

Capacity Restraint in Multi-Travel Mode Assignment Programs

N. A. IRWIN, Vice-President, Traffic Research Corporation, N. Y., and
H. G. VON CUBE, Project Director, Traffic Research Corporation, KCS, Toronto

• **TRAFFIC ESTIMATING METHODS** have been developed primarily for urban areas to help shape and justify road planning decisions. Because peak-hour congestion exists in most cities today on almost all road and rail facilities and will continue to exist to some degree in the future, any realistic method of traffic prediction must recognize the presence of road capacity limitations and its effect on traffic pattern and travel mode.

Developers of traffic assignment programs have realized the necessity of introducing road capacity restraints—the cause of congestion. Moreover, they have also felt it necessary to introduce factors affecting people's choice of travel mode. However, it has been frequently decided for the sake of simplicity (or lack of better methods) that it is sufficient to deal with road users and transit users separately.

It is well known that delay caused by congestion will cause travelers to seek other less congested routes to their destination or other means of travel or will persuade some travelers to cancel their trips entirely. A great number of such decisions may ultimately restrain the functioning of a city or sectors thereof.

Conversely, it is known that the existence of an uncrowded direct route between two urban areas will quickly draw traffic from other more crowded routes and will even create new traffic demand between the two areas. These effects are implicit in the traffic pattern of any city.

Assignment programs that do not take road and parking capacity and choice of travel mode into account in an explicit manner can make no allowance for changes in the pattern of congestion and/or travel mode used which will inevitably occur as the city evolves. In fact, such assignment programs tend to freeze the congestion pattern and travel mode to whatever shape it had when the last O-D survey was made; moreover, they ignore the interaction between road and transit usage which tends to produce a compatible balance between the two travel modes.

The lack of capacity restraint and choice of travel mode in an assignment program can cause considerable error, especially if there are significant differences between the land-use pattern and/or traffic facilities proposed for the future and that in existence when the base survey was made.

Under contract with the Metropolitan Toronto Planning Board the Traffic Research Corporation has developed a systematic model for estimating vehicular and transit flow using high-speed electronic computing techniques. This model contains a direct feedback mechanism by which capacity restraints and the resultant congestion on roads and parking lots are allowed to affect the choice of travel mode, the route selection, and trip volume distribution in successive program blocks.

Thus, the traffic forecasting model described in this paper becomes a powerful tool for the transportation planner. It provides him with a means of estimating traffic flows in an area based on land-use patterns and traffic facilities within the area. It is essentially a working model of the area's transportation system.

Having projected existing land-use patterns to their probable configuration at some given future date, the planner can use the model to analyze and compare alternative proposed transportation systems. For all sections of a proposed system, he can use the model to estimate traffic flows via automobile, commuter train, transit, and truck during an average morning or evening rush hour or during a longer period, if necessary. Using the model he can then estimate the total cost of transportation on the proposed

system during the study period, enabling him to judge how effectively that system serves the given land-use pattern and how well it compares with the other proposed transportation systems. Armed with this tool, the planner is now for the first time able to try out a proposed transportation system, to compare its benefits with its costs, using a model instead of being forced to experiment directly with multi-million dollar investments.

This program has been tested in full-scale studies of the metropolitan Toronto area and has been found to give most satisfactory results when compared with observed data. This paper describes briefly how the forecasting model works and shows how some of its results compare with observed data.

GENERAL DESCRIPTION MODEL

The six basic program blocks comprising the model are

Block 1	Tree generation
Block 2	Time factor
Block 3	Trip distribution
Block 4	Proportional split
Block 5	Assignment
Block 6	Link updating

Two auxiliary program blocks, the grid reduction block (not in operation at the present time) and trip generation block are run beforehand, where necessary, to provide basic data for the model. The order of sequence and flow of information between these units are shown in Figure 1.

Basic Input Information

The area to be studied is divided manually into small contiguous zones, and the road and transit networks are represented by a detailed grid of links and nodes, where a link is the travel path between two adjacent nodes, and a node is the junction point between two or more links. All links are directional, so that the bidirectional travel path between two nodes would be represented by two links. Some nodes are merely intersections in the grid network; others, "centroid nodes" or "home nodes," are taken to represent all properties, except size, of the zone in which they are located.

Input data describing the zones (population, employment, income, car registration, etc.) and the grid (road lengths, flow characteristics, number of lanes, speed limit, transit headways, etc.) are gathered and tabulated in the form of tables and formulas. Traffic prediction is then carried out by means of the six types of program blocks mentioned.

The Flow of Information

The flow of information in the model is briefly as follows:

1. A trip generation is carried out for the land-use patterns and the time period under study. Once completed, this block is never repeated, unless it is desired to start another run based on a different land-use pattern and/or time period.
2. Using the tree generation block (block 1) a first set of routes is determined, one between each pair of zones. Each route is the shortest possible in terms of travel time.
3. The time factor block (block 2) estimates the effect these travel times will have on travelers' propensity (time factors) to travel for various purposes from one zone to another zone, to use one travel mode instead of another (modal split factors) and to follow one route instead of another (assignment factors).
4. The trip distribution block (block 3) takes into account these propensities (time factors) and also opportunities offered for beginning and ending trips in each zone to estimate how many people in total will travel from one zone to another.
5. The proportional split block (block 4) uses these total trip interchange volumes

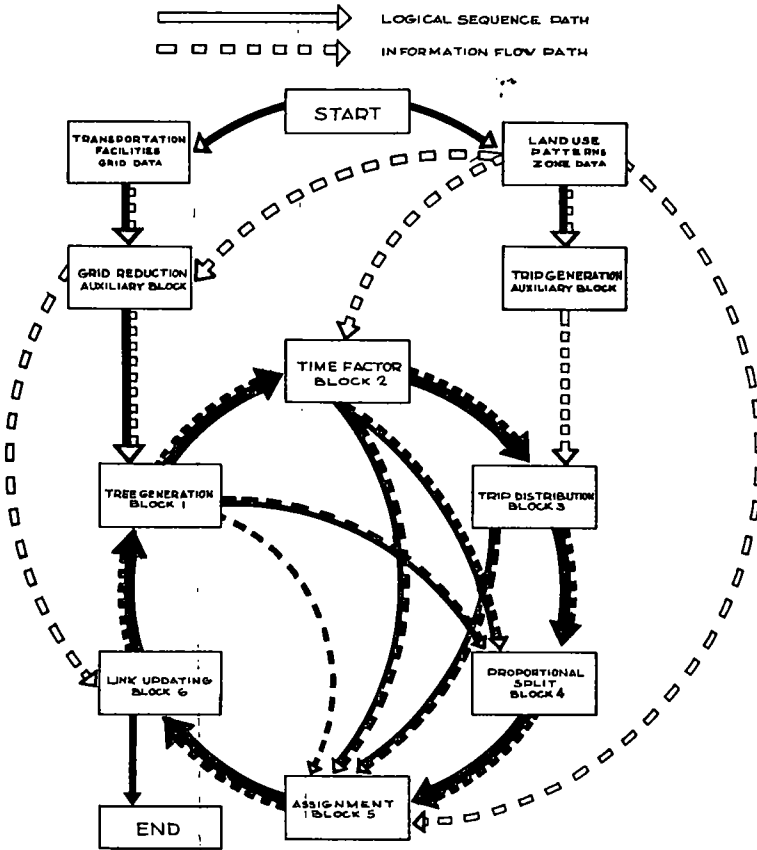


Figure 1. Traffic prediction program block diagram.

and the "split factor," of block 2 to calculate how many people will travel via each mode and via each alternative route from zone to zone (split trips).

6. The assignment block (block 5) assigns these "split trip" figures to the pertinent routes to give estimates of traffic flows, or volumes, via all available travel modes, on each road section or link (link loads) and parking lot of the transportation network.

7. The link updating block (block 6) uses these "link loads" combined with the flow propensities of each link to calculate the link travel times currently applying to all links in the network.

In general, these new link times will be different from those on which the previous determination of routes and travel propensities were based. It is necessary, therefore, to turn again to block 1, using this time the new link times. Weighted averages of the new route times via all available routes between each origin and destination are then used in block 2 to calculate new travel propensities, leading to a new calculation of zone interchange volumes in block 3, and so on around the loop through blocks 4, 5, and 6. It is by this means that feedback occurs. The effects of road traffic congestion and parking lot overflow are fed back within the model to effect the travel patterns as they do in real cities. This feedback procedure is repeated until equilibrium is reached; i. e., until link loads and travel times produced by one cycle are not appreciably different from those produced by a previous cycle.

It can be seen from the foregoing that travel time is the variable that ties all blocks of the model together. It is apparent that the capacity restraints as incorporated in this model play a major part in traffic prediction and are therefore described next.

Capacity Function

The term "capacity function," as used in this paper, means the mathematical formula developed to describe the relationship between traffic volume and travel time for a given road section called link. Flow of vehicles along a road is a very complex phenomenon, depending on many factors. Each road section is probably unique in its combination of factors affecting the flow and should, for precise simulation, have a unique capacity function not only for various speed limits, number of signalized intersections per mile, number of lanes, and presence of public transit but also for different time periods and different weather conditions, etc. Perhaps computer capacity and empirical data will be sufficient someday for this degree of precision.

For the purpose of this paper, however, it is approximated as follows: when large numbers of vehicles are attempting to use a given road section, the speed at which they can travel is materially reduced, a phenomenon that is known as traffic congestion. For any road section it is possible to measure the relationship between the volume or flow, of vehicles traversing the link (cars per hour per lane) and the resulting travel time in minutes per mile. Two typical volume-time relationships of this type are shown in Figure 2, where f is the per-lane flow of vehicle and t is the resulting travel time per mile, known as the per-unit travel time. It can be seen that the expressway can carry a much greater flow than the city arterial street before congestion causes the per-unit travel time to soar. Also, for each of the road sections shown, there is a point, known as the critical flow, f_c , above which the per-unit travel time starts to rise rapidly, and another point, known as the maximum flow, f_m , which is the maximum volume that the road can carry. If user demand continues to climb above f_m cars per hour per lane, then a queue will start forming at some point on the road section and the average travel time per mile experienced by users will increase while the link flow will remain static f_m cars per hour per lane or in some cases drop off to lower values.

To calculate the per-unit vehicle travel time on a given link it is necessary to know in mathematical terms the volume-time relationship for that link. As shown, the shape of this relationship is different for different types of links, such as urban arterial roads and controlled-access freeways. In the present program, 20 different types of links are defined for this purpose, depending on three basic parameters: the speed limit, the number of signalized intersections per mile, and the type of transit vehicles (if any) sharing the road section with automobiles. These 20-link types cover a range from a 30-mph arterial street with 60 streetcars per hour and 10 signalized intersections per mile up to a 60-mph controlled-access freeway.

Data on which these curves are based come from a wide variety of sources. Most prolific of these has been the system of radar detectors mounted at the approaches to several intersections in Toronto as part of the computer automated traffic control project. Input from these detectors is analyzed every two seconds by the computer in such a way that delay at each intersection approach can be calculated as a function of

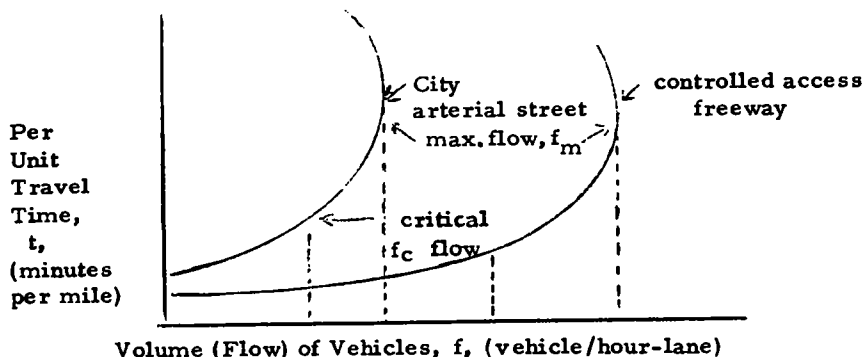


Figure 2. Volume-time relationship for typical road sections.

approach volume. Results from one such location are shown in Figure 3; the observations have been correlated by three straight line segments.

A representative volume-time relationship has been derived for each link type, based partly on empirical data and partly on theoretical considerations. These 20 relationships so derived are called "capacity functions" and are used in block 6 to calculate vehicle link travel times as a function of vehicle flows (see Figs. 4 and 5). Capacity functions are shown in Figure 6.

