Examining the Quality of Synthetic Origin–Destination Trip Table Estimated by Path Flow Estimator

Anthony Chen[1]; Piya Chootinan[2]; and Will W. Recker[3]

Abstract: Path flow estimator (PFE) is a one-stage network observer proposed in the transportation literature to estimate path flows and path travel times from traffic counts in a transportation network. The estimated path flows can further be aggregated to obtain the origin–destination ($O-D$) flows, which are usually required in many transportation applications. In this paper, we examine the capability of PFE in capturing the total demand of the study network as well as individual $O-D$ demands. Numerical examples are provided to show the effects of the number and locations of traffic counts on the quality of $O-D$ estimates. The results indicate that PFE has the potential to correctly estimate the total demand when proper observations, in terms of the number and their locations, are provided. In general, the spatial distribution of $O-D$ demands is difficult to estimate even when traffic counts are available on all network links.

DOI: 10.1061/(ASCE)0733-947X(2005)131:7(506)

CE Database subject headings: Transportation networks; Traffic analysis; Trip distribution; Traffic assignment; Simulation models.

Introduction

The origin–destination ($O-D$) trip table is one of the key components required for traffic assignment and simulation models, which are used to analyze a wide variety of transportation applications. These applications range from long-term transportation planning to short-term transportation management on a daily basis. Traditionally, the $O-D$ trip table is estimated through large-scale surveys such as household survey, roadside interviews, and license plate matching, etc. These survey-based techniques are expensive, time consuming, and labor intensive. In addition, if there is a rapid change in land-use pattern, the estimated trip table will soon become outdated. Therefore, the survey-based techniques, sometimes, are not considered appropriate due to financial and time constraints. Accordingly, the need to develop fast and inexpensive methods using readily available traffic counts for estimating the $O-D$ trip table has motivated researchers to devote significant effort in this important topic during the past several decades.

Research approaches on the $O-D$ trip table estimation problem from traffic counts can be broadly classified based on network configurations and route choice assumptions to: (1) simple networks with no route choice; (2) networks with route choice but no congestion; and (3) general networks with both route choice and congestion. These categories can further be classified to as static and dynamic $O-D$ estimations in which different treatments of time dependencies are applied. Examples of simple networks without route choice include finding turning fractions at an intersection (Van Zuylen 1979; Bell 1984) and determining split ratios for a freeway system with several on and off ramps (Ashok and Ben-Akiva 1993; Nanne et al. 1994; Madanat et al. 1996; Sherali et al. 1997). For uncongested networks with route choice, proportional assignment could be used for estimating the $O-D$ trip table. Route choice proportions are assumed to be independent of the estimation process, and can be obtained based on the observed travel times (or travel speed) and appropriate route choice assumptions. The statistical methods (distance measures) used in the estimation include the generalized least squares (Cascetta 1984; O'Neil 1987; Bell 1991; Ashok and Ben-Akiva 2000, 2002; Halzeton 2000; Bierlaire and Crittin 2004), entropy maximization (Van Zuylen and Willumsen 1980), maximum likelihood (Spiess 1987), and Bayesian inference (Maher 1983). For general networks with both route choice and congestion, the proportional assignment assumption is no longer valid because route choice and $O-D$ trip tables are interdependent (Bell and Iida 1997). It is necessary to incorporate a route choice model into the estimation process. An approach based on bilevel programming by endogenously determining the route choice proportions while estimating the $O-D$ trip table is one of the possible solutions to ensure this interdependency (Fisk 1988, 1989; Yang et al. 1992; Maher et al. 2001). In the bilevel programming approach, the upper-level problem uses one of the statistical techniques mentioned earlier (e.g., generalized least squares) to select the most proper $O-D$ trip table, whereas the lower-level problem endogenously determines the route choice proportions (e.g., deterministic user equilibrium or stochastic user equilibrium) compatible with the estimation. Although the interdependency issue is resolved, the bilevel programming approach could pose computational difficulties when estimating $O-D$ trip tables for large-scale networks.

The analytical and computational difficulties of bilevel programming can be relaxed by making the following assumption. If a complete set of traffic counts constitutes an equilibrium flow pattern, the corresponding optimal trip table then can be described by an underspecified system of linear equations. With this assumption the bilevel program can be transformed into a single-

[1]Associate Professor, Dept. of Civil and Environmental Engineering, Utah State Univ., Logan, UT 84322-4110.

[2]Graduate Student, Dept. of Civil and Environmental Engineering, Utah State Univ., Logan, UT 84322-4110.

[3]Professor, Dept. of Civil Engineering, Univ. of California at Irvine, Irvine, CA 92697-3600.

