Generating Partial Origin-Destination Tables for Streamlined Application of Corridor Models

SAM YAGAR

A method is described for generating origin-destination (OD) matrices for the subset of drivers who are likely to divert in response to any changes to the traffic network, such as ramp metering for freeways. These ODs are generated automatically from responses to survey questionnaires handed out at strategic points in the network. The responses are factored up to represent 100% of the observed flow on critical links. The procedure also preloads onto the network any link volumes that are not represented by factored survey responses, and are presumably insensitive to any control schemes that might be considered. Preloading of the less sensitive flows, rather than the existing procedure of creating and assigning pseudo-ODs, should reduce the work required of the analyst and lead to improved predictions of flows and queues.

Assignment-based traffic models require origin-destination (OD) information in the form of OD tables or matrices. Further, dynamic assignment models such as CORQ (*1*) require the input of set OD matrices to represent the trip demands in each of a series of contiguous time slices (*2*).

Field data are obtained by surveying drivers and must include at least the trip path and trip OD for each surveyed driver. Usually, only information from drivers whose trip paths are sensitive to the effects of any proposed traffic controls are required. However, most assignment models require that this information be presented to them in the form of OD matrices.

One of the most time-consuming tasks in traffic studies is the conversion of raw field data into these OD matrices. This is complicated by the fact that full OD matrices must generally contain some pseudo-ODs in order to represent all of the flows on the network. These pseudo-ODs represent trips which are generally not sensitive to any control strategies that might be tested and therefore need not really be represented by the user survey that is conducted to obtain the OD information.

This paper describes the structure of a preprocessor for producing OD matrices for use by corridor assignment models such as CORQ2 (*3*).

RATIONALE

Kuwahara and Sullivan (*4*) have identified two primary problems which are encountered when one attempts to convert roadside survey data into OD matrices:

1. Double-counting: Trips which pass through more than one survey location may lead to errors when scaling the survey responses to represent the total flow.
2. Leaky screenlines: It is often physically impractical to set up a survey location at every possible crossing point of a screenline.

In their paper, Kuwahara and Sullivan proposed five methods for overcoming these problems. All of their solutions, however, involve estimating the probability that a trip between one origin and one destination will take a specified path. These probabilities will generally be difficult to generate accurately in real situations. Since the outputs of their procedure depend so heavily on these uncertain probabilities, so will the final assignment. This paper outlines a different approach.

Historically, the ODs obtained directly from survey responses have been factored up to match the flows and queues observed on the network, as all trips on the network could not be sampled. Inevitably, there are flows remaining on the network which are not explained by the factored survey responses, yet the models require full ODs. The traditional approach to this discrepancy has been to create pseudo-ODs traveling between fictitious or irrelevant origins and destinations to compensate for any such flows. By doing this, the analysis program will always be able to perform the assignment to a network which is initially empty.

The CORQ2 program does not require that the network be initially empty before assigning any ODs to it. Instead, it uses a set of "deterministic preloads" (DPs) in addition to the OD matrices. DPs are fixed "deterministic" flows loaded onto the network before the OD demands are assigned. The use of DPs allows the analyst to use OD matrices containing only the network sensitive demands that are of interest, while still representing the effect of the remaining flows on the network.

The procedure of producing ODs and DPs by the preprocessor outlined herein, followed by assigning them by CORQ2, obviates the leaky-screenline problem. The creation of pseudo-ODs added no more relevant information to the analysis. Rather, it merely created additional work, both at the preprocessor stage and at the assignment/analysis stage. Indeed, creating and assigning pseudo-ODs actually confounds the analysis, as the pseudo-ODs are given equal weight with the survey ODs in an automated assignment procedure. The use of DPs allows the more sensitive drivers to be assigned last, as the network assignment approaches its equilibrium. This improves

Department of Civil Engineering, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada.

the modeling realism. If the survey locations have been chosen so that the people surveyed are the ones who will be sensitive to any traffic management strategies, then the proper information required for an analysis has been obtained. In reality, the drivers who create the rational equilibrium are those who are most sensitive to link costs. Therefore, they should be the last ones assigned. This is accomplished by preloading the insensitive drivers.

The concept of using DPs rather than pseudo-ODs was proposed in an earlier paper (*5*). The rationale behind DPs is to permanently load onto the network first those flows which will not change their routes as a result of control strategies. These "permanent" flows are preloaded onto the network. Then, assignment is performed for those trips which are sensitive to any network changes. In this way the assignment procedure is most sensitive to the most sensitive trips.