It can be seen that each capacity function is a simple approximation to the corresponding observed volume-time relationship (shown in Fig. 2) for flows less than f_m cars per hour per lane. However, above f_m , the capacity function differs from the observed volume-time relationship in that the curve does not double back to the left, but continues to slope off to the right. This is because the independent variable in the capacity function (the variable plotted on the horizontal axis) is the demand flow rather than the actual flow. The actual flow cannot, of course, rise above the maximum flow capacity of a given road section; however, the demand flow can rise above the maximum flow capacity, giving rise to higher average per unit travel times as users wait in queues for their chance to travel along the section. The slope of the capacity function for demand flows greater than f_m cars per hour per lane can be calculated by means of simple queueing theory. Figure 6 shows there is a "zero-volume travel time," t_0 (also known as the "ideal time"), which is the average per unit travel time experienced by a vehicle when there are no other vehicles using the road. Similarly, there is a "critical travel time," t_c , which is the average per unit travel time experienced when the flow on the road is f_c , the critical flow; and a "maximum flow travel time," t_m , corresponding to a flow on the road of f_m , the maximum flow. The slope of the capacity function for flows between 0 and f_c (known as the "free-flow" region) is d_1 ; the slope of the capacity function for flows between f_c and f_m (known as the "turbulent" region) is d_2 ; and the slope of the capacity function for demand flows greater than f_m (known as the "overloaded" region) is d_3 . Each of the 20 types of links has a unique set of the seven parameters t_c , f_c , t_m , f_m , d_1 , d_2 , and d_3 , which describe fully its capacity function (see Table 1).

Mathematically, the general equations describing the capacity functions are as follows:

$$\text{For } 0 \leq f(V) < f_c : t(V) = t_c + d_1 [f(V) - f_c] \quad \text{min per mi} \quad (1)$$

TABLE 1
CAPACITY TABLE

Type	Speed Limit	Signal. Intersec. per Mi	d_1	d_2	d_3	t_0	t_c	t_m	f_c	f_m
Cars	30	10	0.0013	0.0188	0.0563	4.4	4.9	7.4	400	533
		5	0.0011	0.0167	0.0500	3.4	3.9	6.4	450	600
		3	0.0010	0.0150	0.0450	3.0	3.5	6.0	500	667
		1	0.0008	0.0125	0.0375	2.3	2.8	5.3	600	800
Buses	30	10	0.0013	0.0188	0.0563	4.4	4.9	7.4	400	533
		5	0.0011	0.0167	0.0500	3.4	3.9	6.4	450	600
		3	0.0010	0.0150	0.0450	3.0	3.5	6.0	500	667
		1	0.0008	0.0125	0.0375	2.3	2.8	5.3	600	800
Streetcars	30	10	0.0016	0.0242	0.0726	4.4	4.9	7.4	310	413
		5	0.0014	0.0208	0.0625	3.4	3.9	6.4	360	480
		3	0.0012	0.0183	0.0548	3.0	3.5	6.0	410	547
		1	0.0010	0.0147	0.0442	2.3	2.8	5.3	510	680
Cars	40	2	0.0007	0.0100	0.0300	1.9	2.4	4.9	750	1,000
		1	0.0006	0.0083	0.0250	1.7	2.2	4.7	900	1,200
	50	1	0.0005	0.0068	0.0205	1.5	2.0	4.5	1,100	1,467
		0	0.0004	0.0058	0.0173	1.2	1.7	4.2	1,300	1,733
	60	0	0.0004	0.0054	0.0161	1.0	1.5	4.0	1,400	1,867

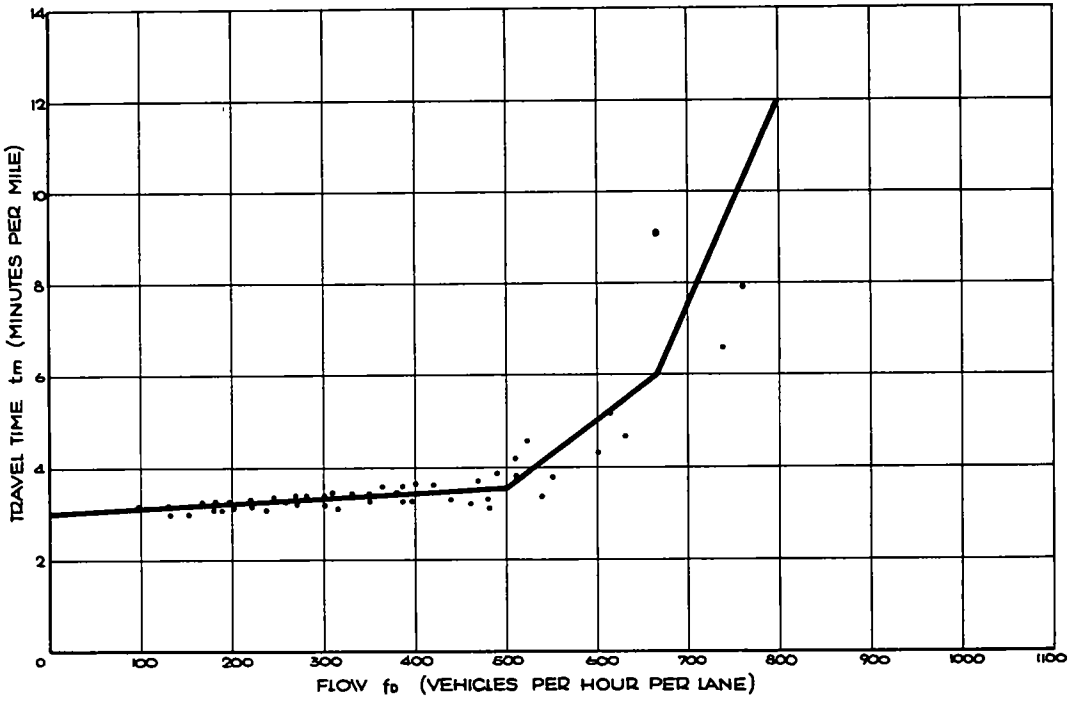


Figure 3. Capacity function for roads with cars only; 30 mph, $S = 3$.

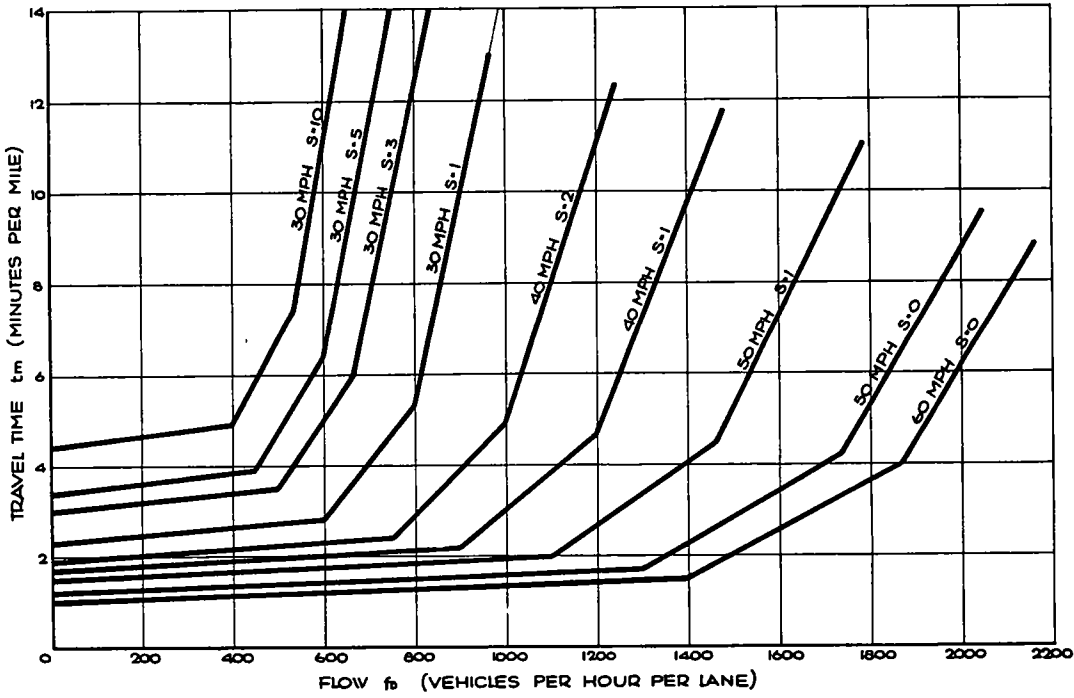


Figure 4. Capacity function curves for roads with cars only, traffic prediction model, Toronto, 1956.

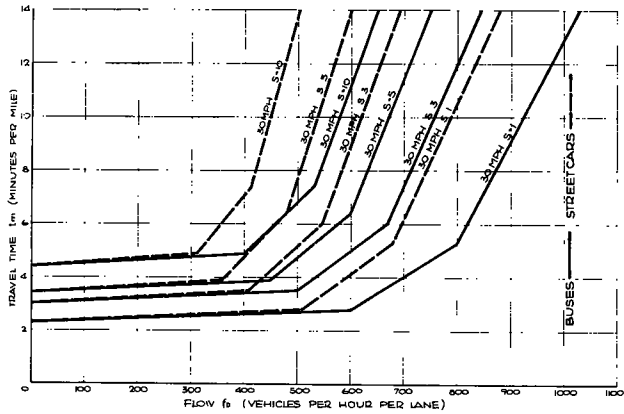


Figure 5. Capacity functions for roads with buses and streetcars, traffic prediction model, Toronto, 1956.

For $f_c \leq f(V) < f_m$: $t(V) = t_c + d_2 [f(V) - f_c]$ min per mi (2)

For $f_m < f(V)$: $t(V) = t_m + d_3 [f(V) - f_m]$ min per mi (3)

in which

$f(V)$ = vehicle demand flow in vehicles per hour per lane, and

$t(V)$ = average per unit vehicle travel time in min per mi.

It can be seen that if the seven parameters t_c , f_c , t_m , f_m , d_1 , d_2 , and d_3 are known for a given road section, and the demand vehicle flow, $f(V)$ (cars per hr per lane) is also known, then the average per unit vehicle travel time, $t(V)$ (min per mi) can be calculated using Eq. 1, 2, or 3, depending on whether the given value of f puts the link into the free-flow, turbulent, or overloaded region.

Equivalent Vehicle Flow

The demand flow used in Eqs. 1, 2, and 3 does not have to be simply the automobile flow in cars per hour per lane. For links on which transit vehicles and trucks also travel, it is possible to calculate an "equivalent vehicle flow," f_e (cars per hr per lane) which takes into account the effects on traffic flow of these other types of vehicle. To obtain the relationship between transit vehicle flow and car flow, conditions at the point of maximum congestion (f_m) were analyzed using observations taken on several major streets in Metropolitan Toronto. The relationship was derived in terms of the number of equivalent cars per transit vehicle (NVPQ) and expressed in

$$F(V) = C + (NVPQ) F(Q)$$
 (4)

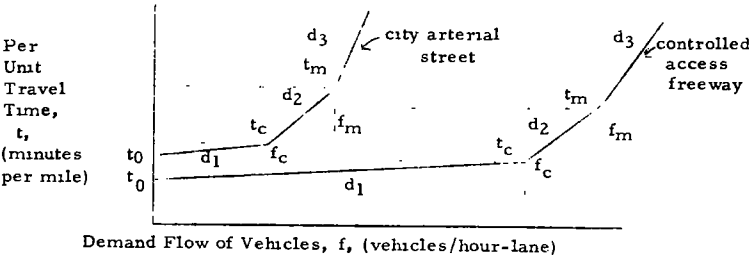


Figure 6. Capacity function for two typical road sections.

in which

$F(V)$ = number of cars per hour;

C = a constant equal to difference between extrapolated value of car flow on a transit route with no transit flowing and theoretical value of a street with cars only at maximum congestion;

$NVPQ$ = ratio of cars per transit vehicle and is equal to inverse of slope of graph (Fig. 7);

$F(Q)$ = number of transit vehicles per hr.

