The study of the impact of policies and programs on travel patterns that result from individual activities that take place from place to place over a transportation network require the collection of survey and census data for individuals, households, firms, etc. that are location specific. While it is not possible to examine the behavior of all individual actors in detail, we can still obtain a reasonable overall picture in terms of appropriate spatial aggregations of activities information of which is usually collected at different geographic levels. Such aggregations are also used for confidentiality purposes of administrative data.

Traffic Analysis zones (TAZs) is a related geography with a particular use in transportation planning. This is because TAZs are the basic areal units of analysis for all input and output data in the urban travel forecasting procedure. TAZs are usually developed to follow physical (e.g., lake, river, etc.), political (e.g., township, county, etc.) or administrative boundaries (e.g., census block/tract, ZIP code, etc.) and contain populations with similar socioeconomic characteristics. The choice of an appropriate TAZ partition of an area involves an implicit tradeoff between achieving homogeneity and obtaining area units that are sufficiently large to allow meaningful statistical sampling.

TAZ geographies would not be very useful for transportation planning studies if they could not relate to other location-referenced databases. This relationship is established with the development of translation tables or crosswalks between TAZs and data from different geographies (e.g., census tracts, ZIP codes, etc.). Nowadays this process has been considerably simplified with the use of Geographic Information Systems.

In travel demand forecasting, analysts need to relate TAZ-specific socioeconomic attributes with transportation network attributes. In this instance, TAZs are represented by their respective centroids, a geographic point inside each TAZ polygon, where travel activities are implicitly assumed to originate from or terminate at. TAZ centroids also represent the location all TAZ-specific socioeconomic activities are concentrated at. The connection of TAZ centroids with the rest of the coded transportation network requires that ‘dummy’ links connect the centroid node with the network links that represent a proper aggregation of the physical network.
Occasionally, transportation planners need to make use of origin-destination data estimated for a metropolitan area, e.g., from a Metropolitan Planning Organization (MPO). In these cases, there is a need to modify the data to fit a geographic area with a particular TAZ system that is different from the one the data were developed for. In other words, suitable aggregation and disaggregation procedures are needed to convert the original TAZ data to the new TAZ system. It turns out that such procedures are available using the gravity model.

Aggregation of Origin-Destination Flows

Assume that an MPO has estimated origin-destination flows at a more refined TAZ system corresponding to an \((IJ)\) origin-destination table with \(I = (i_1, i_2, \ldots)\) origins and \(J = (j_1, j_2, \ldots)\) destinations. Assume, further, a study needs to aggregate these flows at a coarser TAZ system to estimate an \((KL)\) origin-destination table with \(K = (k_1, k_2, \ldots)\) origins and \(L = (l_1, l_2, \ldots)\) destinations. In cases there is a need to aggregate origin-destination flows \((i_1, j_1), (i_1, j_2), \ldots\) from small TAZs to origin-destination flows \((k_1, l_1), (k_1, l_2), \ldots\) the compressibility property of the gravity model can be used. For example in the figure below, the origin-destination flow \((k_1, l_1)\), in the large TAZ will be aggregated from the origin-destination flows \((i_1, j_1), (i_1, j_2), (i_2, j_1), (i_2, j_2)\).

The procedure assumes that a gravity model, \(T_{ij} = A_i B_j F_{ij} = A_i B_j \exp \left( \sum_{k=1}^{K} \theta_{ij} e_{ij}^{(k)} \right)\) has been estimated for the small-zone TAZ system (Metaxatos, 2009; Metaxatos, 2004; Sen and Smith 1995). Then, by simple aggregation, it follows that \(T_{k_1 l_1} = \sum_{i_1}^{i_2} \sum_{j_1}^{j_2} T_{ij} = \sum_{i_1}^{i_2} \sum_{j_1}^{j_2} A_i B_j F_{ij}\).
and \( F_{k_1l_1} = \frac{\sum_{i=1}^{l_1} \sum_{j=1}^{k_1} A_iB_jF_{ij}}{\sum_{i=1}^{l_1} \sum_{j=1}^{k_1} A_iB_j} \). Similar calculations will obtain estimated origin-destination flows for the other TAZs \((k_1, l_2), (k_2, l_1), (k_2, l_2)\).

