

# Mathematical Programming Formulations of Transportation and Land Use Models: Practical Implications of Recent Research

STEPHEN H. PUTMAN

The next generation of transportation, location, and land use models will most probably emerge from mathematical programming formulations. Presented are simple numerical examples of trip assignment and population location, both described as optimization problems, in mathematical programming formulations. A trip assignment model with constant link costs is described first, and then the same model is modified to show the consequences of a flow-dependent link cost formulation. In similar fashion, a linear model of least cost population location is transformed into a nonlinear model that incorporates dispersion of location due to differences in locators' preferences or perceptions. It is then shown how the trip assignment model and the location model can be combined into a single nonlinear programming formulation that solves both problems simultaneously. In the final section of the paper, the theoretical advantages and practical disadvantages of this approach are briefly enumerated. This is followed by suggestions about the likely resolution of practical problems to allow use of these techniques in applied planning situations.

There has been considerable refinement of practical methods of forecasting urban location and transportation patterns during the past 10 to 15 years. Although there is continuing discussion and development, and even the best of forecasts are far from perfect, there appears to be a greater consensus on what methods are clearly outmoded and in what directions future efforts should move. This author's views on general progress in the field have already been published (1). Among the most sophisticated practical methods of transportation and land use forecasting are the extended spatial interaction models, especially when they are included in comprehensive integrated model systems [see paper by Bly and Webster in this Record and Putman (2)].

In addition to these practical developments there have also been important theoretical developments. On the transportation side these include the development of discrete choice models, especially for travel demand and mode choice (3), and the development of mathematical programming formulations of the traffic assignment problem (4). On the location side the development of utility theory as a basis for location models (5) and the general discussion of mathematical programming models as alternate or underlying structures for spatial interaction models (6) were major developments. Some of these developments are important principally because they provide an improved underpinning of existing practical methods; some have

shown the existence of clear errors in prior practice; and others may offer substantial improvements for future applications.

Past experience suggests that there is a lag of 10 years, sometimes more, between the initial development and subsequent practical application of new techniques in transportation and land use forecasting. Thus, although there have been some attempted applications of these methods (7, 8), they are far from being the accepted norm. The purpose of this paper is to present some illustrations of the mathematical programming formulations along with some simple numerical examples. The intent is to show some of the benefits, both practical and theoretical, of these formulations and to provide the practical planner with an introduction to this promising new area.

## NETWORKS AND TRIP ASSIGNMENT

In the discussion that follows, extensive use will be made of the data describing Archerville, a simple five-zone numerical example. Table 1 give the land use and socioeconomic data for Archerville. Figure 1 shows the Archerville highway system.

### Shortest Path Problem

A frequently encountered problem in transportation and location analyses is that of finding the shortest path from one node to another over the links of a network. This is a problem that can be considered as a linear programming problem. The equation form is

$$\text{Min: } Z = \sum_i \sum_j c_{ij} X_{ij} \quad (1)$$

subject to

$$\sum_j X_{ki} - \sum_k X_{ki} = \begin{cases} 1 & \text{if } i = \text{origin} \\ -1 & \text{if } i = \text{destination} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$X_{ij} \geq 0 \quad (\forall i, j) \quad (3)$$

where  $c_{ij}$  is cost of traversing link  $i, j$  and  $X_{ij}$  is flow (trips) on link  $i, j$ .

The objective function is straightforward: simply to minimize the sum of the link cost times the trips incurring that cost. The principal constraint equations (Equations 2) are a set of flow-balance relationships that ensure that the flows, at each

TABLE 1 ARCHERVILLE—LAND USE AND SOCIOECONOMIC DATA

Zone	Land Use Data					Socioeconomic Data					
	Residential	Commercial	Industrial	Vacant	Total	Commercial Employment	Industrial Employment	Total Employment	LI Households	HI Households	Total Population
1	2.5	1.0	1.0	0.5	5.0	150	150	300	200	100	860
2	3.0	1.0	1.5	0.7	6.2	200	150	350	300	50	1,050
3	1.0	2.0	3.0	0.8	6.8	100	400	500	150	50	585
4	2.5	1.0	0.0	4.5	8.0	200	0	200	100	300	550
5	1.5	0.5	0.0	6.6	8.6	100	0	100	50	150	515
Total						750	700	1,450	800	650	3,560

Employee-Household Cross Tabulation

	LI	HI	Total
Commercial	400	350	750
Industrial	400	300	700
	800	650	1,450

Employee-Household Conversion Matrix

	LI	HI	Total
Commercial	.533	.467	1.00
Industrial	.571	.429	1.00

NOTE: LI = low-income and HI = high-income.