The analysis yielded the following results for

$$\text{Buses} \quad F(V) = 0 + 4.5 F(Q) \quad (5)$$

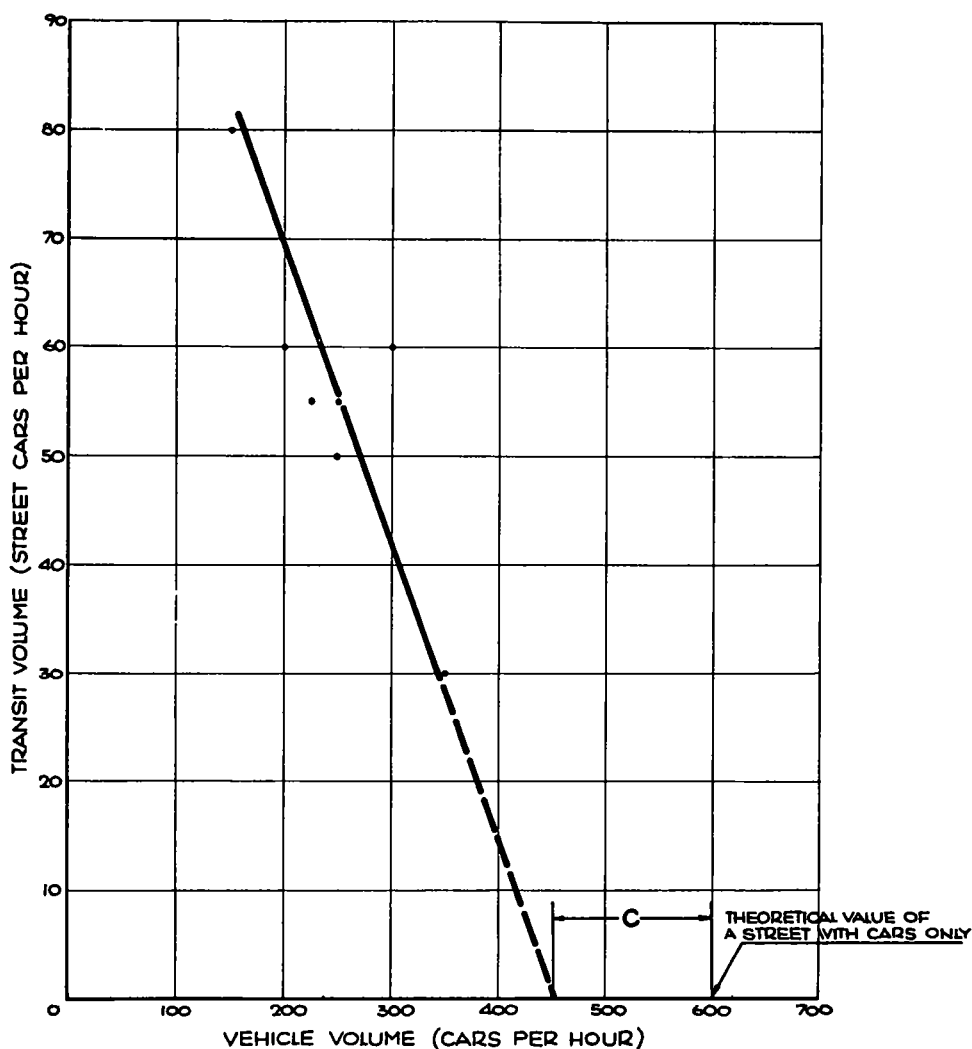


Figure 7. Transit volume (streetcars) vs vehicle volume (cars) at maximum congestion on roads with $S = 5$ and $SL = 30$ mph.

$$\text{Streetcars} \quad F(V) = 150 + 3.5 F(Q) \quad (6)$$

C is zero for bus routes (in all cases the agreement was within 5 percent). Consequently, the capacity function curves as prepared for streets with cars only were directly applicable to streets with buses.

The constant term in the equation for streetcars is explainable by the fact that streetcar tracks themselves impede vehicular flow as well as loading and unloading of passengers. Therefore, it was necessary to develop separate capacity function curves for streets used by streetcars and cars incorporating this constant (see Fig. 5).

In the present program these intermodal relationships are described by two parameters: NVPQ (the number of equivalent cars per transit vehicle) and NVPT (the number of equivalent cars per truck).

The equivalent vehicle flow, for a link that is used by all three modes, may then be calculated using

$$F_e(Q) = (NVPQ) \times F(Q) \quad (7)$$

$$F_e(T) = (NVPT) \times F(T) \quad (8)$$

$$F_e(V) = F(V) + F_e(Q) + F_e(T) \quad (9)$$

in which

$F_e(Q)$, $F_e(T)$ = equivalent flows (in terms of equivalent automobiles) of transit vehicles and trucks, respectively;

$F(V)$, $F(T)$ = link loads of automobiles and trucks, respectively, as produced by block 5;

$F(Q)$ = transit flow as given by transit schedule and listed in a "link table";

$F_e(V)$ = equivalent vehicle flow taking into account effects of transit vehicles and trucks; and

NVPQ, NVPT = intermodal parameters described earlier.

(In this context, f is used to represent flows in cars per hr per lane, and F to represent cars per hr on all lanes of a given link; similarly, t always represents min per mi, whereas T represents min taken to travel from start to end of a given link.)

The equivalent vehicle flow per lane is then calculated from

$$f_e(V) = \frac{F_e(V)}{(NULA)} \quad (10)$$

in which

NULA = number of lanes on the link in question.

When the average per unit vehicle travel times have been calculated the average vehicle travel times, $T(V)$, for a given link is calculated by means of

$$T(V) = t(V) \times L \quad (11)$$

in which

L = link length in miles.

DESCRIPTION OF PROGRAM BLOCKS

Trip Generation (Auxiliary Block)

The Metropolitan Toronto Planning Board carried a home interview survey in 1956. This indicated how many trips for each purpose and mode of travel were generated in each zone for all time periods during the day. Regression analyses were carried out to correlate the data for each time period, trip purpose, and travel mode. Workable

relationships were found between automobile trips generated and three land-use parameters: population, dwelling units, and car registration. Multiple correlation coefficients for the time periods studied exceeded 0.95.

Using these relationships the following categories of trips emanating from each zone are calculated:

1. Primary trips. — Trips having one terminus at the place of residence. These are broken down into three purpose categories:

- (a) Work,
- (b) Business-commercial,
- (c) Social-recreational.

2. Secondary trips. — Trips having neither terminus at place of residence. These are subdivided into the same three categories of trip purpose. The distinction between primary and secondary trips is necessary because different generating relationships are required for each type.

As input for the trip distribution block later in the program, it is also necessary to determine how many trips of each purpose will be attracted to each zone during the period under study. Again, the home interview survey was used to determine relationships between land-use parameters and trip attractor figures for the three purposes. It was found that the work attractors (i. e., the number of work-trips arriving at each zone) could best be correlated by the variable total employment; business-commercial attractors were correlated by retail employment, and social-recreational attractors by population.

When 1956 census data are used, the preceding calculations yield values of trips generated and attracted for each zone in good agreement with the survey figures, as would be expected. Relationships established for 1956 will not necessarily remain valid for future years, and a study is being carried on to determine what trends exist, if any. For present purposes, however, the established relationships must suffice. Quantities such as population, employment, and car registration can be estimated for the future on the basis of past trends, zoning restrictions, and economic forecasts. If desired, ranges of such quantities can be studied. Having made these estimates, the trips generated in and attracted to each zone can be calculated for the future time period under study. This is the function of the trip generation block.

Once, the trips generated and attracted are determined, these "generator" and "attractor" figures are considered fixed for each centroid node. They are changed, by a new run of the trip generation block, only if it is desired to start a new study of a different time period or based on different land-use data.

Before being used by the main program, each set of generators and attractors is adjusted, so that for each trip purpose

$$\sum_{j=1}^{NUHN} A_j = \sum_{i=1}^{NUHN} G_i \quad (12)$$

in which

G_i = generator for i th home node;

A_j = attractor for j th home node; and

NUHN = number of home nodes.

This adjustment is necessary to ensure convergence of the trip distribution calculation for each trip purpose. Usually, Eq. 12 is very nearly met for each purpose before adjustment; one would expect this because the equations linking trip generation and attraction to land-use variables are in general quite accurate. Experience has shown that generators at place of residence can be more accurately determined from source information than can attractors. Therefore, it is always the attractors that are modified to make their sum equal that of the generators; this is done by applying to each attractor

for each trip purpose,

$$A_j = \frac{\sum_{i=1}^{NUHN} G_i}{\sum_{j=1}^{NUHN} A'_j} A'_j \quad (13)$$

in which

A'_j = attractor before adjustment; and

A_j = attractor after adjustment.

Tree Generation (Block 1)

As mentioned earlier, the road network under study is represented by a grid of nodes and links. Each link is fully described for purposes of traffic prediction by the following variables: road capacity function applicable, number of lanes, length in tenths of a mile, transit facility available, and time headway between transit vehicles.

Given the volume of cars per hour using a link at any time during the prediction procedure, it is possible, using the appropriate capacity function, to calculate the travel time required to traverse that link. This information is required by the tree generation block, which determines the shortest route in terms of travel time between every pair of centroid nodes in the grid.

The algorithm used is based on that developed by Dantzig (7) and Moore (8) but has been modified to minimize the number of times a given node must be queried. This means that routes going from an origin to nodes close by are minimized before longer routes are built onto them.

Given the travel time for each link the tree generation block determines for any or all travel modes (automobile, transit, mixed, trucks) the shortest route from each zone to every other zone. The set of all such routes via a given mode emanating from a given centroid node is called the "minimum time tree" for that node and mode.

During one pass of the tree generation block, up to three sets of routes may be found: one set for each of the travel modes (V, Q, and VQ or QV) described later. A separate pass of the tree generation block is necessary to calculate a set of truck routes, should this be desired.

One pass of this block is not capable of determining more than one set of routes for any one mode; if the analyst wishes to have more than one, say, vehicle route in use between any O-D pair, he must determine the second, third, etc., vehicle routes in subsequent passes of the tree generation block, based on different link travel times. These different link travel times are determined by subsequent passes of block 6 as previously described.

There are four different types of trees generated by this block:

1. **Vehicle (V) Trees.**—These are trees in which all routes are via private automobile only.
2. **Transit (Q) Trees.**—These are trees in which all routes are via transit only. In this context transit includes buses, trolley buses, streetcars, subways, elevated trains, and commuter trains.
3. **Mixed (VQ or QV) Trees.**—These are trees in which all routes are via a combination of private automobile and transit. The type of travel mode that these routes are designed to simulate is that of the suburbanite who drives his car in the morning to the local commuter train station, parks it there, and takes the train into the city center. He may or may not take some other form of transit from the downtown station to his office. This type of mixed route is called a VQ route because the private automobile portion of it takes place first and the transit portion last. The return trip in which urban traveler

takes a commuter train from the city center to a suburban station and thence an automobile to his home, requires a QV route for its simulation. In any urban region there is a daily ebb and flow of trips to and from the city center as workers carry out their daily duties. In considering mixed mode trips, this means that VQ trips will predominate in the morning rush hour and QV trips will predominate in the evening rush hour. Because this traffic prediction model has been designed to deal mainly with rush-hour conditions, two versions of the model have been created: one to handle the AM rush hour and one to handle the PM rush hour. In the AM program, VQ trees are the only type of mixed tree generated; in the PM program, QV trees are the only type of mixed tree generated.

4. **Truck (T) Trees.**—These are trees in which routes are via truck only; that is, they follow roads on which trucks of the particular class in question are allowed to travel.

The second function of the tree generation block is to update trees that have been found by previous passes of the tree generation block. Because these trees were found under different conditions of link travel times, it is necessary to trace each previously generated tree, substituting current link travel times for old ones, to determine the current route travel time pertaining to each route in these trees.

The program generates other sets of routes in subsequent iterations. A given set of routes tends, therefore, to avoid areas that are congested at the time it is generated. Because up to nine routes for any O-D pair can be retained, this allows travelers from an origin to a destination a reasonable choice to follow. This is shown in Figure 8, where four alternative auto routes are shown for one O-D pair: the first route follows a roundabout course to utilize 60 mph expressways for most of its length; route 2 also makes use of an expressway, and routes 3 and 4 are forced by expressway congestion to use slower arterial roads.

Time Factor Block

Given the travel time from each origin to each destination via each route generated between them, together with other pertinent land-use and trip behavior data, the time factor block is capable of calculating 3 sets of factors.

Time Factors.—The time factor between an origin and a destination describes the effect that travel time between them has on the propensity of travelers to travel from the origin to the destination. This factor is calculated by means of a negative exponential function:

$$TF = e^{-t} \quad (14)$$

in which

TF = time factor;

t = weighted average travel time via all route-modes available between origin and destination;

β = time factor exponent, determined empirically.

In general, β differs for different trip purposes; for instance people seem to be willing to travel further to work than to shop and this is reflected in a smaller value of β for work trips than for shopping trips. The time factors calculated in the time factor block are used in the trip distribution, therefore, new time factors are calculated only when a new trip distribution is desired.

