Advances in Origin-Destination Trip Table Estimation for Transportation Planning and Traffic Simulation

Ramachandran Balakrishna*
Daniel Morgan
Andres Rabinowicz
Howard Slavin

Caliper Corporation

*rama@caliper.com

Abstract

Dynamic transportation models are becoming widespread owing to the ready access to powerful computing resources. This has led to the increasing use of Dynamic Traffic Assignment (DTA) and traffic simulation for a variety of planning applications. Time-varying origin-destination (OD) trip tables are key inputs to such models. However, they are generally not directly observed and must be estimated from other traffic measurements such as counts. We discuss several dynamic OD estimation approaches found in the literature and describe one of the more recent methodologies in detail. This method was tested in the TransModeler traffic simulation package using two examples: a small case study with simulated data, and a real network and count data from the I-5 corridor in California. The results confirm that dynamic OD estimation is hard in practice, but can provide significant improvements over the current state of the practice.

1. Introduction

Dynamic traffic assignment (DTA) and traffic simulation are becoming increasingly popular for transportation planning. These models are perceived to be superior to static approaches due to their potential ability to capture the spatial and temporal evolution of traffic on the network. Capacity constraints further add to their realism so that the minute-by-minute formation and dissipation of queues and spillbacks can be studied. In research investigations, dynamic models have been used for diverse applications such as High Occupancy Vehicle (HOV) and High Occupancy Toll (HOT) lane analyses, congestion pricing, workzone management, evacuation planning and vehicle routing.

While dynamic models have many potential benefits and applications, setting up their numerous inputs and parameters is generally a costly process. Critical among these quantities are time-varying network demand matrices (or trip tables). Each such matrix or table represents the origin-destination (OD) flows departing the respective origin zones during a single time interval. In contrast to static models, however, these time intervals are short, typically of 5, 10 or 15 minute duration. The trips begin and end in different time intervals so that the OD flows from various departure time intervals interact and share the roadway capacity over space and time. In reality, it is impossible to directly measure all the OD flows even in a static trip table, and it is more difficult to measure dynamic tables.

Travel surveys contain useful OD information that can be expanded from the individual trips sampled. However, due to sample size limitations, inferring an OD matrix from travel surveys is a daunting task even with static assumptions, and the addition of the time dimension complicates the process further. Surveys are time-consuming and expensive and cannot be repeated often. Also, only a very small fraction of the total trips in any interval can be reliably sampled this way. Sampling also results in biases when expanding the results to the population level. These issues have motivated numerous studies to infer static or dynamic OD matrices from link count data.

Various sensor technologies can count vehicles, while others attempt to measure their speeds. Loop detectors embedded in the pavement are currently the most popular counting method. Two such loops installed a known distance apart can be used to observe a vehicle’s travel time between the loops, thereby measuring its speed. Other methods include roadside sensors such as Remote Traffic Microwave Sensors (RTMS) and acoustic sensors, as well as point-to-point technologies based on Automatic Vehicle Identification (AVI). If sensors are located on a sufficient number of network links, it may be possible to infer the OD flows that generated the counts on these links. In other words, OD estimation involves the identification of time-varying OD matrices that, when assigned through a dynamic traffic model, accurately mirrors the time-varying observed link counts.

While it seems conceptually simple, dynamic OD estimation is notoriously difficult in practice. There are several reasons for this. The count data are generally prone to measurement errors, and may be inconsistent across space and time as sensors may malfunction for part of the day. Even if these errors were minimal, the spatial coverage of sensors is likely to be very sparse so that the vehicles between several OD pairs are not captured by any of the sensors. There could also be network coding errors that result in unrealistic paths in the assignment model. Perhaps the most significant challenge is the presence of multiple OD solutions that provide similar fit to the observed count data. This has led to significant academic research on OD estimation under a variety of assumptions.

Cascetta (1988) proposes a least squares version of the static OD estimation problem, later expanding the treatment to congested conditions in Cascetta and Postorino (2000). Cascetta et al. (1993) extend the static least squares method to the dynamic case, introducing simultaneous and sequential estimators. While the simultaneous approach estimates OD flows for all departure time intervals at once, it was found to be computationally expensive. The sequential approach was proposed for practical applications, in which only one interval (the current interval) would be estimated at any time. All previous intervals’ estimates, once obtained, were maintained constant. Both approaches rely on the calculation of dynamic assignment matrices that map OD flows to link flows.

© Association for European Transport and contributors 2008

Spiess (1990) presents an iterative, gradient-based procedure for adjusting static OD matrices using the output of a stochastic equilibrium assignment. This approach minimizes the sum of squared errors between the observed sensor flows and the assigned equilibrium link flows. The OD flow variables are updated iteratively by moving along a descent direction by some step size, with the descent direction derived from an analytical gradient evaluated at the current equilibrium solution.