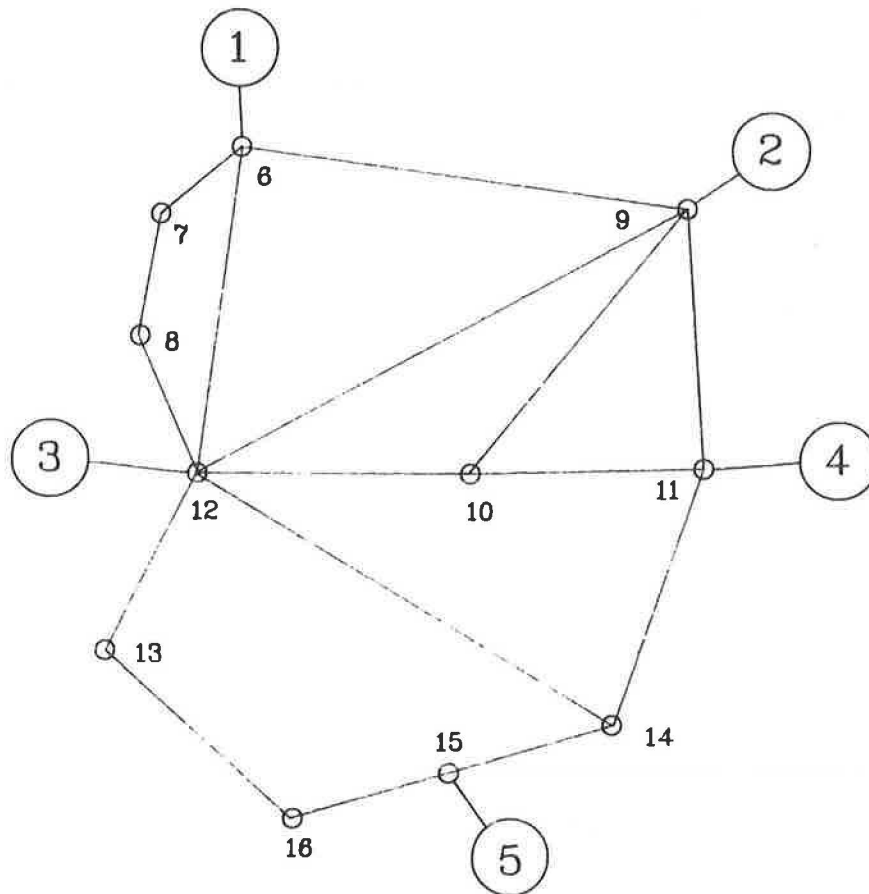


FIGURE 1 Archerville highway system.

network node, balance. Thus for each node the total flow of trips into the node minus the total flow of trips out of the node must equal the net trips supplied (or demanded) at the node. The last constraint equation (Equation 3) is simply the non-negativity requirements that prohibit negative flows. It can readily be seen that writing the shortest path problem in this form yields a rather good sized problem. In the objective function the number of terms equals the number of links in the network. There must be a constraint equation (Equation 2) for each network node and a constraint equation (Equation 3) for each network link.

For practical applications there are many fast algorithms for solving this problem. However, it is worth noting here that, because this is just a simple linear program, it can be solved by the well-known simplex method. If it were done that way, it would be necessary to solve the problem once for each origin-destination path that was wanted. Further note that this problem only addresses the situation for fixed link costs and flow volumes, which must, of course, be known in advance of any attempt to solve for the minimum path or paths.

### Minimum-Cost Flow Problem

Another type of linear programming problem is that known as the minimum-cost flow problem. The Archerville data may be used as an example. Assume that there are a known number of households of each income group in each zone and a known number of employees of each type working in each zone. Implicitly there is a zone-to-zone matrix of home-to-work trips that can be estimated by standard techniques. Suppose now that the location of households in Table 1 and the location of industrial employment in the same table are taken as given. By first assuming that there will be one employee per household and then applying the Employee-Household Conversion Matrix given in Table 1, the number of industrial employees residing in each zone may be calculated. These were 146, 173, 98, 188, and 95, respectively. Note that the number of industrial employees working in each zone is 150, 150, 400, 0, and 0, respectively.

It is possible to consider the proposal that each employee is to choose a place of work such that the total travel cost for all employees is minimized. If network link capacities, and the consequent congestion, are ignored for this illustration, this problem may be stated as a linear program

$$\text{Min: } Z = \sum_i \sum_j c_{ij} X_{ij} \quad (4)$$

subject to

$$\sum_j X_{ij} - \sum_k X_{ki} = \begin{cases} O_i & \text{if } i = \text{origin} \\ -D_i & \text{if } i = \text{destination} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$$X_{ij} \geq 0 \quad (\forall i, j) \quad (6)$$

where  $O_i$  is net trips leaving node  $i$  and  $D_i$  is net trips arriving at node  $i$ .

Equations 4, 5, and 6 are the general minimum-cost flow

problem. If there were specified link capacities ( $V_{ij}$ ) for each link, which could not be exceeded, then another set of constraints would be substituted for Equation 6. This new set of constraints would be of the form

$$V_{ij} \geq X_{ij} \geq 0 \quad (\forall i, j) \quad (7)$$

Note that the objective function here is the same as that for the shortest path problem given in Equation 1. The constraints, Equations 5, are a set of flow-balance relationships similar to those of Equations 2.

Because the intrazonal travel costs are 1.0 for each zone, clearly the first consideration is that all employees residing in any zone first be assigned to jobs in that zone. Given the Archerville data, Zone 1 requires 4 workers and Zone 3 requires 302. Zones 2, 4, and 5 have 23, 188, and 95 surplus workers, respectively. The link flows produced by the simplex algorithm to solve this problem are shown in Figure 2. Note that if it were desired that some of the network links have a maximum allowable flow, then some constraints of the form of Equation 7 would have to be added.