Modal Split Factors.—MSFQ, the transit modal split factor, indicates what proportion of the total trips going from an origin to a destination will make the trip via public transit (Q) as opposed to private vehicle (V). MSFV, the vehicle modal split factor, indicates what proportion will travel by private vehicle. Because travelers treated in this model must take one or another of these travel modes (mixed trips, the VQ and QV trips already described are defined as transit trips in this context), it can be seen that $MSFQ + MSFV = 1$ for each O-D pair. The modal split factors are used in block 4

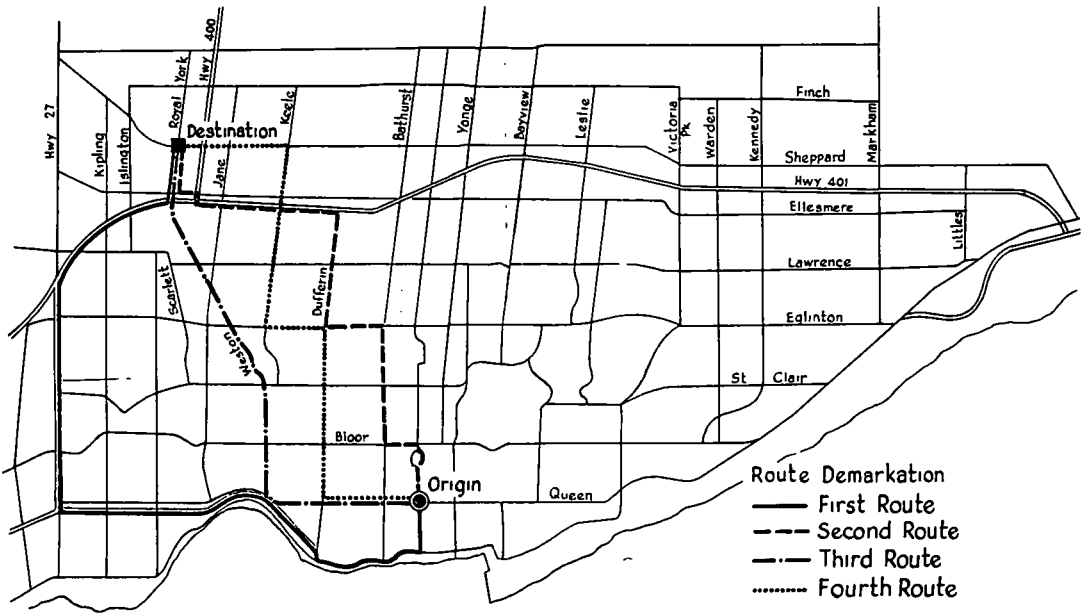


Figure 8. Four routes from a CBD origin.

where they and the assignment factors are multiplied by the total trip interchanges to provide the split trips for each O-D pair. They need be determined only when both modes of travel are available within the model; that is, runs in which only private vehicle flow is being studied would, of course, omit calculation of the modal split factors. A description of the modal split calculation has been reported elsewhere (30, 31).

Observations show the factors influencing choice of travel mode is strongly interdependent. One example of this is the observation that relative travel times via transit and automobile do not influence the choice of low-income travelers as much as they do that of high-income travelers. Another example is that relative comfort and convenience are more important in the traveler's eye if travel costs are roughly equal than they are if one mode costs much more than the other. In short, the effect of one factor depends on the strength of all the other factors.

Based on observed data, diversion curves have been produced showing percentage transit usage as a function of each of the important factors influencing choice. Contours have been developed for each of these curves to show how it changes in shape as the other factors are varied. Each contour of each curve has been reduced to tabular form and fed into the computer memory. Equations have been derived to describe some of the interaction between factors. To calculate the modal split for any O-D pair, the factors (time, cost, convenience, income) applicable to that O-D are determined. Some of these are modified by equations linking them to the values of others. Then the correct tables are entered into in turn, using as argument the modified factor from the previous step, until the modal split factor is determined.

Assignment Factors.—(AF)₁, the first assignment factor, indicates what proportion of the trips going from an origin to a destination via a particular mode will travel via the first route available between that origin and destination for the mode in question. For example, if it is assumed that a given prediction run has made 3 V routes, 2 Q routes, and 1 QV route available between each O-D pair, then (AF)₁, (AF)₂, and (AF)₃ for a given O-D pair would indicate what proportion of the private vehicle travelers going from the origin to the destination would use, respectively, each of the 3 vehicle routes available for that O-D pair, and (AF)₄, (AF)₅, and (AF)₆ would indicate what proportion of the public transit travelers would use the first Q route, the second Q route, and the QV route, respectively, for the O-D pair in question. It can be seen that assignment

factors are required only for a mode that has two or more routes available for any O-D pair; for such a mode there is an assignment factor for each route to specify the proportion of travelers within that mode using the route. However, in the initial stages of the program an arbitrary assignment factor can be used even though there may be just one route available for the mode in question. The assignment factors are used in block 4 together with the modal split factors, to determine the split trips from each origin to each destination in the area under study. Assignment factors are calculated in block 2, using the route travel times for each O-D pair obtained from block 1, by means of

$$(AF)_1 = \frac{\left(\frac{1}{T_1}\right)^{a(V)}}{\left(\frac{1}{T_1}\right)^{a(V)} + \left(\frac{1}{T_2}\right)^{a(V)} + \dots + \left(\frac{1}{T_n}\right)^{a(V)}} \quad (15)$$

in which

$(AF)_1$ = assignment factor for route 1 (specifying what percentage of private vehicle travelers are using the first vehicle route for the O-D in question);

T_n = travel time via nth route from origin to destination (there is a total of n routes for the O-D pair in question);

$a(V)$ = assignment factor exponent for vehicles which is empirically determined and specified by the analyst.

At present this proportional assignment is carried out with $a(V)$ set equal to 1. Further investigation is proceeding to determine a more representative value, if necessary.

It can be seen that the proportional assignment is another means by which capacity restraints are taken into account in this prediction model. Although the shortest route from an origin to a destination will have the shortest travel time under non-loaded conditions, its popularity may lead to traffic congestion which increases its travel time to a higher value than that for the other available routes. The proportional assignment allows this effect to be simulated in a reasonable manner.

For determining assignment factors within the transit mode, $a(Q)$ would replace $a(V)$ in Eq. 15. For a given mode the sum of the calculated assignment factors for a given O-D pair is always equal to 1; i.e., if there are four V routes and two Q routes available for each O-D pair, then for each O-D pair, for the V mode,

$$(AF)_1 + (AF)_2 + (AF)_3 + (AF)_4 = 1 \quad (16)$$

and for the Q mode,

$$(AF)_5 + (AF)_6 = 1 \quad (17)$$

Trip Distribution

Given the attractor and generator corresponding to each centroid node for each trip purpose and given the time factor for each O-D pair for each purpose, the trip distribution block determines the total trips via all modes and for all purposes combined from each origin to each destination. These trips are first determined separately for each purpose and then added together to provide the "total trip interchange" for each O-D pair.

The number of trips between any two points for a particular purpose is dependent on the total number of trips generated for distribution at the origin for that purpose, G_i , the total number of trips attracted to the destination for the same purpose, A_j , and the time factor, $(TF)_{ij}$, describing the "friction" between the origin and destination for the particular purpose in question.

The following is used to calculate trip interchange volumes:

$$J_{ij} = G_i A_j (TF)_{ij} \quad (18)$$

in which

- J_{ij} = number of trips going from origin i to destination j for purpose in question;
 G_i = total trips generated for this purpose at origin i ;
 A_j = total trips attracted for this purpose at destination j ; and
 $(TF)_{ij}$ = time factor for trip between origin i and destination j for this purpose.

This equation is the well-known "gravity formula," so called because of its similarity to the equation derived by Newton to describe gravitational attraction between two masses.

There are two basic differences, however. One is that Newton's equation replaces $(TF)_{ij}$ by $1/r_{ij}^2$, in which r_{ij} is the distance between the two masses. As previously described, the trip interchange formula used in this model employs instead a negative exponential function of the travel time, because this describes best the observed trip behavior of urban travelers.

The other difference is more fundamental. If Eq. 18 is applied to a city in which all zones are equally spaced and have equal generators and equal attractors, and in which there are no edge effects, then the trip interchanges so calculated for each trip purpose would be such that

$$\sum_{j=1}^{NUHN} J_{ij} = G_i \quad (19)$$

and

$$\sum_{i=1}^{NUHN} J_{ij} = A_j \quad (20)$$

That is, the sum of the trips leaving each zone would equal the generator at that zone and the sum of the trips arriving at each zone would equal the attractor at that zone.

However, real cities are not homogeneous as to zone size and spacing, and some zones are always on the edge rather than in the middle, so that there are many zones for which conditions expressed by Eqs. 19 and 20 are not met if Eq. 18 is used. For example, a zone closely surrounded by many large generator zones would probably tend to receive more trips than warranted by the size of its attractor. The fact that the attractor (the number of trips that can actually be received at the zone) is smaller than the arrivals that the unmodified gravity model would lead to it is due to factors that the gravity model does not attempt to take into account. These could be space or accommodation limitations in the zone, or the fact that wages in the zone have been driven down by competition among the large number of workers who live in neighboring zones and wish to work there.

Rather than trying to take these factors into account explicitly, the gravity formula is modified on an empirical basis to match the departures with the generator and the arrivals with the attractor for each trip purpose at each zone. This is done by the following repetitive process.

Eq. 21 is used in a first pass of the distribution algorithm (a method of calculation following a systematic set of rules; the term often being used to describe specific calculation routines used in a computer program) to obtain the first adjusted generators

$$G_i^{(1)} = \frac{G_i^{(0)}}{\sum_{j=1}^{NUHN} A_j^{(0)} (TF)_{ij}} \quad (21)$$

During this first pass of the distribution algorithm, the generator and attractor figures are further adjusted for each zone by means of

$$A_j^{(1)} = \frac{A_j^{(0)}}{\sum_{i=1}^{NUHN} G_i^{(1)} (TF)_{ij}} \quad (22)$$

and

$$G_i^{(2)} = \frac{G_i^{(0)}}{\sum_{j=1}^{NUHN} A_j^{(1)} (TF)_{ij}} \quad (23)$$

in which the superscript (0) refers to the unadjusted value of G or A, and superscripts (1) and (2) refer to adjusted values produced by the first pass of the distribution algorithm.

If after this first iteration of the distribution algorithm it is felt that matching is still insufficient as determined by the convergence criterion, a second iteration is carried out, (Eq. 26), during which the generators and attractors are adjusted again. Successive iterations can be carried out as many times as necessary to achieve matching of desired accuracy.

At the end of this iterative procedure the generators are adjusted (n + 1) times, whereas the attractor is adjusted only n times. This ensures that the departures will exactly match the generator at each zone, and the arrivals will approximately match each attractor with an accuracy depending on the number of distribution algorithm iterations carried out.

The adjusted generators and attractors produced by the nth iteration of the distribution algorithm are given by the following generalized forms of Eqs. 22 and 23:

$$A_j^{(n)} = \frac{A_j^{(0)}}{\sum_{i=1}^{NUHN} G_i^{(n)} (TF)_{ij}} \quad (24)$$

and

$$G_i^{(n+1)} = \frac{G_i^{(0)}}{\sum_{j=1}^{NUHN} A_j^{(n)} (TF)_{ij}} \quad (25)$$

in which the superscripts $(n + 1)$ and (n) refer to adjusted values of G or A produced by the n th iteration. The table of adjusted generators and attractors, $G_i^{(n+1)}$ and $A_j^{(n)}$, produced by the last iteration of the trip distribution algorithm for each trip purpose, is written on tape by the trip distribution block and can be used as input for a subsequent run of block 3 if desired. This could be done to save calculation time in the subsequent block 3 run if congestion patterns in the study area have not changed appreciably during the intervening cycle.

Use of this trip distribution algorithm has shown that two iterations (i.e., the initial pass plus one repetition) are usually enough to match arrivals and attractors to within 5 percent for the majority of zones in the study area.

There are two criteria by which the analyst can control the accuracy with which arrivals will be matched to attractors for each trip purpose. First, he can specify the value of ϵ (also called EPSI) below which a mathematical expression called the "epsilon convergence criterion" must drop before he will be satisfied. That is, he gives ϵ a value such that as soon as the condition

$$\frac{\sum_{j=1}^{NUHN} \left[1 - \frac{A_j^{(n)}}{A_j^{(n-1)}} \right]^2}{NUHN} \leq \epsilon \quad (26)$$

has been met the desired degree of accuracy will have been reached. Examination of Eq. 26 shows that the epsilon convergence criterion approaches zero as the arrivals match the attractor at more and more zones because the ratio $A_j^{(n)}/A_j^{(n-1)} = 1$ for each j at which such matching has occurred.

Second, he can specify NUIT, the maximum number of iterations of the distribution algorithm to be allowed for the trip purpose in question.

Both NUIT and EPSI are specified by the analyst for each trip purpose. As soon as either criterion is met for a particular purpose, no further iterations are carried out for that purpose. When the adjusted generators and attractors have been so calculated for each trip purpose, the trip interchanges for each O-D pair are calculated for each purpose by means of Eq. 18 and then summed over all purposes to produce the total trip interchange volumes for each O-D pair.