Note. Discussion open until December 1, 2005. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on February 24, 2004; approved on June 11, 2004. This paper is part of the ***Journal of Transportation Engineering***, Vol. 131, No. 7, July 1, 2005. ©ASCE, ISSN 0733-947X/2005/7-506–513/$25.00.

level optimization problem (Nguyen 1977; Yang et al. 1994). Sherali et al. (1994) reformulated the $O-D$ estimation problem as a single-level linear program to estimate path flows; hence, $O-D$ flows. This method, however, requires traffic volumes on all network links to be measured and conformed to the deterministic user equilibrium (DUE) flow pattern. The fact that path flows are not uniquely defined under the DUE assumption and the requirement of observations on all network links make this approach less practical. Sherali and Park (2001) also extended this path-based approach to estimate a dynamic $O-D$ trip table. The constrained generalized least square model was used instead to accommodate the missing observations on some network links. Bell and Shield (1995) developed a nonlinear path flow estimator (PFE) based on the stochastic user equilibrium (SUE) assumption, which gives unique path flows and does not require traffic volumes on all links to be measured. The theoretical advantage of PFE lies in the assumption of SUE, which allows the selection of nonequal travel time paths due to imperfect knowledge of network travel times.

Though the nonlinear PFE has been successfully applied and tested in a number of projects in Europe (see the detail of each project in Bell and Grosso 1998), to the best of our knowledge, no systematic approach for quantifying the quality of $O-D$ estimates has been reported. Only the estimated link flows and observed traffic counts were compared. This study considers the total demand as an additional measure of quality and examines the effects of the number and locations of traffic counts on the estimation results. The remainder of this paper is organized as follows. "Path Flow Estimator" provides a brief description of PFE along with its solution procedure. The total demand scale (TDS) proposed by Bierlaire (2002) is described as a quality measure of the $O-D$ estimates in "Estimation of Total Demand." In "Numerical Examples," numerical experiments are first performed on a small hypothetical network to gain some insights on the effects of the number and locations of traffic counts. Real-world application of the PFE on a medium size network in the City of Irvine, California is also reported. Finally, findings and conclusion of the study are presented in "Findings and Conclusions."

Path Flow Estimator

The nonlinear PFE was originally developed by Bell and Shield (1995) as a one-stage network observer. It is able to estimate path flows and path travel times from traffic counts obtained from detectors and other types of detection devices in transportation networks. This one-stage approach circumvents the analytical and computational difficulties of the bilevel programming formulation. The core component of PFE is a logit-based path choice model, which interacts with link cost functions to produce a SUE traffic pattern. The interaction between travel times and route choices is modeled in a similar way to that of supply and demand interaction in the market place. The "market clearing" price and the quantity consumed are equivalent to the set of flows and the set of travel times equilibrated to a SUE traffic pattern (Bell and Iida 1997).

Mathematical Formulation

The PFE model is very similar to the formulation of the logit-based SUE problem proposed by Fisk (1980) with the exception of the constraint set. The formulation of PFE is given as follows: Minimize

$$\frac{1}{\theta}\sum_{rs}\sum_{k} f_k^{rs}(\ln f_k^{rs}-1)+\sum_{a}\int_0^{x_a} t_a(w)dw \qquad (1)$$

subject to

$$(1-\varepsilon_a)\cdot\nu_a \leqslant x_a \leqslant (1+\varepsilon_a)\cdot\nu_a \qquad \forall a\in M \qquad (2)$$

$$x_a \leqslant C_a \qquad \forall a \in U \qquad (3)$$

$$(1-\varepsilon_{rs})\cdot z_{rs} \leqslant q_{rs} \leqslant (1+\varepsilon_{rs})\cdot z_{rs} \qquad \forall rs\in RS \qquad (4)$$

where

$$x_a=\sum_{rs}\sum_{k} f_k^{rs}\delta_{ka}^{rs} \qquad \forall a\in A \qquad (5)$$

$$q_{rs}=\sum_{k\in K_{rs}} f_k^{rs} \qquad \forall rs\in RS \qquad (6)$$

where θ = dispersion parameter in the logit model; f_k^{rs} = flow on path k connecting $O-D$ pair rs; $t_a(\cdot)$ = travel time on link a; x_a = estimated traffic volume on link a; δ_{ka}^{rs} = path-link indicator: 1 if link a is on path k between $O-D$ pair rs and 0 otherwise; ν_a = observed traffic volume on link a; ε_a = measurement error allowed for traffic count on link a; C_a = capacity of link a; q_{rs} = estimated travel demand between $O-D$ pair rs; z_{rs} = a priori travel demand for $O-D$ pair rs; ε_{rs} = measurement error allowed for the target trip table; M, U, and A = sets of observed links, unobserved links, and all network links ($A=M\cup U$); and RS = set of $O-D$ pairs.

