1964, approximately $275 per space (5). Again assuming a total of three spaces must be provided for each space required during the one peak hour, and that these facilities are paid for in 30 years at 6 percent interest, we have an annual parking cost of $60 per peak pass./hr.

Including the parking costs, the final cost coefficients for a 10-mi long transit line are as follows:

C_{rt} = $2,870,000,
P_i = $253 per pass./hr, $i = 1, 2, \ldots, 5$, and
P_i = $313 per pass./hr, $i = 6, 7, \ldots, 15$.

THE CONSTRAINTS

This section examines the set of constraints which characterize our problem. The equations related to capacity, modal choice, level of service, and the calculation of vehicle-miles are considered separately.

Linearity

Our constraints as well as our objective function are, of course, in linear form. The real world is obviously not that neat. We do not feel, however, that any unjustifiable liberties were taken in achieving linearity. For one thing, all the relations did, in fact, closely approximate linearity, at least in the range which was relevant for the problem. Furthermore, no sharp discontinuities are apparent, suggesting that a linear approximation will not give vastly unrepresentative results. Finally, and most important, nonlinearity has not proved destructive of linear programming in the past due to the existence of techniques such as piecewise linear approximation. Though the problem would no doubt become substantially more complicated, there is no reason to believe that such techniques could not be used here. It is left to critics to show that even a complicated problem is not superior to the next best technique available for the solution of such an urban transportation problem.

Capacity

The capacity constraints are of the form

$$Ec_j + \sum_{i=j}^{n} p_i \geq \sum_{i=j}^{n} D_i \qquad j = 1, \ldots, k$$

$$Ec_j \geq \sum_{i=j}^{n} D_i \qquad j = k + 1, \ldots, n$$

where

c_j = capacity of the road in zone j, vph;
E = 1.5 persons per vehicle, average automobile occupancy;

p_i = peak hour transit passengers originating in zone i, pass./hr;

D_i = total number peak hour passenger trips generated in zone i, persons/hr; and

k = the last zone served by transit.

The D_i are generated from the estimated 1980 population figures in each radial ring as estimated by CATS (4). The ratio of trips to population for 1956 is given as 0.163. This figure is retained for 1980. Since we are dealing with a corridor representing one-seventh of the population, the D_i are given by applying a coefficient of 0.0023 = (1/7) (0.163) (0.1) to the CATS figures. The 0.1 is to convert daily into peak hour passengers given that approximately one-tenth of daily trips are made during the peak hour. The specific D_i are presented in Table 2.

TABLE 2

PEAK PERIOD TRIP GENERATION

Zone i	Trips D_i (persons/hr)	Zone i	Trips D_i (persons/hr)
1	356	9	1040
2	1281	10	1076
3	2139	11	1006
4	1976	12	1027
5	1868	13	1093
6	1391	14	584
7	1325	15	720
8	951		

This set of constraints stipulates that the total transportation capacity of the zone must at least equal the number of passengers who will pass through that zone during the peak hour. Obviously, if the capacity of the system is sufficient to meet peak hour demand, it will, because of the definition of peak hour, be able to meet all remaining demand as well. We have taken the peak hour demand to represent 10 percent of the daily total, and have estimated that 40 percent of total daily demand will occur under peak hour conditions, i.e., there are four "equivalent" peak hours. This latter consideration is important in estimating the modal split since the split will, in general, be different during the peak and off-peak periods. While c_i represents a true capacity (since it will not, in general, be reached except during peak periods), p_i is both the demand and the capacity for rail transit. Thus we assume that the rapid transit line is operated with no excess seating capacity at the location of maximum loading during the peak hours. This set of constraints, then, because of this identity, can be seen as stipulating the road capacity in each zone.

Modal Choice

The modal choice constraints are of the form

$$p_j = A_j \sum_{i=1}^{h} m_i + B_j \qquad h = j \text{ if } j \leq k \text{ and } h = k \text{ if } j > k$$

where

A_j, B_j = empirical constants related to modal choice characteristics and D_j.

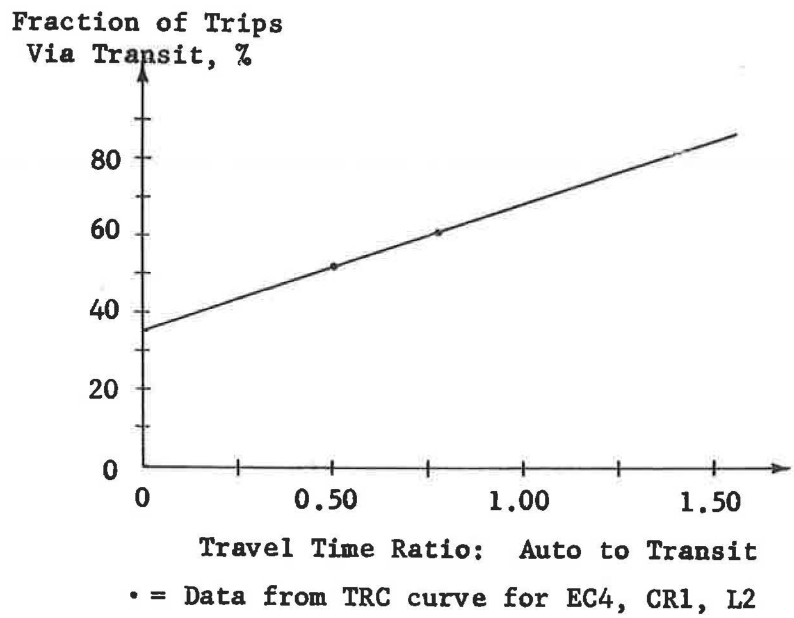

Figure 8. Example modal choice curve.

As previously explained m_i denotes the slowness of the road system in zone i, during peak hour conditions. The basis for these equations is the Deen, Mertz and Irwin report (8), in which the percentage of passengers going by transit is formulated as a function of the ratio of travel times of the two modes with certain cost and service characteristics entering as parameters, along with the income class of the group. Their curves, which express transit travel as a function of travel time by transit to that via automobile, were not, in general, of linear shape. However, we found that by plotting

TABLE 3

MODAL CHOICE EQUATIONS

Income Range ($/family/yr)	Zones i	Equation
3100 or less	2	$\frac{P_i}{D_i} = 0.65 + 0.125\,(TTR)$[a]
3100-4700	3	$\frac{P_i}{D_i} = 0.60 + 0.100\,(TTR)$
4700-6200	4	$\frac{P_i}{D_i} = 0.35 + 0.350\,(TTR)$
6200-7500	1, 5, 6	$\frac{P_i}{D_i} = 0.35 + 0.333\,(TTR)$
7500 or more	7-15	$\frac{P_i}{D_i} = 0.17 + 0.500\,(TTR)$

[a]TTR = Travel time ratio, auto to transit

the percentage of rail passengers against the inverse of their travel time ratio, we achieved a very nearly linear relation. An example curve is shown in Figure 8, and the equations are given in Table 3.

