How To Code the Generalized Cost of Accessing a Transit System

ANTTI TALVITIE

A method of coding generalized cost transit system access links is described, and the advantages of using large zones in transportation planning are demonstrated. Use of large zones is motivated by the expense (in time and money) of making travel demand forecasts. The method was developed concurrently with the models for predicting the values of the access link times for rail and bus trips. These supply models estimate the in-vehicle times on automobile and bus and the out-of-vehicle times for bus and rail access links using zonal and modal characteristics. Use of different zone sizes is evaluated by a correlation analysis between the predicted and the actual number of rail riders. The results indicate that there is a high correlation between the predicted and actual number of users; the error increases with the zone size, but the increase is small. Error sources independent of zone size also exist; these errors are discussed in detail. It appears that large zone sizes can profitably be used in transportation planning. The access supply models used are very simple. A more complex set of models has been developed. The reason for reporting this simple model is validation. Data collected in extant transportation studies focus on the demand side and made the validation of supply side models impossible. In the present study the data are rich enough but date back to the late 1960s. Good methods are timeless, however.

Today it is a common planning practice to use zone sizes of 1 mi² or less in area. The principal reason for using small zones is to reduce the inaccuracies deriving from the access links, to reduce the within-zone variance. However, the access links remain a large error source in travel demand forecasting (1). The purpose of this study is to describe a systematic method to calculate the values of the access links and show that larger zones can be used effectively in the planning process. The soundness of using larger zones is made possible by supply models (2,3). The supply models are based on the characteristics of the zone and the transport system serving it and make explicit the intrazonal transportation system, which is needed to enable the use of large zones without losing accuracy. In fact, accuracy may be gained by the explicit modeling of intrazonal transportation system, not by using smaller zones.

The use of large zones brings with it many advantages, including quicker coding of networks, less chances to make errors in network coding, less expensive traffic assignments (4-6), interpretable assignment outputs, more accurate land use projections (7), better travel forecasts, and visual control of input data and other data errors. By using the supply models together with access mode/station choice models, reliable forecasts are also possible for access mode and station usage. This information is important for the design of public trans-

portation systems. There has been renewed interest in modeling transportation access networks and access mode choices. Recent work (8,9) in modeling transportation access choices and access systems is similar to the present work.

METHOD AND DATA

Consistent Calculation of Access Link Values

The method to be described for calculating the values of the access links, and the subsequent decomposition of volumes on these links by mode and station, is based on a model system that is estimated in stages with each stage affecting the following stage (the nested logit model).

A joint decision for choosing to travel on access mode a via station s on line (or path) l, by priority mode m, during time h, to destination d with frequency f, is a function of the level-of-service L provided by the system, the activity system attributes A, and socioeconomic attributes of the traveller S. It can be expressed as

$$P(f, d, h, m, l, s, a) = F(L, A, S)$$
 (1)

This model can be broken into any number of sequences using the theorem of total probability. For example, it can be expressed as a multiplication of models:

$$P(f, d, h, m, l, s, a) = P(a|s, l, m, h, d, f)$$

 $\times P(s|l, m, h, d, f) \times P(l|m, h, d, f)$
 $\times P(m|h, d, f) \times P(h|d, f) \times P(d|f) \times P(f)$ (2)

where

$$P(a|s, l, m, h, d, f) = access mode choice,$$

 $P(s|l, m, h, d, f) = station choice,$
 $P(l|m, h, d, f) = line choice,$
 $P(m|h, d, f) = main mode choice,$
 $P(h|d, f) = hour-of-day choice,$
 $P(d|f) = destination choice, and$
 $P(f) = trip frequency.$

This general model system provides a sound framework for estimating access mode, station loadings, transit line choice, choices of mode and destination, and so forth. If some elements are not present in the analysis (e.g., choice of travel hour), that segment of the model system can be dropped without ruining the model system. This paper focuses on the first two models: access mode and access station choices. Other

choices are dropped for clarity. The access mode and station selection models are logit models.

