A Robust Algorithm of Estimating Dynamic O-D Matrix for Large Freeway Networks with Measurement Errors

Pei-Wei Lin and Gang-Len Chang
Dep. of Civil and Environmental Engineering, Univ. of Maryland, College Park, MD 20742,
pwlin@wam.um.edu

This paper is submitted to the 9th International Conference on Applications of Advanced Technology in Transportation

Abstract

This study presents a robust algorithm that can deal with the incomplete volume information so as to significantly improve the estimation accuracy. To tackle the inevitable measurement data error or only partially available information, the proposed robust algorithm converts each model input data into one interval with its upper and lower bounds best approximated from historical data or/and prior knowledge. A simulated system, the I-95 freeway corridor between I-495 and I-695, has been created to generate example data and to perform the numerical evaluation of the developed robust algorithm.

Introduction

Due to the increasing needs of using the O-D information for a variety of planning and operation studies, transportation researchers over the past decades have devoted considerable efforts in developing effective approaches for reliably estimating O-D demands either in a static or a time-varying format. Most existing advanced approaches [1-7, 10-13, 15-18, 20-21] for such a need require the use of a set of initial/prior O-D and ramp/mainline volumes as model inputs. Depending on the available information and the network structure, the O-D patterns estimated with those approaches may result in a large variance, and insufficient reliability to use in practice. Besides, many of those essential data may not be available in a real world network. Hence, it is imperative that any developed system for such applications be sufficiently robust to accommodate to real world operational constraints.

In general, link volume data are assumed to be available and do not contain measurement errors. The quality of such information, however, may impact the estimation accuracy. The assumption of having accurate link volume data is often subject to challenge, as a large amount of traffic volumes from detectors for dynamic O-D estimation are constantly suffering from the hardware quality deficiency. The issue of having unreliable link volume data is especially critical for a large network as its formulations for O-D estimation generally contain a significant number of unknown system parameters. Depending on the information availability and the network structure, the O-D patterns estimated with those approaches may result in a large variance, and insufficient reliability for use in practice [15]. To contend with such deficiencies, this study develops a dynamic O-D model that employs an interval-based estimation algorithm to account for measurement variances in network link volumes, which may vary significantly from their mean volumes used in the existing O-D estimation models.

The remaining of the paper is organized as follows. A basic dynamic O-D estimation model for a freeway corridor is presented in the next section. Section 3 illustrates an interval-based solution algorithm to contend with input volume variance. Extensive numerical analyses for evaluating the effectiveness of the proposed solution algorithms are presented in Section 4. Conclusions and further enhancements are summarized in the last section.

A Basic Model Formulation for Dynamic O-D Estimation

Consider a typical freeway corridor with link count information as shown in Figure 1, where detectors are deployed at on-ramps, off-ramps and mainline links. The information that is readily available for estimation of its time-dependent O-D flow proportion or dynamic O-D distribution is the time series of entering flow, $q_i(k)$, exiting flow, $y_j(k)$, and mainline flow, $U_l(k)$.

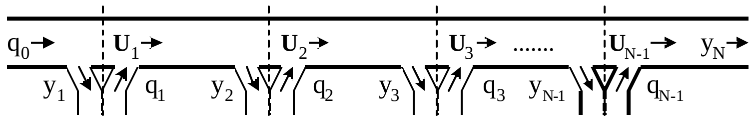

Figure 1. A Typical Freeway Corridor

Let $b_{ij}(k)$ denote the proportion of vehicles entering from origin i to destination j during time interval k. By definition, it is subject to the following two natural constraints:

$$0 \le b_{ij}(k) \le 1, \qquad 0 \le i < j \le N \tag{1}$$

$$\sum_{j=i+1}^{N} b_{ij}(k) = 1, \qquad i = 0, 1, ..., N-1 \tag{2}$$

As shown in Figure 1, the freeway corridor consists of N segments from 0 to N-1 and the set of variables used in modeling the dynamic traffic flow and O-D relationships is defined as follows:

