Tests of a Family of Trip Table Refinements for Long-Range, Quick-Response Travel Forecasting

Alan J. Horowitz

Alan J. Horowitz, Professor, Center for Urban Transportation Studies, University of Wisconsin – Milwaukee, PO Box 784, Milwaukee, WI 53201, voice: 414-229-6685, fax: 414-229-6958, e-mail: horowitz@uwm.edu

Abstract: This paper addresses the problem of using traffic counts to ascertain zonal trip generation characteristics when performing quick response travel forecasts. A family of OD (origin-destination) trip table estimation methods containing three unexplored members (biproportional, uniproportional and dynamic biproportional) are proposed to solve this problem. The family is tested on static planning networks for Tallahassee, Florida, Northfield, Minnesota and Fredericton, New Brunswick. The tests show that travel forecasting models can be made to better match ground counts by simple factoring of origins, destinations or both. The three methods that directly solve for origin or destination factors have computational and statistical advantages over full-matrix OD trip table estimation procedures and the results are qualitatively and quantitatively interpretable.

INTRODUCTION

A travel forecast is considered to be quick response if local survey data is unavailable and model parameters are transferred from other places. A recurring problem in quick response travel forecasting is validating and adjusting trip generation parameters or time-of-day parameters that have been adopted from other locations or from national defaults *(1, 2)*. A common ad hoc method adopted by many planners is the manual adjustment of trip generation rates until a good agreement is achieved between the base-year forecast and existing ground counts *(3)*. This method often involves weeks of tedious model runs, lacks rigor, is subject to human frailty and is incomplete. Furthermore, planners have had trouble within this method of accommodating zone-to-zone variations in trip generation characteristics due to mixes in land use or other location-specific issues.

Although there has been very little theoretical work done on fitting trip generation rates or time-of-day parameters to ground counts, there has been substantial progress over the past three decades on estimating OD (origin-destination) trip tables to ground counts, see *(4)* for a comprehensive review. Such work could be extended to trip generation by estimating row and column factors for a seed trip table (an imperfect existing trip table) rather than estimating a whole new trip table. The row and column factors so obtained could then be used to ascertain if there are significant problems with trip generation rates or other characteristics for individual zones or for the network as a whole. In addition, an improvement in trip generation, as indicated by row and column factors, might ameliorate the need for further adjustments to the trip distribution step.

This paper explores three different means of obtaining row and column factors for full-sized travel forecasting networks and compares them to a full-matrix method and an optimal single expansion factor method. Altogether, these methods form a family of OD estimation techniques that allow comparisons of results and flexibility in approach.

- Biproportional. This method directly estimates separate row and column factors to be applied to an OD table that has been created in the usual way, such as from a gravity model or from existing data sources. The biproportional method is analogous to finding optimal Fratar factors. This method is suitable for peak period networks where the origin and destination patterns might differ substantially from each other. This method is tested on a full-sized planning network from Tallahassee.
- Uniproportional. This method estimates a single set of factors that would be applied to both the rows and columns of an OD table. The uniproportional method is suitable for 24-hour networks where the pattern of origins is nearly identical to the pattern of destinations and time-of-day issues are irrelevant. This method is tested on full-sized planning networks from Northfield, Minnesota and Fredericton, New Brunswick.
- Dynamic Biproportional. This method obtains optimal Fratar factors in multiple time periods and would be suitable for any planning situation where dynamic traffic assignment is employed and where better time-of-day factors are needed. This method is presented here to illustrate how static methods can be extended to dynamic networks. It has not yet been subjected to a full-scale test, so it will only be briefly described here.

This paper primarily emphasizes computational experience with these methods. The dynamic biproportional technique has many potential applications well beyond the needs of long-range travel forecasting, but this paper focuses on static methods within the travel forecasting process.

Surprisingly, this author has been unable to find any evidence of the development or the use of these techniques in the literature and there does not seem to be any readily available software (commercial or experimental) to implement these techniques on full-scale regional travel forecasting networks. It is possible, of course, to emulate the static biproportional method by estimating a full-matrix by some appropriate technique, then summing the rows and columns. Such an approach is overly computational and difficult to control. In very large networks with many traffic counts and zones, estimating a full OD table by statistically rigorous methods may require a prohibitive amount of computational resources.

It should be immediately obvious that uniproportional and biproportional methods cannot provide as good a fit to ground counts as full-matrix methods because there a much fewer variables that can be estimated. However, it is unclear whether a full-matrix method can provide results that are statistically better in any given network, considering that the trip distribution step of a travel forecasting model should already have a considerable amount of validity. Furthermore, planners often need to protect the results of the trip distribution step from unwarranted empirical manipulation so as to retain the behavioral underpinning of the model when performing a forecast. Coupled with the difficulty of their use, this is likely the reason why heavy-handed, full-matrix methods have not become essential to the long-range travel forecasting process. It seems that methods with a lighter touch might make more sense in many situations, if their usefulness can be demonstrated.