Nielsen (1997) describes single-path and multi-path methods for updating static OD matrices by comparing assigned and observed link flows along the path(s) used by each OD pair. The ratios of observed to assigned flows for each OD pair are collectively used to calculate an updating factor for the flow between that OD pair. Nielsen treats the counts as random variables that can include measurement errors. Hazelton (1999, 2000) suggests a maximum likelihood approach for static OD estimation.

Tavana and Mahmassani (2000) apply the method in Cascetta et al. (1993) to a small, synthetic network and use a simulation-based DTA to calculate assignment matrices. Balakrishna (2002) adopts a similar approach to calibrate OD flows using a network and count data from Irvine, California. This work also jointly calibrates error covariance terms and a route choice model that impacts the DTA's assignment matrix calculations.

Ashok (1996) formulates the dynamic OD estimation problem as a state-space model and solves it using Kalman filters. Antoniou et al. (2007) use Kalman filters to solve for OD flows and other DTA parameters in real-time, while Balakrishna et al. (2007) present a general framework for simultaneously estimating OD flows and other DTA parameters across all time intervals.

The research described above is almost entirely in the realm of academic research. In this paper, we discuss the application of the approach in Balakrishna et al. (2007) to a real-world dynamic OD estimation situation. The rest of the paper is organized as follows. The methodology including a scalable solution approach is summarized in section 2. Numerical test results on a small example and a real-world network are described in sections 3 and 4 respectively. The main conclusions of the estimation work are summarized in section 5.

2. Methodology

Let the study period be divided into *H* departure time intervals, $h = 1,2,...,H$. Let x_h be the column vector of OD flows departing during time interval *h*. Such a vector would be obtained by stacking the columns of an OD matrix. Let y_h denote the column vector of observed link counts measured at the end of interval *h*. Further, let $y = \left[y_1' \quad y_2' \quad \dots \quad y_H' \right]'$ and $x = \left[x_1' \quad x_2' \quad \dots \quad x_H' \right]'$.

The general dynamic OD estimation problem with count data can thus be expressed as an optimization problem:

$$\begin{aligned}&\text{Minimize } z(x) = z_1(\hat{y} - y) + z_2(x - x^a)\\&\text{subject to:}\\&\hat{y} = Assign(x)\\&x \geq 0\end{aligned} \quad (1)$$

where $z_1(\ )$ is a measure of fit between the observed counts y and their modeled counterparts $\hat{y}$ and $z_2(\ )$ measures the "distance" between the OD flow estimates x and some target flows x^a. Target flows are required because the number of equations (i.e. the number of independent sensors) is generally far less than the number of variables (i.e. the number of unknown OD flows). The model counts are obtained as a function of the optimization variables x, from a DTA or a simulator denoted by *Assign*(x). Non-negativity (and other constraints such as loader capacity) must also be imposed on x.

The above formulation is hardly new, yet its solution in this general form can be very difficult. For instance, $\hat{y}$ could be the output of a simulation model. As an example, vehicle counts and speeds in the simulation may be collected at exactly the same link locations where real-world counts are available. Such simulation outputs generally bear complex and non-linear relationships with their input parameters. Consequently, $\hat{y}$ and the objective function $z(x)$ may not be amenable to many of the classical minimization algorithms that require convexity. Analytical gradients will also be hard to obtain in such cases. The solution is further challenged by scalability requirements: the size of the OD estimation problem grows rapidly with the number of zones in the network and the number of time intervals we wish to estimate.

In light of the above challenges, the solution method most often encountered in the literature is based on assignment matrices. An assignment matrix is a linear approximation of *Assign(x)* that maps OD flows (from the current and previous departure time intervals) to the counts in the current interval h:

$$y_h = \sum_{p=h-p'}^{h} a_h^p x_p + v_h \quad (2)$$

where a_h^p maps x_p (the departures in interval p) to y_h (the counts in interval h), and p' is the longest trip length in terms of time intervals. Since the mapping is linear, the solution of an unconstrained least squares formulation is conceptually simple and involves the inversion of a matrix. However the simultaneous estimation of OD flows for all (or even multiple) intervals requires the inversion of a massive block-diagonal assignment matrix (formed from individual a_h^p matrices) that has been shown to be significantly more expensive than inverting just one assignment matrix (Bierlaire and Crittin, 2004; Cascetta and Russo, 1997; Toledo et al., 2003).

The strategy thus has been to estimate the OD flows one interval at a time, or sequentially (rather than simultaneously), starting from h=1. The contribution of previous departure intervals to the current counts are assumed to be fixed, and subtracted from y_h:

$$y_h - \sum_{p=h-p'}^{h-1} a_h^p \hat{x}_p = a_h^h x_h + v_h \qquad (3)$$

Equation (3) assumes that a major share of x_h reaches the sensors within the same (current) interval *h*. This is unrealistic on heavily congested networks, especially when the time intervals are very short.