$q_0(k)$: The number of vehicles entering the upstream boundary of the freeway section during time interval k.
$q_i(k)$: The number of vehicles entering freeway from on-ramp i during time interval k, i = 1, 2, …, N-1.
$y_j(k)$: The number of vehicles leaving freeway from off-ramp j during time interval k, j = 1, 2, …, N-1.
$y_n(k)$: The mainline volume at the downstream end of the freeway section during time interval k.
$U_i(k)$: The number of vehicles crossing the upstream boundary of segment i during time interval k, i = 1, : 2, …, N-1.
$T_{ij}(k)$: The number of vehicles entering the freeway from on-ramp i during time interval k that are destined to off-ramp j (i.e., the time-dependent O-D flow), where $0 \le i < j \le N$.
t_0 : The length of one unit time interval.
$t_{ij}(k)$: The average travel time from on-ramp i to off-ramp j departing during time interval k.
$\sigma_{ij}(k)$: The standard deviation of the travel time for vehicles traveling from on-ramp i to off-ramp j : departing during time interval k.
$b_{ij}(k)$: The proportion of $q_i(k)$ heading toward destination node j during time interval k.
$\alpha^m_{ij}(k)$: The fraction of $T_{ij}(k-m)$ vehicles departing from entry node i during time interval k that takes m time intervals to exiting node j.
$\alpha^m_{ilj}(k)$: The fraction of $T_{ij}(k-m)$ trips from entry node i during time interval k that takes m time intervals to mainline node *l*.

Lin and Chang [15] assumed that the travel time of drivers departing from node i during time interval k to node j follows a normal distribution, i.e., $N[\bar{t}_{ij}(k), \sigma^2_{ij}(k)]$. Since the travel time for an O-D pair departing during the same time interval follows a normal distribution, $\alpha^m_{ij}(k)$ can be replaced with the following cumulative density function of a time interval, m.

$$\alpha^m_{ij}(k) = \int_{m t_0}^{(m+1)t_0} f_{ij}(x)dx, \; 0 \le \alpha^m_{ij}(k) \le 1, \; m = 0, 1, ..., M \tag{3}$$

$$\sum_{m=0}^{M} \alpha^m_{ij}(k-m) = 1, \; 0 \le i < j \le N, \; m = 0, 1, ..., M \tag{4}$$

where $f_{ij}(x)$ is the density function of the travel time distribution with mean $\bar{t}_{ij}(k)$ and standard deviation $\sigma_{ij}(k)$. By applying the above travel time distribution concept, the relationships between ramp volumes and O-D proportions, and the relationships between mainline volumes and O-D proportions can be written as:

$$y_j(k) = \sum_{m=0}^{M}\sum_{i=0}^{j-1} q_i(k-m)\alpha^m_{ij}(k)b_{ij}(k-m) \tag{5}$$

$$U_l(k) = q_l(k) + \sum_{m=0}^{M}\sum_{i=0}^{l-1}\sum_{j=l+1}^{N} q_i(k-m)\alpha^m_{ilj}(k)b_{ij}(k-m) \tag{6}$$

$$t_{ilj}(k) \sim N[?_{ilj}\, ?\bar{t}_{ij}(k), ?^2_{ilj}\, ??^2_{ij}(k)] \tag{7}$$

$$?_{ilj}\ ?\ d_{il} / d_{ij} \tag{8}$$

where $?_{ilj}$ is the ratio of the distance d_{il}, to the distance d_{ij}. In addition to the use of a normal distribution to capture the variation of travel times among drivers of the same departure time, one can also estimate the average $\bar{b}_{ij}(k)$ for consecutive intervals, instead of solving the O-D flow distribution matrix for each small interval [7]. This is proposed to reduce the number of unknown parameters. With this enhancements, one can estimate the average travel time for each O-D pair with data provided by a surveillance system, and let the O-D proportions, $\bar{b}_{ij}(k)$, and standard deviations,$?_{ij}(k)$ be the unknown set of parameters.