Each model has a utility function that includes variables relevant to the choice being modeled. For example, function U-a describes the utility to travel to a station on Access Mode a. For the sake of example, this function may be made up of the trip time (T-a), cost or fare (C-a), and service headway (H-a). Mathematically,

$$U-a|s, l = b_1(T-a) + b_2(C-a) + b_3(H-a)$$
 (3)

Once the coefficients b and the values of the explanatory variables are known, *U-a* is a number, commonly called a "generalized price" of Mode a. The aggregate, "logitly" consistent generalized price to Station s by all modes is given by the following expression:

$$U|s, l = \log(\sum \exp U - a|s, l)$$
 (4)

Similarly, the access station model may be expressed as follows:

$$U-s|l = c_1(PKG-s) + c_2(T-ss') + c_3(U|s, l)$$
 (5)

where

PKG-s = parking cost (or parking availability) at Station
s, and

T-ss' = travel time on line haul from Station s to a competing station, s'.

The aggregate "logitly" consistent inclusive price to access a rail (transit) line is

$$U-l = \log(\sum \exp U-s|l) \tag{6}$$

Use of Access Mode and Station Selection Models in Network Coding

The objective of the access mode/station selection models is to give a number to be assigned to an access link in coding networks and provide information on access mode and station usage in the zone where travelers reside. This information can be obtained if the traffic zone is connected to the network in a manner consistent with the travel demand models used to characterize not only travelers' access mode/station choice behavior but also the other travel dimensions such as mode and destination choices. In fact, the connection of traffic zones to the network is critical from the point of view of obtaining reliable traffic forecasts by line and mode. It is precisely the inclusive price, developed earlier, that should be used to connect the zone centroids to bus lines or rail stations (*U-s*|*I*) or lines (*U-l*). The difficulty introduced by the egress attributes will be addressed later.

The following examples illustrate how in practice traffic zones should connect to the network in a systematic and consistent manner. Consider a traffic zone served by rail and bus modes. Access to rail stations is by walk, automobile, and bus, whereas walk is the only access mode for bus (this restriction is solely for illustration). The connection of a zone

to each rail station (or rail line) is via an access link whose value is a weighted average that is a proper aggregation of the available access modes (and stations). The connection of the zone to bus lines is via a walk access link. In order not to confuse the two modes, presentations of bus and rail are worked separately to point out how *U-l*, the generalized line price, is calculated for coding the access links.

The bus network is considered first because it is simpler (see Figure 1). The example zone is served by three bus lines, 1, 2, and 3, which are connected to the zone centroid by values U-1, U-2, and U-3. Because there is only one access mode, these are the average zonal walk times to these lines and will be denoted by U-1. These values depend on such factors as zone area, bus spacing, and bus frequency and are obtained directly from the supply models, to be explained shortly.

The calculation of the rail access links requires both access network supply and access mode/station choice models. Consider the rail network in Figure 2. The two rail lines serving the zone can be accessed by walk, automobile, and bus. Bus Lines 1 and 3 can also be used for access purposes. The example zone is thus connected to Rail Lines 10 and 11 by Access Links *U*-10 and *U*-11, respectively. To calculate the value of these access links, the access mode/station selection models (Equations 3 and 5) and the values of *T-a*, *C-a*, *H-a*, *PKG-s*, and *T-ss'* are needed. The latter are calculated for each zone as a function of the transportation system serving or planned for the zone. In the present example application a simple parametric access network model is used (10) as explained next.

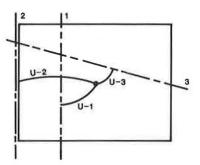


FIGURE 1 Bus network.

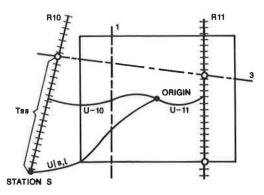


FIGURE 2 Rail network.

Access Supply Models

The access supply models give the mean values for automobile and bus in-vehicle times and bus out-of-vehicle (walk) time. Models for within-zone variances of these variables were also developed for other uses with explicit aggregation procedures, such as sample enumeration (6). Only the former models are used here. A summary of the models appears in Table 1. The variables used in these models are shown and defined in Figure 3.

The walk time to a bus stop depends on such variables as distance between stops, spacing between bus lines, and zone size. The model logically shows that increases in all these values increase the walk time to bus. The farther the station is from the zone centroid, the longer are bus ride and car drive times to station. The greater the speed of the vehicle, the shorter are bus ride and car drive times. Depending on whether lines run parallel or perpendicular, one of two equations is used in finding the bus out-of-vehicle and in-vehicle times.