Deen, Mertz and Irwin (8) further divide their split estimates into work and non-work groups, but the data from which they derived the non-work estimates were so sparse that we ignored the minor differences between them and used their work trip relationships for all trips. There are still two modal split equations for each zone, however, since the peak hour travel time ratio and demand will, in general, be different from the non-peak ratio. Only the peak hour mode choice equations will affect the capacity constraints because of the definition of road capacity.

Three points should be made concerning the modal split equations. First, the relative income status of each zone was assigned on an intuitive basis aided by the author's experience in the Chicago area. A more rigorous method of assignment is, of course, desirable, but, unfortunately, not readily available on a zone-by-zone basis. Secondly, the Deen, Mertz and Irwin equations were derived primarily from data for Washington, D. C. It is assumed that these relations are valid for Chicago, partly because such features as income status are separated out of the equations as parameters. The ultimate reason for the assumption is, as always, the lack of such data for Chicago. Finally, it should be noted that as the equations stand there is nothing in the mathematics which would prevent the percentage of passengers going by rail from exceeding 100. Ideally one would like to be able to formulate the equations so as to prevent this possibility without putting restrictions on the travel time ratio. Unfortunately there seems to be no easy way of doing this without substantially cluttering up the model or resorting to nonlinear equations. However, the use of upper and lower bounds on slowness for both technological and service reasons not only serves these primary purposes but these constraints should also in general act to retain the percentage going by rail well within the 0-100 range. An additional check is present in the constraints which limit total travel time.

The travel time ratio for zone j which appears in the modal split equations of Table 3 is of the following form

$$\frac{\sum_{i=1}^{h} L_i m_i + H_h}{\sum_{i=1}^{h} L_i R + W_j} \qquad \begin{matrix} h = j \text{ if } j \leq k \\ h = k \text{ if } j > k \end{matrix}$$

where

L_i = length of zone i (two miles for all zones);

H_h = average time required for the traveler to go from his home to the main corridor highway plus the time required to travel from the highway in the CBD to his destin[illegible]n when h = j (i. e, j ≤ k) or simply the average time from the highway to his downtown destination when h = k (i. e., j > k) since the travler in this instance is already on the highway and is considering the benefits of transferring to rail at zone k—thus, in our modal choice equation, only the time still left to be spent traveling is relevant;

R = uniform average slowness of transit, min/mi; and

W_j = average walking and waiting time to and from the transit station when zone j is served by transit (both ends of the trip), or the average transfer and waiting time for transit when zone j is not served by transit, plus the walking time at the downtown end.

When numbers are chosen for the H_h and W_j and the value of R is entered, the modal choice equations take the form shown in the constraints. For our example we use H_h = 15 min for j ≤ k, H_h = 8 min for j > k, W = 5 for all zones, and R is a uniform 1. 89 min per mi.

The third set of constraints in this group is of the form

$$d_j = X_j \sum_{i=1}^{h} m_i + Y_j \sum_{i=1}^{h} s_i + Z_j \qquad h = j \text{ if } j \leq k \text{ and } h = k \text{ if } j > k$$

where

d_j = daily travelers from zone j traveling via transit, pass./day; and

X_j, Y_j, Z_j = empirical constants related to modal choice characteristics and D_j.

Thus the sum of peak and off-peak demand, d_j, is obtained from a set of equations identical in form to those appearing in the previous set of constraints with the exception that s_i, slowness in the off-peak periods, is used in addition to m_i. This set of constraints is, as seen, composed of equalities which are used in the next set to determine total vehicle-miles for the purpose of deriving operating costs of automobile travel.

Vehicle-Miles

The constraints which determine total daily vehicle movement are

$$v_j = \frac{2}{E}\left(\sum_{i=1}^{j} L_i + F_j + G\right)(10D_j - d_j) \qquad j \leq k$$

$$v_j = \frac{2}{E}\left(\sum_{i=1}^{k} L_i + G\right)(10D_j - d_j) + \frac{2}{E}\left(\sum_{i=k+1}^{j} L_i + F_j\right)(10 \cdot D_j) \qquad j > k$$

where

v_j = total daily automobile movement due to trips generated in zone j, veh-mi;

F_j - average distance from the home to the main corridor road, mi; and

G = average distance from the main corridor road to the downtown parking location, mi.

In this problem we took F_j as 3 mi for all zones and G as 2 mi.

These last two constraint sets, giving daily transit travel and vehicle-miles of automobile movement, could have been collapsed into one set. The reason for this is that daily transit travel does not enter directly into the criterion function and there is a unique relationship between daily transit travel and vehicle-miles of travel for each zone. But these were left separate so that the solution would include the important statistic of total daily transit travel for each zone.

Level of Service

The constraints on level of service imposed by current highway technology, reflected in the upper and lower bounds on slowness in this program, have already been discussed. In addition, a set of overall travel time constraints is included:

$$\sum_{i=1}^{j} L_i m_i \leq T_j \qquad j = 1, 2, \ldots, n$$

and

$$\sum_{i=i}^{j} L_i s_i \leq T_j' \qquad j = 1, 2, \ldots, n$$

where

T_j = maximum peak period travel time from zone j to the downtown, min; and

T'_j = maximum non-peak period travel time from zone j to the downtown, min ($T'_j \leq T_j$).

These specify that the total travel time from any particular zone to the CBD be not greater than an externally defined number. In general any model which attempts to specify facilities to meet a target demand should first be able to satisfy the total travel times assumed, since that factor is an important element in determining future demand. However, in other models of this type, it is often found that the travel time which is generated by the facilities which are planned is different from that which was specified in order to predict demand. It is then necessary to start the problem again with a different total travel time, correspondingly different demands and facilities, etc. It is hoped that this procedure will lead eventually to a solution which is consistent with the assumptions. It is clear that this problem is avoided in our model. The total travel time and therefore the demand can be specified with certainty and entered as a constraint.

In our problem we chose to include only three sets of travel time constraints, feeling that 30 mph travel was satisfactory for zones within 18 mi of downtown. These constraining travel times are given in Table 4.

RESULTS

The Matrix

At this point, we can present the complete matrix of the linear programming problem:

$$\min \sum_{i=1}^{n} \left(C_i c_i + M_i \left(1 - \frac{m_i}{M_i}\right) + S_i (m_i - s_i) + V_i v_i + P_i p_i + 0 \cdot d_i\right) + C_{rt}$$

Subject to

$$Ec_i + \sum_{j=i}^{n} p_j \geq \sum_{j=i}^{n} D_j \qquad i = 1, 2, \ldots, k$$

$$Ec_i \geq \sum_{j=i}^{n} D_j \qquad i = k+1, k+2, \ldots, n$$

$$p_i = A_i \sum_{j=1}^{i} m_j + B_i \qquad i = 1, 2, \ldots, k$$

$$p_i = A_i \sum_{j=1}^{k} m_j + B_i \qquad i = k+1, k+2, \ldots, n$$

$$d_i = X_i \sum_{j=1}^{i} m_j + Y_i \sum_{j=1}^{i} s_j + Z_i \qquad i = 1, 2, \ldots, k$$

$$d_i = X_i \sum_{j=1}^{k} m_j + Y_i \sum_{j=1}^{k} s_j + Z_i \qquad i = k+1, k+2, \ldots, n$$

$$v_i = \frac{2}{E}\left(\sum_{j=1}^{i} L_j + F_i + G\right)(10 \cdot D_i - d_i) \qquad i = 1, 2, \ldots, k$$

$$v_i = \frac{2}{E}\left(\sum_{j=1}^{k} L_j + G\right)(10 \cdot D_i - d_i) + \frac{2}{E}\left(\sum_{j=k+1}^{j} L_j + F_i\right)(10 \cdot D_i)$$