An Interval-Based Algorithm for Dynamic O-D Estimation

This section presents an enhanced solution algorithm that can estimate an O-D matrix under the circumstance that its traffic volumes are subjected to some level of measurement errors. In most existing approaches, the unknown parameters, $b_{ij}(k)$ and $?_{ij}(k)$, are assumed to follow the random walk process: $b_{ij}(k+1) = b_{ij}(k) + w_{ij}(k)$ and $?_{ij}(k+1) = ?_{ij}(k) + v_{ij}(k)$, where the random terms, $w_{ij}(k)$ and $v_{ij}(k)$ are independent Gaussian white noise sequences with zero mean and covariance.

To facilitate the formulations, the variables are defined as: $\mathbf{b(k)} = [b_{ij}(k)]$, $?\mathbf{(k)} = [?_{ij}(k)]$, $\mathbf{W(k)} = [w_{ij}(k]$, and $\mathbf{V(k)} = [v_{ij}(k]$. Thus, the matrix forms are as follows:

$$\mathbf{b(k+1)}=\mathbf{b(k)}+\mathbf{W(k)} \text{ and } ?\mathbf{(k+1)}=?\mathbf{(k)}+\mathbf{V(k)} \tag{9}$$

With the above refinements for $\mathbf{b(k)}$ and $?\mathbf{(k)}$, Equations (5) and (6) can be restructured into the following matrix form:

$$\mathbf{Z(k)}=\mathbf{H}[?_{\mathbf{ij}}\mathbf{(k)}]\cdot\mathbf{b(k)}+\mathbf{e(k)} \tag{10}$$

where $\mathbf{Z(k)}\ ?\ [y_1(k), y_2(k),..., y_N(k); U_1(k)\ ?\ q_1(k),..., U_{N?1}(k)\ ?\ q_{N?1}(k)]^T$, $\mathbf{H}[?_{\mathbf{ij}}\mathbf{(k)}]\ ?\ [h^k_{rs}]$ is a matrix with its entries given by the corresponding coefficients in Equations (5) and (6), and $\mathbf{e(k)}$ is an observation noise vector, where $\mathbf{R}$=Var[e(k)]=diag[r_1,…,r_{2N-1}] is a diagonal positive definite matrix. Through Equations (9)-(10), a canonical state-space system model has been established. Such a system equation can be solved with any solution algorithm, such as the Kalman filtering or the minimum least squares estimation algorithms.

The matrices $\mathbf{Z(k)}$ and $\mathbf{H(k)}$ in Equation (10) consist of the time-series measurement of traffic volumes, $q_i(k)$, $y_i(k)$ and $U_i(k)$. If the traffic volume information is not precisely known due to measurement errors and progressively changes over time, most existing solution algorithms will not be applicable. The interval-based solution algorithm presented hereafter is proposed for contending with such a critical issue. Suppose that these two uncertain matrices of $\mathbf{Z(k)}$ and $\mathbf{H(k)}$ are only known to be bounded, one then can rewrite their relations as follows:

$$\mathbf{Z^I(k)}=[\mathbf{Z(k)}-|?\,\mathbf{Z(k)}|, \mathbf{Z(k)}+|?\,\mathbf{Z(k)}|] \tag{11}$$

$$\mathbf{H^I(k)}=[\mathbf{H(k)}-|?\,\mathbf{H(k)}|, \mathbf{H(k)}+|?\,\mathbf{H(k)}|] \tag{12}$$

where $|?\,\mathbf{Z(k)}|$ and $|?\,\mathbf{H(k)}|$ are the positive constant bounds for those unknowns. Hence, Equation (10) can be written as an interval representation:

$$\mathbf{Z^I(k)}=\mathbf{H^I(k)}\cdot\mathbf{b(k)}+\mathbf{e(k)} \tag{13}$$

Equation (13) along with Equation (9) forms a generalized state space system, in which the volumes are treated as intervals rather than constants. To apply the interval-based solution algorithm for the dynamic O-D estimation, the most challenge part is to ensure that the natural constraints of the system model as shown in Equations (1) and (2) can be held during the recursive computing process. As reported in the literature [16], most studies employ the methods of truncation and normalization to tackle this difficult issue. To incorporate the above concepts in developing the algorithm for the proposed formulations, one needs to modify the truncation and normalization process as follows:

✍ Modified Truncation Process

Note that the purpose of the truncation is to find the largest step of improvement for the unknown parameters b and ? for the next time interval so that the O-D proportions, b_{ij}, can still satisfy Equation (1). To incorporate this concept in developing the solution algorithm, it is essential that the truncation process be restated as follows:

$$?\begin{matrix} b^I_n \\ ?^I_n \end{matrix}? \; ? \; ?\begin{matrix} b^I_{n?1} \\ ?^I_{n?1} \end{matrix}? \; ? \; ? ??^I_n \, ?g^I_n, \text{ where } ? ?? \underset{0??1}{MAX}[? \mid 0 ? [b^I_{n?1}] ? \; ? \; ??^I_n \, ?g^I_n \; ? \; 1] \qquad (14)$$

where ? is a value, which represents the largest step of improvement so that all the possible values of b_{ij} can satisfy Equation (1).

✍ Modified Normalization Process

The normalization step is employed to satisfy Equation (2), which means that the sum of the O-D proportions from the same origin is equal to one. This constraint is not applicable to the interval formulations. Since b^I_{ij} is an interval, many possible combinations could satisfy this constraint. It is not necessary that either the sum of the lower bound or the sum of the upper bound be equal to one as there exists several possible combinations. In such a case, the estimated O-D proportions are valid as long as there exist combinations that can satisfy the constraint. Hence, one can formulate such relationships with the following equations:

$$0 \; ? \; \sum_{j?m?1}^{N} \underline{b}^{n?1}_{mj} \; ? \; 1, \qquad i = 0, 1, \ldots, N\text{-}1 \qquad (21)$$

$$\sum_{j?m?1}^{N} \overline{b}^{n?1}_{mj} \; ? \; 1, \qquad i = 0, 1, \ldots, N\text{-}1 \qquad (22)$$

The Equation (15) is to ensure that the sum of the lower bound, $\underline{b}^{n?1}_{mj}$, for O-D proportions with the same origin is equal to or less than one, and Equation (16) is to let the sum of the upper bound, $\overline{b}^{n?1}_{mj}$, for O-D proportions with the same origin be equal or larger than one. If all the O-D proportions satisfy these two constraints, it is guaranteed that there exists at least one combination that can satisfy Equation (2).

With Equations (9) and (13), one can construct a dynamic model for O-D estimation that takes into account potential measurement errors. This formulation system treats the traffic volumes as intervals so it consists of the upper and lower bound in the system equations. To solve such formulations, one may simply employ the standard solution algorithm, such as the standard Kalman filtering algorithm, for these two system equations. However, the results with those existing estimation algorithms do not encompass all possible optimal solutions for the interval system formulations. Hence, this study employs the *interval Kalman filtering scheme* [8] to develop an interval-based solution algorithm.

Due to the nonlinear nature of the formulations and the concern of computing efficiency, this study has employed sequential extended Kalman filtering algorithm [9] and Gumbel distribution (an approximation of normal distribution) to develop a solution algorithm. A major part of the interval-based solution algorithm for the formulations of Equations (9) and (13) with the Interval Kalman Filtering method is presented below:

For j=1, 2, …, 2N-1

✍ $g^I_j \; ? \; P^I_{j?1}(f^I_j)^T[f_j P_{j?1}(f_{j?1})^T \; ? \; r_j]^{?1}$

✍ $P^I_n \; ? \; P^I_{n?1} \; ? \; g^I_n f^I_n P^I_{n?1}$

✍ $?^I_{n?1} \; ? \; y^I_n(k) \; ? \; f^I_n b^I(k \; ? \; 1)$

✍ $? \; ? \; ? \underset{0??1}{MAX} ?? \mid (14) \sim (16)?$, set $?\begin{matrix} b^I_n \\ ?^I_n \end{matrix}? \; ? \; ?\begin{matrix} b^I_{n?1} \\ ?^I_{n?1} \end{matrix}? \; ? \; ? ??^I_n \, ?g^I_n$

Numerical Examples

This section presents the numerical evaluation results of the proposed algorithm using the I-95 northbound freeway corridor between two major beltways, I-495 and I-695. For computing efficiency, each interchange is represented with a pair of on-ramp and off-ramp, and the network is thus reduced to 7 pairs of on-ramps and off-ramps, and 36 O-D sets as shown in Figure 2.