The available data did not permit the use of the more advanced versions of the models here; the use of the more sophisticated access supply models in a different modeling environment can be found elsewhere (4-6).

Data Source

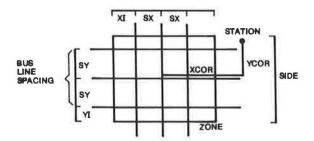
The data come from the origin-destination survey conducted in 1969 by W. C. Gilman Company for the Southward Transit Area Coordination Committee (STAC) in Chicago. Three data sets with varying zone sizes are formed from the STAC study. The first data set consists of sixty-eight 1-mi² zones and twenty-nine 4-mi² zones. This data set is denoted as A(1-4); these were the actual sizes in the original STAC study. The second data set combines 42 square mile zones into thirteen 4-mi² zones. This data set is denoted as A(4). The third set of data is the combination of thirty-six 1-mi² zones into four 9-mi² zones and sixteen 4-mi² zones into four 16-mi² zones. This data set is denoted as A(16). The combined zones were

TABLE 1 Parametric Supply Models

MEAN T = 10.55 + .37 AREA - .52 SPEED + 2.67 DUMMY + 1.95(XCOR + YCOR)

Variable definitions:

AREA XCOR,	the area of the zone in square miles the coordinates of the station from the centroid
YCOR	of the zone in miles
SIDE	the side of the zone in miles
SPEED	speed on arterials in miles per hour
SX, SY YI or XI	spacing of the bus lines in miles the distance from zone boundary to the nearest bus line in miles; it is negative if the bus line is outside the zone
STOPS DUMMY	the number of bus stops per mile a variable to identify whether or not the station is inside the zone, it equals 0 if station is inside the zone, and 1 if station is outside the zone



Definition variables:

AREA the area of the zone in square miles

XCOR, the coordinates of the station from the centroid

YCOR of the zone in miles

SIDE the side of the zone in miles SPEED speed on arterials in miles per hour

SX, SY spacing of the bus lines in miles

YI, XI the distance from zone boundary to the nearest bus line

in miles; it is negative if the bus line is outside the zone

STOPS the number of bus stops per mile

DUMMY a variable to identify whether or not the station is inside the zone, it equals

0 if station is inside the zone, and 1 if station is outside the zone

FIGURE 3 Variables used in supply equations.

picked at random with the criteria that the zones were connected to one another. Also taken from the STAC report are the number of people going to each rail station by mode (walk, automobile, or bus).

ESTIMATION OF MODELS AND EVALUATION OF RESULTS

In this section the travel demand model system and the supply models just described are evaluated in various ways to gauge how well equations portray supply and, in particular, whether zone sizes can be increased without compromising accuracy. First, the access mode/station choice models are developed. Second, the travel demand model system is applied; this model incorporates the supply models and the access mode/station choice models. Third, several indices are used to evaluate the results with the three data sets.

Access Mode Model

Logit access mode models were developed for choices between walk, automobile, and bus access modes. Alternative model specifications were considered even though no socioeconomic data were available. The most satisfactory model resulted when it was assumed, as suggested by data, that persons who reside within a ½-mi radius of a rail station walk to that station. This three-mode model is given by

Mode choice:
$$\begin{cases} \text{Walk} & \text{if distance to station} < \frac{1}{2} \text{ mi} \\ P(m = \text{automobile, bus } | s, l) \\ & \text{if distance to station} > \frac{1}{2} \text{ mi} \end{cases}$$
 (7)

Two of the estimated models are given in Table 2. The difference between the models is that in Model I there is a constant bus fare of 30 cents, whereas in Model II the bus fare is zero. The values of the time variables, with the exception of the automobile out-of-vehicle time (which was set to a constant 2.5 min), were generated by the supply equations reported earlier.

Statistical tests indicate that all the variables in Models I and II are significant at a .99 level of confidence. Similarity is also found in comparisons of the residuals and the implied values of time for out-of-vehicle and in-vehicle times from the two models. Model I was used in this research.