$j = k + 1, k + 2, \ldots, n$

$$\sum_{j=1}^{i} L_j m_j \leq T_i, \qquad i = 1, 2, \ldots, n$$

$$\sum_{j=1}^{i} L_j s_j \leq T'_i, \qquad i = 1, 2, \ldots, n$$

$$m_i \geq \underline{M}_i, \qquad i = 1, 2, \ldots, n$$

$$m_i \leq \overline{M}_i, \qquad i = 1, 2, \ldots, n$$

$$s_i \leq m_i, \qquad i = 1, 2, \ldots, n$$

$$s_i \geq \underline{M}_i, \qquad i = 1, 2, \ldots, n$$

The tableau is shown in Figure 9.

TABLE 4

MAXIMUM TRAVEL TIMES

Zone i	Maximum Travel Time to Inner End of Zone 1	
	Peak Period T_i (min)	Non-Peak Period T'_i (min)
10	35	30
13	41	36
15	45	40

TABLE 5

FRACTION OF FLOW VIA TRANSIT IN ZONES 1 TO 5

Zone i	Fraction of Flow via Transit	
	Peak Period (%)	Daily (%)
1	78.2	78.2
2	77.8	77.8
3	77.2	77.2
4	77.8	77.8
5	76.8	76.8

TABLE 6

FRACTION OF TRIPS VIA TRANSIT

Zone i	Fraction of Trips Via Transit	
	Peak Period (%)	Daily (%)
1	99.0	99.0
2	85.2	85.2
3	74.1	74.1
4	79.1	79.1
5-6	75.5	75.5
7-15	78.0	78.0

TABLE 7

ROAD DESIGN SPEEDS

Zone i	Design Speed	
	Peak Period (mph)	Non-Peak Period (mph)
1-5	30	30
6-7	30	60
8	40	60
9-15	60	60

$c_1..c_i..c_n$	$m_1..m_i..m_n$	$s_1..s_i..s_n$	$v_1..v_i..v_n$	$p_1..p_i..p_n$	$d_1..d_i..d_n$	Variables
$C_1..C_i..C_n$	$...S_i-\frac{M_i}{\bar{M}_i}...$	$-S_1.-S_i.-S_n$	$V_1..V_i..V_n$	$P_1..P_i..P_n$	0...0...0	Costs
E_1 . . E_i . . E_n				1...1...1 . . 1.1...1	$\geq\sum_{j=i}^{n} D_i$	Capacity
A_1. $A_i.A_i$ $A_n.A_n$				1 . . 1 . . 1	$=B_1$ $=B_i$ $=B_n$	Peak Mode Choice
X_1. $X_i.X_i$ $X_n.X_n$	Y_1. $Y_i.Y_i$ $Y_n.Y_n$				1 $=Z_1$ . . 1 $=Z_i$ . . 1 $=Z_n$	Daily Mode Choice
			1 . . 1 . . 1		U_1. $=W_1$ $U_i.U_i$ $=W_i$ $U_n.U_n$ $=W_n$	Vehicle-Miles
	L_1. $L_1..L_i$. $L_1..L_i..L_n$				$\leq T_1$ $\leq T_i$ $\leq T_n$	Peak Travel Time
		L_1. $L_1..L_i$. $L_1..L_i..L_n$			$\leq T'_1$ $\leq T'_i$ $\leq T'_n$	Non-Peak Travel Time
	1 . . 1 . . 1				$\leq\bar{M}_1$ $\leq\bar{M}_i$ $\leq\bar{M}_n$	Upper Bound, Peak Slowness
	1 . . 1 . . 1				$\geq\underline{M}_1$ $\geq\underline{M}_i$ $\geq\underline{M}_n$	Lower Bound, Peak Slowness
	1 . . 1 . . 1	-1 . . -1 . . -1			≥ 0 ≥ 0 ≥ 0	Upper Bound, Non-Peak Slowness
		1 . . 1 . . 1			$\geq\underline{M}_1$ $\geq\underline{M}_i$ $\geq\underline{M}_n$	Lower Bound, Non-Peak Slowness
$c_i, v_i, p_i, d_i \geq 0,\ i=1,2,...,15$						

Figure 9. The tableau.

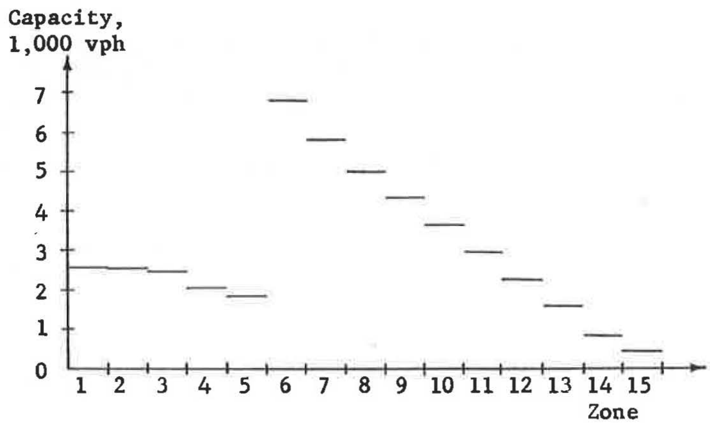

Figure 10. Road capacity at peak period speeds.

Solution and Interpretation

The main results of the computer program are presented in Tables 5 to 7 and Figure 10 which indicate, respectively, the percentage of traffic handled by rail in zones 1 to 5, the percentage of passengers from each zone who go by rail (within the last 10 miles), the peak and non-peak speeds, and the prescribed road capacity for each zone. Of course, the program is designed to yield normative rather than descriptive results, but one is comforted when the prescribed results are in the same vicinity as observations of current conditions. This seems to be the case with our particular example. For instance, the values in Table 5 coincide well with the empirical observation that 87 percent of Chicago's current peak hour downtown oriented traffic goes by transit.

The optimal choice is to build the more expensive high speed facilities in the outer zones, provided the travel time constraints can be met. This is in contrast to the planning in many areas where full freeways are called for even in the heart of downtown areas. However, if in our problem lower travel times in the zones near the downtown areas are desired, higher type facilities would have to be constructed in those areas also.

The dual to the road plus transit capacity constraints appears to be amenable to interpretation. As expected, the constraints are satisified with equality and hence the duals exist and are positive. They range from $5300 in zone 1 to $.027 in zones 7 to 15. This we interpret representing the maximum amount one could profitably bribe a passenger in a given zone to do his traveling through that zone during an off-peak hour.