Through the use of DPs, the preprocessor and CORQ2 reduce the number of steps and the amount of computational effort required to simulate the operation of a network. Rather than expending the effort of removing the additional flows, creating pseudo-ODs for them, and then reassigning them to the network, CORQ2 simply preloads them onto the network before any ODs are assigned. Thus the effects of these flows on the network will be present regardless of the control strategies tested.

The preprocessor also deals with the problem of double-counting. The rest of this paper is devoted to the description of the preprocessor and the method that it uses to scale up the survey responses to create OD matrices that represent all of the assignable trips once and only once. As discussed above, the DPs will always be present on the links and need not be assigned.

PROCEDURE

The preprocessor attempts to factor up the survey responses at each survey location in order to represent all of the counted trips passing through that location. If the drivers passing through a survey location are likely to have to respond to a traffic management strategy that will be tested using the ODs, such as the potential metering or closure of an on-ramp to a freeway, then it is necessary to have the sensitive flows at that location represented by ODs. Since all of the vehicles on a link would have to change paths if the link were closed, it is customary to factor survey responses for a link up to 100% of the link's flow. When the factoring is completed, these flows are removed from the network and placed into the OD matrices. Any flows remaining on the network, and not represented by the assignable OD matrices, are considered to be deterministic preloads.

The preprocessor requires the following input data:

- network topology (how links connect)
- link cost vs. flow relationships
- list of origin nodes and destination nodes
- observed flows and queues on links
- placement of survey handout locations
- first and last survey numbers handed out at each location
- total flow passing through each surveyed link
- traces of paths followed by representative samples of the drivers using the survey links

A returned survey questionnaire provides a trip trace from which can be gleaned the origin, trip path (nodes or links), destination, time of departure, and a list of all survey locations passed. The map from a typical questionnaire for this purpose is shown in Figure 1.

Given the above data, the preprocessor examines each questionnaire response. It determines the survey location and the handout time based on a coded number assigned to each individual handout. Starting at this location, it traces backward and forward along the stated path and generates a list of the links traveled. It also searches for the appropriate origin and destination nodes. A coded sequence of nodes/links for the trip traced in Figure 1 is illustrated in Figure 2. The latter represents the form in which the driver-traced trip of Figure 1 is fed to the preprocessor.

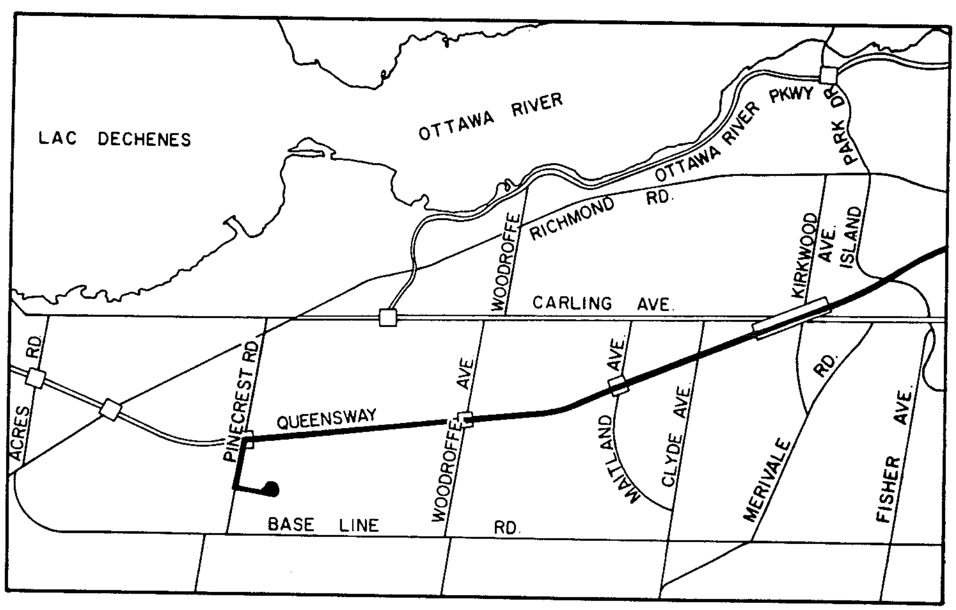

FIGURE 1 Driver's trace as shown on the survey questionnaire.