Station Choice Model

The functional form of the station choice model is given by Equation 5. The results in Table 3 indicate two models whose difference is in the value of U|s, l. In Model I, U|s, l is the composite of automobile and bus modes, and in Model II U-s|l is the composite of all three modes (walk, automobile, and bus). For Model II, the walk utility was computed by using the relationship

$$P(w) = \exp(U-w)/\sum \exp(U-a)$$

$$a = \text{walk, automobile, bus}$$
(8)

$$\exp(U-w) = (1 - \operatorname{Cov})[\exp(U-a) + \exp(U-b)]/\operatorname{Cov} \tag{9}$$

where

P(w) = the probability that walk was chosen as the access mode,

TABLE 2 Two Estimated Access Mode and Station Choice Models

Coefficients and Relevant Information of the Access Mode Models

Variable	-	Standard Error		Standard Error
Out-of-vehicle time	-0.115	0.011	-0.108	0.011
In-vehicle time	-0.027	0.007	-0.045	0.008
Cost	-0.082	0.007	-0.094	0.009
Auto bias	-0.293	0.228	2.855	0.200

of observations for both models 291

Model I

L(0) = -3890 (log likelihood for coefficient of zero) L(0) = -1310 (log likelihood for estimated coefficients)

Model II

L (0) = -3890 (log likelihood for coefficients of zero)

L (0) = -1280 (log likelihood for estimated coefficient)

Coefficients and Relevant Information of the Station Choice Models

Variable	Model I Coefficient	Standard Error	Model II Coefficient	Standard Error
PKG	0.367	0.037	0.638	0.029
T-ss'	-0.047	0.002	-0.040	0.002
ln p ^m	1.000		1.000	

of observations for both models 291

Model I

L (0) = -7510 (log likelihood for coefficients of zero)

L (0) = -6770 (log likelihood for estimated coefficients)

Model II

L (0) = -15100 (log likelihood for coefficients of zero)

L (0) = -14200 (log likelihood for estimated coefficients)

TABLE 3 Error Measures for Predicted Access Mode and Station Volumes

	R(t-value)	Ave. Vol.	AE	AE	SD	RMSE
Walk						
A(1-4)	.98 (90.1)	19.0	-0.01	4.2	10.5	10.5
A(4)	.89	20.7	0.01	10.3	20.1	20.1
A(9-16)	.94	20.6	0.00	14.6	29.8	29.8
Auto						
A(1-4)	.94 (49.3)	18.5	-0.23	4.8	9.4	9.4
A(4)	.74 (10.8)	24.6	-0.33	7.4	14.8	14.8
A(9-16)	.88 (16.9)	62.2	-0.29	14.6	27.4	27.4
Bus						
A(1-4)	.71 (17.9)	.6	-0.01	1.3	3.0	3.0
A(4)	.43 (4.6)	1.1.	-0.00	3.2	7.2	7.2
A(9-16)	.38 (3.8)	.9	0.00	3.7	7.0	7.0
Station						
A(1-4)	.83 (25.9)	38.0	-0.01	20.9	41.1	41.1
A(4)	.52 (5.9)	46.4	0.02	32.7	52.7	52.7
A(9-16)	.82 (13.1)	83.7	-4.36	42.1	73.5	73.6

R(t-value) correlation coefficient between actual and predicted volumes and its t-value

AE average error

the average absolute error |AE|

standard deviation SD

 $SD^2 + A^2$ = the root mean square error **RMSE**

U-a = the utility of an access mode, andCov = the percentage of the zone covered by the walk access area.

Statistical tests indicate that the coefficients in both models are significant at .99 level of confidence. Model I was chosen over Model II primarily because of its consistency with the access mode model, or

Station choice:
$$\begin{cases}
\text{nearest} & \text{if distance to station} \\
< \frac{1}{2} \text{ mi} & \text{otherwise}
\end{cases}$$
(10)

Application of the Travel Demand System

Now that the access supply models and the access mode/ station choice models have been developed, the travel demand model can be evaluated by applying it in stages with each stage affecting the following stage.

The first stage of the model is the prediction of the values of the supply variables. Parametric models of Table 1 are used to get the mean values for the supply variables to each station serving the zone. The operating cost for the automobile was a function of the distance from the zone centroid to the station.