From Minimum to Minimum Minimorium

At this point we attempt to explain the meaning to the solution we have found. We have set the length of the transit system at 10 miles (extending through zone 5) and have set up the program to select those values of the choice variables which minimize the total annual cost of building and operating the multi-mode system subject to the constraints, given that k = 5. The value of the objective function at the optimum is approximately $52.3 million to which must be added the capital cost of the given transit system (since this is constant for a given length transit) to arrive at the total annual cost of the given program—$55.2 million. This, of course, is only a minimum. To find the overall solution we must select that value of k, the length of the transit line, which yields the minimum minimorium; in other words, that complete system for which total annual cost (including capital costs) is minimized. The complete problem, then, requires additional runs of the program in which some (but not all) of the coefficients will change.

We feel that we are justified in restricting k to a few values because of the real world observation that transit stations and the corresponding parking lots are feasible at only certain points. Further, if the curve relating total costs to transit length represents any kind of a smooth function we can be reasonably confident that allowing k to take on continuous values will not have any significant effect on the solution.

Extensions

The model oversimplifies the real world in two important respects which we feel should be the main targets for additional research and refinement. First, the model deals with a single corridor, one of seven rays emanating from the CBD. The ideal program should treat the entire network, including the radial and circumferential facilities. Second, the notion of planning for a target year, although frequently employed, is patently unrealistic. The program should be dynamized to find that sequence of construction of new facilities and extension of existing ones which would optimize some cost criterion while providing transportation for a population which is expanding year by year rather than in discrete jumps of twenty years. However, although the solution we have presented does not fully represent the needs of the real world, we feel that it provides a good point from which to begin.

SUMMARY

This paper presents a linear programming model of an urban transportation problem, viz, the design of a two-mode transportation system in an urban corridor for a target year. The specific example used to test the feasibility and efficiency of the model employed data for the city of Chicago for a target year of 1980. The model differs from other methods of solution in that it selects that system which is optimal among all possible systems of a given type rather than merely examining a small number of alternatives.

The objective function to be minimized represents the total annual cost of the entire system. The standards of service are specified in the constraints. The primary choice variables are the transit capacity, the highway capacity, and the peak and off-peak highway speeds. The length of the transit route enters parametrically but by a finite number of runs of the program this also becomes a choice variable.

The model does not force any persons to a particular mode of transportation against their will, except insofar as the transit line extends only a certain length into the corridor. Rather, each individual makes his choice on the basis of several parameters, the most significant of which is the relative travel time of the two modes. It is these travel times which are the primary operational variables of the planners.

The model proved computationally feasible and appeared to yield reasonable results. Certain caution is urged, however, in the use of the model without the proper data. Furthermore, the model represents only a first (but important) step in the approximation of reality; the usual trade-off between model validity and operational ease still remains.

We hasten to assert that much of the application of linear programming to real world problems represents a learning process. One starts with a basic model and tests for validity and feasibility. This is what we have done. Moreover, one gains insights into what must be added to the model and how it might be changed by examination of the results of the simple problem. The theoretical work is not yet completed. But perhaps the most vital area of work which remains is the empirical. The model, as a tool for practical policy, is a function of its coefficients. Without the proper coefficients or at least reasonable approximations the model remains an abstraction.

ACKNOWLEDGMENTS

This research is based on work completed at Northwestern University. For their support and contribution of ideas and constructive criticism, we would like to thank J. E. Snell, F. J. Wegmann, D. S. Berry, G. T. Satterly and P. W. Shuldiner. We would also like to express our appreciation to Michael J. L. Kirby and Charles Mylander of Research Analysis Corporation for their aid in running the program.

This paper was originally issued as Systems Research Memorandum No. 134, by the Systems Research Group, The Technological Institute, The College of Arts and Sciences, Northwestern University. Part of the research underlying this report was undertaken for the Office of Naval Research, Contract Nonr-1228(10), Project NR 047-021.

REFERENCES

1. Aitkin, P. W. Cost of Constructing Transportation Facilities in the Chicago Standard Metropolitan Area. CATS Research News, Vol. 1, No. 14, pp. 8-11, Aug. 2, 1957.
2. Beckmann, Martin. A Continuous Model of Transportation. Econometrica, Vol. 20, No. 4, pp. 643-660, Oct. 1952.
3. Carter, E. C., and Stowers, J. R. Model for Funds Allocation for Urban Highway Systems Capacity Improvements. Highway Research Record 20, pp. 84-102, 1963.
4. Chicago Area Transportation Study. Final Report, Vol. II, Transportation Plan. April, 1962.
5. Chicago Transit Authority. Skokie Swift Progress Report No. 2. pp. 8-13, July 1964.
6. City of Chicago, Bureau of Parking. 1960 Annual Report. pp. 4-5, 1961.
7. Creighton, Roger L., Gooding, David I., Hemmens, George C., and Fidler, Jere E. Optimum Investiment in Two-Mode Transportation Systems. Highway Research Record 47, pp. 23-45, 1964.
8. Deen, Thomas B., Mertz, William L., and Irwin, Neal A. Application of a Modal Split Model to Travel Estimates for the Washington Area. Highway Research Record 38, pp. 97-123, 1963.
9. Garrison, W. L., and Marble, D. F. Analysis of Highway Networks: A Linear Programming Formulation. HRB Proc., Vol. 37, pp. 1-17, 1958.
10. Haikalis, George. Economic Analysis of Roadway Improvements. Chicago Area Transportation Study, Chicago, Dec. 1962.
11. Lang, A. Scheffer, and Soberman, Richard M. Urban Rail Transit; Its Economics and Technology. The M.I.T. Press, Cambridge, 1963.
12. Quandt, Richard E. Models of Transportation and Optimal Network Construction. Jour. of Regional Sci., Vol. 2, No. 1, pp. 27-45, Summer 1960.
13. Roberts, P. O., and Funk, M. L. Toward Optimum Methods of Link Addition in Transportation Networks. Presented to the Committee on Highway Programming of the Highway Research Board, Washington, D. C., Jan. 1965. Unpubl.
14. Satterly, Gilbert S. Spacing of Interchanges and Grade Separations on Urban Freeways. Unpubl. Doctoral Diss., Northwestern Univ., p. 105, Dec. 1964.
15. Wilbur Smith and Associates. Future Highways and Urban Growth. Automobile Manufacturers' Association, New Haven, Feb. 1961.

Discussion

EDWARD F. SULLIVAN, Tri-State Transportation Committee—Mr. Morlok and Mr. Hay under Dr. Charnes' direction have made an interesting attempt at applying the power of linear programming to development of urban transportation plans. As the authors are quick to point out, their work is just a beginning. But they have pointed the way towards development of mathematical programming methods which might provide assistance to transportation planners and decision-makers. Although the complexities of transportation system capacities and demands defy simple formulation, mathematical programming (linear, dynamic, etc.) holds sufficient promise to warrant encouragement of further development of such methods for transportation planning.

What may we expect of mathematical programming? It is not simply another set of formulas. Rather, it is a conceptual approach in which an entire system is described in a comprehensive (even though simplified) way, from the viewpoint of how to allocate resources most effectively for a whole enterprise—in our case, the transportation system.

An integrated transportation system offers a wide choice of alternatives. Different costs are associated with different operational and investment alternatives. Resources are allocated to the component parts of this integrated enterprise. In a business the ultimate objective is to allocate resources to maximize overall profitability. This objective applies equally well to a public enterprise (such as transportation) if instead of "profit" we say "difference between gains (benefits) and costs."