Proportional Split

Given the modal split factors and assignment factors from block 2 and the total trip interchange for each O-D pair from block 3, the proportional split block calculates the number of person trips that will proceed via each route and each mode for each O-D pair in the area under study. These split trips are obtained quite simply: the trips proceeding from an origin to a destination via a given route and mode are calculated by multiplying the total trip interchange for the O-D in question by the modal split factor and assignment factor pertaining to the mode and route in question for the given O-D pair.

The split trips (that is, the number of person-trips that will proceed via each route and each mode for each O-D pair) are then used as input for the assignment block to calculate the passenger and vehicle loads on the various links in the area under study.

Assignment

Given trees describing a number of routes from each origin to each destination, and given the trips that correspond to each of these routes, the assignment block traces each tree and assigns the trips using each route to the links comprising it. The different "bundles" of trips using each link (resulting from the different routes that traverse that link) are summed for each link to give the flow of traffic along each link via each mode, known as the link roads.

A given pass of the assignment block can handle either all non-truck trips or truck trips only. It is not possible for one pass of this block to handle both non-truck trips and truck trips.

Up to this point in a program cycle, trips via all non-truck modes (i.e., V, Q, VQ, or QV) have been handled in terms of person-trips; that is, in terms of people per hour traveling respectively by private vehicle, transit vehicle, and commuter train from zone to zone. The main function of the assignment block is to translate interzone trip volumes via the various routes and modes into link flows via the various modes. For transit and commuter trips it is useful to have these link flows expressed in terms of people per hour; however, for private vehicles the authors are more interested in the number of vehicles per hour traversing a given link than in the number of people per hour traversing the link by car. The flow in vehicles per hour is needed, for example, to estimate the amount of congestion existing on each link and to calculate the average speed and travel time that result from the vehicle flows.

Consequently, a second function of the assignment block is to translate person-trips via automobile into automobile trips. If the truck assignment option is used, there is no need to translate person-trips into vehicle trips, because truck trips are generated, distributed, and assigned in terms of trucks per hour.

A third function is to calculate the parking cost for automobiles in each zone as a function of parking supply and demand. For this purpose the computer is fed a table showing the number of parking places available in each zone and another table relating the cost of parking to the present utilization of these parking places in each zone.

The link loads calculated by the assignment block are output in so-called load tables. There is one load table for each basic mode: V, Q, and T. These load tables can be added to produce a composite load table as output from that pass of block 5, and input for block 6.

Link Updating

The purpose of the link updating block is to determine the link travel times, via any or all modes, for the flow conditions (link loads) obtained in block 5. In addition, when desired, the block will punch out link data (i.e., flows, speeds, travel times, and link characteristics). The input data required are car flows, transit vehicle flows and travel times, truck flows, and transit person trips on each link; link data; and inter-modal information. The general methods of calculation used in this block are described previously; further details are explained here.

Special provision is made for overloaded links; i.e., for those on which the flow on a link exceeds the maximum capacity. As in the case of other links, the value of the maximum flow difference $f_e(V) - f_m$ is, in fact, punched out, to show the analyst that the link in question is overloaded; that is, that the demand flow of equivalent vehicles per lane, $f_e(V)$, is greater than the maximum flow capacity of equivalent vehicles per lane, f_m .

Clearly, for such overloaded links, the throughput actually being achieved is

$$F_e(V)_m = f_m(\text{NULA}) \quad (27)$$

in which $F_e(V)_m$ is the maximum possible flow of equivalent vehicles which the link can accommodate, this being equal to the maximum possible flow of equivalent vehicles per lane, f_m , multiplied by the number of lanes, NULA.

The excess of demand flow over maximum possible flow, $[F_e(V) - F_e(V)_m]$, represents road users who will have to wait in a queue and will not actually negotiate the link until sometime during the succeeding hour. Because $F_e(V)_m$ is the actual flow of equivalent vehicles on an overloaded link, the actual flow of vehicles (automobiles only) on an overloaded link, will be

$$F(V)_m = F_e(V)_m - F_e(Q) - F_e(T) \quad (28)$$

in which $F_e(Q)$ and $F_e(T)$ are the flows of transit vehicles and of trucks, respectively, in terms of equivalent vehicles, as described.

It is this value, $F(V)_m$, which is punched out as the vehicle flow on overloaded links. It should be emphasized, however, that the demand vehicle flow $F(V)$, is always put into the load table for such links, and that the link time for overloaded links is always calculated on the basis of $F_e(V)$ rather than $F_e(V)_m$.

AVERAGE TRANSIT TRAVEL TIME $T(Q)$

As mentioned, there are three types of transit link considered in this program. Types 2 and 3 (subway and commuter train links, respectively) are unimpeded by automobile flow and are, therefore, able to operate to given schedules, having fixed average travel speeds and fixed time headways between successive trains. It is, therefore, possible to list the average per unit transit travel time, $t(Q)$, for each type 2 and type 3 transit link as an input parameter that is unchanged during a given prediction. This parameter is listed in the link table.

However, type 1 transit links represent lines on which the transit vehicles run on roads and, therefore, proceed at an average speed that is strongly dependent on the degree of automobile traffic congestion in existence; i.e., on the average speed of the automobiles using the same road. It is, therefore, necessary to calculate the average travel time of transit vehicles on type 1 transit links as a function of the average automobile travel times on the corresponding vehicle links. This is done by means of the transit travel time table, which lists a value of the transit travel time for each of 60 per unit auto travel times (covering the range $t(V) = 1$ min per mi to $t(V) = 60$ min per mi) and is calculated on the basis of transit vehicle acceleration and deceleration rates, loading and unloading times, and interstop distances, in the area under study. The type of relationship described by the transit travel time table is shown in Figure 9.

In general, the maximum inter-stop speed achieved by surface transit vehicles cannot exceed that of the surrounding vehicle flow. Because the transit vehicle is further slowed down by having to stop for passengers, its travel time is therefore higher than that of surrounding vehicles. If transit vehicles had infinite deceleration and acceleration rates and could load passengers instantaneously, they would go as fast as vehicles, and the dotted line in Figure 9 would describe the relationship between $t(V)$ and $t(Q)$. However, the sort of relationship found in practice is shown by the solid line, which shows $t(Q)$ always slightly greater than $t(V)$. The value of $t(Q)$ corresponding to each pertinent value of $t(V)$, as shown by this curve, is listed in the transit travel time table.

AVERAGE TRUCK TRAVEL TIME $T(T)$

The acceleration rate of most medium sized and heavy trucks is in general less than that of private vehicles, which results in a different average speed for trucks and is accounted for by a truck travel time parameter. This can be found from

$$TTTP = \frac{t(T)}{t(V)} \quad (29)$$

in which $t(T)/t(V)$, the ratio of truck travel time per mile to private vehicle travel time per mile, is measured under typical urban conditions for the type of truck being considered in the study. Because trucks travel more slowly than cars, the TTTP is greater than one; it is used in the program to factor up the average travel time on each link for the truck grid.

LINK TRAVEL SPEEDS

For each link, the travel speed via a given mode is the average speed in miles

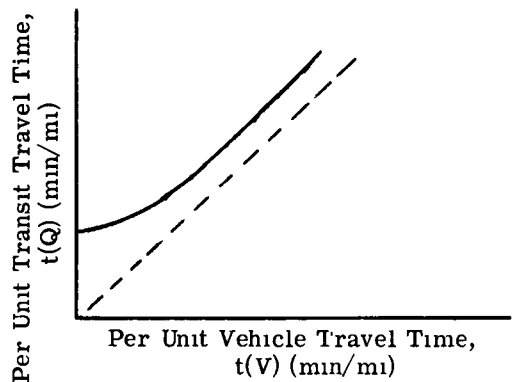


Figure 9. Transit travel time as a function of vehicle travel time.

per hour at which travelers using that mode will traverse the link in question under prevailing flow conditions. These travel speeds are calculated for each link in block 6 from the pertinent travel times using the following equations:

For cars,

$$S(V) = \frac{60}{t(V)} \text{ (mph)} \quad (30)$$

For transit,

$$S(Q) = \frac{60}{t(Q)} \text{ (mph)} \quad (31)$$

For trucks,

$$S(T) = \frac{60L}{T(T)} \text{ (mph)} \quad (32)$$

in which

$S(V)$, $S(Q)$, $S(T)$ = link travel speeds via car, transit, and truck, respectively (mph);

$t(V)$, $t(Q)$ = per unit travel time via car and transit, respectively (minutes per mile);

$T(T)$ = truck travel time (minutes); and

L = link length (miles).

NUMBER OF TRANSIT TRAVELERS PER TRANSIT VEHICLE (NPPQ)

The number of transit travelers per transit vehicle (NPPQ) is of interest to the transportation planner, because it indicates what degree of compatibility exists between the demand for transit facilities along a given route and the number of seats per hour which are being provided on that route. For each transit link, NPPQ is calculated in block 6 by means of

$$NPPQ = \frac{P(Q)}{F(Q)} \quad (33)$$

in which

$P(Q)$ = flow of transit passengers (people per hour); and

$F(Q)$ = flow of transit vehicles (vehicles per hour) on link in question.

SYSTEM TIME COST

A useful yardstick for comparing alternative proposed transportation systems that are being tested by the traffic prediction model is the system time cost of both systems. The system time cost is the number of hours spent traveling by all system users within the time period under study. For instance, if one proposed system had a system time cost of 50,000 person-hours and another had a system time cost of 60,000 person-hours, both systems serving the same number of trips in, say, an evening rush hour, then it is probable that the first system is the more desirable one. Of course, many other considerations must be weighed in comparing one proposed transportation system with another; however, the system time cost provides an over-all quantitative measure by which the user-benefits of proposed systems can be estimated.

The system time cost is also useful in indicating the degree of convergence reached by the traffic prediction model during a prediction run. If, for example, the system time costs for the various travel modes remain stable during two or more successive iterations of block 6, this is a good indication that equilibrium has been reached by the model. A separate system time cost is calculated for each travel mode, as follows:

For autos,

$$SCTV = \sum_{\text{all V links}} \frac{F(V) \times T(V)}{60} \text{ vehicle-hr} \quad (34)$$

For travelers,

$$SCTP = (SCTV) \times (NPPV) \text{ person-hr} \quad (35)$$

For auto travelers,

$$SCTQ = \sum_{\text{all Q links}} \frac{P(Q) \times T(Q)}{60} \text{ person-hr} \quad (36)$$

For trucks,

$$SCTT = \sum_{\text{all T links}} \frac{F(T) \times T(T)}{60} \text{ truck-hr} \quad (37)$$

in which

SCTV, SCTP, SCTQ, SCTT	= system time costs for automobile vehicles, automobile travelers, transit travelers, and truck vehicles, respectively;
F(V), P(Q), F(T)	= link loads of autos (cars per hour), transit travelers (persons per hour), and trucks (trucks per hour), respectively;
T(V), T(Q), T(T)	= link travel times via auto, transit, and truck, respectively (minutes); and
NPPV	= average number of travelers per automobile.

All four of these system time costs are punched out on a card automatically during each pass of block 6.

INTERACTION OF BLOCKS 1 TO 6

Referring to Figure 1, the two basic sets of input information—transportation facilities data and land-use data—are collected and coded to produce grid data for subsequent use in block 1 (for determining trees) and block 6 (for calculating link travel times). Census data are processed by the trip generation block to produce trip generator and attractor figures for subsequent use in block 3 (for calculating interzone trip volumes). Census data are also of direct use in block 2 (for calculating the proportion of trips made by public transportation and by private car between each pair of zones—the so-called "modal split factors") and in block 5 (for calculating vehicle parking costs in each zone). With the preparation and processing of data, the stage is now set for the main part of the traffic prediction program, blocks 1 to 6, to begin functioning.

As shown in Figure 1, the six main program blocks form a sequential loop: when blocks 1 to 6 have been run in that order it is possible to go back to block 1 and start the same sequence again. It is by this means that feedback occurs—the effects of road traffic congestion and parking lot overflow are fed back within the model to affect travel patterns as they do in actual cities.

Many of the cycles carried out in a run of the traffic prediction program will not

contain all six program blocks. Possible block sequences in a given cycle are shown in Figure 1. The figure shows the sequential loop formed by blocks 1 to 6 can be short-circuited as follows: block 3 can be eliminated by going directly from block 2 to block 4; blocks 3 and 4 can be eliminated by going directly from block 2 to block 5; and block 4 can be eliminated by going directly from block 3 to block 5.