FRAMEWORK OF METHODS OF OD TRIP TABLE ESTIMATION

Fourth Level, Whole Trip Table GLS

The single factor, uniproportional and biproportional methods can all be implemented using generalized least squares (GLS) and are directly related to the conventional GLS method of estimating all the cells of an OD table:

$$\min P = \sum_{a=1}^{A} w^a \left(V^a - \sum_{i=1}^{N} \sum_{j=1}^{N} p_{ij}^a T_{ij} \right)^2 + \sum_{i=1}^{N} \sum_{j=1}^{N} z_{ij} \left(T_{ij}^* - T_{ij} \right)^2 \qquad (1)$$

where V^a is a ground count for link direction a, T_{ij} are the trips between origin i and destination j to be estimated, T_{ij}^* is the seed trip table, p_{ij}^a is the proportion of trips between zones i and j that use link direction a (as determined by a static equilibrium traffic assignment), N is the number of zones and w^a and z_{ij} are weights. Each direction of a two-way link, a, may have a separate ground count and may be considered the same as a "link" when only directional links are present in the network. Nonnegativity constraints apply to T_{ij}.

Framework Overview

A family of OD table estimation methods can be conveniently described by introducing a set of "k-factors" to Equation 1. K-factors, often referred to as socioeconomic adjustment factors, are sometimes used by practicing planners in attempts to improve the performance of their travel forecasting models. In Equation 2, the matrix of z's has been replaced by a single parameter, z, because there is no obvious justification to having differing weights across OD's for the tests in this paper.

$$\min P = \sum_{a=1}^{A} w^a \left(V^a - \sum_{i=1}^{N} \sum_{j=1}^{N} p_{ij}^a k_{ij} T_{ij}^* \right)^2 + z \sum_{i=1}^{N} \sum_{j=1}^{N} T_{ij}^{*2} \left(1 - k_{ij} \right)^2 \qquad (2)$$

Except for single factor GLS, which can be solved in closed form, all of the methods are solved in this paper with the gradient projection method (GPM) augmented with PARTAN acceleration steps. The implementation of this minimization technique will be briefly described later.

Third Level, Fratar Biproportional GLS

The biproportional method introduces two vectors of factors, one for origins, x_i, and one for destinations, y_j, and an overall scale factor, s.

$$k_{ij} = sx_i y_j \tag{3}$$

The scale factor s is optional and might be set prior to the optimization in an attempt to keep x_i's and y_j's close to 1. In order to be consistent with the philosophy of linear regression analysis, s is arbitrarily chosen for the case studies described here such that the average of the ground counts is the same as the average estimate of the ground counts when all x_i's and y_j are equal to 1 (i.e., the seed OD table is otherwise unmodified). Other means of setting s might be more appropriate in different situations.

Nonnegativity constraints on both x_i and y_j are required to assure reasonable results. In addition, it might be appropriate to place upper and lower bounds of the values of x_i and y_j to control the amount of distortion of the original trip generation or time-of-day results.

It should be immediately evident from Equation 3 that the solution is not unique if it is unconstrained. However, early computational tests indicated that the values of x_i's and y_j's tend stay centered on their initial values when constraints are imposed, so no effort was made to further enhance the case studies presented here in order to assure uniqueness.

Second Level, Uniproportional GLS

The uniproportional method is similar to the biproportional method, but all y_j's are replaced by x_j's. So

$$k_{ij} = sx_i x_j \tag{4}$$

Again, s is optional and may be set in an attempt to keep the values of the x's close to 1. Bounds may also be applied to the x's.

First Level, Single Factor GLS

Single factor GLS can be obtained by replacing all the k's by a single number, s. That is,

$$k_{ij} = s \tag{5}$$

Because there is only one variable in Equation 2 as modified by Equation 5, it can be readily solved in closed form.

DYNAMIC EXTENSIONS OF THE FRAMEWORK

Dynamic Family Overview

The dynamic framework is similar to the static framework, but addresses situations where different OD tables are needed for different intervals of time and trip durations are longer than a single time interval *(7)*. For instance, such a method might be useful for ascertaining time-of-day factors for statewide travel forecasts. There is a separate set of k-factors for each starting time interval, d. They can be found by solving:

$$\min P = \sum_{c=F}^{L}\sum_{a=1}^{A} w^a \left(V^{ac} - \sum_{d=1}^{L}\sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij}^{acd} k_{ij}^{d} T_{ij}^{d*} \right)^2 + z \sum_{d=1}^{L}\sum_{i=1}^{N}\sum_{j=1}^{N} T_{ij}^{d*2} \left(1 - k_{ij}^{d}\right)^2 \tag{6}$$

where V^{ac} is a ground count for link direction a as measured over time interval c, T_{ij}^{d*} is the seed trip table for starting time interval d, p_{ij}^{acd} is the proportion of trips between zones i and j and start in interval d that use link direction a during interval c (as determined by a dynamic equilibrium traffic assignment) and F and L are the first and last time intervals over which traffic is counted. Although it is possible to have different seed tables for different intervals, the method could also be implemented with just one seed table for all intervals.

Biproportional and Uniproportional Dynamic

The origin and destination factors, x_i^d or y_j^d, for any starting time interval d may be found by replacing k_{ij}^d with:

$$k_{ij}^d = s x_i^d y_j^d \tag{7}$$

As before, nonnegativity and bounding constraints apply to x_i^d or y_j^d. Note that only one value of s is used. The uniproportional method could be achieved by replacing y_j^d with x_j^d. This replacement is not simple computationally because the partial derivatives of P are considerably different. The dynamic unipropotional method was not investigated further because it is difficult to conceive of many travel forecasting situations where the dynamic uniporoportional method would be appropriate.