In linear programming an objective (cost) function describes an economic objective in terms of which the system is described. Operational variables describe the interdependence of various activities of the system corresponding to physical conditions. Investment variables establish the physical configuration of the system and introduce the effect of additional capacity on operations. The value taken by the cost function depends on the values assigned to all of the operational and investment variables. Determining the best plan of action or the "optimum solution" for a transportation system consists of finding a set of values for all the variables so as to maximize (or minimize) the cost function while, at the same time, satisfying all the relationships which describe the physical operation of the system. This set of interrelationships includes not only equations describing interdependency, but also inequalities which describe limitations imposed on the system.

Sets of optimum solutions can readily be generated corresponding to various levels of demand and facility investment. Likewise, the results of different policies can be tested by restatement of the objective function. For example, we might examine optimal solutions based on minimizing total transportation costs, minimizing public costs, or maximizing benefits minus costs. Useful by-products are generated with each solution which make it possible to study the sensitivity of each variable, and costs imposed by each constraint. Coefficients can be checked to see how far their values might be changed before changing the strategy indicated by the solution.

Thus, mathematical programming is a potentially powerful tool for transportation system development and evaluation. It deals efficiently with large amounts of information and can explore systematically a great number of alternatives and restrictions characterizing the functioning of a complex transportation system.

The paper under discussion meets some of these expectations, but falls short of others.

In the model, transportation costs (both capital and operating) are minimized. Limits to be satisfied include demand volumes, minimum speeds and maximum times. Cost formulas reflect capital and operating costs corresponding to facility demand levels. Another formula determines mode choice as a function of auto vs transit travel time.

A noteworthy feature inherent in this approach is that the formulas describe all feasible possibilities within a broad range. Within this range of feasible solutions, the most economical combination is found by systematically converging calculations. The authors have set out to provide a means for assuring that alternatives considered in the transportation planning process are within the optimal range, taking into account the cost and performance characteristics of the various elements of the highway and transit systems.

The general applicability of the method as presented is severely limited by the simplified assumptions, such as dealing only with CBD trips through one corridor. Recognizing that in this first effort such assumptions were necessary to keep the problem manageable, the question remains whether a more comprehensive description of the system might be achieved.

The only operational variable is highway speed. Within the model the percent using transit varies only with the travel time ratio (auto/transit). Person trip demands are held fixed, and only CBD trips within a single corridor are considered. Thus, there is no provision for changes in magnitude or orientation of demand with changes in capital investment. Whereas the model does reflect changes in mode usage in response to system investment, it does not reflect changes in trip orientation, which may be of equal significance. Therefore, the optimal linear programming solution would have to be checked by more explicit system-wide trip distribution and assignment.

Likewise, the only investment variable relates highway capital cost to highway speed. Here, the formulation seems needlessly oblique, expressing the increments in cost incurred to provide sufficient capacity to maintain levels of service. The off-peak term seems an unnecessary and unlikely provision, since it is difficult to conceive of saving significant costs by reducing geometric and traffic design standards. The other investment variable, transit capital cost, is actually introduced as a constant.

The fixed set of person trips implies the same average length of trip, regardless of mode. Recent evidence seems to point to longer CBD trips by transit than by auto. This greater length is a counterbalance to the small increment of transit trip cost with distance (Fig. 7).

Only one set of highway capital costs is employed (Fig. 5). It appears that existing facilities can be handled by the present model simply through appropriate cost coefficients. Further development of the model might well incorporate highway cost as a function of area characteristics such as development density.

The assumption of one radial road competing with a transit line in each corridor is troublesome, leading to gross assumptions, such as an average of 3 miles from home to the radial road and 2 miles from the road to downtown parking. Such a constant assumption may well dictate the solution more than the variables. It also neglects to reflect alternatives within the highway system, such as sharing the traffic load between arterial streets and freeways. In other words, the description of the system is not explict enough to describe its operation properly.

The model does not consider whether the optimum solution is fiscally feasible. Calculations of highway and transit revenues, however, could readily be made from the outputs, along with the assumptions regarding fares and tax revenues. If the optimum solution were too costly, re-orientation of demand might be indicated, or constraints would have to be relaxed, such as lowering minimum speeds. Conversely, the cost of providing better service could be assessed by tightening the constraint limits.

In summary, these remarks are intended to encourage further explorations into applying mathematical programming to transportation system planning.

Reference

16. Rapoport, L. A., and Drew, W. P. Mathematical Approach to Long-Range Planning. Harvard Business Review, May-June, 1962.

KENNETH J. SCHLAGER, Southeastern Wisconsin Regional Planning Commission—It is important to comment that the relative lack of previous interest in design models as opposed to forecasting models is indicated by the fact that only one paper presented herein deals with a design model. The other models are related to forecasting and policy formulation problems. The lack of plan design models, or at least conceptual plan design frameworks, has severely limited the determination of requirements for data collection and analysis in urban transportation studies. The largest costs in urban transportation planning relate to the collection and analysis of data. The great majority of these data are used for describing the current state of the system and for forecasting probable future development. Very little data are collected that allow for the consideration of alternative plan designs. Since most studies do not provide for such data, much less a model framework for evaluating plan designs, the degree to which real alternatives are considered in a final plan is open to questions. Transportation planning seems to be in the same situation as defense planning was in this country before the planning-programming-budgeting approach to planning that for the first time allowed for the consideration of alternatives to meet stated objectives.

In the area of model formulation, one question that might be directed at the paper is the treatment of a transportation plan design without regard to land use. It has become an accepted concept in urban planning that transportation and land use interact. At the very least, transportation models should be constrained by land-use requirements because one of the possible solutions to a transportation model so constrained is that no new transportation facilities will be needed at all. Difficulties exist sometimes in quantifying certain land-use constraints in a transportation model, but an imperfect quantification of such constraints is usually better than ignoring such constraints altogether.

Many of the practical problems raised by such a model relate to the estimation of costs used in the model. Previous estimations of transportation costs have not usually been in a form suitable for use in design models. Much work remains to be done on the estimation of capital costs and costs relating to the operation of the transportation system. It is also important that such costs be developed so as to allow the consideration of real transportation alternatives. Transportation should be treated as a system with technological alternatives and not as a commodity to serve an aggregate travel demand.

The use of linear programming has some limitations as a framework for a transportation design model in that some of the constraints are discrete rather than linear in nature, and it is difficult to express these in a linear programming algorithm. Integer programming models have not proved practical for transportation networks of any size. Linearity also presents problems in the statement of cost relationships, but these linear limitations are probably still small compared to the errors in the cost parameters themselves. At the present state of the art of design model development, much may still be gained through the use of linear programming with all its limitations.

The principal suggestion that this commentary would make for the improvement of the subject model would be a model modification that would allow for joint consideration of the existing as well as the proposed two-mode transportation system. The revised model would consider a transportation system using a primal linear programming model to represent the loading of the present network and would study the benefits and cost of alternatives to this basic network through parametric analysis of the model of the existing system. Such an application would allow for long-run changes in the light of the optimal short-run use of the existing system. In this way, a better relationship between the alternatives of improving the existing system versus the construction of new facilities may be weighed. Such a model may indicate that funds could be better spent for the development of command and control systems to improve the efficiency of the existing system rather than the construction of new facilities.