### Nonlinear Minimum-Cost Flow Problem

The minimum-cost flow problem described here can be reconsidered as a nonlinear programming formulation. In the previous formulation the linear objective function, Equation 4, was simply the sum of the trips (flows) on each network link times the travel cost of the link. The link costs were fixed, remaining constant regardless of link flows (though it was shown how the link flows could themselves be restricted by use of additional constraint equations). Suppose the more realistic view was taken that link costs depend on link flows. For the sake of illustration consider the following function, where link cost varies with link flow

$$C_{ij} = C_{ij}^0 (1.0 + \delta X_{ij}^2) \quad (8)$$

where

$$\begin{aligned} C_{ij} &= \text{"congested" or flow-related link travel cost,} \\ C_{ij}^0 &= \text{free-flow link travel cost,} \\ X_{ij} &= \text{link flow volume (trips), and} \\ \delta &= \text{a parameter.} \end{aligned}$$

With this function the link travel cost is equal to the free-flow cost when the link flow volume is zero. As link flow volume increases, link travel cost increases too.

In the linear version of this problem the solution involved only the finding of the minimum paths and the subsequent routing of trips along those paths. If there were specific link flow volume constraints, then the excess trips would be re-routed to the second shortest path. When link flows determine link costs the essential nature of the problem changes. The solution becomes a matter of adjusting volumes, observing the resulting costs, and then adjusting the volumes again. Thus, in a very real sense, even for the small problem size of the Archerville data the complexity of the problem begins to defy

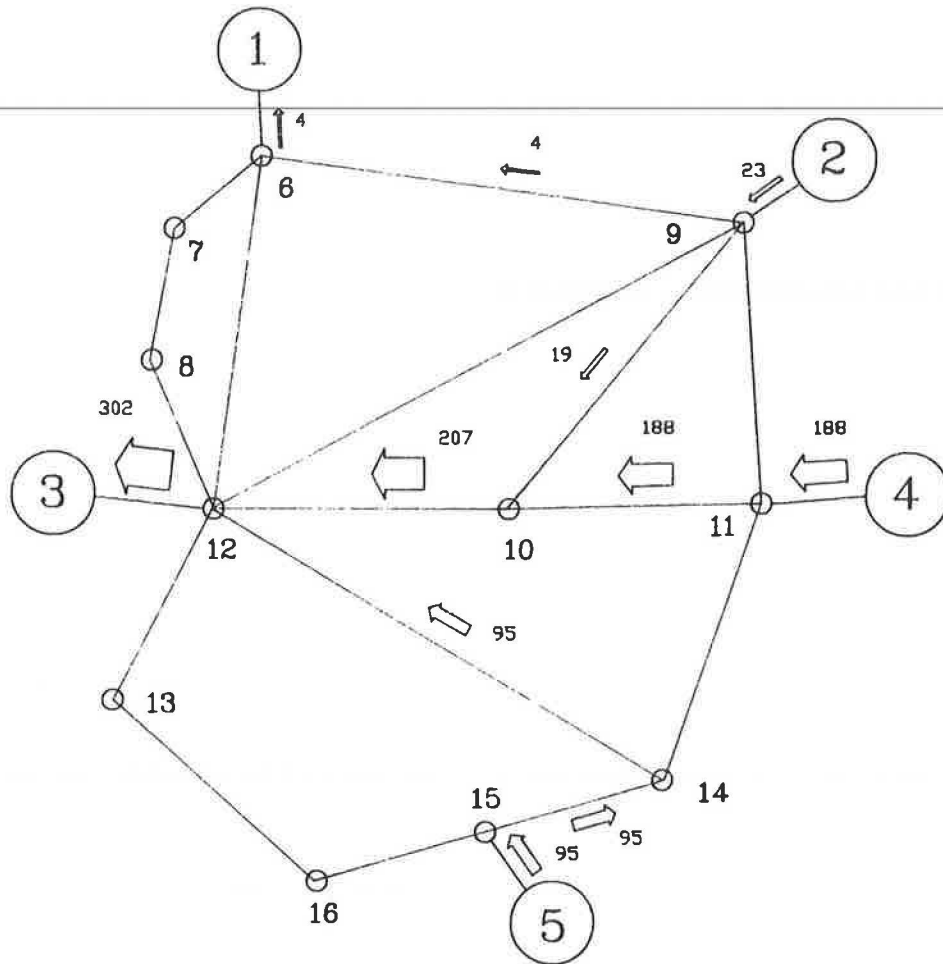


FIGURE 2 Archerville trip flows resulting from minimum-cost flow assignment algorithm with constant link costs.

solution by inspection. The introduction of simultaneity or nonlinearity, or both, to a problem often transforms the problem into one that lies beyond intuitive solution.

Incorporating Equation 8 into the original objective function of Equation 4 yields

$$\text{Min: } Z = \sum_i \sum_j [C_{ij}^0 (1.0 + \delta X_{ij}^2) X_{ij}] \quad (9)$$

and thus the objective function becomes a cubic equation. The same linear constraints as before (Equations 5, the flow-balance relationships) still hold true, as do the nonnegativity constraints of Equation 6. This new set of equations is a nonlinear programming (NLP) problem with a nonlinear objective function and linear constraints.

It was necessary to set a value for the parameter  $\delta$ . A value of 0.0002 was selected so that link flow volumes on the order of 100 trips would result in a tripling of link cost. At this scale, link costs increase significantly, but not astronomically, with link flows on the order of those observed in the linear form of the problem described previously. The flows on the network that result from this new nonlinear problem are shown in Figure 3, and the link volumes, free-flow link costs, and congested link costs are given in Table 2 (only those links that have flows are included). The results hold no great surprises but do

show a clear response to the reformulation of the problem so that link costs are a function of link flows.