Single Factor

Finally, a single factor dynamic OD trip table estimation may be achieved by estimating a separate factor for each starting time interval d, such that:

$$k_{ij}^d = s^d \tag{8}$$

The single factor values cannot be obtained in closed form, as was the situation with the static single factor method.

COMPUTATIONAL ISSUES

Solution Algorithm

The solution algorithm was held constant throughout these tests in order to maintain some consistency between the methods and to assure that each run attained about the same degree of convergence. The gradient projection method with PARTAN seemed to provide an acceptable speed of convergence for all but one of the tests. Both algorithms are textbook ways of solving nonlinear programming problems. The gradient projection method was required so that feasibility could be guaranteed throughout the search process and PARTAN was required to reduce the number of derivatives to be calculated. Analytical differentiation was required in all cases to save computation time; however, the derivatives for biproportional and uniproportional methods were very complex – much more so than for full-matrix GLS – and time consuming to calculate. It is entirely possible that a faster solution algorithm could be devised, particularly for the case of the biproportional method.

Convergence Criteria

Searches in the gradient projected direction are stopped when the step size, η, decreases beyond:

$$\eta < \phi\sqrt{N} \tag{9}$$

where ϕ is a suitably small number, chosen to be 0.000001 for all tests in this paper and N is the number of variables. The optimization is terminated when the relative change in the objective function between PARTAN steps is smaller than an arbitrary number, taken to be 0.0001 for the tests presented here. The static single factor, of course, is determined exactly, so no convergence criteria are required.

Discussion

With all three methods, it is possible to exclude certain classes of origins or destinations (e.g., external stations or special generator centroids) from the optimization.

Although there are typically far fewer variables in either the biproportional or uniproportional methods than in a full-matrix GLS problem for any meaningful problem, the input data requirements are identical. The p_{ij}^a array is particularly difficult to work with because of potentially huge number of nonzero elements, so a serious amount of attention is required to find efficient ways of holding the array in memory, accessing array elements, computing summations and performing equilibrium averages. Otherwise, computation times would be prohibitively long.

TRAVEL FORECASTING FRAMEWORK

The OD trip table estimation problem with equilibrium traffic assignment requires a bilevel (or equivalent) solution procedure, for example *(8, 9)* among others. In the case studies presented here, the minimization problem is embedded within a conventional method of solving for an elastic-demand equilibrium traffic assignment using the method of successive averages (MSA). The flow diagram of Figure 1 illustrates the procedure.

The figure shows two arrows looping back to earlier steps. These loops occur essentially simultaneously at the end of a single equilibrium iteration. The loop on the right assures that congestion found during traffic assignment is incorporated within the gravity model (should a gravity model be needed). The loop on the left carries the trip table “refinements”, as determined by the OD table estimation procedure, to the next equilibrium iteration. The refinements are found by subtracting the current equilibrium-averaged seed trip table from the estimated trip table, keeping in mind that the seed trip table can change as the simulation progresses. It is important to note that the minimization problem, a very time consuming process, must be solved once for each MSA average.

The dynamic biproportional OD table estimation procedure is similar, except another loop across time intervals is required to find a dynamic traffic assignment. Adding the second loop gives the algorithm the same dynamic properties as Dynasmart, a well respected dynamic traffic assignment algorithm developed at the University of Texas, Austin. Nonetheless, a minimization problem is only run once for each MSA average, as with the static method.

CASE STUDIES: TALLAHASSEE, FREDERICTON AND NORTHFIELD

Three case studies have been selected in order to draw conclusions that would pertain to quick response networks, in general. Although the networks differ radically in size and detail, they share common aspects. All three traffic networks were developed by local agencies (or their consultants) following the guidelines and parameters found in NCHRP #187 or NCHRP #365 or both. This author was not involved in the creation or calibration of these networks. Each network used a gravity model for at least part of the trip distribution step. Each network received a large amount of ad hoc calibration until the base-case forecast provided acceptable results for the purposes of the agency. The networks contained explicit traffic controls, with delays at intersections being calculated by methods equivalent to those found in the 2000 Highway Capacity Manual *(10)* and with delays on interrupted facilities and between intersections by the BPR curve using parameters found in NCHRP #365.

Available data were limited in all three cities to that used in the base year-calibration runs. The complete family of methods were run on each network. In addition, the entropy maximizing method was implemented on each network using 500 multiproportional iterations.

Tallahassee, Florida

Tallahassee is the largest of the test networks. The network is used by both the City of Tallahassee and Leon County for site impact assessment, and it is highly refined for short-term, strategic forecasts. The network is continuously updated. It covers all of Leon County (population of 240,000) and contains 861 zones, 13 external stations, 2699 two-way street links, 285 one-way street links and 1553 centroid connectors. This network normally is run through 13 (user-optimal) equilibrium iterations, but only 11 were used for the tests described in this paper. Traffic counts on this network were comprised of tube counts and aggregations of manual turn movement counts. Some older turn movement counts were discarded for the tests described here, leaving 1784 link directions with counts – slightly more than one count for each origin and each destination. An effort was made by the planners in Tallahassee to scale trip production rates from NCHRP #365 to better match ground counts. The network used a gravity model exclusively for trip distribution; external-to-external trips were preloaded to the network.