The meaning of these operations becomes clearer if one realizes the purpose of each block. For example, the main purpose of the first few cycles is to produce the desired number of alternative routes between each O-D pair. Experience has shown that reasonable routes can be obtained based on arbitrarily estimated modal split factors and assignment factors, and that it is not necessary to carry out a new trip distribution during each of these route-generating cycles. Consequently, blocks 3 and 4 can be omitted from most of these cycles with a consequent saving of machine time. Similarly, in the final "settling" cycles when equilibrium is being reached, no new routes are being generated, so it is possible to leave part of block 1 out of each cycle. Experience has also shown that the interzone trip volume figures reach equilibrium before the link loads and travel time have completely settled down so that it is possible to "freeze" the trip distribution and leave block 3 out of the final few cycles.

The flexibility resulting from these alternative sequences is enhanced by a comparable flexibility of input information. For instance, if estimates have already been made of interzone travel volumes (by means of a scaled-up O-D survey, say), it is possible to feed these volumes directly into the model, by-pass the time factor, distribution, and proportional split blocks and find the assigned traffic flows that would result. This procedure would result in a saving of computing time but would be useful only for short-term predictions; over long periods of time, patterns of land use and traffic congestion in the area could be expected to change considerably, therefore requiring a method of estimating interzone travel volumes which will take these things into account. Blocks 2, 3, and 4 in this model have been designed to do so.

The modal split factor percentages, route assignment factor percentages, interzone trip volumes, link flows, and link travel times produced by the final cycle of a given run describe the predicted traffic pattern for the study area and the time period in question, based on the specified land-use patterns and transportation facilities.

Having produced this traffic prediction, it is now possible by inspection to determine which links are most heavily overloaded, which areas are least efficiently served by roads and rail, etc. with a view to proposing new land-use configurations and/or new transportation facilities. Having made these proposals the planner can then make use of the model again, with the new proposals as input data, to test them for efficiency of operation and to compare them with the original proposals. By this means he is able to make planning decisions based on systematic appraisal of the various proposed combinations of land use and transportation facilities.

Various other sorts of information describing the predicted traffic patterns can be produced by the model to aid in planning decisions. One such item, as mentioned previously, is the total time spent traveling by all tri-makers during the time period in question, the so-called "system time cost," which is usually measured in terms of person-hours and vehicle-hours spent on route. Another such item, called "link usage data" is a list, for any link desired, of the number of travelers from each zone who are traversing that link during the time period in question. Such data can be very useful in establishing the relative importance of various links. Another useful set of information shows the turning movements at each interchange along a given facility and the trip lengths along the facility of all travelers entering at each interchange. This information is useful in deciding at what intervals interchanges should be located to serve adequately the adjacent corridor and yet prevent congestion from building up as a result of too much local traffic.

EFFECTS OF CAPACITY RESTRAINTS

The effects of capacity restraints on travel time make themselves felt at four points in this prediction model:

1. In finding of routes. Route generations are carried out under differing conditions of congestion to provide several reasonable routes from every origin to every destination.

2. In choice of destinations. Trip distributions are carried out under prevailing traffic patterns to simulate the effect of congestion on travelers' choice of destinations.

3. In choice of route. Confronted with several possible routes from an origin to a destination simulated travelers are allowed to choose among them so that more take the shortest route than take the longest.

4. In choice of travel mode. The ratio of travel time by car and by transit is the most important factor affecting modal split.

Many of the relationships can be improved given more accurate source information. Nevertheless, the prediction model in its present stage is capable of producing meaningful results as is shown in a test run of the 1956 traffic pattern in Metropolitan Toronto.

Under contract with Metropolitan Toronto Planning Board the Traffic Research Corporation has used the traffic prediction model in a study of the Toronto area (Fig. 10) for the time period of a morning rush hour in 1956.

The year 1956 was chosen for three main reasons:

1. For the year 1956, most comprehensive source information was readily available.

2. The relationships developed and incorporated in this model had to be tested to establish their validity by reproducing a situation that could be compared with observed data.

3. The sequence of operating the various program blocks had to be established in such a way that minimum amount of time was spent in estimating future traffic flows for which no other check is available.

In the following graphs it can be seen that the traffic prediction model reproduced the 1956 traffic situation in close agreement with the observed data for the same period of the day.

Figure 11 shows the system time cost as previously described. In the first iteration of the program, the system time cost has been recorded for the first route time multiplied by the interchange volume for each O-D pair. Because at this stage the roads are not used by any road-user, the system time cost represents the total time of all road-users if the road-user would be restricted in his movements by speed limitations only. At the end of the first iteration only one vehicle route and one transit route were available; consequently, the system time cost reached its highest level. As more routes are generated in the subsequent iterations the system time cost decreased to a level where no further decrease could be gained by new route generation. The slight oscillation of the system time cost curve in the second half of the prediction run is noteworthy. As mentioned earlier, interzone trip volumes reach equilibrium before the link loads and travel time have settled down.

Figure 12 shows the behavior of the frequency distribution of the trip length in terms of time. Here again, reasonable agreement has been achieved between estimated and observed data. In the initial stage of the prediction run, all trips estimated are temporarily shorter than those observed: i.e., the capacity restraints of traffic facilities have not made themselves felt. To demonstrate the inherent oscillation as shown in the system time cost an intermediate frequency curve has been shown in this graph. At equilibrium, the estimated frequency curve tends to follow closely the curve for observed data.

Figures 13, 14, and 15 show the volumes of vehicles and transit passengers crossing a cordon line west of the CBD of Toronto. The total volumes estimated agree very well with those observed, although some of the individual road sections vary considerably. In general, it can be observed that the variation is relatively large for small volumes. In Figure 14 the deviation of the estimated volume on Lakeshore Boulevard is caused by an error in coding the number of lanes available on that road which was not discovered before the run was completed. In Figure 13 the deviation of the estimated volume on Davenport, St. Clair, etc., is caused by an overestimate of the level of service the Toronto Transit Commission had in operation in 1956.

Figures 16 and 17 show the volumes of vehicles and transit passengers crossing a

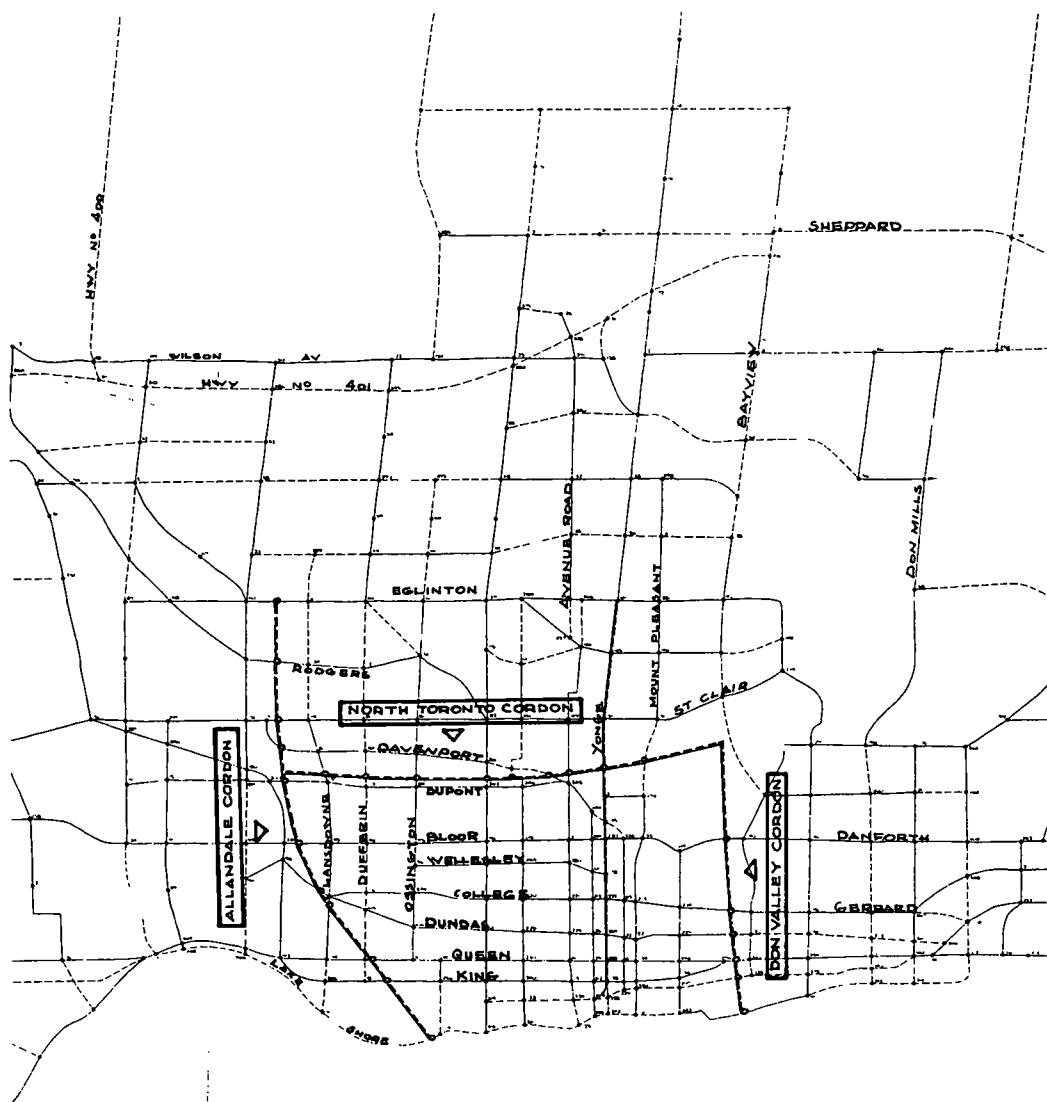


Figure 10. Transportation facility grid, Toronto, 1956.

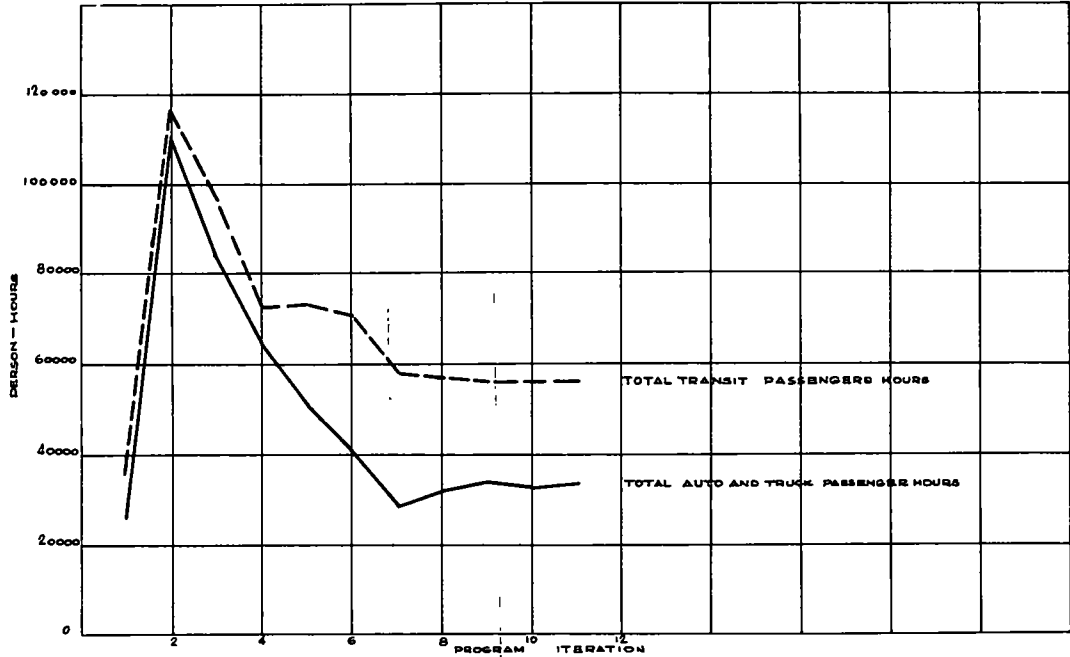


Figure 11. System cost chart for total time on roads.