There are other limitations to the assignment matrix formulation. The solution is far from easy if constraints (non-negativity, lower and upper bounds on the OD flows, etc) are imposed. Next, the assignment matrix is calculated by assigning the current OD solution and tracking vehicle crossings at each sensor. Thus the estimates from each optimization step can be inconsistent with the assignment matrices used in that step, requiring costly fixed-point iterations. Heuristic fixed-point solutions have been attempted by averaging (smoothing) over OD flows, travel times or assignment fractions with mixed results that are difficult to generalize. Further, the estimates of the assignment fractions can be unreliable for low-flow OD pairs.

The preceding discussion highlights the many theoretical advantages of the general formulation. We now present an efficient way to solve the same.

2.1. The SPSA algorithm

The Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm (Spall, 1998, 1999) provides a direct solution to the general formulation in Equation (1). The modeler can thus simultaneously estimate the OD flows of several departure time intervals, while preserving the complex connections between the data (i.e. link counts) and the variables (i.e. OD flows). SPSA is similar to steepest descent in that a new set of OD flows is obtained by moving by a step size along a search direction. However, the search direction is based on a stochastic approximation of the gradient at the current solution.

Each SPSA iteration (denoted by the symbol *i*) consists of two steps.

- First, a search direction at the current solution x^i is calculated as:

$$\hat{g}(x^i) = \frac{z(x^i + c^i \Delta_i) - z(x^i - c^i \Delta_i)}{2c^i} \begin{bmatrix} \Delta_{i1}^{-1} \\ \Delta_{i2}^{-1} \\ \vdots \\ \Delta_{in}^{-1} \end{bmatrix} \qquad (4)$$

  where $\hat{g}(x^i)$ is the gradient approximation at the current solution, Δ_i is a vector of independent random draws, and c^i is a gradient step size. The Bernoulli ± 1 distribution for Δ_i satisfies the theoretical requirements for convergence (Spall, 1999).

- Once a search direction (gradient approximation) is available, all the OD flow variables (from all departure intervals) are updated in a single step by moving a distance of a^i along $\hat{g}(x^i)$:

$$x^{i+1} = x^i - a^i \quad \hat{g}\left(x^i\right) \tag{5}$$

The SPSA algorithm is a variant of the standard Finite Difference Stochastic Approximation (FDSA) approach (Kiefer and Wolfowitz, 1952) in which the gradient is evaluated component-wise. But instead of perturbing each variable in turn and re-running an assignment, SPSA perturbs all components simultaneously (though independently) and approximates the gradient from just two evaluations of the function irrespective of the number of variables, *n*. In comparison, FDSA requires *n*+1 function evaluations for a single gradient calculation (or 2*n*, if each variable is perturbed once on either side). This represents an *n*-fold improvement in per-iteration efficiency, which is significant when each (simulation) assignment can take hours. It should be noted, however, that a few (generally far fewer than 10) gradient replications must be averaged to obtain reliable search directions.

The efficiency of SPSA has another dimension: the number of iterations to convergence. Spall (1998) shows that SPSA and FDSA converge in a similar number of iterations (Figure 1), which preserves SPSA's computational superiority.

SPSA offers other advantages for OD estimation:

- The relationship between the data and the variables is directly captured through *Assign*(*x*), and there are no linear approximations (such as the assignment matrix).
- The assignment method i.e. the function *Assign*(*x*) may be replaced with any model, of any fidelity.
- The objective function can be supplemented with any available measurements, and not just counts.

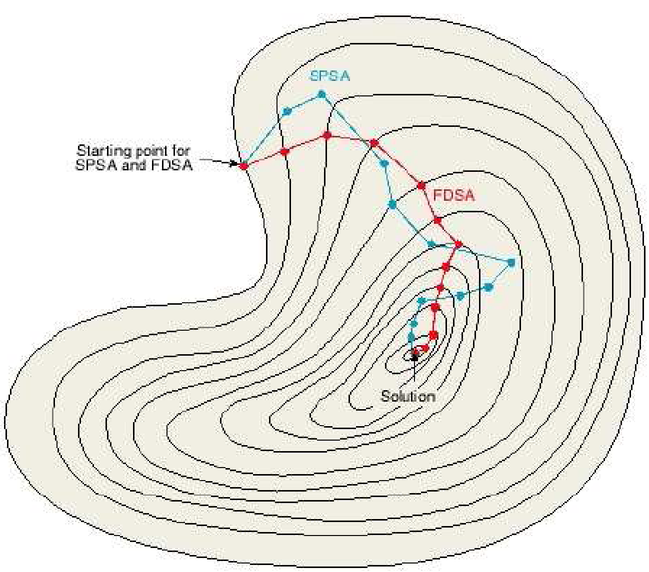

Figure 1: Comparing SPSA and FDSA Convergence [Spall, 1998]

3. Case study I

The network used in the first case study is shown in Figure 2, and includes freeways, signalized arterial/urban streets and ramps. The numbers on Figure 2 indicate the locations of link-wide sensors whose traffic data were used in estimating the demand.