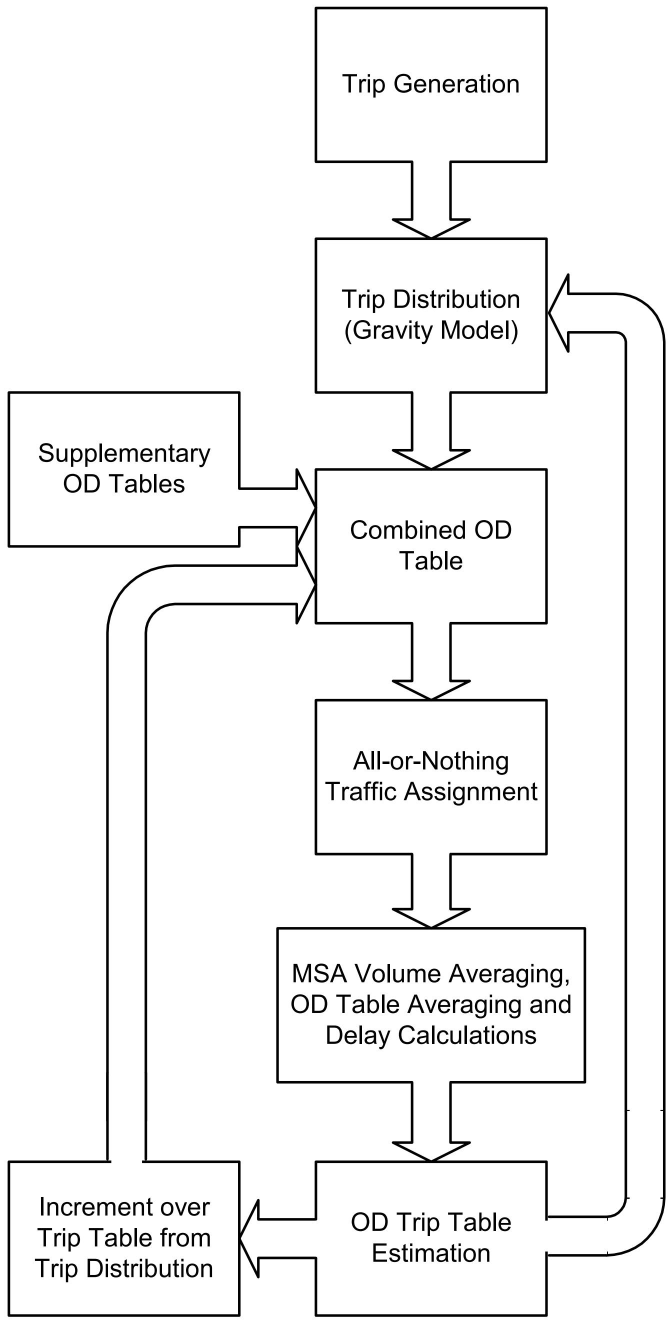

FIGURE 1 A Bilevel Algorithm for OD Trip Table Estimation

The Tallahassee network is for the PM peak hour, so the biproportional method was initially considered most appropriate. With this method the number of nonzero elements in the p_{ij}^a array grows to about 26 million (out of a possible 1.4 billion or 1.9%) by the last equilibrium average. The weights on ground counts, w^a, were set equal to 1 on all links with counts, except those leading directly to external stations that were given a weight of 5. The trip table weight, z, was 100. OD factors were constrained only to be greater than zero. The average (weighted) ground counted was 542 vehicles per hour. The average value of an OD cell was 0.118 trips per hour.

Northfield, Minnesota

The Northfield network is the smallest of the three. Northfield has a population of just 17,000 people and the network covers the city limits and a small amount of surrounding land. The network was developed for the sole purpose of evaluating the short-term traffic impact of a new bridge. Thus, the city was most interested in obtaining an excellent fit of the base-case model to ground counts in the immediate vicinity of the bridge. The Northfield network is unusual in that every street in the city is coded into the network to GIS standards and there are no centroid connectors. Traffic assignment is done between almost all the intersection nodes using a technique called "area-spread" assignment *(11)*. The network contains 29 zones, 12 external stations, 816 two-way street links, 3 one-way street links and 711 intersection nodes where trips can begin and end. There are 60 link directions with traffic counts, or slightly less than 2 for each zone. Traffic assignment used 20 equilibrium averages. Trip generation parameters were adopted verbatim from NCHRP #365. The gravity model was supplemented with an external-to-external trip table.

Because the Northfield network does forecasts for a 24 hour time period, the uniproportional method was initially considered most appropriate, because the number of origins in a zone should equal the number of destinations in the same zone. With this method the number of nonzero elements in the p_{ij}^{a} array grows to about 10,500 (out of a possible 100,860 or 10.4%) by the last equilibrium average. The results from this method are compared with full-trip table GLS, entropy maximization, and the biproportional method. The weights on ground counts, w^{a}, were set equal to 1 on all links with counts. The trip table weight, z, was also 1. When the OD factors for the biproportional and uniproportional techniques were constrained, they were limited to be between 0.2 and 5. The average ground count was 3840 vehicles per day. The average OD cell had 41.5 trips per day.