The second stage of the model is the estimation of the modal splits to each station. Again, it was assumed that everyone living in the zone within a ½-mi radius of the station walks to that station. It was, furthermore, assumed that people walking to the station have chosen their housing premises on the basis of proximity of the rail station. Therefore, people living within walking distance of the station are more likely to ride the rail system than people who must use another access mode. Accordingly, it was assumed that the walkers are 85 percent more likely to use the rail mode than people who must use automobile or bus to reach the station. This value was based on a Skokie Swift mode choice study where a logit coefficient was estimated for a similarly defined variable (11).

In the third stage the station selection model was applied. Each station was evaluated on its own merits and compared with the characteristics of competing stations.

The final stage of the system, as applied here, is the computation of the "generalized" access price to line, U-l. As explained, the "logitly" consistent aggregation of modes and stations amounts to computing the expression $\log(\Sigma \exp U$ -s|l). The information used in computing U-l can be used to parcel the travel volume by station and access mode.

The recursive model system was applied to each of the three data sets having different zone sizes, and the predicted shares of the access modes and station usage were obtained. Volumes were found by multiplying the predicted shares from the recursive model by the actual total number of station users (known from the STAC report).

Evaluation of Results

Four measures of predictive accuracy are calculated for each of the three data sets A(1-4), A(4), and A(16). The measures are the correlation coefficient (R) between actual and predicted volumes and its t-value, the average error (AE), the

average absolute error (|AE|), standard deviation of the error (SD), and the root mean square error (RMSE = $SD^2 + AE^2$). The results are given in Table 3.

The observed and predicted access mode shares are in good agreement for all the modes in all three data sets. However, the walk and automobile models have much better results than the bus. Both walk and automobile have very high correlation coefficients and *t*-values at each level of aggregation.

The bus mode gave good results for the first data set but deteriorated for the larger zones. After examining the two larger data sets, A(4) and A(16), it was found that the bus volumes were well predicted in the core areas where the bus lines were closer and the service more frequent. However, in the fringes of the study area, some zones that had no bus service, or at least very poor service, were combined with zones having either satisfactory or good bus service. It was these combined zones in the fringe areas that caused the results to be not quite as good as in the other zones. This fact points toward the need for explicit aggregation procedures to maintain high accuracy in forecasts; however, see later comments on other sources of error.

The station choice results are also very good. The first and third data sets, A(1-4) and A(9-16), have high correlation coefficients, whereas the second data set has only average results. However, each of the three correlation coefficients is significantly different from zero with .99 level of confidence. They also have low standard errors.

The examination of other error indicators (AE, |AE|, SD, and RMSE) gives rise to the following observations. First, bus volume errors are low, contrary to what we would expect from the correlation coefficients. Second, prediction errors increase with increasing zone size. This is especially true for the automobile and walk modes, and it is also true for the station choice. It would be easy to declare that either zone sizes must be kept small or that travel forecasts need to include specific aggregation measures if nominal forecasting errors are to be kept reasonable, at least when using large zones. However, such a conclusion is not supportable by these data. There are other hitherto unmentioned sources of error that act precisely in the same direction as zonal aggregation, that is, increasing errors with increasing zone size.

In addition to model error and the lack of aggregation procedures, the most outstanding of the so far unmentioned error sources are the following. First, the percentage of people living within ½ mi of the station, and thus the percentage of people having walk access to the station, was simply approximated by the percentage of the zonal areas that fell within ½ mi of the stations. It is well known that development densities near stations are often higher than further from them; vacant land is also more likely farther from the stations. Better knowledge of the distribution of the residences within traffic zones would definitely have increased the accuracy of the results. Second, in several zones the stations were on competing rail lines, either on the two branches of the Illinois Central or Rock Island Railroads or on the South Shore & South Bend Railroad. Because no model was developed for line choice and because these railroads, particularly the Rock Island RR versus the other two, had distinctly different egress attributes that directly affected line choice and indirectly affected station and access mode choices, one would expect noise in the access mode access station predictions. The Chicago network shows vividly why the inclusion of egress analyses, in exactly the same way as the access analyses in this paper, is necessary for reliable travel forecasts in transportation studies involving rail lines. Third, in many zones express bus service provided by the Suburban Safeway Company competed vigorously for the rail traveler. Even though this is a bus service, its attributes are more in line with the rail service and should definitely be taken into account in complete analyses. Fourth, it must be kept in mind that the model access mode choice was estimated (in part) by using the data generated at the finest aggregation level. Thus, if the model coefficients are "contaminated" by data aggregation, they will perform best at the same level of aggregation. At a minimum it can be said that the demand side model favors the finest level of aggregation. Thus, even though the model system was by necessity applied in an incomplete manner, the results are strikingly good and suggest that it is a useful planning tool. In addition, the results provide indirect evidence that choice of mode to work is closely tied to residential location decisions. This fact, which was observed in the Skokie Swift study cited earlier and the fact that coefficients estimated in that study proved useful in the present study, indicates that the relationship between mode to work and residential location is subject to regularities that can be modeled.