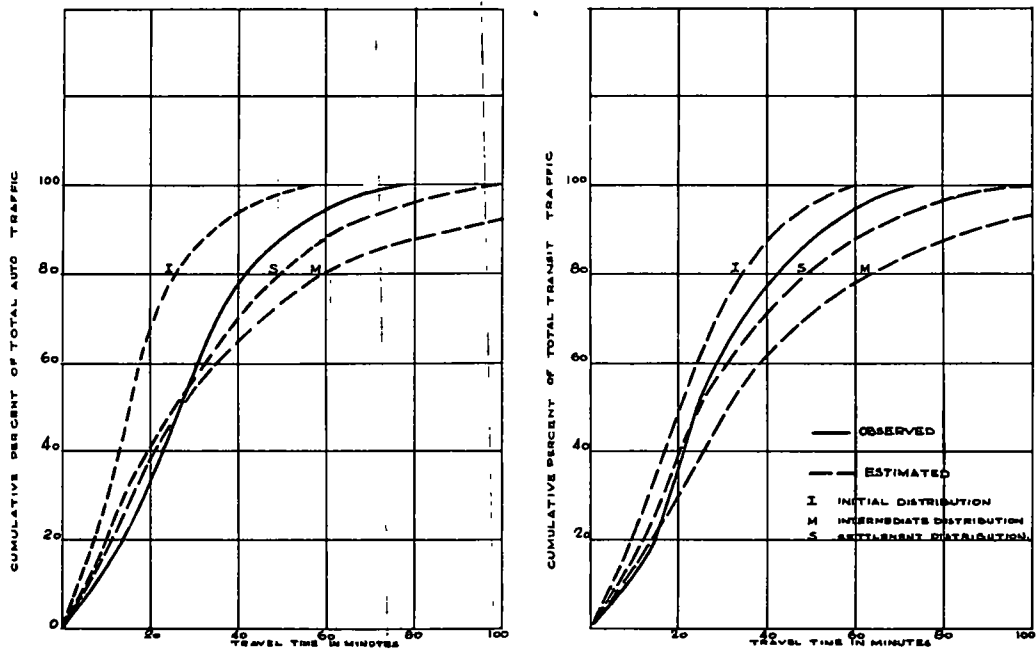


Figure 12. Trip distribution by travel time.

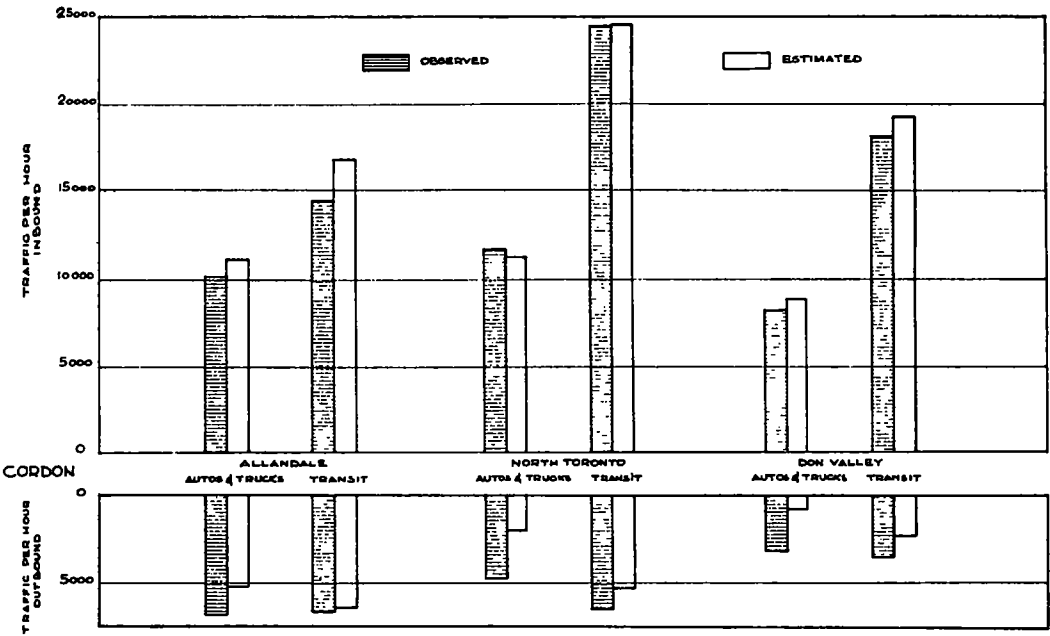


Figure 13. Total auto, truck, and transit passengers crossing cordon lines during AM rush hour, 1956.

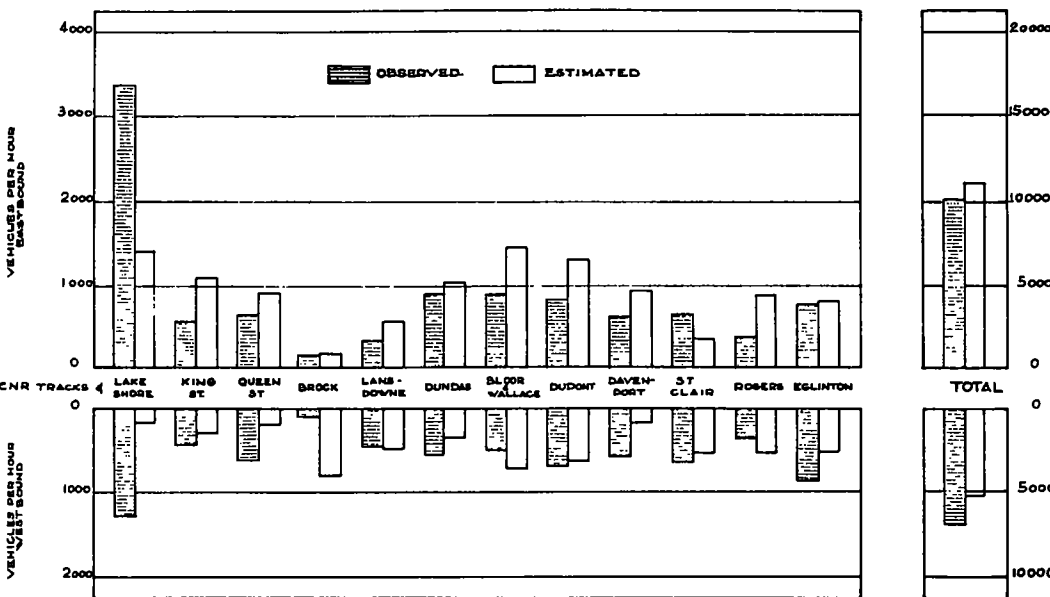


Figure 14. Automobile and truck flows crossing Allandale cordon during AM rush hour, 1956.

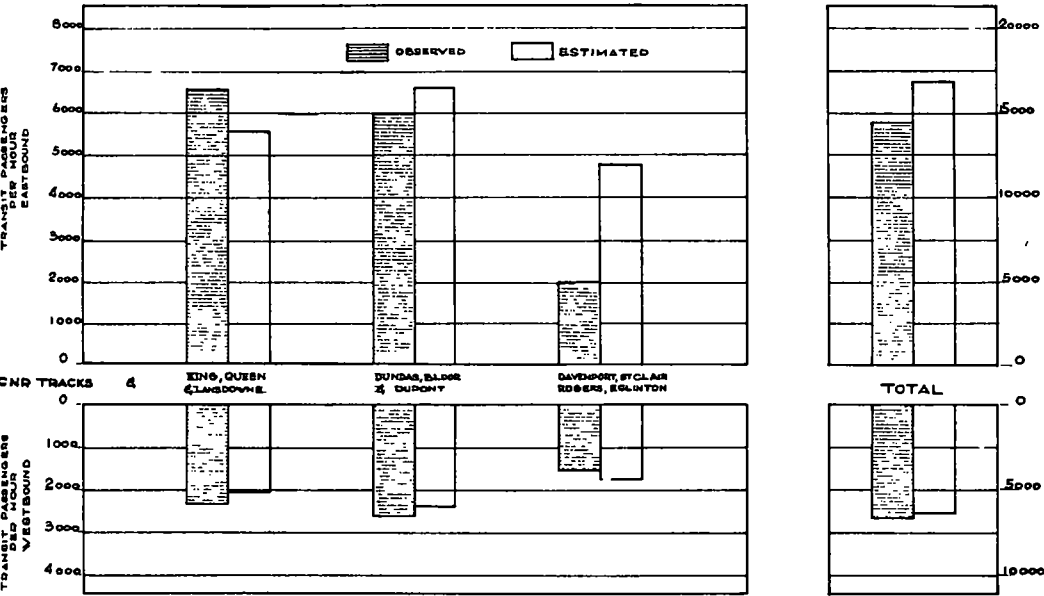


Figure 15. Transit passenger flow crossing Allandale cordon during AM rush hour, 1956.

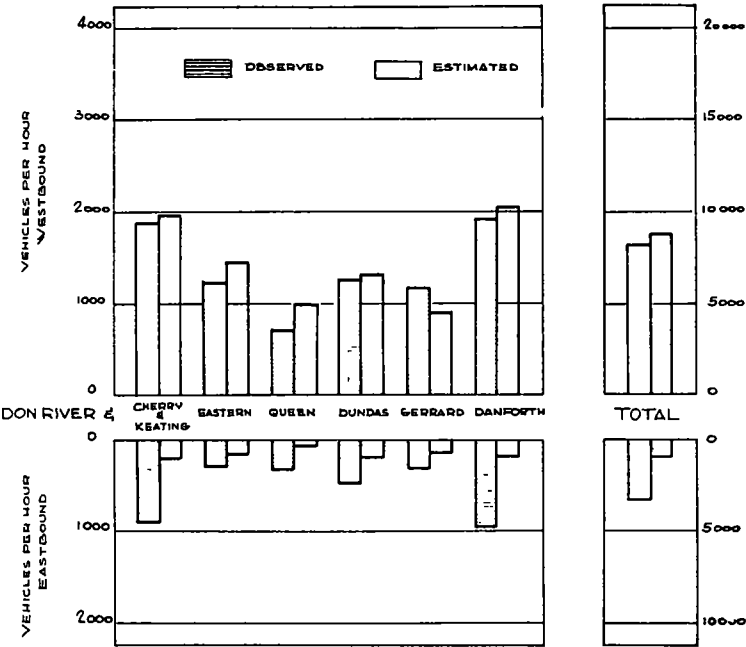


Figure 16. Automobile and truck flow crossing Don Valley cordon during AM rush hour, 1956.

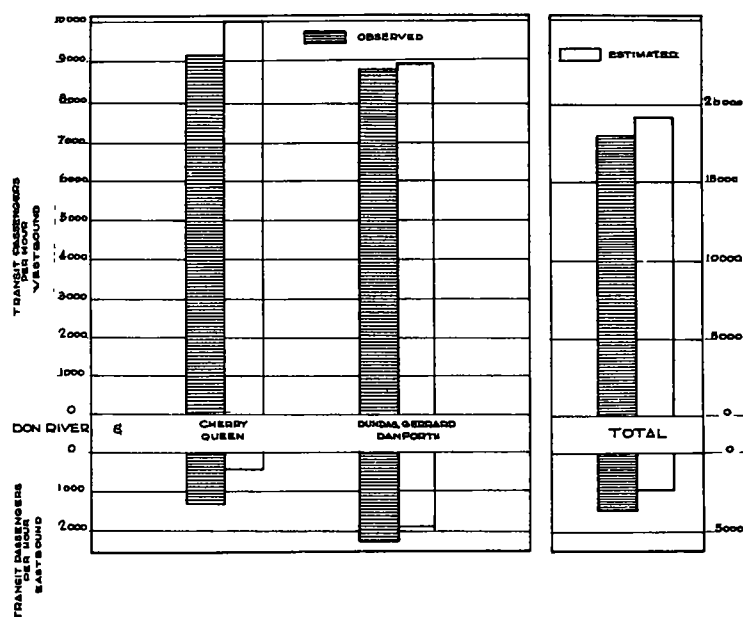


Figure 17. Transit passenger flow crossing Don Valley cordon during AM rush hour, 1956.

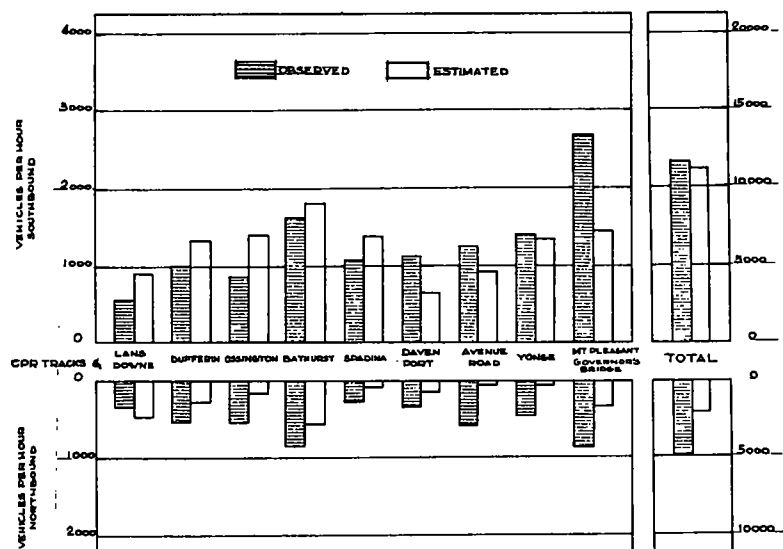


Figure 18. Transit passenger flow crossing North Toronto cordon during AM rush hour, 1956.