The two-hour study period was arbitrarily designated as 7:00 – 9:00 AM. "True" dynamic OD demand profiles using 15-minute time intervals were created to generate realistic congestions levels on the network. TransModeler's (Yang and Morgan, 2006; Caliper, 2007) microscopic simulation assignment was used to generate a single realization of traffic counts, which were then archived as the observed data.

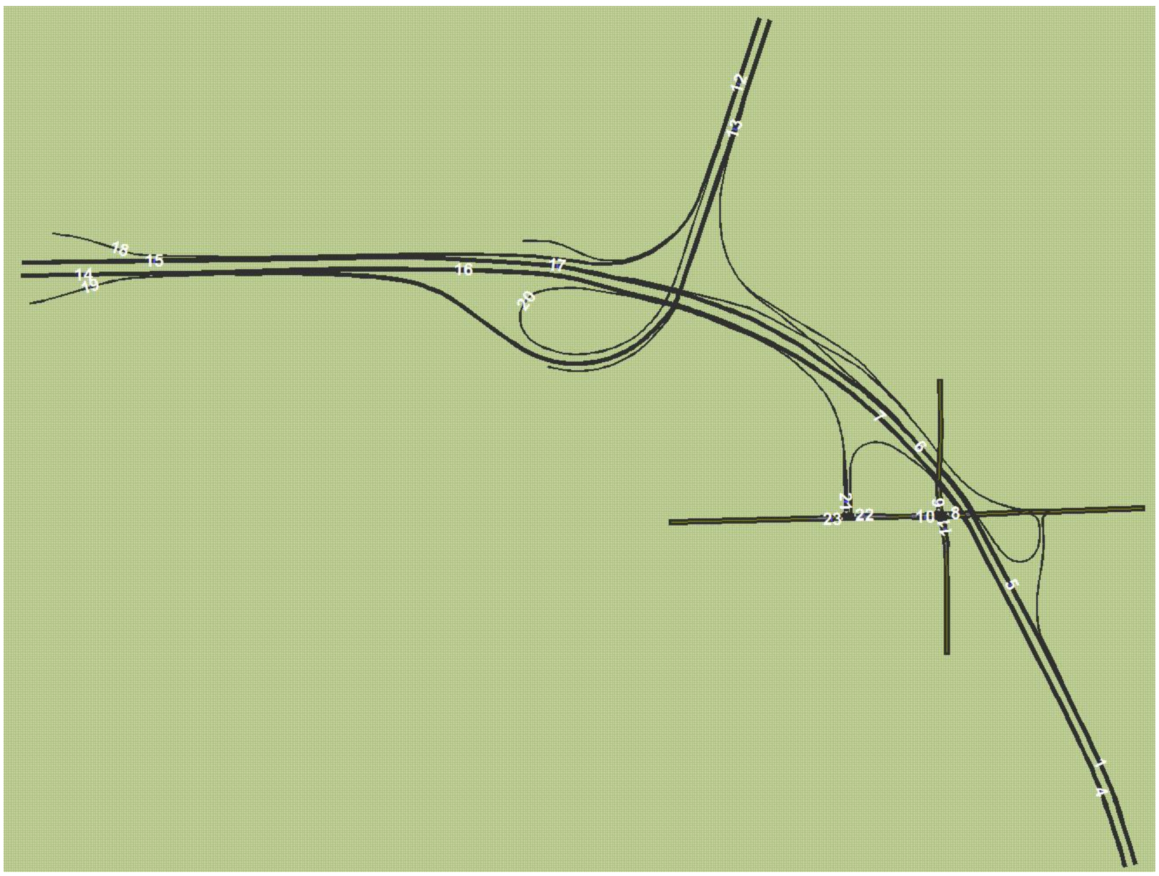

Figure 2: Study Network

The "true" time-dependent OD matrices were perturbed randomly to generate seed matrices to initialize the calibration process. OD flows for all eight departure time intervals were estimated simultaneously. Microscopic assignments were used to evaluate the objective function *z*( ) used by the SPSA-based solution algorithm. SPSA parameters of $a = 0.1$ and $c = 2.0$ were employed. Figure 3 illustrates the progress of the algorithm when each gradient approximation involved 7 replications, with the objective function improving by nearly 83%.