It is interesting to compare the results shown in Figure 3 from the NLP solution with those in Figure 2 from the linear programming (LP) solution. The NLP solution, due to the effects of congestion on link travel costs, shows much greater utilization of network links. For the LP solution only 11 links were used whereas for the NLP solution 20 were used. As a result, the trips on links  $X_{11,10}$  and  $X_{10,12}$ , which were 188 in the LP solution, are only 112 in the NLP solution.

These examples only hint at the substantial additional work that has been done with user equilibrium and stochastic user equilibrium formulations of the traffic assignment problem as a mathematical program. Yet, they do give a clear way of seeing the assignment problem, as well as networks in general, expressed in equation form. This will be particularly helpful in analyzing ways of linking transportation and location models. This insight also provides a much easier way of comprehending the problems of traffic assignment than did the "black-box" approach of traditional all-or-nothing assignment procedures. In the next section of this paper simple examples will be presented of location models presented as mathematical programs.

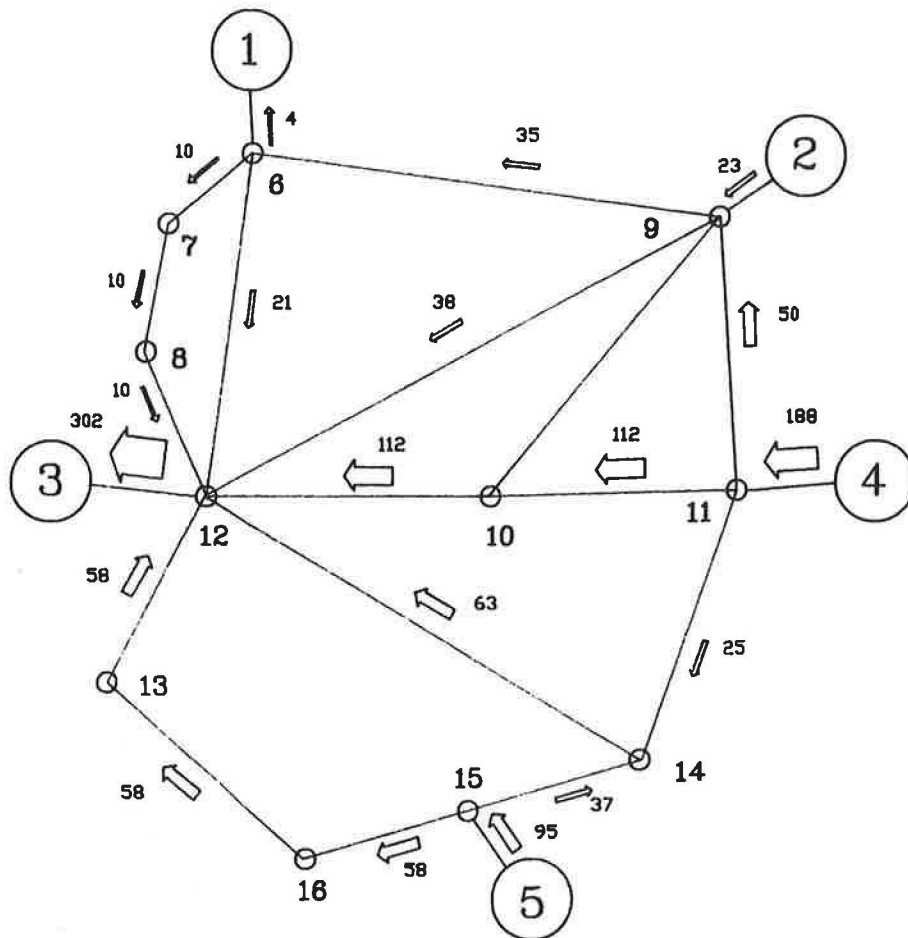


FIGURE 3 Archerville trip flows resulting from minimum-cost flow assignment algorithm with variable link costs.

**TECHNIQUE FOR THE OPTIMAL PLACEMENT OF ACTIVITY IN ZONES: TOPAZ**

TOPAZ is a mathematical programming technique that was originally proposed in the late 1960s and early 1970s (9, 10). The most complete discussion of the applications of the model is to be found in the book by Brotchie et al. (11). The model was originally proposed as a method for determining least cost allocations of activities to zones. Perhaps the most recently published work on TOPAZ is by Sharpe et al. (12) from which the formulation used here is adapted.

To begin, the model was abbreviated to a form for residence location only. Further, it was assumed, as is customary in these examples, that there is one employee per household and thus one work trip per household. The Archerville data show 0.55 low-income households per 1.0 employee and 0.45 high-income households per 1.0 employee. Thus the new, simplified, problem formulation becomes

$$\text{Min: } Z = \sum_i \sum_j \sum_l T_{ijl} c_{ijl} + \sum_i \sum_j b_{ij} X_{ij} \quad (10)$$

subject to

$$\sum_l T_{ijl} - s_i X_{ij} = 0 \quad (11)$$

$$\sum_j T_{ijl} - r_i X_l = 0 \quad (12)$$

$$\sum_j X_{ij} = A_i \quad (13)$$

$$\sum_i X_{ij} \leq Z_j \quad (14)$$

$$T_{ijl} \geq 0, X_{ij} \geq 0 \quad (15, 16)$$

where

- $A_i$  = regional total of activity  $i$ ,
- $b_{ij}$  = unit cost less benefit of locating activity  $i$  in zone  $j$ ,
- $c_{ijl}$  = unit cost less benefit of interaction for activity  $i$  between zone  $j$  and zone  $l$ ,
- $s_i$  = level of interaction (trips generated) per unit of activity  $i$ ,
- $r_i$  = trips attracted by employment per unit of activity  $i$ ,
- $T_{ijl}$  = level of interaction (trips) of activity  $i$  between zone  $j$  and zone  $l$ ,
- $X_{ij}$  = amount of activity  $i$  to be allocated to zone  $j$ , and
- $Z_j$  = capacity of zone  $j$ .