The objective function (1) has two terms, which are the entropy and user equilibrium terms. The entropy term seeks to spread trips onto multiple paths, while the user equilibrium term tends to cluster trips on the minimum-cost paths. The confidence levels (ε_a and ε_{rs}) are introduced to Eqs. (2) and (4) to account for measurement error of traffic counts and the confidence associated with the target trip table, respectively. More reliable information will use a smaller tolerance to constrain the estimated flows within a narrower range, while less reliable information will use a larger tolerance to allow a larger range of estimated flows. The introduction of confidence levels in Eqs. (2) and (4) also allows for more flexible estimation of the $O-D$ trip table. For the unobserved links, Eq. (3) constrains the estimated link flows to be less than or equal to the links' capacity. Similar to the logit-based SUE model, thanks to the logarithmic term, path flows can be derived analytically as a function of path cost and dual variables associated with constraints (2), (3) and (4), as follows:

$$f_k^{rs}=\exp\Bigg(\theta\cdot\bigg(-\sum_{a} t_a(x_a)\delta_{ka}^{rs}+\sum_{a\in M}(u_a^{-}\delta_{ka}^{rs}+u_a^{+}\delta_{ka}^{rs}) + \sum_{a\in U} d_a\delta_{ka}^{rs}+o_{rs}^{+}+o_{rs}^{-}\bigg)\Bigg)$$

$$\forall k\in K_{rs} \qquad rs\in RS \qquad (7)$$

There are two dual variables associated with both constraints (2) and (4): the lower $(u_a-, o_{rs}-)$ and upper $(u_a+, o_{rs}+)$ limits. These dual variables are zero if the estimated link flows and $O-D$ flows are within an acceptable range defined by the measurement error bound, and nonzero if they are binding at one of the limits. The dual variables associated with constraint (3) can be interpreted as queuing delay (Bell and Iida 1997). Queuing delays exist if the estimated link flows for the unobserved links reach the available capacity, or zero otherwise. The remaining dual variables can be interpreted as the corrections to the cost functions (e.g., link travel time or $O-D$ travel time), which are used to generate the estimated flow pattern to match the observed one.

With the relationship of flows given in Eqs. (5)–(7), the iterative balancing technique (see Bell and Shield 1995; Bell and Iida 1997) can be applied to solve the problem. The basic idea is to scale path flows to fulfill one constraint at a time by adjusting the associated dual variables. For a specific path set, the iterative balancing technique is iterated until it converges (i.e., insignificant adjustment of flows and dual variables). If the algorithm diverges (i.e., the dual variables of certain constraints tend to be positive or negative infinity), new paths need to be generated to better match these constraints (e.g., generating paths that pass through the links with traffic counts). Link costs are updated according to the current link flow and link delay (d_a—dual variables of the capacity constraints). This process is known as column generation, which is integrated into PFE to circumvent the need of path enumeration when applied to general networks.

Estimation of Total Demand

In the literature, statistical measures are often used to quantify the quality of $O-D$ estimates. Mean absolute error and root mean square error (RMSE) are examples of the statistical measures. They indicate the closeness between the observed (true) and estimated values, which could be link flows or $O-D$ flows (if known). However, they may not be sufficient to quantify the quality of $O-D$ estimates since the true $O-D$ trip table is often unknown in practice. Recently, Bierlaire (2002) proposed the use of TDS to quantify the intrinsic underdeterminate nature of the $O-D$ estimation problem (e.g., the number of $O-D$ pairs to be estimated is often much larger than the number of links with traffic counts). The TDS measure, which is independent of the $O-D$ estimation method, quantifies the quality of $O-D$ estimates based on the route choice proportions, network topology, and traffic counts. It can be computed by solving two linear programs as follows:

$$\text{TDS} = \phi_{\max} - \phi_{\min} \tag{8}$$

where

$$\phi_{\max} = \underset{q_{rs}}{\text{Max}} \sum_{rs} q_{rs}$$

and

$$\phi_{\min} = \underset{q_{rs}}{\text{Min}} \sum_{rs} q_{rs} \tag{9}$$

subject to

$$\sum_{rs} \sum_{k} q_{rs} p_k^{rs} \delta_{ka}^{rs} = v_a \qquad \forall a \in M \tag{10}$$

$$q_{rs} \geqslant 0 \qquad \forall rs \in RS \tag{11}$$

where p_k^{rs} = proportion of travelers using path k between $O-D$ pair rs; and v_a = observed flow on link a. Using the TDS measure, the quality of $O-D$ estimates using PFE can be quantified by the following three possibilities:

1. If the TDS is zero, it indicates that this set of observations has correctly captured the total demand of the network;
2. If the TDS is greater than zero, it indicates that with this set of observations, the total demand of the network cannot be captured precisely, but it is only known within a specific range; and
3. If the TDS approaches infinity, it indicates that the demand of at least one of the $O-D$ pairs is not totally captured by this set of traffic counts.

Possibility 1 is the most desirable since the only concern left is the determination of the spatial pattern of the $O-D$ estimates. Possibility 2 requires additional work for dealing with both the uncertainty of total demand and the spatial $O-D$ distribution. Possibility 3 is the most problematic, and may often arise in practice. Under the third case, the unobserved $O-D$ pairs are revealed by the maximization problem (e.g., $q_{rs} \Rightarrow \infty$) and additional observations are needed in order to improve the quality of $O-D$ estimation.