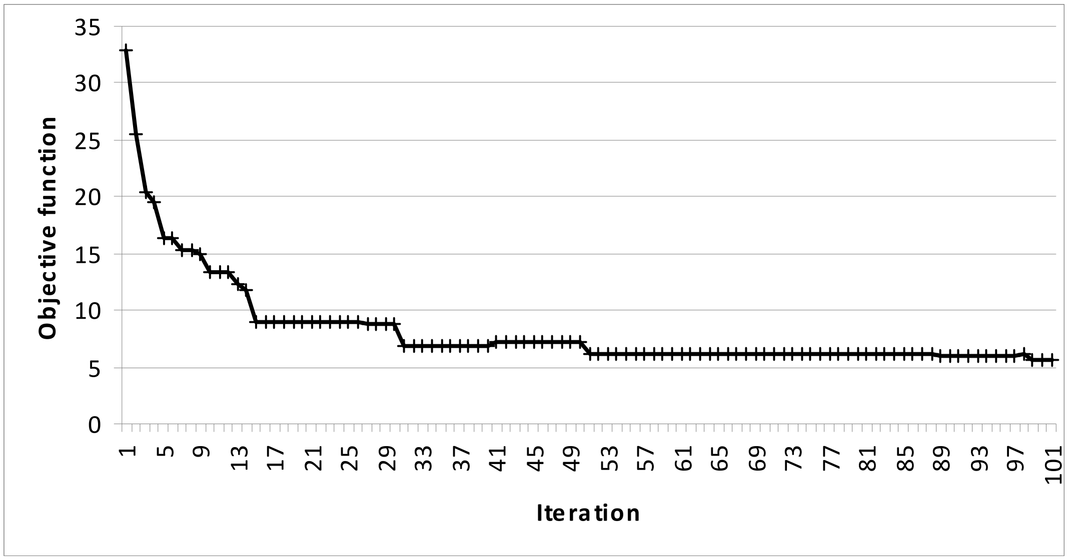

Figure 3: Sample Calibration Progress

The Root Mean Square Normalized (RMSN) statistic was used to evaluate how well the solution approach replicated the time-varying data:

$$RMSN = \frac{\sqrt{N\sum_{n=1}^{N}\left(\hat{y}_n - y_n\right)^2}}{\sum_{n=1}^{N}\hat{y}_n}$$

where N is the number of sensor measurements for each time interval.

Table 1 compares the initial and final fit statistics for each of the eight departure time intervals. It can be seen that the solution approach achieves very good results for most of the intervals (the first couple of intervals must be treated as a warm-up from an empty network).

Interval	1	2	3	4	5	6	7	8
Initial	12.69	21.96	12.16	16.95	9.05	20.48	14.86	11.62
Final	7.07	11.29	5.53	5.37	4.85	5.14	6.88	7.89

Table 1: RMSN Fit Statistics by Time Interval

Multiple calibration runs starting from the same set of inputs showed similar overall performance as depicted in Figure 3 and Table 1. However, it was observed that using fewer gradient replications (5 and 3 instead of 7) showed progressively poorer final solutions, though the RMSN values still remained below 12%.

4. Case study II: I-5 corridor, California

Figure 4 illustrates the I-5 subarea that was used to further study the dynamic OD estimation problem. Figure 4 (a) indicates the region from which the simulation model in Figure 4 (b) was extracted. The subarea is a part of the I-5 corridor through Irvine, California.