Note that the second term in the objective function is simply a minimum-cost location term. The first term is the linear

TABLE 2 ARCHERVILLE—COMPARISON OF FREE-FLOW AND CONGESTED LINK COSTS FOR NONLINEAR OBJECTIVE FUNCTION

Link	Volume	Free-Flow Cost	Congested Cost
$X_{2,9}$	23	1.00	1.11
$X_{4,11}$	188	1.00	8.07
$X_{5,15}$	95	1.00	2.81
$X_{6,1}$	4	1.00	1.00
$X_{6,7}$	10	2.00	2.04
$X_{6,12}$	21	5.00	5.44
$X_{7,8}$	10	2.00	2.04
$X_{8,12}$	10	2.00	2.04
$X_{9,6}$	35	4.00	4.96
$X_{9,12}$	38	7.00	9.07
$X_{10,12}$	112	2.00	7.06
$X_{11,9}$	50	5.00	7.52
$X_{11,10}$	112	1.00	3.53
$X_{11,14}$	25	4.00	4.52
$X_{12,3}$	302	1.00	19.24
$X_{13,12}$	58	2.00	3.33
$X_{14,12}$	63	6.00	10.74
$X_{15,14}$	37	2.00	2.56
$X_{15,16}$	58	2.00	3.33
$X_{16,13}$	58	4.00	6.65

“transportation” problem. Taking the location problem first, it is clear that developing the necessary “data” raises some difficult issues. The net benefits ( $b_{ij}$ ) are supposed to be the net of the costs and benefits of locating a unit of activity type  $i$  in zone  $j$ . In many TOPAZ applications the virtual impossibility of measuring benefits resulted in the  $b_{ij}$  being simply a cost of location, to be minimized. For the Archerville example several possibilities existed. The easiest way was simply to create an average annualized house cost variable, realizing that then the model would attempt to locate all households in the zone with the lowest house cost. All that prevents this location are the constraints on the amount of activity that can be accommodated in each zone.

Raising the issue of zonal constraints raises, in turn, the issue of converting activity types into land consumed. Here again, there were several possible ways to proceed: (a) regional land consumption rates by activity type, (b) zonal land consumption rates by activity type, or (c) exogenously developed housing stock estimates. For this illustration of the model a set of regional land consumption rates was assumed, and their values were set so as to allow all households to be accommodated by existing residential land in the region. The rates were 0.00525 land units per low-income (LI) household and 0.00646 land units per high-income (HI) household. The cost of location by zone was taken to be the average annualized house cost by zone, which was set to \$7,800, \$7,200, \$7,600, \$6,800, and \$9,668, respectively, for the five zones.

Next the transportation, or interaction, cost term in the objective function (Equation 10) was examined. This required the specification of data for the trip end constraints (Equations 11 and 12) as well. As mentioned previously, it was assumed that there was only one employee per household and that there were 0.55 LI and 0.45 HI households per employee. Recalling that, in this example, only home-to-work trips are being dealt with and there is only one employment type, then both  $s_i$  will equal one, so that the constraint equations (Equation 11) will be

$$\sum_l T_{ijl} - X_{ij} = 0$$

or, in words, the sum over all possible destination zones  $l$  of all trips of type  $i$  leaving (i.e., produced in) zone  $j$  must equal the trips of type  $i$  generated in zone  $j$ . In this example the trips of type  $i$  (i.e., household type  $i$ ) generated in zone  $j$  equal the number of households of type  $i$  living in zone  $j$ .

Following the same reasoning, the  $r_i$  will be equal to the household-type-per-employee ratios, so that the constraint equation (Equation 12) will be as given earlier, with  $r_1 = 0.55$ ,  $R_2 = 0.45$ , and  $X_l$  = total employment in zone  $l$ . Then, in words, the sum over all possible origin zones  $j$  of all trips of type  $i$  terminating in zone  $l$  must equal the total households of type  $i$  attracted to employment in zone  $l$ . The total households of type  $i$  attracted to employment in  $l$  is simply the conversion rate times the employment.

An extensive series of test runs was done with this model with varying weightings multiplied times one or another of the two components of the objective function. At the extremes, a location-cost-only solution and a transportation-cost-only solution were found. The minimum transport cost solution gave somewhat more dispersion of households to zones than did the minimum location cost solution. This was due in large part to the exogenously determined location of employment. Were employment to be less dispersed, then, subject to the land use constraints, residential location would be less dispersed as well.

Another matter that will have interesting consequences for the next set of tests, in which both location and interaction costs are used along with a dispersion term to determine residential location, was that even though the location cost component was almost four times larger than the transport cost component, the transport cost portion of the objective function completely dominated the model solution. In several more test runs the weighting of the location cost term in the objective function was varied from 0.01 to 2.00.