An interesting implication of "area-spread" equilibrium assignment is that it has many more used paths than a conventional equilibrium assignment, thereby creating a larger number of nonzero cells in the p_{ij}^{a} array, which increases the computational burden.

Fredericton, New Brunswick

Fredericton has a population of about 47,000. Its network was created for the purpose of long range transportation planning, so an attempt was made by the planners who created it to achieve a reasonably good fit to ground counts everywhere. The Fredericton network has the largest amount of ground counts per TAZ. The network covers all heavily populated areas and a considerable amount of rural land in 56 TAZs and 12 external stations. The network has 415 two-way street links and 134 one-way street links. 549 link directions have counts, or about 8 for each zone. Traffic assignment used 25 equilibrium averages. Trip generation parameters were adopted verbatim from NCHRP #365. External-to-external trips were preloaded to the network. The average ground count was 4934 vehicles per day. The average cell in the OD table was 61.2 trips per day.

The Fredericton network forecasts 24-hour volumes, so the uniproportional method was initially considered most appropriate. All conditions of the uniproportional and biproportional methods were the same as for Northfield. With the uniproportional method the number of nonzero elements in the p_{ij}^{a} array grows to about 88,300 (out of a possible 2.5 million or 3.5%) by the last equilibrium average.

RESULTS

Improvements in Assigned Volumes

Table 1 contains root-mean-square errors of base-case forecasts to all ground counts. Applying optimal OD factors to the assigned trip table reduces the RMS error substantially – about 50% in the case of the biproportional technique in Northfield. The biproportional method offers only a small improvement over the uniproportional method in all three networks.

Table 1 also shows that the biproportional and uniproportional methods have between two and three times the RMS errors of the full-matrix GLS method. The reason for the higher RMS error for the entropy maximization method over the full-matrix GLS in Fredericton is likely related to inconsistency in ground counts. In Fredericton, more than 45% of the RMS error is not explainable by a better OD table (entropy or full-matrix GLS). The corresponding RMS errors to the seed trip table are shown on Table 2. It is difficult to draw distinctions between the amount of distortion caused the uniproportional, biproportional or full-matrix method. The entropy maximization method caused a considerable amount of distortion and the single factor method caused hardly any.

TABLE 1 RMS Errors to Ground Counts Before and After Applying Refinement to Trip Table

	Tallahassee	Northfield	Fredericton
No Factoring	197	1466	2511
Single Factor	188	1445	2390
Uniproportional	165	874	2193
Biproportional	146	727	2069
Full-Matrix GLS	61	248	1151
Full-Matrix Entropy	36	1	1374

TABLE 2 RMS Errors to Seed Trip Table Before and After Applying Refinement to Trip Table

	Tallahassee	Northfield	Fredericton
No Factoring	0	0	0
Single Factor	0.036	6.12	16.27
Uniproportional	0.357	97.72	90.58
Biproportional	0.280	93.11	101.07
Full-Matrix GLS	0.239	75.35	159.65
Full-Matrix Entropy	1.727	189.79	334.92

Distortion of Trip Lengths

Table 2 contains data that indicate the degree of distortion of the seed OD table; however, it is difficult to interpret these numbers. Table 3 shows the average trip length (in miles) from each method and from the baseline without any refinement. The origin-to-destinations distances were held constant for each city and were calculated from the baseline run. For the Northfield runs, the OD distances were the weighted average of all node-to-node distances between the pair of zones.

TABLE 3 Average Trip Lengths (Miles) Before and After Applying Refinement to Trip Table

	Tallahassee	Northfield	Fredericton
No Factoring	5.9	2.7	5.9
Single Factor	5.9	2.7	6.1
Uniproportional	6.0	2.6	6.5
Biproportional	6.1	2.6	6.5
Full-Matrix GLS	5.5	2.6	5.7
Full-Matrix Entropy	4.5	2.3	5.3

The only method to cause serious distortions of trip length is entropy maximization. The much shorter trip lengths in all cases are counterbalanced by a much larger number of trips. For example, in Tallahassee the baseline run involved about 93,000 vehicle trips, but the entropy maximizing run involved about 120,000 vehicle trips. Trip length frequency diagrams in all cases (not presented here) confirm these findings.

F-Test of Significance

Traditionally, the significance of variables in a nonlinear regression is judged by the F-test. The F-test has certain limitations for OD table estimation, as it is not possible to incrementally add variables to the model and it is difficult to assess the effective number of degrees of freedom due to the constraining effect of the second (trip table) term of the objective function. The choices are limited to zero variables for no estimation, one variable for the single factor method, n variables for the uniproportional method, 2n variables for the biproportional method, and n^2 variables for the full-matrix GLS method. Table 4 lists the F values between successive models for each case study. The F-values between the biproportional and full-matrix GLS methods could not be ascertained because the number of variables (OD cells) exceeds the number of data points (counts) in all cases.

TABLE 4 F-Values between Models

	Tallahassee	Northfield	Fredericton
No Factoring to Single Factor	174.8	1.72	56.9
Single Factor to Uniproportional	0.3	0.8	1.3
Uniproportional to Biproportional	0.0	*	0.7

*Cannot be computed.