Therefore, if we aggregate the small-zone TAZ system to a larger-zone TAZ system we obtain a new gravity model that predicts the same total number of trips between aggregated zones as the initial gravity model for the more detailed zone system (compressibility property). The practitioner would need to consider the possibility that when the aggregated zones are large, the functional form of the separation factor, \( F \), may not remain the same when moving from the smaller to the larger TAZ system.

**Disaggregation of Origin-Destination Flows**

Frequently, the reverse problem must be considered. Origin-destination flows may only be available for a large-zone TAZ system, e.g. on a state-by-state, or a county-by-county basis but are needed at a small-zone TAZ system, e.g. for metropolitan-level travel forecasting. The procedure below attains this objective using methodology used to estimate the gravity model parameters. The procedure, further, can be thought of as a small-area estimation technique for bi-directional flow data.

Assume that an \( K \times L \) dimensional table of origin-destination flows, \((N_{kl})\), is available for the large TAZ system. A gravity model, \( T_{kl} = A_kB_lF_{kl} \), can them be fit and its parameters \( A_k \)'s, \( B_l \)'s, and \( \theta \)'s estimated. This can be done by using the so-called Modified Scoring procedure to solve the following system of linear equations

\[
\sum_T T_{kl} = \sum_N N_{kl}, \quad \sum_T T_{kl} = \sum_N N_{kl}, \quad \sum_k c_{kl}^{(q)} T_{kl} = \sum_k c_{kl}^{(q)} N_{kl} \quad (\text{Metaxatos, 2004; Sen and Smith 1995}).
\]

The second step involves obtaining exogenous estimates of the row and column sums, \( \sum_T T_{ij} \), \( \sum_N T_{ij} \) (e.g., using small-area estimation techniques), as well as the separation measure(s) \( c_{ij} \) for the small zones. Using the same \( \theta \)'s from the large zones the balancing coefficients \( A_i \)'s, \( B_j \)'s for the small zones can then be estimated using a two-dimensional balancing procedure (known in the literature as the DSF procedure, iterative proportional fitting procedure, and RAS procedure). The iterations for this procedure in each odd (even) iteration are:

\[
A_i^{(2r-1)} = \frac{T_{ij} + \sum_j c_{ij}^{(q)} T_{ij}}{\sum_j B_j^{(2r-2)} F_{ij}} B_j^{(2r)} = \frac{T_{ij} + \sum_i A_i^{(2r-1)} F_{ij}}{\sum_i A_i^{(2r-1)} F_{ij}} \quad \text{with} \quad F_{ij} = e^{\theta c_{ij}}.
\]

For more separation measures, \( c_{ij}^{(q)} \),

\[
F_{ij} = \sum_{q=1}^{Q} e^{\theta c_{ij}^{(q)}}.
\]

The following convergence criterion may be used: \( \sum_{i=1}^{l} | N_{i+} - T_{i+}^{(2r)} | + \sum_{j=1}^{l} | N_{+j} - T_{+j}^{(2r)} | < 10^{-12} \) that requires that the sum of the differences between the observed and estimated row and column marginal sums is close to zero. Upon convergence, \( T_{ij}^{(2r)} = A_i^{(2r-1)} B_j^{(2r)} F_{ij} \), that is the estimated small-scale flows, \( T_{ij} = A_i B_j F_{ij} \).
Traffic Analysis Zones and Scale Issues

Although small-area estimation techniques for locational and bi-directional data exist to allocate data from larger-scale to smaller-scale geographies with a certain level of confidence, Traffic Analysis zone data, as all other locational data, are bound by similar scale issues when they enter into multivariate relationships. Scale issues in relationships have been recognized in sociology (ecological fallacy), statistics (Simpson’s Paradox), geography (Modifiable Areal Unit Problem (MAUP)), and geostatistics (Change of Support Problem (COSP)). The MAUP, in particular, includes the scale effect and the zoning effect. The scale effect occurs when the original data are aggregated and the values for the various univariate, bivariate, and multivariate parameters change because of a loss of information. The zoning effect occurs when each partitioning results in different values for the aggregated statistics.

References