Over the range from 0.045 to 1.00 the location cost multiplier resulted in a more or less gradual shift from the transport-cost-only solution toward the location-cost-only solution. From a value of slightly less than 1.0 to a value of 1.1852 the multiplier causes no change in the model solution. At a value of 1.1855 the multiplier results in a model solution identical to that of the location-cost-only solution. Thus at some critical point where the multiplier of the location cost term in the objective function is between 1.1852 and 1.1855, there is a sudden shift in the model solution from an apparently stable intermediate solution to the location-cost-only solution. Although it is presented as an aside here, clearly the matter of model sensitivity and solution stability is an area for future research. In any case, for a midrange weighting of 0.70, the results of the TOPAZ model solution were as given in Table 3.

### Numerical Example Incorporating Dispersion

The location patterns produced by the linear programming version of this model, particularly when examined at the zone-to-zone trip level, are rather lumpy. The addition of a nonlinear dispersion term to the objective function can make a noticeable difference. This is achieved by substituting a constrained gravity model for the linear “transportation” model portion of the objective function.



TABLE 3 ARCHERVILLE—LOCATION OF HOUSEHOLDS BASED ON LOCATION COST PLUS TRANSPORT COST COMPONENTS OF MODEL ( $\lambda_2 = 0.70$ )

Zone	Households			Households, Showing Place of Work			
	LI	HI	Total	LI	HI		
1	0	294	294	0	1-135	3-69	4-45 5-45
2	359	173	532	1-166	2-193 4-16		
3	0	155	155	0	3-155		
4	441	28	469	3-276	4-110	5-55 4-28	
5	0	0	0	0	0		

NOTE:  $\lambda_2$  = location cost multiplier in objective function.

The standard doubly (or fully) constrained spatial interaction model has the form

$$T_{jl} = A_j B_l O_j D_l \exp(-\beta c_{jl}) \quad (17)$$

where

- $T_{jl}$  = number of trips between zone  $j$  and zone  $l$ ,
- $O_j$  = trips generated in (originating from) zone  $j$ ,
- $D_l$  = trips attracted to (terminating in) zone  $l$ ,
- $c_{jl}$  = zone-to-zone travel cost, and
- $\beta$  = a parameter;

and where

$$A_j = \left[ \sum_l B_l D_l \exp(\beta c_{jl}) \right]^{-1} \quad (18)$$

$$B_l = \left[ \sum_j A_j O_j \exp(-\beta c_{jl}) \right]^{-1} \quad (19)$$

It has been shown by Murchland (13) and Wilson (14) that this model can be derived from an equivalent optimization problem. The form of that problem is

$$\text{Max: } S = - \sum_j \sum_l T_{jl} \ln T_{jl} \quad (20)$$

subject to

$$\sum_l T_{jl} = O_j \quad (21)$$

$$\sum_j T_{jl} = D_l \quad (22)$$

$$\sum_j \sum_l c_{jl} T_{jl} = C \quad (23)$$

where the only new value is  $C$ , taken to represent the total system travel cost.

There is a relationship between  $\beta$  in Equations 17–19 and  $C$  in Equation 23. The  $\beta$  in the spatial interaction model produces a dispersion of trips away from the optimum or minimum-cost solution. What human behavior might account for this dispersion is unspecified but presumably includes such factors as variables not in the model, as well as variations in individual perceptions of costs and differences in individual utility functions.

To introduce the spatial interaction model into TOPAZ in lieu of the minimum-cost "transportation" model requires that the model of Equations 20–23 be substituted for the transport cost term in the TOPAZ objective function. Following algebraic manipulations, this yields (after changing signs to allow minimization)

$$\begin{aligned} \text{Min: } S = & \frac{1}{\beta} \sum_i \sum_j \sum_l T_{ijl} \ln T_{ijl} \\ & + \sum_i \sum_j \sum_l T_{ijl} c_{ijl} + \sum_i \sum_j b_{ij} X_{ij} \end{aligned} \quad (24)$$

subject to

$$\sum_l T_{ijl} - s_i X_{ij} = 0 \quad (25)$$

$$\sum_j T_{ijl} - r_i X_l = 0 \quad (26)$$

$$\sum_j X_{ij} = A_i \quad (27)$$

$$\sum_i X_{ij} \leq Z_j \quad (28)$$

$$T_{ijl} \geq 0, X_{ij} \geq 0 \quad (29, 30)$$

The Archerville data were again used for tests of the model, which now required a numerical value of  $\beta$ . This, however, raises an interesting question, which relates directly to the previously described experiments with weightings or multipliers of the terms in the objective function. To simplify the coming discussion it will be convenient to think of the objective function, Equation 24, as having three components:

$$\text{Min: } Z = \lambda_3 U_3 + \lambda_1 U_1 + \lambda_2 U_2 \quad (31)$$

where  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  are arbitrary weights, and where  $U_1$  is the transport cost term,  $U_2$  is the location cost term, and  $U_3$  is the entropy term. The no-dispersion solution given in Table 3 was for  $\lambda_1$  equals one,  $\lambda_2$  equals 0.70, and  $\lambda_3$  equals zero.