Given that the F-test is overly strict, it can at least be observed that the single factor method is justified (F values greater than one) in all three networks and the uniproportional method is justified in the Fredericton network. It should be readily apparent that all the methods (except those without any degrees of freedom) would have done better had the weight (z) on the second term of the objective function been set lower. The pattern of F-values are consistent with the notion that the networks were already calibrated.

Linear Regression Results

Perhaps a better test of significance can be created by determining whether the OD factors from the biproportional or uniporportional methods, which at this time have only a mathmatical interpretation, can be explained behaviorally. For example, it might be hypothesized that the origin and destination factors obtained through the trip table estimation process may be explained by demographic and socioeconomic characteristics of zones. If the same demographic characteristics that are used in trip generation are then used as independent variables in a regression analysis of the origin or destination factors, then any significant coefficients would suggest that the trip generation rates are in error and the OD estimation has validity. Table 5 shows how well the traditional quick response demographic variables used for trip attraction calculations explain OD factors by linear regression.

TABLE 5 "t" Statistics from Regressing OD Factors against Traditional NCHRP #187 Quick Response Trip Generation Variables

	Tallahassee, Origin	Tallahassee, Destination	Northfield	Fredericton
Retail Employment	-4.5	-0.6	-0.1	-0.5
NonRetail Employment	-4.6	0.9	-1.5	-1.6
Dwelling Units	3.6	5.3	0.3	-2.5

The "t" statistics for Tallahassee are the most interesting, with all demographic variables being significant for origin factors and dwelling units being significant for destination factors. There is also one significant "t" value for Fredericton (dwelling units). The t-test of linear regression coefficients seems like a good way to overcome the problem of too many variables found with the F-test.

Intuitive and Spatial Results

The existence of a clear spatial pattern in the OD factors would further indicate that they contain useful information. Figures 2 and 3 shows the biproportional factors on a network graphic for the central portion of the Tallahassee network. Larger green (gray) dots show factors greater than one and smaller black dots show factors less than one. The figures show that there are highly recognizable spatial patterns to the distribution of factors. Some fairly wide areas of the city are overpredicted by the trip generation model (shown by smaller black dots) and other areas are underpredicted (shown by larger green or gray dots). For example, the central business district in Tallahassee is split down the middle for destinations, with the eastern portion being overpredicted and the western portion being underpredicted. The pattern of overpredictions and underpredictions differ considerably between the two figures, although a few places were consistent. For example, the far northeast side was overpredicted for both origins and destinations. When these figures were shown to planners in Tallahassee, they were immediately able to identify reasons for the spatial patterns. This knowledge would be quite helpful is reformulating their trip generation model.