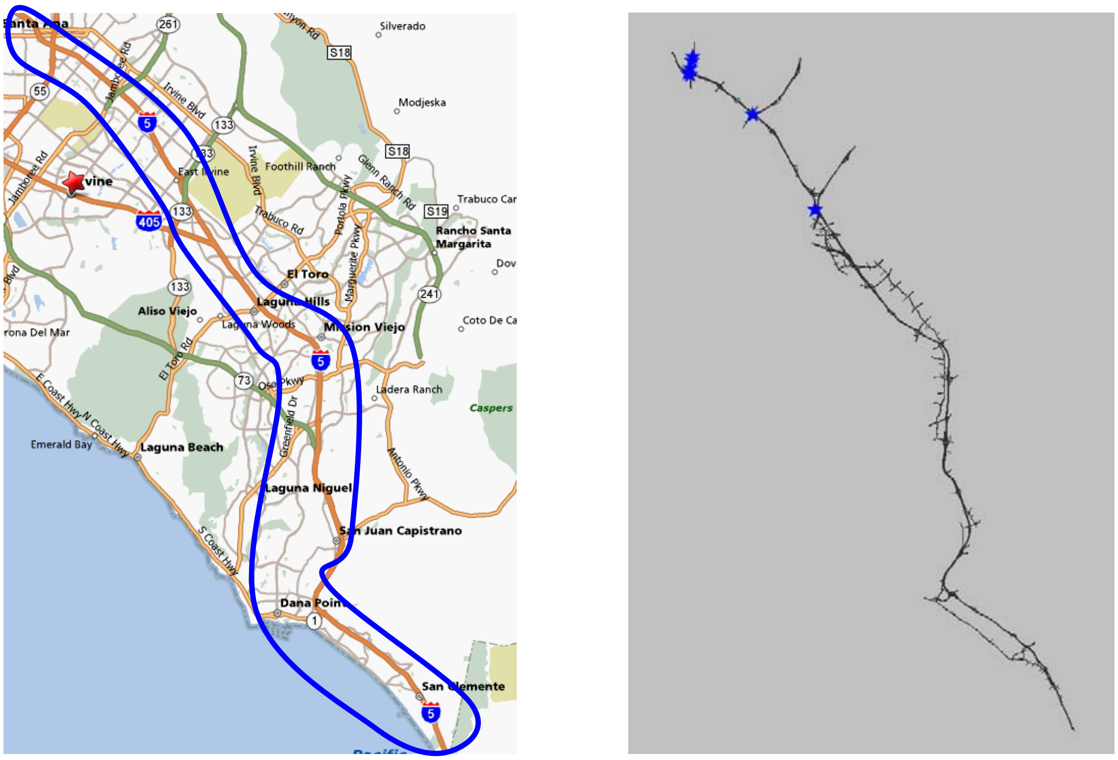

Figure 4. I-5 Subarea from (a) Mapquest and (b) TransModeler

TransModeler's GIS representation of the subarea consists of 760 nodes and 972 links. The links were further divided into 1790 segments to accurately depict linear changes in cross-section (including lane drops and acceleration/deceleration lanes). Demand was loaded between 82 centroid zones and time-varying traffic count data were available from 323 link-wide loop detectors.

The sensor data were extracted for a whole day from the PeMS database maintained by California's PATH program. Time-dependent counts and detector occupancies were obtained for each of the 323 detectors and the counts aggregated to 15-minute values. Speeds at select detectors were imputed from the fundamental diagram for validating certain simulation parameters.

Traffic observed on the road network at any given time is likely to have originated in time intervals prior to the current interval. It is therefore important that the study period's start time coincide with little or no traffic so that boundary effects are minimized. The start time was selected as 3:00 AM based on temporal analyzes of vehicle counts across the study area. The same analyses also revealed that the AM peak was achieved around 7:00-8:00 AM, beyond which the counts typically began to drop. The end of the study period was therefore set at 9:00 AM. This 6-hour period was divided into 15-minute intervals to capture dynamic OD patterns and departure effects. OD estimation was performed with starting matrices derived from a static planning matrix using observed count profiles from sensors located close to trip origins. Two matrices (for single-occupancy and HOV trips) were estimated simultaneously.

Figures 5 through 8 visualize the results of OD estimation using the general formulation and SPSA. It can be seen that the counts at 5:00 AM were reproduced with high accuracy, ensuring a good loaded network for the subsequent estimation. The fit to the data for all subsequent intervals are also adequate, though there is scope for further improvement.