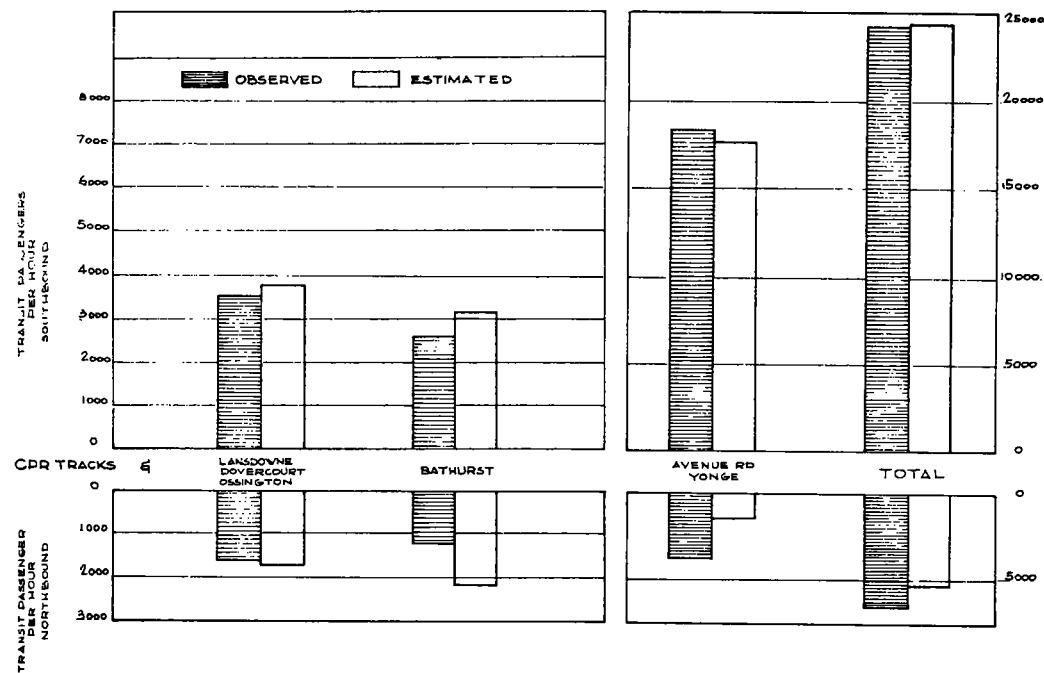


Figure 19. Transit passenger flow crossing North Toronto cordon during AM rush hour, 1956.

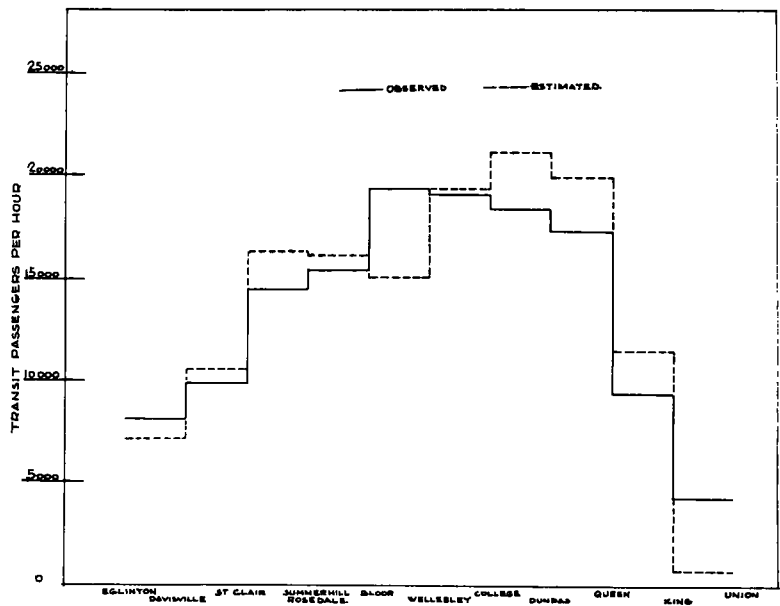


Figure 20. Total passenger flow of Toronto subway, northbound and southbound, during AM rush hour, 1956.

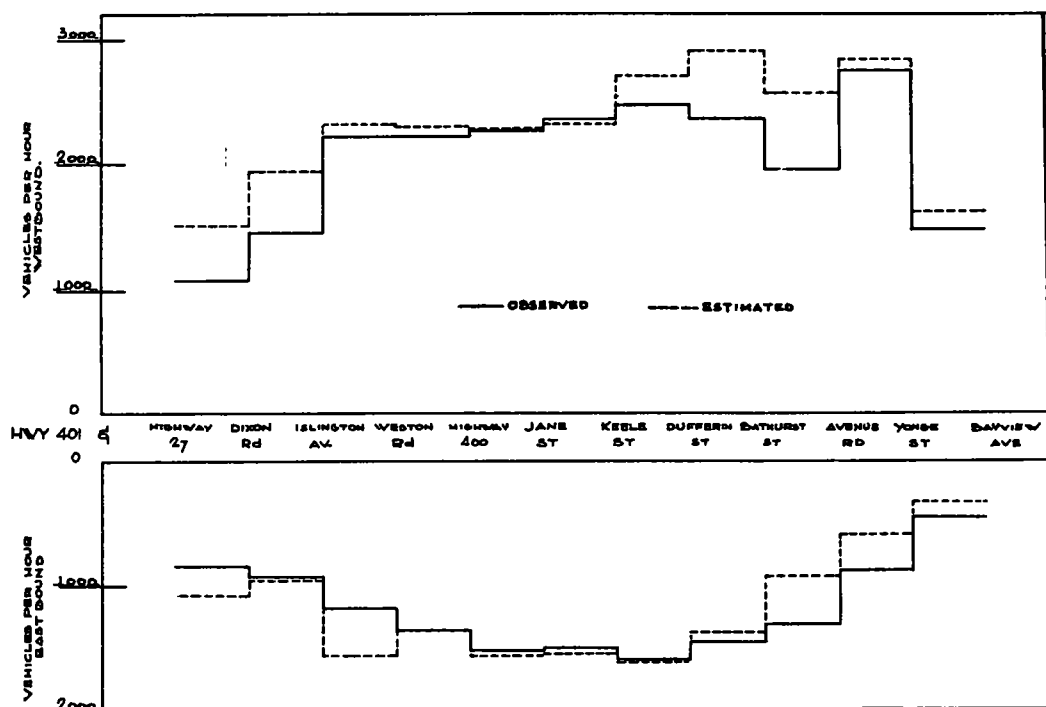


Figure 21. Eastbound and westbound flow of vehicles on Toronto bypass, Highway 401, during AM rush hour, 1956.

cordon line east of the CBD of Toronto. These graphs show excellent agreement for the large volumes inbound. Also, the modal split produced acceptable estimates for the inbound transit volumes. However, the outbound volume was greatly underestimated. This is apparently caused by errors in the input information for trip generation and attraction rather than in the modal split. Further analysis is pending. Figures 18 and 19 show the volumes of vehicles and transit passengers crossing a cordon line north of the CBD of Toronto. This cordon line is also crossed by the Toronto Subway for which the flow of passengers is shown in Figure 20. The discrepancies at Wellesley and Union Station are caused by the zoning of the area under study and could be rectified by subdividing the adjacent zones. These instances clearly show an unexpected sensitivity for coarseness in zoning and coding of street capacities. Figure 21 shows the vehicular flow on the Toronto Bypass, Highway 401, in both directions. The estimates are in reasonable agreement with the observed data.

CONCLUSION

The application of capacity restraints and the resultant feedback in a large-scale traffic forecasting model has given promising results in reproducing historic data. Although further research is necessary to improve the various functions incorporated, the traffic forecasting model can realistically produce the travel behavior of the population in a given transportation system and given land-use plan.

REFERENCES

1. Cherniak, N., "Measuring the Potential Traffic of a Proposed Vehicular Crossing." ASCE Trans. Vol. 106 (1941).
2. "Highway Capacity Manual." U. S. Bureau of Public Roads (1950).
3. Trueblood, D. L., "Effect of Travel Time and Distance on Freeway Usage." HRB Bull. 61, 18-37 (1952).

4. Reilly, W. J., "The Law of Retail Gravitation." Filsburg Publishers, N. Y. (1953).
5. Beckmann, M., McGuire, C. B., and Winsten, C. B., "Studies in the Economics of Transportation." Yale University Press (1956).
6. "Highway Traffic Estimation." Eno Foundation (1956).
7. Dantzig, G. B., "The Shortest Route Problem." *Operations Res.*, 5:270-3 (1957).
8. Moore, E. F., "The Shortest Path Through a Maze." *Internat. Symposium on the Theory of Switching*, Harvard Univ. (1957).
9. Bevis, H. W., "A Model for Predicting Urban Travel Patterns." *Jour. Amer. Inst. of Planners*, 25:No 2, (May 1959).
10. Booth, J., and Morris, R., "Transit vs Auto Travel in the Future." *Jour. Amer. Inst. of Planners*, 25:No. 2 (May 1959).
11. Calland, W. B., "Traffic Forecasting for Freeway Planning." *Jour. Amer. Inst. of Planners*, 25:No. 2 (1959).
12. Voorhees, A. M., "A General Theory of Traffic Movement." *Inst. of Traffic Engineers* (1955).
13. von Cube, H. G., Desjardins, R. J., and Dodd, N., "Assignment of Passengers to Transit Systems." *Traffic Engineering*, 28:No. 11 (Aug. 1958).
14. Griffith, B. A., and von Cube, H. G., "Evaluation of Alternative Subway Routes." *Proc. ASCE, Jour. City Planning Div.* (May 1960).
15. Sagi, G., "Theoretical Traffic Volume and Timing Studies." *Traffic Engineering*, 30:No. 8 (May 1960).
16. Hall, E. M., and George, S., Jr., "Travel Time—An Effective Measure of Congestion and Level of Service." *HRB Proc.*, 38:511-529 (1959).
17. Berry, D. S., "Field Measurement of Delay at Signalized Intersections." *HRB Proc.*, 35:505-522 (1956).
18. Charlesworth, G., and Paisley, J. L., "The Economic Assessment of Returns from Road Works." *Proc., Inst. of Civil Engineers*, Vol. 14 (Nov. 1959).
19. Smeed, R. J., "Theoretical Studies and Operational Research on Traffic and Traffic Congestion." *Bull. Internat. Inst. Statistics*, 36:No. 4 (1958).
20. May, A. D., Jr., "A Friction Concept of Traffic Flow." *HRB Proc.* 38:493-511 (1959).
21. Rothrock, C. A., and Keefer, L. E., "Measurement of Urban Traffic Congestion." *HRB Bull.* 156, 1-13 (1956).
22. May A. D., Jr., and Wagner, F. A., Jr., "Headway Characteristics and Interrelationships of the Fundamental Characteristics of Traffic Flow." *HRB Proc.*, 39:524-547 (1960).
23. Webb, G. M., "Freeway Capacity Study 1955." *Progress Report*, State of California, Department of Public Works, Division of Highways (1955).
24. Trautman, D. L., et al., "Analysis and Simulation of Vehicular Traffic Flow." *Inst. Transportation and Traffic Engineering*, Univ. of Calif., *Research Report* 20 (Dec. 1954).
25. Gardner, E. H., "The Congestion Approach to Rational Programming." *HRB Bull.* 249, 1-22 (1960).
26. Keese, C. J., Pinnell, C., and McCasland, W. R., "A Study of Freeway Traffic Operation." *HRB Bull.* 235, 73-132 (1959).
27. Malo, A. F., Mika, H. S., and Walbridge, V. P., "Traffic Behavior on an Urban Expressway." *HRB Bull.* 235, 19-37 (1959).
28. Campbell, E. W., Keefer, L. E., and Adams, R. W., "A Method for Predicting Speeds Through Signalized Street Sections." *HRB Bull.* 230, 112-125 (1959).
29. Wynn, F. H., "Tests of Interactance Formulas Derived from O-D Data." *HRB Bull.* 253, 62-85 (1960).
30. Hill, D. M., and von Cube, H. G., "Notes on Studies of Factors Influencing Peoples' Choice of Travel Mode." *A report prepared for Metropolitan Toronto Planning Board* (July 1961).
31. Hill, D. M., and Dodd, N., "Travel Mode Split in Assignment Programs." *HRB Bull.* 347, 290-301 (1962).

32. Irwin, N. A., Dodd, N., and von Cube, H. G., "Capacity Restraint in Assignment Programs." HRB Bull. 297, 109-127 (1961).
33. Dieter, K. H., "Analysis of Traffic Flow in Large Source-Sink Systems." Paper, 3rd Annual Meeting, Canadian Operational Res. Soc. (May 1961).