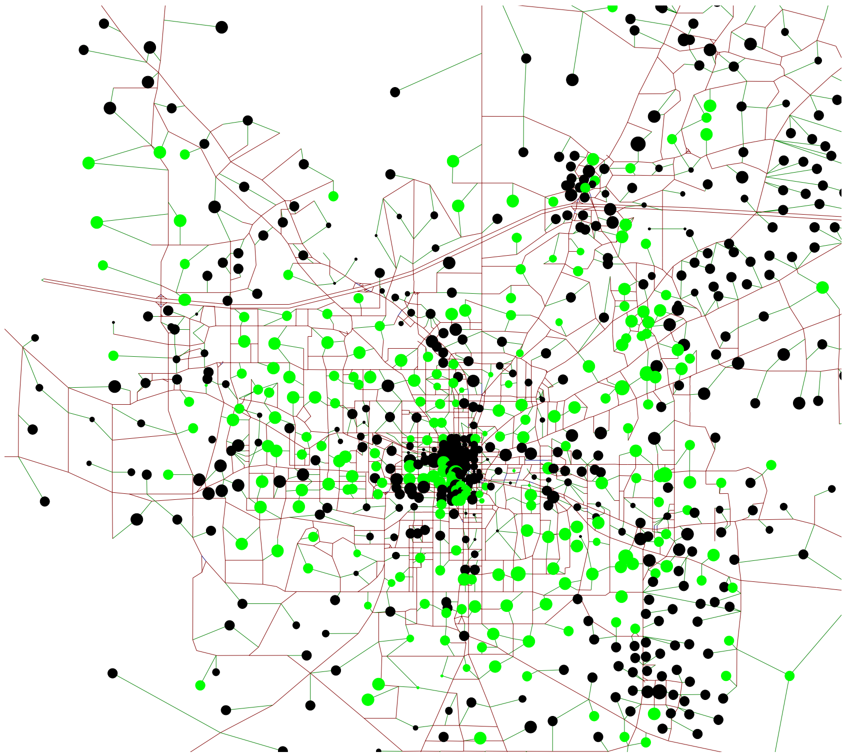

FIGURE 2 Origin Factors by TAZ for Central Tallahassee

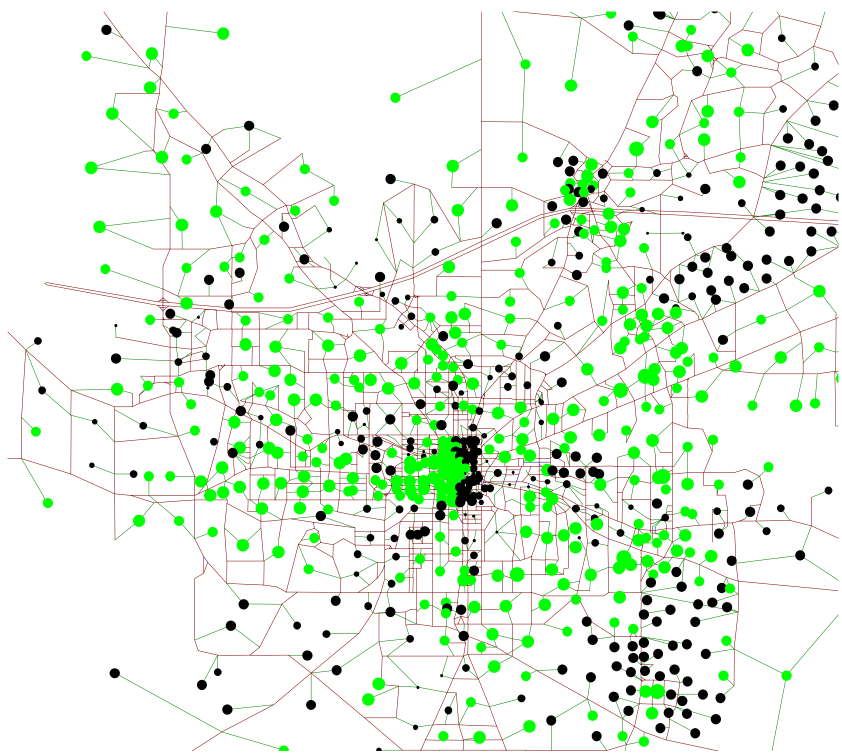

FIGURE 3 Destination Factors by TAZ for Central Tallahassee

Distribution of Origin and Destination Factors

Figure 4 shows the frequency distribution of OD factor values in Tallahassee. Separate bars are shown for origins (on the left) and destinations (on the right). It is seen that the OD factors are spread between 0 and 4, with about half of the factors clustered near one. That origin factors tended to be less than one and destination factors tended to be greater than one is largely an artifact of the optimization procedure. The large number of origin or destination factors greater than 1.6 or less than 0.4 indicates that the trip generation model is not in good agreement with ground counts for many zones. Similar results were seen in both Northfield and Fredericton.

For the Northfield and Fredericton networks the uniproportional and biproportional methods were also run with the OD factors constrained to be within the range of 0.2 to 5. The RMS errors obtained were almost identical to those from the optimizations where the OD factors were constrained only to be nonnegative.

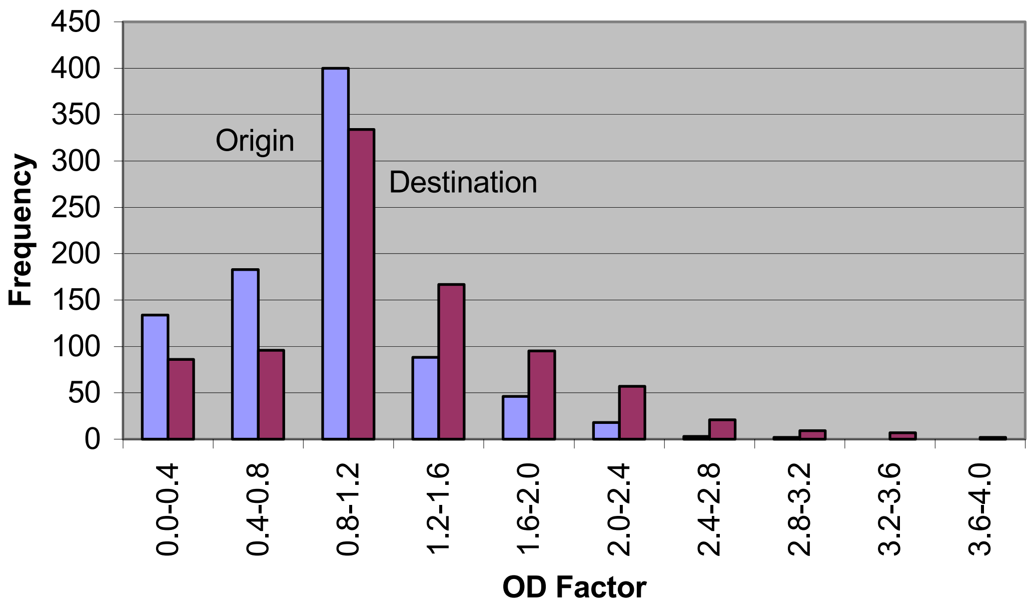

FIGURE 4 Distribution of OD Factors in Tallahassee

Computational Effort

It is entirely possible that considerably faster algorithms could be devised for each of the methods, but the relative speeds are indicative of the amount of computation required. For Northfield, the uniproportional method took almost exactly the same amount of time as the full-matrix GLS method because the time required to minimize the objective function was negligible compared to the other parts of the model. However, for Fredericton the uniproportional method took only 20% of the time and in Tallahassee the biporportional method took only about 8% of the time. The computational advantage of the OD factor estimation seems to quickly increase with the number of zones. Since for large networks computation is measured in days or weeks, these differences in computation time are significant.