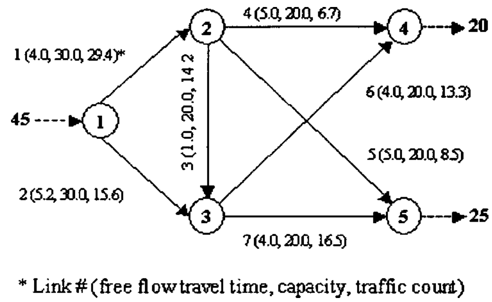

Fig. 1. Link characteristics and synthetic traffic counts for hypothetical network

Numerical Examples

Hypothetic Network

For illustration purposes, a simple network, as shown in Fig. 1, is used. It consists of five nodes, seven links, and two $O-D$ pairs. Traffic counts, which are also displayed in Fig. 1, were obtained by assuming that there are 45 travelers leaving from Node 1, 20 and 25 of them are, respectively, destined to Nodes 4 and 5, and they follow the SUE principle with a dispersion parameter of 0.10. Link cost function is based on the standard Bureau of Public Road function, given in the following equation, with 0.15 for α and 4.0 for β. This cost function is used throughout this study

$$t_a(x_a) = t_a^0 \cdot [1 + \alpha \cdot (x_a/C_a)^{\beta}] \tag{12}$$

In order to examine the effects of the number and locations of traffic counts, we do not use the full formulation of PFE. We assume that there is no measurement error in the traffic counts (i.e., $\varepsilon_a = 0$ for all links). Hence, constraint (2) becomes equality. Estimated link flows must exactly match the observed traffic counts. In addition, we further assume that the target trip table is not available; hence, constraint (4) is not used. In order to examine the closeness of the estimation, the RMSE given below is used to compare the estimated values to the observed (true) values

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (x_{\text{est}}^n - x_{\text{obs}}^n)^2} \tag{13}$$

where N = number of observations; and x_{est} and x_{obs} = estimated and observed values.

Table 1. Origin–Destination ($O-D$) Demand Estimates Using Different Sets of Traffic Counts

Number of observations	Combination of observed links	$O-D$ demand		
		$O-D$ (1,4)	$O-D$ (1,5)	Total
7	1,2,3,4,5,6,7	20.00	25.00	45.00
1	1	15.10	15.10	30.20
1	2	8.61	8.61	17.23
2	1,2	22.50	22.50	45.00
2	1,3	15.10	15.10	30.20
2	4,6	20.00	1.21	21.21
3	1,2,4	22.07	22.93	45.00
3	1,2,6	21.00	24.00	45.00
3	1,4,6	20.00	10.75	31.75
4	1,2,4,6	20.00	25.00	45.00
4	1,2,5,7	20.00	25.00	45.00
4	4,5,6,7	20.00	25.00	45.00

Effects of Number and Locations of Traffic Counts

This section investigates the minimum number of link observations required for PFE to correctly capture the total demand and/or individual $O-D$ demand (spatial distribution). It should be noted that such a requirement is dependent on the network topology and the number of $O-D$ pairs to be estimated. Table 1 presents the results using different sets of observations to estimate the $O-D$ trip table for the small network in Fig. 1. When all seven observations are available, both total and individual $O-D$ demands can be obtained correctly. This provides a benchmark for comparing with other sets of traffic counts. It should also be noted that this might not be the case for other network topologies where the number of independent links in the network is less than the number of $O-D$ pairs. In such a case, the individual $O-D$ demand estimates are difficult to be obtained uniquely. The estimation problem is generally solved using an optimization procedure with a distance measure as the objective to select the most proper $O-D$ trip table that is consistent with the under-specified linear constraints (observation constraints). The number of independent links can be defined by the total number of network links minus the number of intermediate nodes (Bell and Iida 1997). Here, all nodes other than the origins and destinations are referred to as the intermediate nodes. For this network, it is easy to verify from Fig. 1 that there are five linearly independent links. However, only four of them are required to construct two linearly independent equations for the two unknowns ($O-D$ pairs). As we shall see next that for this network any combinations of the four linearly independent observations will produce the same estimate with the correct total and individual $O-D$ demands.

Using only one observation is clearly not sufficient to capture the total demand for this network. However, each link contains a different amount of information. For example, using Link 1 produces better estimates of the total demand as well as individual $O-D$ demands compared to those produced using Link 2. This suggests that Link 1 contains a better quality of information. In addition, it can be explained by considering the path set for this network, which is provided in Table 2. Since Link 1 is used more often (by four paths) and it contains higher flows compared to other links, it has a higher contribution to the total demand estimate. When two observations are used, only the combination of Links 1 and 2 can correctly capture the total demand. It is easy to see from the network topology why other combinations (e.g., Links 1 and 3 or Links 4 and 6, etc.) cannot estimate the total demand correctly. Since there is only one origin (Node 1) and when traffic volumes are observed on Links 1 and 2, travel demand originated from Node 1 is totally captured. As can be seen from Table 2, with these two observations (Links 1 and 2), traffic flows on all paths are also observed. However, the spatial distribution of the total demand is not correct. Similar results are also observed using three observations.