The value of  $\lambda_3$  as discussed here is the inverse of the  $\beta$  of Equation 17 and thus will directly affect the extent to which location is dispersed from the no-dispersion case where  $\lambda_3 = 0$  and thus  $\beta = \infty$ . With a  $\lambda_3$  of 20 and thus  $\beta = 0.05$ , the dispersion shows quite clearly in the results given in Table 4. A further increase in  $\lambda_3$ , to 30, giving  $\beta = 0.033$ , yields even greater dispersion. Numerous tests were run with different combinations of values for  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$ . In all cases increasing values of  $\lambda_3$  produced increased dispersion of household location. The most dispersion was achieved with values of  $\lambda_1$  equal to one,  $\lambda_2$  in the vicinity of 0.3, and  $\lambda_3$  equal to 30 or more. The conclusion here is simple, the addition of a dispersion term to the objective function of a mathematical programming model of residence location results in a more even distribution of residents to zones and is probably a more realistic representation of actual human behavior.

TABLE 4 ARCHERVILLE—LOCATION OF HOUSEHOLDS BASED ON LOCATION COST PLUS TRANSPORT COST COMPONENTS OF MODEL ( $\lambda_2 = 0.70$ ), INCORPORATING DISPERSION, WITH  $\beta = 0.05$

Zone	Households			Households, Showing Place of Work	
	LI	HI	Total	LI	HI
1	45	258	302	1-42 3-3	1-128 2-21 3-84 4-14 5-11
2	319	205	524	1-105 2-181 3-25 4-4 5-4	1-6 2-135 3-17 4-32 5-15
3	20	138	158	1-1 3-19	3-117 4-12 5-9
4	416	49	465	1-18 2-11 3-230	3-6 4-33 5-10 4-106 5-51
5	0	0	0	0	0

NOTE:  $\lambda_2$  = location cost multiplier in objective function.

COMBINED ACTIVITY LOCATION AND TRIP ASSIGNMENT MODEL

Having shown in the preceding two sections of this paper how both a traffic assignment model and an activity location model could be formulated as mathematical programming problems, it is now possible to consider linking them together into a single model to solve both problems simultaneously.

First, note that the original transport cost term in the location model objective function is simply the sum of all trips between origins  $j$  and  $l$ , times the cost of those trips. It is reasonable to assume that, regardless of whether a congested or uncongested network is being used in the model, the costs used will be those of the minimum paths through the network. In the location model the zone-to-zone costs used are the exogenously determined minimum costs given as input to the calculations. The minimum-cost flow problem determines the set of minimum cost paths through the network. Given that these minimum paths are the same, the flow cost produced by the minimum-cost flow problem solution should be identical to the transport cost term in the location model objective function.

In the nonlinear version of the location model the entropy term of the cost function takes care of keeping the zone to zone trips in the objective function, thus the transport cost term can be replaced by the objective function from the minimum-cost flow problem. It will, of course, be necessary to add the flow balance constraints in order to describe the network, and to set them equal to what is now a variable set of trip origins and destinations. These are now variable because the activity locations, and thus the trip matrix, are to be determined as part of the model solution. Thus the combined model has the following form:

$$\text{Min: } S = \sum_i \sum_m \sum_n d_{mn} F_{imn} + \frac{1}{\beta} \sum_i \sum_j \sum_l T_{ijl} \ln T_{ijl} + \sum_i \sum_j b_{ij} X_{ij} \quad (32)$$

subject to

$$\sum_i \sum_n F_{imn} - \sum_i \sum_n F_{inm} = \begin{cases} \sum_i (s_i X_{im} - r_i E_m) & m \leq J \\ 0 & m > J \end{cases} \quad (33)$$

$$\sum_l T_{ijl} - s_i X_{ij} = 0 \quad (34)$$

$$\sum_j T_{ijl} - r_i E_i = 0 \quad (35)$$

$$\sum_j X_{ij} = A_i \quad (36)$$

$$\sum_i X_{ij} \leq Z_j \quad (37)$$

$$T_{ijl} \geq 0, X_{ij} \geq 0, d_{mn} > 0 \quad (38, 39, 40)$$

where the variable names  $d$  and  $F$  (link cost and link flow) are substituted for the  $c$  and  $X$  used in the minimum-cost flow problem equations.  $J$  is the number of zones (load nodes). In addition employment in zone  $l$  is represented by  $E_l$  in Equation 35 to avoid confusion.

The objective function now has three terms, minimum-cost flows, entropy or trip dispersion, and minimum-cost location. A new set of constraints (Equations 33) is added to the previous set of location model constraints to incorporate the flow-balance portion of the minimum-cost flow problem. This being done it is possible to solve the combined problem for both link flows and household location. Now it is possible to include congestion in this problem too. The first term in the objective function (Equation 32) is modified as per Equation 9 to make link cost a function of link volume (flow). The remainder of the objective function and all of the constraints are unchanged.

CONCLUSIONS

The work described in this paper is only a brief introduction to the topic, yet certain conclusions can now be clearly drawn. The first of these is that numerous model tests using the kinds of models described here indicate that linear mathematical programming models of location are inherently unrealistic. The least-cost zone will get all possible locators even if the next-to-least-cost zone is only marginally more expensive. The objective function component weighting problem implies that an arbitrary difference in units of measurement (e.g., between hundreds of dollars or thousands of dollars for annualized rent) can result in one component of a model solution's being dominant over another.