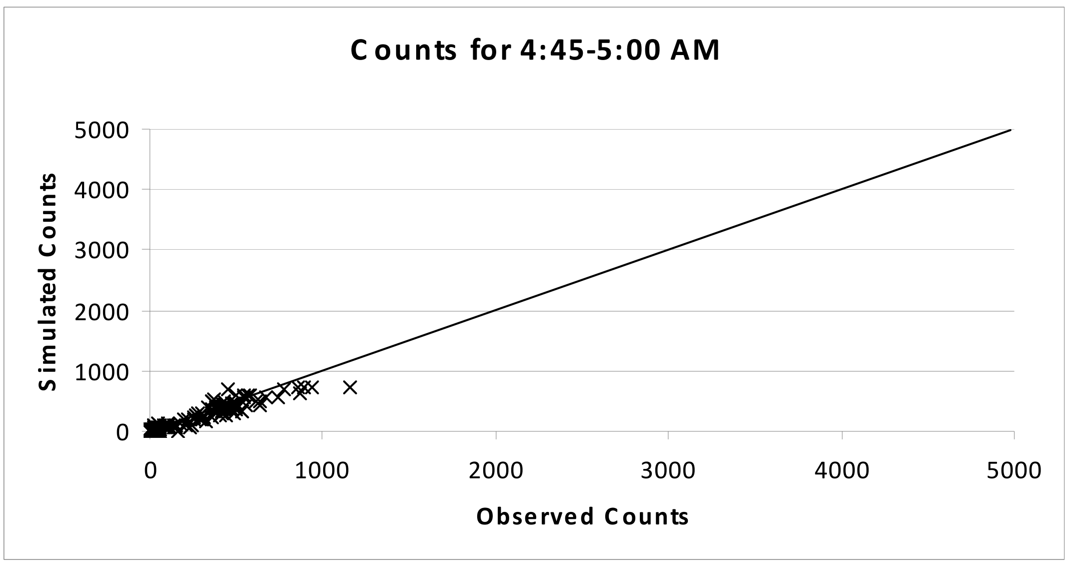

Figure 5. Counts Comparisons at 5:00 AM

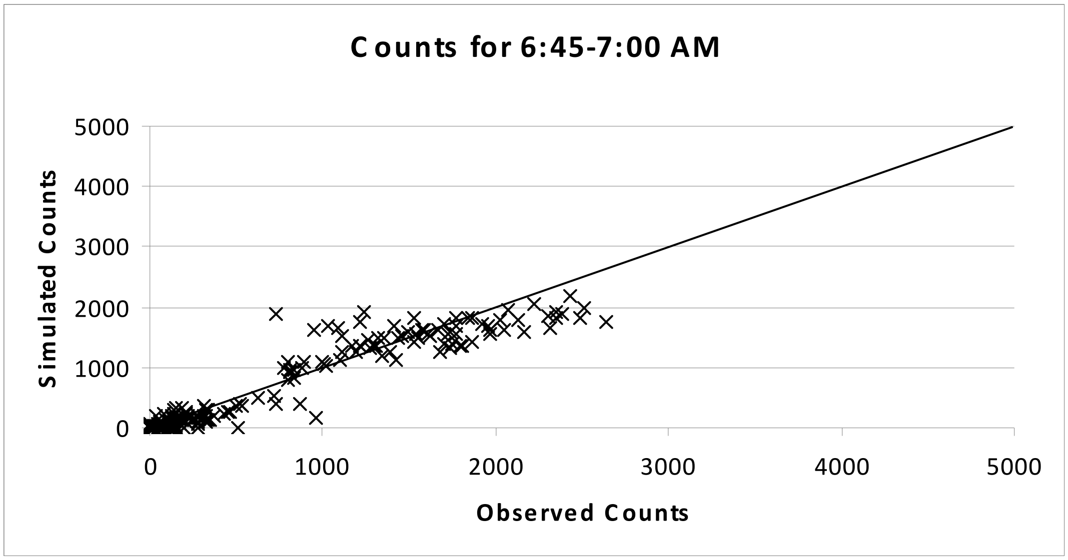

Figure 6. Counts Comparisons at 7:00 AM

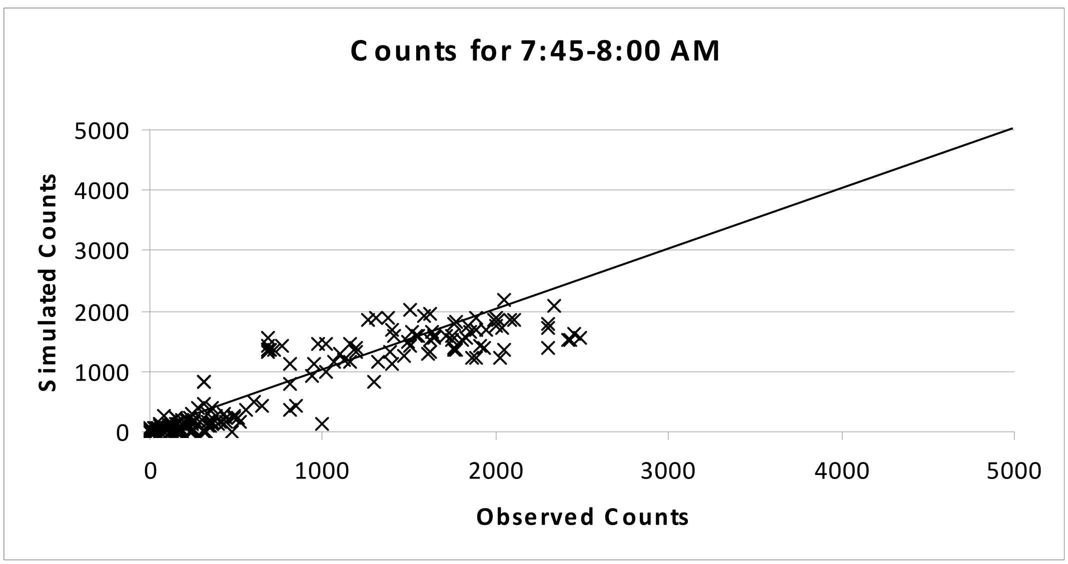

Figure 7. Counts Comparisons at 8:00 AM

© Association for European Transport and contributors 2008

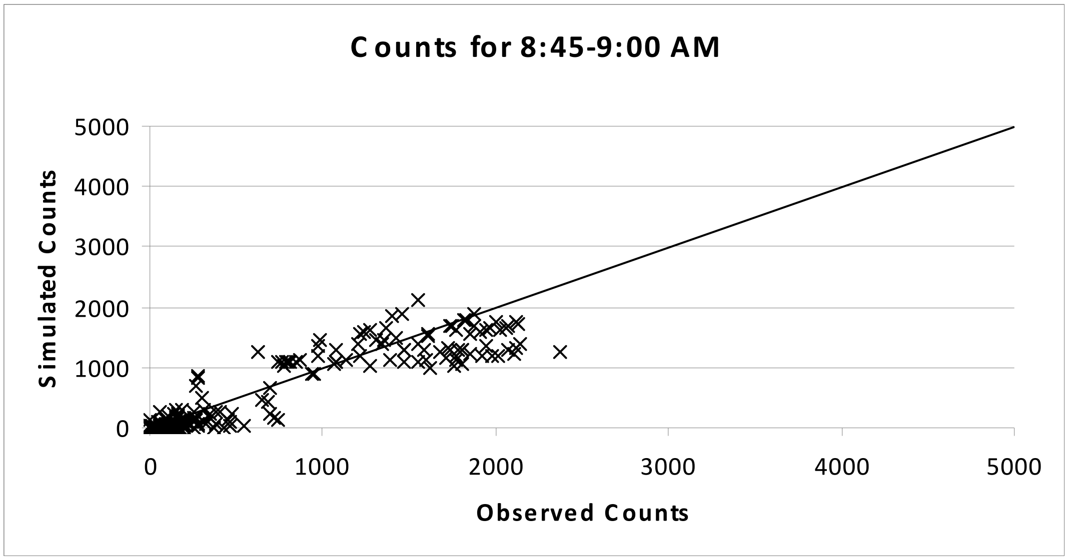

Figure 8. Counts Comparisons at 9:00 AM

5. Conclusion

While sophisticated micro-simulation and dynamic traffic assignment (DTA) models are being developed, their practical use requires the specification of OD demand matrices for each short departure time interval of 5, 10 or 15 minute duration. The estimation of such matrices from traffic data is non-trivial owing to its large-scale and non-linear nature and the numerous sources of errors in the traffic data. While several OD estimation studies are reported in the literature, they remain in the realm of academic research. In this paper we review a recently developed methodology and solution approach for estimating the OD matrices for many time intervals simultaneously using sensor count data, and illustrate the same through two case studies. The results indicate the potential of the approach, while further work is required to ascertain performance with sparser sensor coverage and more complex networks.

References

1. Antoniou, C., M. Ben-Akiva and H. N. Koutsopoulos (2007) Nonlinear Kalman Filtering Algorithms for On-Line Calibration of Dynamic Traffic Assignment Models. IEEE Transactions on Intelligent Transportation Systems, Vol. 8, No. 4.
2. Ashok, K. (1996) Estimation and Prediction of Time-Dependent Origin-Destination Flows. PhD thesis, MIT.
3. Balakrishna, R. (2006) Off-line Calibration of Dynamic Traffic Assignment Models. PhD thesis, MIT.
4. Balakrishna, R., M. Ben-Akiva and H. N. Koutsopoulos (2006) Time-Dependent Origin-Destination Estimation without Assignment Matrices. Second International Symposium on Transport Simulation (ISTS 2006).