An interesting result of the model application is the correspondence between the model output and the existing modal split between highway and transit in the city of Chicago. Such a correspondence indicates that the transportation market is performing admirably well, and it makes one wonder if the market is working so effectively, whether at our present level of understanding of the activities we are modeling that we should not leave the market alone.

DANIEL BRAND, Senior Project Engineer, Traffic Research Corporation—This paper proposes a linear programming solution to an important transportation planning problem. The problem is that of providing a minimum cost combination of two modes of transportation service from a corridor to a downtown area. The solution includes demand for the transportation service as well as the cost of supplying the service. Hence, the method gives a solution which is both optimal from the standpoint of supplying and which is capable of being achieved in practice, i. e., being utilized to the extent planned for.

The major points where the paper needs discussion are (a) the lack of mutual independence of several variables in the cost function, (b) the inability to calculate properly vehicle operating costs in the cost function, and (c) the assumption of fixed total demand for transportation service.

Critique

Interdependence of Terms in the Linear Road Capital Cost Model. —In formulating the first three terms of the objective function (the linear road capital cost model for a road spanning zone i) the assumption is made that money may be spent to add peak period road capacity, peak period speed, and additional off-peak period speed, independently of each other. This is contrary to the fact that design measures to increase peak period capacity (additional lanes, grade separations, etc.) are highly correlated with measures to increase peak period speeds, as the traditional speed-capacity curves would indicate. The same independence is also largely true for increasing off-peak speeds. Examples are given in the paper only for design measures to increase off-peak period speeds independently of the other two variables. Of the examples given, the longer acceleration and deceleration lanes normally increase ramp capacities as well as off-peak speeds by reducing relative speeds of merging vehicles and increasing gap acceptances. Another measure given, more adequate signals, is perhaps the only independent design measure, since signals are not fixed in their effect on different traffic flow patterns. They can be made to vary in their response to traffic at various times of day. Thus, additional money may be spent to add off-peak progressive timing or detailed traffic responsive control to increase speeds of off-peak period traffic. However, this additional money will be small compared to the cost of building new physical facilities.

Contrasted to this, an assumption that additional money may be spent for lower off-peak transit travel times (lower waiting times) may be appropriate, since the costs of running additional trains to shorten headways are the primary moneys involved. Studies of transit operating costs in the Boston area show these additional costs to be quite important.

An inability to provide an optimal mix of capacity and speeds eliminates the ability to calculate optimal speeds, and hence to predict transit trips, remaining auto trips, and vehicle miles of auto travel. This is a blow to the model as presently formulated.

Calculation of Vehicle Operating Costs. —In the calculation of vehicle operating costs, the assumption is made that one-half of the drivers using transit would get rid of the car which would otherwise be used for the trip and, therefore, one-half of the auto users should be charged the marginal operating costs of driving while the other half should be charged with the full costs.

The model uses the same fraction of one-half in two calculations, even though the fraction is computed with different bases, i.e., transit riders in the first instance, and auto users in the second. In addition, there is no provision to use the proportion of trips using transit, predicted by the model, to calculate the fractions of auto trips to charge full and marginal costs to.

Other difficulties in calculating vehicle operating costs are that these should be calculated using average interzonal car occupancy rates, which rates may vary from 1.1 to 2.0 or more, depending on the origin and destination of the trip, the trip purpose, the time of day, etc. Also, the assumption that one-half the drivers using transit would sell their car is a very difficult assumption to make. This number would vary with the location of the trip origin because of varying compositions of transit trip purposes and income of trip makers at the different origins.

Assumption of Fixed Total Demand for Transport Service. —The authors state: ". . . in other models of this type (the type treated in the paper), it is often found that the travel time which is generated by the facilities which are planned is different from that which was specified in order to predict demand. It is clear that this problem is avoided in our model. The total travel time and therefore the demand can be specified with certainty and entered as a constraint."

The contention that their model avoids the stated problem may be contended. Only the range in which travel times are generated by the model is limited. Demand is fixed but peak period trunk line travel time is allowed to vary on individual links (in their example) from 30 mph to 60 mph. (In the example, the model does in fact additionally restrict travel speeds over and above the 30 mph to 60 mph range by limiting overall travel time to a certain maximum from the six zones farthest out; about 30 percent of the trips are made from these six zones.) This is not an abnormal speed range to find in traffic models (gravity models with capacity restrained assignments) which vary travel demand by iterating over demand prediction and travel time calculation. It would appear, therefore, that total demand cannot be specified beforehand in this model with much more accuracy than in other models. Hence, it is not clear that the problem has been avoided.

To carry the discussion one step further, a reduction in the allowable range of speeds would enable the fixing of total demand in this model with more certainty. However, this would narrow the range of alternatives which could be tested by the model. It also may be possible that the model application yielded "reasonable" modal splits, because within the speed ranges given, the modal splits are reasonable. Hence, the setting of allowable speed ranges has important meanings for the model as presently formulated.

Model Application

Solutions Tending Toward Boundary Values of Variables. —Are the authors disturbed that many of the variables solved for, yielded values on the boundaries of the region of possible values of the variables? For example, the optimal results for peak and off-peak speeds (m_i and s_i) are either 30 or 60 mph for 29 out of the 30 solutions. In particular, the question may be asked, does the propensity of linear programming methods to yield values on the boundaries of possible solution space affect the ability of this model to yield reasonable results?

Consistency of Results for Modal Splits and Speeds. —The similarity of the peak and daily fraction of trips via transit yielded by the model (Table 6) does not appear consistent with the solution for peak and off-peak speeds (Table 7). The latter vary between the two periods of the day. It is the varying speeds which are used in the determination of the similar peak and daily fractions of trips via transit.

A Possible Extension of the Model

The application of the model in the paper yields optimal values of 30 mph for both peak and off-peak speeds in downtown and neighboring zones. The authors' comment: "This is in contrast to the planning in many areas where full freeways are called for even in the heart of downtown areas."

It must be noted that only trips to the single downtown destination zone are being considered. Through trips and trips to intermediate destinations are not being considered.

An extension of the model to be origin-destination specific rather than origin specific is needed if real planning problems are to be solved. This would complicate certain aspects of input data preparation, in particular the modal choice constraints and their associated parameter values. Also the notion of how to interpret capacity between many origin-destination pairs is of interest.

A discussion by the authors of whether such an origin-destination formulation of their model could be solved would be useful.

Conclusion

Despite the criticisms in this discussion, I feel this is a very important paper. The present linear programming solution may fall short of being meaningful to the problem-oriented planner; however, with additional work and reformulation, linear programming may be capable of providing efficient low-cost solutions to meaningful transportation planning problems.

EDWARD K. MORLOK, Jr., and GEORGE A. HAY, Closure—The discussions can be divided into two broad categories: (a) comments about applications of mathematical programming in general and (b) comments about the specific application in our paper. We shall concentrate on those comments specifically about our paper, since we are in agreement with virtually all of the comments in the other category. However, we would like to add two general but relevant observations about modeling and decision-making to the general comments of the discussants.