Table 2. Path Set for Hypothetical Network

$O-D$ pair	Path	Link sequence
(1,4)	1	1-4
	2	1-3-6
	3	2-6
(1,5)	1	1-5
	2	1-3-7
	3	2-7

Note: $O-D$=origin–destination.

To obtain the correct individual $O-D$ demands, PFE requires at least four observations. There are three combinations that can achieve this as reported in Table 1. Since the network contains only one origin (Node 1) and two destinations (Nodes 4 and 5), it is clear that if all entry flows to both destinations (Links 4, 5, 6, and 7) or the total demand (Links 1 and 2) with all entry flows to one of the destinations (e.g., Links 4 and 6 or Links 5 and 7) are observed, both total and individual $O-D$ demands would be estimated correctly.

Application of Path Flow Estimator for Real Transportation Network

In this section, PFE is applied to estimate the $O-D$ trip table for the Irvine network in Orange County, Calif., as depicted in Fig. 2. This network consists of three major freeways (I-5, I-405, and SR-133), and several arterials in the City of Irvine. The network and demand data were extracted from the Orange County Transportation Analysis Model (OCTAM), which contains data for the whole county. The *OCTAM 3.0* was developed and validated for Base Year 1991 conditions, and revalidated for the Year 1998 to better reflect the current highway and transit data. The extracted network is composed of 163 nodes, 496 links, 39 traffic analysis zones, 28 external stations, and 1,547 $O-D$ pairs.

For this study, traffic counts were generated by assigning travel demands from OCTAM to the network according to the SUE principle with a dispersion parameter of 0.01. Three sets of traffic counts were constructed to test the PFE (see Fig. 2). The first set contains traffic counts available on all centroid connectors (links directly connected to origin or destination nodes); the second set contains traffic counts available on all freeway links; and the third set contains traffic counts available on both centroid connectors and freeway links. As a result, we have 202 observations (41% of network links) for the first set, 125 observations (25%) for the second set, and 251 observations (51%) for the third set. Again, we assume that there is no measurement error in traffic counts. Using the findings in the previous section, the first set of traffic counts represents the case that all travel demands from all origins to all destinations are completely observed. Thus, one can expect that the total demand of network should be correctly estimated. However, since centriod connectors do not

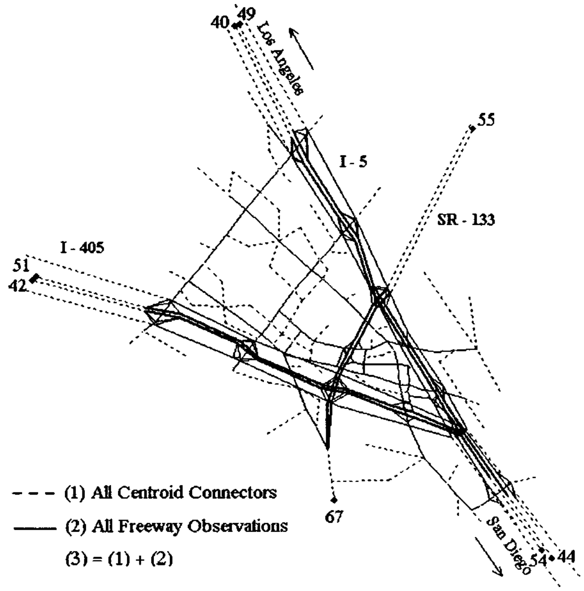

Fig. 2. Irvine Network, Orange County, California

physically exist (hence neither do these traffic counts), it is necessary to find ways to establish a set of actual network links that can capture the total demand of the network.

The *O–D* pairs of interest, for this network, are the through traffic on three major freeways. For comparison purposes, the *O–D* trip table extracted from OCTAM is used as the true trip table to compare with the estimated *O–D* trip table obtained by PFE. To further investigate the effects of including the target trip table on the estimation results, we incorporate a subset of the target trip table that consists of interchanges of 17 freeway *O–D*s into observation Set 1 and named it as observation Set 4. Therefore, constraint (4) for the target trip table is included in this experiment. The estimation results are summarized in Table 3 in terms of the total demand estimates, estimation errors, RMSE for the target *O–D* demands, and the TDS measure presented in "Estimation of Total Demand." It should be noted that when link counts (constraints) could be reproduced exactly by PFE, the corresponding RMSE is zero.

As the number of observations increases, the quality of estimates is generally improved as shown by the lower RMSE and lower percentage error of the total demand estimate. Since traffic counts on all centroid connectors are used in Cases 1, 3, and 4, the total demand of the study area can be captured perfectly. This is also indicated by the TDS measure (i.e., zero for all three cases). For observation Set 2, it is not surprising that the total demand is underestimated (2.56% error and a diverging TDS) since only the traffic counts on freeway links are used. Although the total demand can correctly be estimated by observation Sets 1 and 3, the spatial distribution of *O–D* demands is somewhat different from the true *O–D* demands. This is indicated by the *O–D* RMSE measure: 47.34 for Case 1 and 39.42 for Case 3. Such estimation results are due to the under-specified nature of the problem (i.e., the number of unknowns is more than the number of observations) and the way that PFE chooses the most likely traffic flow pattern among all possibilities through its objective function. By incorporating the target trip table information related to the freeway *O–D*s (observation Set 4), the quality of *O–D* estimates is generally improved compared to using link observations alone. The *O–D* RMSE is reduced from 47.34 in Case 1 to 36.24, which is also lower than the *O–D* RMSE value in Case 3.