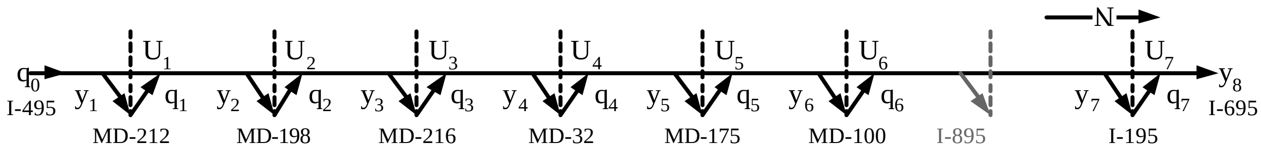

Figure 2. A Graphical Illustration of the Main Interchanges for the I-95 Corridor

To generate a meaningful data set for numerical analysis, the example freeway system under the assigned time series O-D percentages was simulated with AIMSUN 4.0 [19], to produce time-dependent link traffic volumes. The simulation was executed for one hour using the dynamic O-Ds at an interval of 2 minutes. Following the simulation, one can obtain volume information for each link and the average travel time information for each O-D pair. In this example, the ramp/mainline volumes are added with measurement errors for up to 10%. To conduct an evaluation for the proposed algorithms, a series of time-varying O-D sets estimated with the extended Kalman filtering algorithm (i.e., Algorithm-1) presented by Lin and Chang [14] is set as a basis to compare with the results from the proposed interval-based algorithm (i.e., Algorithm-2). Figure 3 illustrates the graphical results from Algorithm-1 and Algorithm-2 along with the true O-D set. As shown in Figure 3, the interval-based algorithm (Algorithm-2) outperforms Algorithm-1.

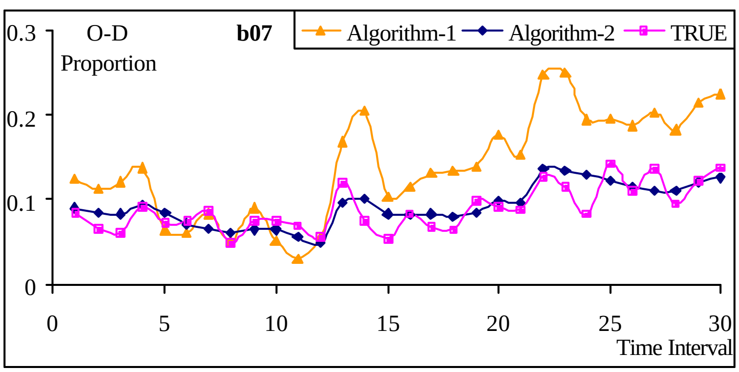

Figure 3. The Estimation Results for Scenario-2 (b07)

Table 1 shows the statistical results, of which the absolute error is computed for each O-D pair in every time interval. In addition to the time-average absolute error for every O-D pair, the maximum and minimum are chosen as the evaluation criteria. As shown in Table 1, the proposed algorithm (Algorithm-2) outperforms the commonly used algorithm (Algorithm-1) with the improvements of more than 30%.

Table 1. The Statistical Results and Improvements

Algorithm-1			Algorithm-2			Improvement		
Average	Max.	Min.	Average	Max.	Min.	Average	Max.	Min.
0.0679	0.1433	0.0225	0.0432	0.0968	0.0121	**36.38%**	**32.45%**	**46.22%**

Conclusions and Recommendations

This study has developed a robust algorithm to contend with the inevitable measurement errors embedded in most traffic sensors. This study has also performed the evaluation of the proposed algorithm using a large freeway network, the I-95 freeway corridor between I-495 and I-695 in Maryland. The results have revealed that the estimated time-varying O-D proportions for this example freeway corridor are more reliable due to the use of the interval-based algorithm. One of the critical issues remains to be investigated in the area of dynamic O-D estimation is to compute the reliability level a model can provide with the available information.