There seems to have been considerable misunderstanding of the road capital cost function. Mr. Brand makes the comment "the assumption is made that money may be spent to add peak period road capacity, peak period speed, and additional off-peak period speed, independently of each other." We have not made this assumption nor even intended it, at least with respect to peak period values. We fully recognize the interdependence of these variables. In fact, we are considering a single class of improvements, whose benefit may be taken in additional speed, increased capacity, or some combination of the two. The interdependence, and therefore, the combinations which are achievable are defined by the speed capacity tradeoff diagram (Fig. 2). Additional expenditure need not be directed specifically toward speed or specifically toward capacity, but can be thought of as yielding an outward shift in the whole speed capacity frontier. This frontier is defined by the functional relationship $f(m, c) = k$ where k is the expenditure, and it is this frontier which is represented in Figure 2. In deriving our capital cost function $\left[C_i c_i + M_i \cdot (1 - \frac{m_i}{\overline{M}_i})\right]$ we have simply given a specific form to this functional relationship, a linear one. There may be objections to this form of the relationship, but they are not those to which Brand has referred.

Among the possible objections are the following: (a) we have approximated a set of non-linear curves with a set of linear ones; (b) we have assumed that the relationship $f(m, c) = k$ is homogeneous of degree one. This, together with the linearity, implies not only constant returns to scale but also that the marginal costs of increasing one variable (e.g., speed), is independent of the level of the other variable, capacity. The first implication is probably acceptable within the range of acceptable alternatives. The second should be accepted or rejected on technological rather than theoretical grounds.

Both Mr. Brand and Mr. Sullivan mentioned that the additional cost of increasing off-peak period speed over that for the peak period probably would be small in comparison to the cost of building a new facilities. We were unable to find any definitive evidence on this. We decided to include the third term in the cost function, because we felt that this cost could be significant in some situations.

Turning to another aspect of road costs, Sullivan states that the only investment variable relates highway capital cost to highway speed and that this formulation seems oblique. Highway costs are related to capacity, peak period speed, and non-peak period speed (where capacity is defined as that volume at which the specified speed is achieved). We related cost to measures of output capability (capacity and speed) rather than to the physical road itself and then to output capability because the former is more efficient. The physical road its speed-volume characteristics are referred to in developing the cost function, but once this is developed there is no reason to return to the road itself. Mr. Sullivan also states that only one set of highway capital costs are used and suggests that in future applications these costs might be a function of development density. Actually this was done in the application given in the paper. The road capital cost coefficients used decrease with increasing distance from the CBD (Table 1).

In addition, Sullivan says that the other investment variable, transit capital cost, is actually introduced as a constant. This, of course, is not true. While introduced as a constant in the first run of the sample program (the only run presented in the paper) it must be remembered that this run is one of several which must be performed (in each run the transit system is extended one zone further with transit capital costs increasing correspondingly) according to the minimum minimorium technique outlined in the paper. The run which yields the lowest total cost of all those considered will be the true optimum solution.

Brand's points concerning the difficulty of accepting our assumptions about average automobile occupancy rates and the fraction of drivers who would sell their autos if they used transit are well taken. The reason for our assumption was the absence of more detailed data for the region we considered.

It is important to note that both of these parameters can depend on the residential zone of the travelers in question simply by approximately subscripting the relevant parameters. In the case of auto ocupancy, the capacity constraints become

$$c_j \geqslant \sum_{i=j}^{n} \frac{D_i}{E_i} - \sum_{i=j}^{n} \frac{P_i}{E_i} \qquad j = 1, \ldots, k$$

$$c_j \geqslant \sum_{i=j}^{n} \frac{D_i}{E_i} \qquad j = k+1, \ldots, n$$

Since each vehicle-mile constraint calculates the total daily vehicle-miles generated by trips from a single zone, the occupancy rate E and the cost parameter V only need be subscripted in the present formulation to take care of zonal differences.

Also, a further distinction between peak and non-peak periods can be made very simply. Since the capacity constraints refer to only the peak periods, they present no problem. In order to distinguish between peak and non-peak values of E_i, V_i, and v_i, we might add a second subscript, p for peak period and n for non-peak period. The constraints which determine vehicle-miles of travel then become

$$v_{jp} = \frac{2}{E_{jp}}\left(\sum_{i=1}^{j} L_i + F_j + G\right) \cdot (2D_j - 2p_j) \qquad j \leq k$$

$$v_{jp} = \frac{2}{E_{jp}}\left(\sum_{k=1}^{k} L_i + G\right) \cdot (2D_j - 2p_j)$$

$$+ \frac{2}{E_{jp}}\left(\sum_{i=k+1}^{j} L_i + F_j\right) \cdot (2D_j) \qquad j > k$$

$$v_{jn} = \frac{2}{E_{jn}}\left(\sum_{i=1}^{j} L_i + F_j + G\right) \cdot (8D_j - d_j + 2p_j) \qquad j \leq k$$

$$v_{jn} = \frac{2}{E_{jn}}\left(\sum_{i=1}^{j} L_i + G\right) \cdot (8D_j - d_j + 2p_j)$$

$$+ \frac{2}{E_{jn}}\left(\sum_{i=k+1}^{j} L_i + F_j\right) \cdot (8D_j) \qquad j > k$$

Brand's conclusion that in this model one cannot specify interzonal travel times at the outset with much certainty of achievement appears to be based on a limited examination of the example application rather than the general form of the model. While in the example we used only three peak and three non-peak period time constraints, the model as given contains two for each zone—one for the peak period and one for the non-peak period. These constraints are upper bounds on travel time.

Because extra speed costs money, minimum cost solutions will generally call for travel times very close to or equal to the maximum allowed. This was verified by experimentation with the model, and is exhibited in the example given. Thus we feel justified in our statement that interzonal travel time expectations will be met (or nearly so), so that interzonal demands can reasonably be taken as fixed.

The reason for no travel time constraints for zones 1 through 9 in the example was that we felt that 30 mph average main road speeds were adequate for travel to points up to about 18 miles from the downtown area. Since this speed is already embodied in other constraints (zonal speeds), there was no reason to add redundant travel time constraints. It was felt that the constraints for zones 10, 13 and 15 sufficiently narrowed the range of travel time choices for zones 11, 12 and 14 that no explicit constraints for these were included. This suspicion was confirmed by the outcome.

If any difficulties with traveltimes were to arise, it would be possible to add a second set of constraints. A second constraint for each one would place a lower limit on travel time. Thus the travel time from any zone could be constrained to a range as small as desirable.

Brand's doubts concerning the consistency of our modal split results are easily cleared up. The set of equations which yields the number of passengers who take transit daily is as follows:

$$P_j = A_j \sum_{i=1}^{h} m_i + B_j \qquad h = j \text{ if } j \leq k \text{ and } h = k \text{ if } j > k$$

Note that when $j > k$, the zone in which transit ends, the modal split depends only on the speeds in the first k zones. In our example $k = 5$ and the peak and off-peak speeds are determined to be the same for those zones. The peak vs off-peak discrepancies in zones 6-8 do not, therefore, affect the modal split.