In Table 4, the estimation results for the freeway *O–D* pairs are reported together with their estimation errors. We do not report the results for observation Set 4 because the *O–D* estimates are exactly the same as the target trip interchanges provided in the second column. In Case 2, even though the number of observations is less than that of Case 1, the spatial distribution of the freeway *O–D* pairs seems much better. For example, the maximum absolute percentage error reduces from 115.16 to 11.11 for *O–D* pair (42,67). This is because observation Set 2 contains better information for estimating the freeway *O–D* demands; traffic flows on all freeways are completely observed and they usually involve high traffic volumes. In addition, when this information is combined with observation Set 1, as in Set 3, the overall estimations, for both the individual *O–D* pairs and the total demand, are improved. The comparison of the true and estimated *O–D* flows is also depicted in Figs. 3(a–d) for each set of observations (in log scale). Each point represents a pair of the true and

Table 3. Summary of Estimation Result for Irvine Network

		Observation			
Description	OCTAM	Set 1	Set 2	Set 3	Set 4
Number of link observations	—	202	125	251	202(+17[a])
Number of paths generated	2,011[b]	2,004	1,982	1,955	2,107
Total demand	47,522.00	47,522.00	46,307.04	47,522.00	47,522.00
Error (%)	—	0.00	−2.56	0.00	0.00
RMSE—Link flow estimates	—	0.00	0.00	0.00	0.00
RMSE—*O–D* flow estimates	—	47.34	44.07	39.42	36.24
Number of *O–D* pairs not covered	—	0	273	0	0
ϕ_{max}	—	47,522.00	∞	47,522.00	47,522.00
ϕ_{min}	—	47,522.00	42,826.29	47,522.00	47,522.00
$\phi_{max}-\phi_{min}$	—	0.00	∞	0.00	0.00

Note: OCTAM=Orange County Transportation Analysis Model; RMSE=root mean square error; and *O–D*=origin–destination.
[a]Number of prior trip interchanges.
[b]Number of paths obtained when assigning the assumed *O–D* trip table according to the stochastic user equilibrium principle.

Table 4. Estimation Results for Freeway Origin–Destination ($O-D$) Pairs in Irvine Network

		Set 1		Set 2		Set 3		
$O-D$ pairs	OCTAM	PFE	Error (%)	PFE	Error (%)	PFE	Error (%)	Remark
(40,54)	5,084.00	4,836.60	−4.87	4,786.67	−5.85	4,756.71	−6.44	On I-5 S to San Diego
(40,55)	780.00	676.60	−13.26	667.09	−14.48	788.68	1.11	From I-5 S to SR-133
(40,67)	419.00	217.95	−47.98	416.29	−0.65	373.79	−10.79	From I-5 S to SR-133
(42,54)	2,829.00	2,443.49	−13.63	2,580.31	−8.79	2,593.98	−8.31	From I-405 S to San Diego on I-5 S
(42,55)	946.00	882.85	−6.68	887.11	−6.23	920.68	−2.68	From I-405 S to SR-133
(42,67)	122.00	262.49	115.16	135.56	11.11	142.88	17.12	From I-405 S to SR-133
(44,49)	3,392.00	3,139.08	−7.46	3,427.74	1.05	3,443.92	1.53	On I-5 N to Los Angeles
(44,51)	1,553.00	1,353.28	−12.86	1,396.28	−10.09	1,448.44	−6.73	From I-5 N to I-405 N
(44,55)	973.00	676.58	−30.46	749.11	−23.01	725.82	−25.40	From I-5 N to SR-133

Note: OCTAM=Orange County Transportation Analysis Model; and PFE=path flow estimator.