5. Bierlaire, M. and F. Crittin. An efficient algorithm for real-time estimation and prediction of dynamic OD tables. *Operations Research*, 2004.

6. Caliper (2007) TransModeler Version 2.0 User's Guide.
7. Cascetta, E., D. Inaudi and G. Marquis (1993) Dynamic Estimators of Origin-Destination Matrices using Traffic Counts. Transportation Science 27(4), pp. 363-373.
8. Cascetta, E. and F. Russo (1997) Calibrating aggregate travel demand models with traffic counts: Estimators and statistical performance. *Transportation*, 24, pp. 271-293.
9. Kiefer, J. and J. Wolfowitz (1952) Stochastic Estimation of the Maximum of a Regression Function. Annals of Mathematical Statistics.
10. Nielsen, O. A. (1997) Multi-Path OD-Matrix Estimation (MPME) based on Stochastic User Equilibrium Traffic Assignment. 76th annual meeting of the Transportation Research Board.
11. Spall J. C. (1998) Implementation of the Simultaneous Perturbation Algoritm for Stochastic Approximation, IEEE Transactions on Aerospace and Electronic Systems, Vol. 34(3).
12. Spall, J. C. (1999) Stochastic Optimization, Stochastic Approximation and Simulated Annealing. In Webster, J. G. (ed) Wiley Encyclopedia of Electrical and Electronics Engineering, Wiley-Interscience, pp. 529-542.
13. Spiess, H. (1990) A Gradient Approach for the O-D Matrix Adjustment Problem. Centre de Recherche sur les Transports de Montreal, Publication 693.
14. Tavana, H. and H. S. Mahmassani (2000) Estimation of Dynamic Origin-Destination Flows from Sensor Data using Bi-level Optimization Method. 80th annual meeting of the transportation research board.
15. Toledo, T., H. N. Koutsopoulos, A. Davol, M. E. Ben-Akiva, W. Burghout, I. Andreasson, T. Johansson and C. Lundin (2003) Calibration and Validation of Microscopic Traffic Simulation Tools: Stockholm Case Study. Transportation Research Record: Journal of the Transportation Research Board, No. 1831, pp. 65-75.
16. Yang, Q. and D. Morgan (2006) A Hybrid Traffic Simulation Model. 85th annual meeting of the Transportation Research Board.

© Association for European Transport and contributors 2008