Beyond issues of computation time, there were no observed convergence problems with any of the algorithms.

Discussion

It would have been useful to have had a more complete set of demographic and socioeconomic data for zones to determine whether a better trip generation model could have been build from existing data source. However, the agencies were fully committed to the quick response approach, so more complete sets of data were not assembled.

Although this paper is not too concerned with computation issues, quicker algorithms are very important, especially for large networks. The time necessary to complete an OD trip table estimation relates most closely to the number of equilibrium averages, the dimensionality of the solution space (e.g., number of variables to be determined), the complexity of the partial derivatives, the network size, number of traffic counts and the required precision. The biproportional and uniproportional methods offer considerable advantages over full-matrix GLS because of the much smaller number of variables typically found on planning networks, in spite of the much more complex set of partial derivatives. A good indicator of the computational effort required for any given method is the number of nonzero elements in the p_{ij}^{a} array, which is influenced by the number of traffic counts, the number of OD pairs and the number of links in the network. For example, the Tallahassee network has a nonzero p_{ij}^{a} array that is about 2500 times greater than the similar array in Northfield. Tallahassee is near the upper end of networks that would be practical with the current generation of personal computers.

The routines used in this paper was custom built using the author's software. An inspection of Figure 1 would suggest that these OD table estimation techniques might be readily implemented within commercial software platforms that are modular in nature.

CONCLUSIONS

Biproportional and uniproportional OD trip table estimation methods are components of a larger family of OD estimation methods that can be used to refine trip tables that have been created through conventional means (e.g., a gravity model), such that the results can be interpreted as modifications to trip generation. The improved trip table could be directly used in short-range travel forecasts, however it is preferable that the trip generation model be modified to better match the results of the OD trip table estimation. These methods can also be useful in identifying errors in base-year levels of activity within zones. These methods are especially germane to networks that use quick response techniques, because of the lack of survey data to calibrate such networks. The methods could also be useful in other planning contexts to identify problems or to update trip generation models that are based on older surveys. The biproportional and uniproportional methods offer a means of greatly reducing the number of variables to be estimated and offer a clear path to further statistical analysis for additional reductions in the number of variables.

Tests of the biproportional and uniproportional OD trip table estimation methods in already calibrated networks indicate that they can further explain about one-sixth to one-half of the RMS error in ground counts. About one-seventh to about one-third of the RMS error cannot be explained by any reasonable modification of the OD trip table. None of the full-matrix methods could be justified using standard statistical tests because of the large number of variables in the model.

Many of the OD factors obtained in the three test networks differed considerably from one. There are several possible explanations for this finding, but the most obvious explanation is that the trip generation model is appreciably overpredicting or underpredicting the number of trip ends in many zones.

Because of their faster execution times, the biproportional and uniproportional methods may be practical for networks that are too large for full-matrix OD trip table estimation.

A solution method consisting of the gradient projection method with a PARTAN step worked well for these tests, but additional attention needs to be paid to improving computation times and memory requirements for large networks.

ACKNOWLEDGMENTS

The author thanks the cities of Tallahassee, Northfield and Fredericton for providing their networks, ground counts and other data.

REFERENCES

1. A. B. Sosslau, et al., Quick-Response Urban Travel Estimation Techniques and Transferable Parameters: User's Guide, NCHRP Report #187, 1978.
2. William Martin and Nancy A. McGuckin, *Travel Estimation Techniques for Urban Planning*, NCHRP Report #365, 1998.
3. D. Ismart, "Calibration and Adjustment of System Planning Models", Federal Highway Administration, FHWA-ED-90-015, December 1990. (Full text in National Transportation Library, www.bts.gov.)

4. Torgin Abrahamsson, “Estimation of Origin-Destination Matrices Using Traffic Counts – A Literature Survey”, International Institute for Applied Systems Analysis, IR-98-021, May 1998.
5. H. Van Zuylen and L. G. Willumsen, “The Most Likely Trip Matrix Estimated from Traffic Counts”, *Transportation Research*, Vol. 14B, 1980.
6. Y. J. Gur, “Estimating Trip Tables from Traffic Counts: Comparative Evaluation of Available Techniques”, *Transportation Research Record*, #944, 1983, pp. 113-117.
7. H. Tavana and H. S. Mahmassani, “Estimation of Dynamic Origin-Destination Flows from Sensor Data Using Bi-Level Optimization Method”, Transportation Research Board 80th Annual Meeting CD-ROM, 2001.
8. C. S. Fisk, “Trip Matrix Estimation from Link Traffic Counts: The Congested Network Case”, *Transportation Research B*, Vol 23B, No. 5, 1989, pp. 331-336.
9. Y. Chen, “Bilevel Programming Problems: Analysis, Algorithms and Applications”, PhD Thesis, University of Montreal, 1994.
10. Alan J. Horowitz, “Internodal Delay Issues in Long-Range Adaptive Travel Forecasts”, *Transportation Research Record Journal* #1783, 2002, pp. 49-54.
11. Alan J. Horowitz, “Computational Issues in Increasing the Spatial Precision of Traffic Assignments”, *Transportation Research Record Journal*, #1777, 2001, pp. 68-74.