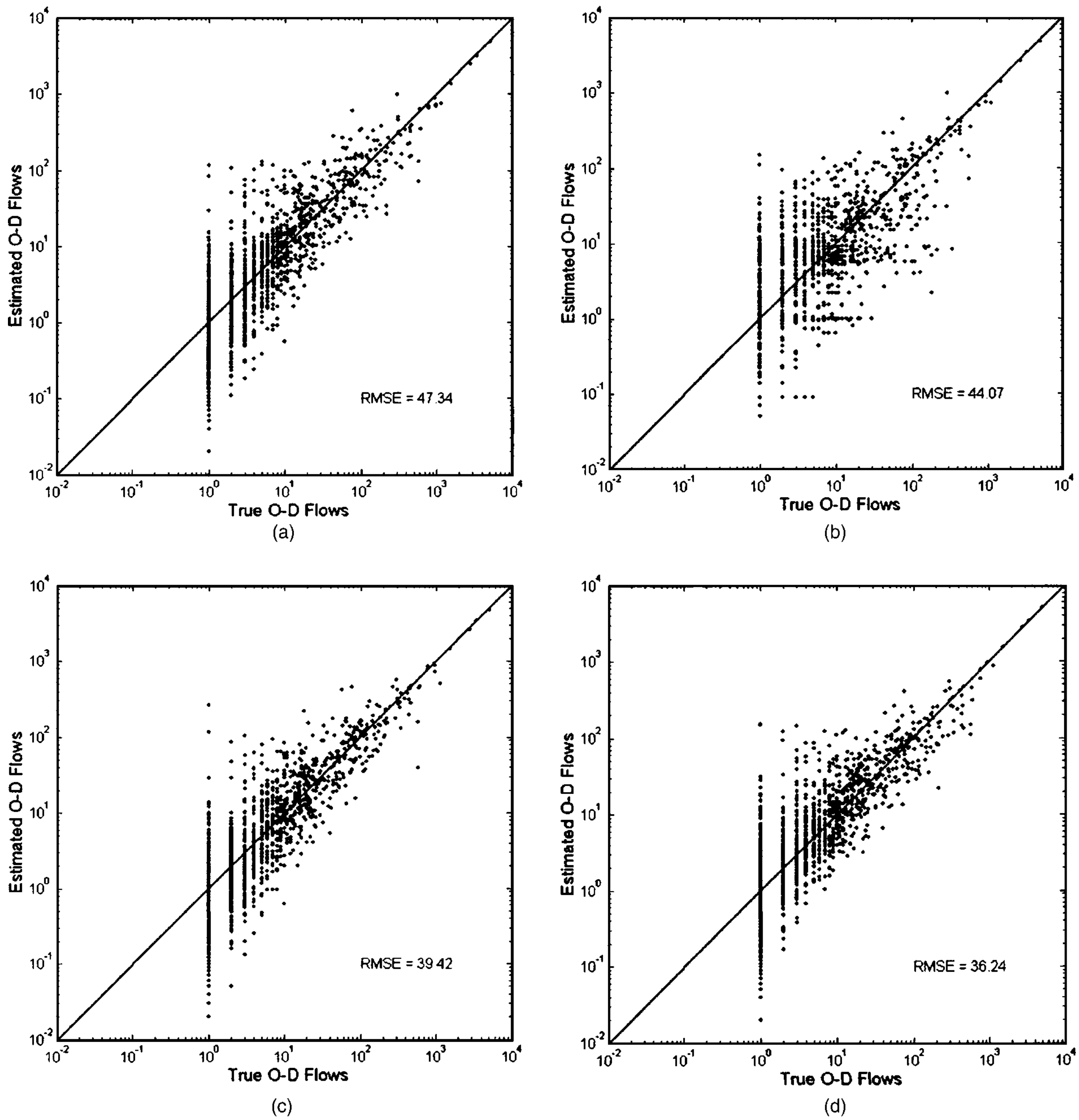

Fig. 3. Comparison of true and estimated origin–destination demands using different sets of traffic counts: (a) Observation Set 1; (b) Observation Set 2; (c) Observation Set 3; and (d) Observation Set 4