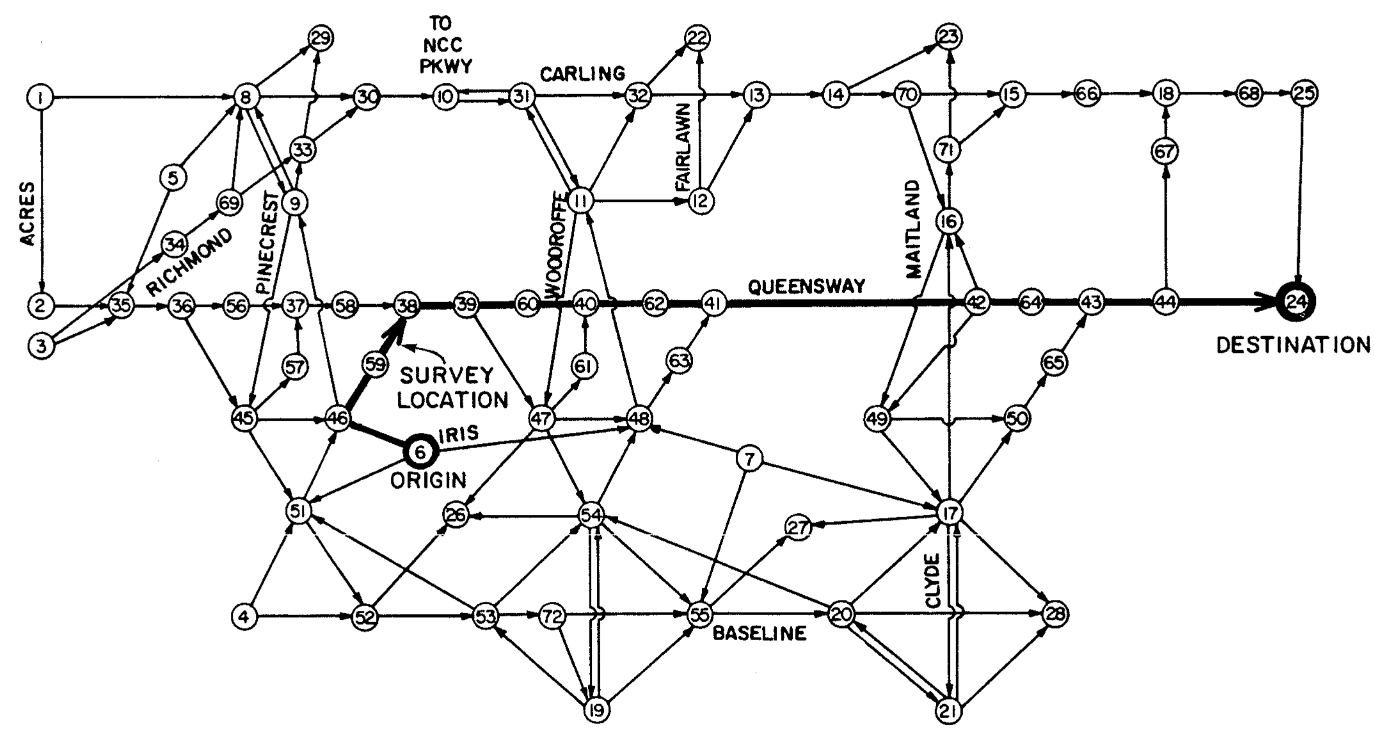

FIGURE 2 Driver's trace on link-node diagram.

Generally, the analyst will have coded the path as starting at one of the prescribed origin nodes and ending at one of the prescribed destination nodes. Since the path specified by the motorist may not exactly coincide with the network as defined, the analyst may have to make some decisions regarding what origin, destination, and intermediate trajectory should be chosen. If the start or end of the trip is not well enough defined that it automatically occurs near a terminal node, then the analyst will have to be careful as to which one is selected.

As an example, if the motorist's returned questionnaire in Figure 1 had only indicated that the start of the trip was at the Pinecrest interchange, the analyst would have to decide whether to assign that trip to origin node 9 or origin node 6. These are the origin nodes closest to the location at which the trip specified by the driver started. The node sequence from origin node 9 would be 9,45,57,37,58,38, . . . , rather than the sequence 6,46,59,38, . . . , from origin node 6 that was illustrated in Figure 2.

The choice of the origin node will clearly be an issue here in terms of the analyses to be performed, especially since this questionnaire will generally be factored up to represent several vehicles in the OD matrix. For example, if the interchange at Pinecrest were to be removed, then motorists from origin node 9 would likely divert along Carling Avenue, while motorists from origin node 6 would likely divert either to Baseline Road or to the Woodroffe on-ramp to the Queensway. The CORQ2 preprocessor has several routines for estimating appropriate origin/destination nodes when the driver has not specified trip-end at origin and destination nodes. These are based on the network connectivity and topology. It is noted that not all network nodes are origin and/or destination nodes.