To a rather considerable degree the constraint equations of a linear programming model can ameliorate some of these dif-



faculties. This is, however, a mixed blessing. That they can ameliorate some of the difficulties also points up the rather considerable extent to which they determine the model solution. The constraints (e.g., available residential land per zone) must be exogenously determined, yet their determination, in and of itself, implies a difficult forecasting problem. The availability of data for constraints during the development of a mathematical programming location model can give a false sense of confidence in the model's predictive power if the problems of forecasting the constraints are not taken into account.

Another major issue is that of obtaining the necessary data. Housing costs are notoriously difficult to estimate. With the TOPAZ model the literature suggests that it has often been so difficult to estimate the benefits of location or interaction that only costs were used. Yet, both versions of TOPAZ would certainly have given different results if some locational advantage variable had been used to yield a "net" location cost variable for the location cost term in the objective function.

Finally there is the computational problem. The GAMS package (15) was used for all of the Archerville tests reported here. The linear formulations were run on an IBM PC with 640K of memory and a hard disk drive. These runs took just a few minutes each. The nonlinear formulations were problematic on the PC; some took 2 or more hours to solve, and others would not run at all. Finally, a version of GAMS for the IBM 3081 GX mainframe was used for the nonlinear formulations. In the combined model there were 622 variables in the objective function. The objective function for the final nonlinear version of TOPAZ, alone, had 110 variables, but, if there were a 30-zone region to be analyzed, the objective function in the nonlinear TOPAZ model would have 3,660 variables. This is a rather sizable problem, and yet a 30-zone spatial interaction model is really too highly aggregated for most policy analysis purposes. Further, the examples presented used only two household types and no housing stock consideration. Current transportation and location modeling applications tend to have from 200 to 300 zones. With no increase in activity types, 250 zones would yield 250,500 variables in TOPAZ. The model formulations that combine location and trip assignment are even larger. Problems of this size are really quite impractical for direct solution. The problems of solving optimizing models of a realistic (from the applications point of view) size can be dealt with by decomposition procedures if it is desirable to maintain the mathematical programming formulations. The possibility of transforming programming models into spatial interaction models as briefly mentioned earlier offers another avenue of approach. Both of these approaches will be examined in future research.

Despite these concerns there are several important points to be learned from these experiments. Perhaps the most important is that developing these model formulations and then testing their behavior gives wonderful insight into various hypotheses about locational behavior. The eminent sensibility of describing locational behavior as an optimizing process is beyond reproach. The effects, and general importance, of constraints in

such formulations became clearly evident in these experiments. At the same time the experiments clearly illustrated the need for inclusion of dispersion terms in such models. The inference is that although locational behavior may be said to be, in principle, an optimizing process, in actuality there are obviously other factors that result in a dispersion of locations around a simple "least-cost" optimum. Yet the optimizing process provides a model-building rationale that can be particularly helpful in understanding the implications of model structure and can thus, in turn, be expected to improve modeling practice as well.

## REFERENCES

1. S. Putman. Future Directions for Urban Systems Models: Some Pointers from Empirical Investigations. In *Advances in Urban Systems Modelling* (B. Hutchison and M. Batty, eds.), North-Holland, Amsterdam, The Netherlands, 1986.
2. S. Putman. *Integrated Urban Models*. Pion Ltd., London, England, 1983.
3. M. Ben-Akiva and S. Lerman. *Discrete Choice Analysis*. MIT Press, Cambridge, Mass., 1985.
4. Y. Sheffi. *Urban Transportation Networks*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1985.
5. A. Anas. *Residential Location Markets and Urban Transportation*. Academic Press, New York, 1982.
6. A. Wilson, J. Coelho, S. Macgill, and H. Williams. *Optimization in Locational and Transport Analysis*. John Wiley & Sons, Chichester, England, 1981.
7. D. Boyce and R. Eash. Application of a Combined Residential Location, Mode, and Route Choice Model to Strategic Planning for the Northeastern Illinois Regional Transportation Authority. Presented at 66th Annual Meeting of the Transportation Research Board, Washington, D.C., 1987.
8. P. Prastacos. Urban Development Models for the San Francisco Region: From PLUM to POLIS. Presented at 64th Annual Meeting of the Transportation Research Board, Washington, D.C., 1985.
9. J. Brotchie. A General Planning Model. *Management Science*, Vol. 16, 1969, pp. 265-266.
10. R. Sharpe and J. Brotchie. An Urban Systems Study. *Royal Australian Planning Institute Journal*, Vol. 10, 1972, pp. 105-118.
11. J. Brotchie, J. Dickey, and R. Sharpe. *TOPAZ*. Lecture Notes in Economics and Mathematical Systems 180. Springer-Verlag, Heidelberg, Federal Republic of Germany, 1980.
12. R. Sharpe, B. Wilson, and R. Pallot. Computer User Manual for Program TOPAZ82. Division of Building Research, Commonwealth Scientific and Industrial, Research Organisation, Highett, Victoria, Australia, 1984.
13. J. Murchland. *Some Remarks on the Gravity Model of Trip Distribution and an Equivalent Maximizing Procedure*. LSE-TNT-38. London School of Economics, London, England, 1966.
14. A. Wilson. A Statistical Theory of Spatial Distribution Models. *Transportation Research*, Vol. 1, 1967, pp. 252-269.
15. D. Kendrick and A. Merraus. *GAMS, An Introduction*. The World Bank, Washington, D.C., 1985.