References

1. Ashok, K., and Ben-Akiva, M.E., "Dynamic Origin-Destination Matrix Estimation for Real-Time Traffic Management Systems," *Trans. and Traffic Theory*, pp. 465-484, 1993.
2. Bell, M.G.H., "The Estimation of Origin-Destination Matrices by Constrained Generalized Least Squares," *Trans. Research B*, Vol. 25, No. 1, pp. 13-22, 1991a.
3. Bell, M.G.H., "The Real-Time Estimation of Origin-Destination Flows in the Presence of Platoon Dispersion," *Trans. Research B*, Vol. 25, No.2/3, pp. 115-125, 1991b.
4. Cascetta, E. "Estimation of Trip Matrices from Traffic Counts and Survey Data: A Generalized Least Squares Estimator," *Trans. Research B*, Vol. 18 No. 4/5, pp. 289-299, 1984.
5. Cascetta, E. Inaudi, D., and Marguis, G., "Dynamic Estimators of Origin-Destination Matrices using Traffic Counts, " *Trans. Science*, Vol. 27, pp. 363-373, 1993.
6. Chang, G-L, and Tao, X., "An Integrated Model for Estimating Time-Varying Network Origin-Destination Distributions," *Trans. Research A*, Vol. 33, No. 5, pp. 381-399, 1999.
7. Chang, G-L, and Wu, J., "Recursive Estimations of Time-Varying Origin-Destination Flows from Traffic Counts in Freeway Corridors," *Trans. Research B*, Vol. 28, No. 2, pp. 141-160, 1994.
8. Chen, G., Wang, J., and Ghieh, L. S., "Interval Kalman Filtering," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 33, No. 1, pp. 250-258, 1997.
9. Chui, C. K, and Chen, G., *Kalman Filtering with Real-time Applications*, Springer-Verlag, 1999.
10. Cremer, M., and Keller, H., "A New Class of Dynamic Methods for the Identification of Origin-Destination Flows," *Trans. Research B*, Vol. 21, No. 2, pp. 117-132, 1987.
11. Cremer, M., and Keller, H., "A Systems Dynamics Approach to the Estimation of Entry and Exit O-D Flows," *the Pro. 9th Inter. Symposium on Trans. and Traffic Theory*, pp. 431-450, 1984.
12. Hazelton, M., "Estimation of Origin-Destination Matrices from Link Flows on Uncongested Networks," *Trans. Research B*, Vol. 34, No. 7, pp. 549-566, 2000.
13. Keller, H., and Ploss, G. "Real-time Identification of O-D Network flow from Counts for Urban Traffic Control," *Trans. and Traffic Theory*, pp. 267-284, 1987.
14. Lin, P. W., and Chang, G-L, "A Decomposition Algorithm for Estimating the Initial O-D Matrix in a Large Freeway Network," Submitted to *the IEEE Transactions on Intelligent Transportation Systems*, September 2005.
15. Lin, P. W., and Chang, G-L, "A Robust Model for Estimating Freeway Dynamic Origin-Destination Matrix," Accepted in *Trans. Research Record*, 2005.
16. Nihan, N.L., and Davis, G. A., "Recursive Estimation of Origin-Destination Matrices from Input/Output Counts," *Trans. Research B*, Vol. 21, No. 2, pp. 149-163, 1987.
17. Sherali, H. D., and Park, T., "Estimation of Dynamic Origin-Destination Trip Tables for a General Network," *Trans. Research B*, Vol. 35, No. 3, pp. 217-235, 2001.
18. Tavana, H. and Mahmassani, H., "Estimation of Dynamic Origin-Destination Flows from Sensor Data Using Bi-level Optimization Method," *the Pro. 80th TRB Annual Meeting*, 2000.
19. TSS-Transportation Simulation Systems, *GETRAM Operation Traffic Simulation Evrironment* – TEDI Version 4.0 User Manual, 2001.
20. Wu, J., and Chang, G-L, "Estimation of Time-Varying Origin-Destination Distributions with Dynamic Screenline Flows," *Trans. Research B*, Vol. 30, No. 4, pp. 277-290, 1996.
21. Zhou, X., and Mahmassani, H. S., "Dynamic Origin-Destination Demand Estimation Using Automatic Vehicle Identification Data," *the Pro. 84th TRB Annual Meeting*, 2005.