Both Mr. Brand and Mr. Sullivan emphasize the importance of extending the model so as to include consideration of trips which neither originate nor terminate in the CBD. We could not agree more fully.

The consideration of trips made solely along the axis of the corridor would not be too difficult: the capacity constraints must be changed so that the combined road and transit capacity in any zone is at least as great as the total flow through that zone. The equations for calculating vehicle operating costs would become much more complex, but these present no problem from the programming point of view. In principle, an equation for modal choice should be included for each origin-destination zone pair which is served (at least for part of the trip) by transit. This is possible, but would tend to make the program unwieldy, and we would suggest consideration of the assumption that trips to some zones served by transit would not be made by transit. This assumption could be defended for zones in which only a small fraction of the zone's total business activity occurs near the transit stations.

As to the extension of the model to consideration of trips with one or more ends outside of the corridor, we feel that this would be much more difficult than the previous extension. While we are certain that the extension to inclusion of all trips solely within the corridor could be made, success in making this further extension without some major (and possibly unacceptable) assumptions is not certain. One such possibility is to fix the point at which trips enter and leave the corridor. This reduces the problem to one very similar to the extension covered in the preceding paragraph.

Sullivan brings up the additional point that there is no provision for changes in the magnitude or orientation of trips with changes in capital investment. To the extent that changes in capital investment correspond to changes in travel time, cost, etc., these changes will cause some changes in trip volumes and orientation. However, in the model we constrain road travel times to a narrow range, and, of course, transit running times are fixed. Pricing is also assumed fixed for each run of the model. Therefore, we feel justified in the assumption of a fixed total demand. Major changes in travel times and pricing are accommodated only with additional runs of the model, in which the total demand and modal choice parameters have been revised to reflect these changes.

In a broader sense, however, we must agree with Sullivan's comment. Over a long period of time the nature of the transportation facilities and services provided in a region undoubtedly strongly influences the pattern of development of the region. An example within the context of our model might be: the provision of rapid transit in the corridor could attract a concentration of dwelling units and business establishments along its route, which would be more widely dispersed throughout the entire region if the rapid transit line were not constructed. Presumably this increase in development in the corridor would result in more travel within the corridor. Thus, if one takes into account the differences in developmental consequences of alternative transportation services, then there certainly is an effect of alternative transport choices upon travel patterns and land use. The question is: How strong are these influences?

The question posed has not been answered, to our knowledge. The urban studies which we have seen do not appear to have actually taken the developmental influences of alternative systems into account in their models. Much more research directed at identifying and quantifying the appropriate relationships is necessary before we would have any justification for inclusion of such relationships in our model. We do earnestly hope that this research will be carried out.

Closely related to Sullivan's remarks, but of a more general nature, is Mr. Schlager's comment that the model should take greater account of the interaction of transportation and land use. To the extent that this interaction is reflected in traveler movements in the corridor, our discussion in the preceding three paragraphs is relevant. Schlager undoubtedly is also referring to environmental constraints on such items as the location of new facilities and the extent to which additional land can be taken for improvements to existing travel arteries.

As to routing, the model in its present form presumes that the routes of new roads and transit lines are specified outside of the model. It assumes that the cost and other coefficients are applicable to routes which are feasible, both from the economic and social standpoint. There is, however, no provision to limit the land area occupied by the new or improved facilities. Similarly there is no means for restricting other design features, such as elevation, which might affect the environment. The inclusion of these types of restrictions might be quite difficult given the present form of the model, although the subject must be investigated in detail before a statement as to the feasibility of adding such restraints could be made.

Sullivan also discusses the problem of fiscal feasibility, which is not explicitly handled by our model. His suggestion that demand and travel time constraints be varied so that estimates can be made of the additional cost of accommodating more travelers or increasing speeds has great merit. In this way one could compare the benefits and cost of improvements to the transportation system.

An alternative, which we do not consider as useful as that described above, is to include a budget constraint in the model. This would limit the capital expenditure to a predetermined amount.

In his discussion of the assumption of demands fixed in magnitude and orientation, Sullivan suggests that if strong assumptions are made in the mathematical programming formulation the solution should be checked by the more complex network and demand simulation models. We doubt that any mathematical programming models could ever rival computer simulation models in their ability to accommodate all the details of a phenomenon. However, they do have the distinct advantage of efficiently finding the optimum of a very wide range of alternatives, while the searching for optimum solutions with computer simulation models usually is extremely expensive. Therefore, we feel that these two types of models can be complementary, with the optimization models being used to narrow the choice to a few distinct alternatives. The simulation models then would be used to explore these alternatives in more detail.

Schlager's suggestion that the model be revised to consider the loading of the present system as the primal programming problem is very interesting. The constraints in this formulation presumably would reflect the characteristics of the existing network. The dual variables then would indicate the value of various marginal changes in the existing network, as these would be reflected as changes in the constraints.

The results of such a program would be very different from the results of the present program. The revised program would yield the most beneficial marginal improvements, whereas the present program yields specifications for a system which is optimal at some future date. Of course, the question as to whether the problem could be formulated in the suggested manner still remains. However, we feel the suggestion has considerable merit and intend to examine the feasibility of a revised formulation before developing our model further.

There are two additional ideas relevant to models and decision-making of the sort considered in this paper which we feel are important but which have not been mentioned.

The first is that a model of some real world phenomenon is necessarily a simplification of that phenomenon. Those who evaluate a model must decide—on partly subjective and partly objective grounds—whether the model includes all of the important or relevant relationships and factors and ignores all others. It is not clear to us, for example, that the most fruitful direction for further development of our model is toward considering an entire region, or toward considering the staging of improvements, or toward considering the developmental consequences of alternative transport decisions.

Moreover, if all of these were included, the model might become so complex and costly to run that it would be of little value to the problem-oriented planner. The complexity might defy comprehension, so that understanding the various solutions is difficult. Costliness would tend to limit the number of alternatives considered, defeating the purpose of the model. Under these conditions, transport decision-making would not be improved by the extensions. We do not claim that these conditions would result from major extensions, just that they could.

The other major idea we wish to transmit is not our own and has been stated often (especially in the writings of William Garrison and Tillo Kuhn), but seems to be heeded rarely. If one is dealing with a transport decision of such magnitude that it will influence travel patterns and the pattern of development of a region, simple economic criteria related to transport phenomena are wholly inadequate. At the least, the criteria used should reflect the broad spectrum of society's benefits and costs (both monetary and non-monetary) resulting from alternative decisions.

We feel that it will be extremely difficult to quantify and transform into the same units (such as dollars) this spectrum of benefits and costs. This is especially difficult in situations where the benefits and costs are non-uniformly distributed over the population and the region. If this cannot be done, it will not be possible to utilize the optimum-seeking capabilities of mathematical programming, for the optimum is not defined. It may not be that broad choices as to transport development are inherently social choices, best left to the citizens and the political process (17). For these broad questions, the value of programming formulations is probably in identifying alternatives—not specifying solutions.

Reference

Haefele, Edwin T. Transport Planning: A Process In Search of a Policy. Symposium on Transportation Decision Making—Its Environment and Methodology, Institute of Management Sciences College on Logistics, Washington, D. C., March 31, 1966.