CONCLUSIONS

The study has demonstrated that large zones in conjunction with parametric supply and demand equations can effectively be used in transportation planning. This can speed planning processes and allow for more reliable and quicker prediction of land use activities. Use of large zones also enables review of input data and land use predictions by expert panels, local interest groups, and others having an interest in the planning process and travel predictions.

Parametric supply models can be developed also for interzonal transportation systems (3), thus freeing the analyst and the planner from coding the networks, which currently is one of the major roadblocks to analyzing systematically a large number of significantly different alternatives. Parametric supply models also facilitate sensitivity analyses because unit changes in supply can be related in a straightforward manner to both demand and resource costs; (marginal) pollution impacts can also be traced in this manner. The implementation of such a model system would be a major step toward more timely and systematic transportation planning.

There is nothing to prevent increasing the zone size indefinitely. In so doing, however, the parametric supply models must be made an order of magnitude more sophisticated. Such zone-independent supply models would be analogous to properly aggregated behavioral travel demand models. They would relate the values of the supply variables to the transportation system attributes and the distribution of economic activities within the region via the travel demand models. Such a model system would find its most rewarding use in sketch planning and comparing alternative city forms as well as in statewide planning, where the zones must necessarily be very large.

REFERENCES

- A. Talvitie and Y. Dehghani. A Comparison of the Observed and Coded Network Travel Time and Cost Measurements. In Transportation Research Record 723, TRB, National Research Council, Washington, D.C., 1979, pp. 46-51.
- A. Talvitie and T. Leung. A Parametric Access Network Model. University of Oklahoma, 1974.
- A. Talvitie and Y. Dehghani. Models for Transportation Level of Service. Transportation Research B, Vol. 14B, 1980, pp. 87– 99.
- A. Talvitie. Transport System Management in San Francisco: An Assessment. Transportation Planning and Technology, 1980.
- A. Talvitie and Y. Dehghani. Comparison and Evaluation of Case Study Alternatives for a Light-Rail System and Its Possible Land Use Impacts in Buffalo, N.Y. Region. Transportation Forum, 1985.
- A. Talvitie, Y. Dehghani, and M. Morris. An Integrated Land Use-Transportation Model System for Corridors Analysis and Its Applications. Presented at the National Conference on Transportation Planning Applications, Florida, 1987.
- A. Talvitie, M. Morris, and M. Anderson. An Assessment of Land Use and Socioeconomic Forecasts in the Baltimore Region. Presented at the 59th Annual Meeting of the Transportation Research Board, Washington, D.C., 1980.
- D. L. Kurth and C. L. Chang. Circulator/Distributor Model for the Chicago Central Area. In *Transportation Research Record* 1328, TRB, National Research Council, Washington, D.C., 1991.
- S. Mukundan et al. Development of a Mode of Arrival Model for the Washington D.C. Metrorail System. Presented at the 70th Annual Meeting of the Transportation Research Board, Washington, D.C., 1991.
- J. Templeton and A. Talvitie. A Model System for Network Access. Presented at TRF meeting, Boston, Mass., Oct. 1976.
- A. Talvitie. A Comparison of Probabilistic Mode Choice Models: Estimation Methods and System Attributes. In *Highway Research Record* 392, HRB, National Research Council, Washington, D.C., 1972.

Publication of this paper sponsored by Committee on Passenger Travel Demand Forecasting.