After the entire path for a trip has been defined, from the origin node through a series of links to the destination node, the time of arrival at each node is calculated so that the preprocessor knows in which timeslice the trip traveled on each link. The timeslice in which the trip originates is also calculated. The preprocessor lists the survey locations through which this trip passed and the timeslices in which the respective survey locations were passed, and stores these on a temporary file for later use as described below. It also updates the number of returned questionnaires from each survey location.

After all of the questionnaires have been processed, there is enough information available to calculate the scaling factors for each response. At this point, the responses are read back in from the temporary file, factored up, removed from the observed flows and queues on the appropriate links (in the appropriate timeslices), and added to the OD matrices. The OD matrices and the remaining deterministic link preloads are then written out to files for use by CORQ2.

SCALING FACTORS

If each surveyed motorist were to pass through only one survey location, then the factoring procedure would be simple. Simply dividing the total flow at each location by the number of responses attributed to location would give the correct scaling factor.

For an idealized situation in which there is no interaction between survey locations, the simple relationship $F = V/R$ holds, where:

F = the factor for a survey location,
V = the total volume at the survey location, and
R = the number of respondees (returned questionnaires) from the survey location.

For example, if there is a flow of 300 vehicles on the surveyed link in a timeslice, and 50 questionnaires are returned, then every respondee is assumed to represent $F = 300/50 = 6$ vehicles. As usual, it is assumed that the returned questionnaires form a representative sample of the trips of all of the motorists using the surveyed link.

In practice, there may well be trips which pass through more than one survey location. This is where the scaling becomes difficult. If a vehicle passes through one survey location with a simple scaling factor of 4, for example, and another location with a scaling factor of 6, then how many vehicles does this respondee really represent on each of the traveled links? This is the basic problem encountered when attempting to factor up the survey responses to the total OD matrix. It cannot be solved exactly in mathematical terms, because it cannot be fully resolved theoretically.

When some trips pass through two or more survey locations, we are not guaranteed an exact solution to the problem. The procedure used in the CORQ2 preprocessor is summarized below. A full description of the procedure is beyond the scope of this short paper. Each trip passing through N survey locations increments counters so that $1/N$th of the effect of the response is assigned to each survey location. This has roughly the same effect as setting the factor for that trip equal to the average of the factors of the N locations. While this procedure is quite robust, it can still be aided by the user through proper selection of survey locations to minimize the number of trips passing through more than one survey location. It also helps if drivers' response rates at the various survey locations are about the same.

DISCUSSION OF RESULTS

Using the preprocessor output as input data to a dynamic corridor assignment model such as CORQ2 should serve to reproduce the flows and queues observed in the field. However, the simulated results will vary somewhat from those observed for the following reasons:

1. The survey responses returned represent only a strategically chosen sample of the users of the network.
2. Flows which pass through more than one survey location may not be factored up precisely as discussed above in the description of the scaling algorithm. Approximation is necessary.
3. Different people have different sets of values and measurements, and they will not necessarily choose the same path between a given origin and destination at a given time, i.e., they will not necessarily utilize the path selected by CORQ2, for example. This is a problem encountered by all assignment techniques. However, the use of a shortest path algorithm in producing the OD matrices for assignment by a shortest path algorithm is a form of pre-calibration, and should help to reduce the error caused by individual driver preferences.

REFERENCES

1. S. Yagar. CORQ—A Model for Predicting Flows and Queues in a Road Corridor. *Transportation Research Record 533,* TRB, National Research Council, 1975, pp. 77–87.
2. S. Yagar. Dynamic Traffic Assignment by Individual Path Minimization and Queueing. *Transportation Research,* Vol. 5, Pergamon Press, 1971, pp. 179–196.
3. S. Yagar. Characteristics of a Freeway Corridor Model. *Procs., Engineering Foundation Conference on Management and Control of Urban Traffic Systems,* Henniker, N.H., Eng. Found. Publication No. 87-14, June 1987, pp. 161–170.
4. M. Kuwahara and E. C. Sullivan. Estimating Origin-Destination Matricies from Roadside Survey Data. *Transportation Research,* Vol. 21B, No. 3, 1987, pp. 233–248.
5. S. Yagar. Partial Synthesis of Traffic Demands to Facilitate Application of Corridor Models. *R.T.A.C. Forum,* Vol. 5, No. 2, 1983, pp. 31–37.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.