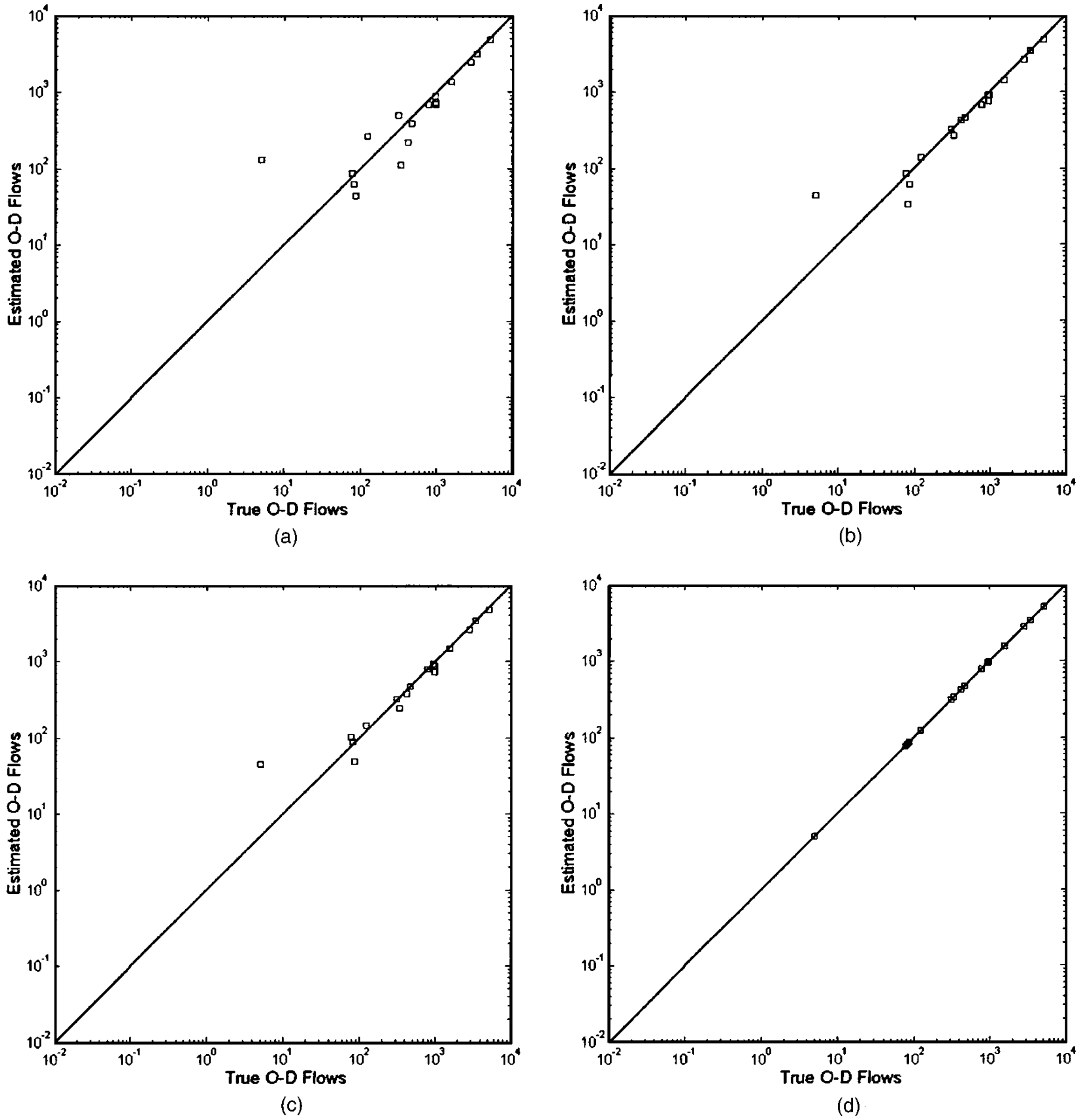

Fig. 4. Comparison of true and estimated freeway origin–destination using different sets of traffic counts: (a) Observation Set 1; (b) Observation Set 2; (c) Observation Set 3; and (d) Observation Set 4

estimated $O-D$ flows. The data points along the 45° line represent perfect matches. Figs. 4(a–d) provide similar comparisons, but focus only on the freeway $O-D$s.

From Fig. 3, it is observed that PFE can estimate high $O-D$ demands quite well, but it has a problem in estimating low $O-D$ demands. The estimates for $O-D$ pairs with low demands could deviate quite a bit from the true values. It is worth pointing out that the RMSE only represents the aggregated quality of $O-D$ estimates (on average). As can be seen, the data points in Fig. 3(b) seem more scattered than those in Fig. 3(a); thus the overall quality of observation Set 1 should be better. However, the resultant RMSE indicates that the $O-D$ estimates obtained from observation Set 2 are slightly better than those obtained from observation Set 1. This is because the RMSE measure penalizes more on the $O-D$ pairs with higher demands.

Findings and Conclusion

Using both simple and real networks, we have demonstrated that PFE has the capability to correctly estimate the total and individual $O-D$ demands when proper information is provided. Based on the preliminary results, it is found that the selection of observed links plays an important role in the $O-D$ estimation problem as each observation contributes different amounts of information. Although the quality of estimates can be improved with the usage of more traffic counts, this may be impractical when the budget for data collection is scarce. From the small network, the results show that if there is at least one observation on each path, the total demand of the network can be correctly captured. In addition, higher observed traffic volume seems to contribute more to the quality of $O-D$ estimation. The most difficult task observed so far is estimation of the spatial pattern of

$O-D$ demands. Even when all network links are measured, the individual $O-D$ demands may not be estimated correctly.

For future research, we need to examine the full formulation of PFE. Additional traffic information such as $O-D$ travel times, subpath information, intersection turning movements, etc., are necessary to better capture the spatial distribution of travel demand. A systematic study is required to examine the effects of measurement errors of traffic counts. In addition, the $O-D$ trip table derived from PFE needs to be validated against real traffic data including traffic counts, path travel times, etc. Insights gained from the validation process may help us to enhance the performance of the PFE model.

Acknowledgment

This research was supported in part by the California Partners for Advanced Transit and Highways (PATH) Program, United States. The contents of this paper reflect the views of the writers who are responsible for the facts and the accuracy of the data presented herein, and do not necessarily reflect the official views or policies of the State of California. This paper does not constitute a standard, specification, or regulation.

References

Ashok, K., and Ben-Akiva, M. E. (1993). "Dynamic origin-destination matrix estimation and prediction for real-time management systems." *Proc., 12th International Symp. on Transportation and Traffic Theory*, Berkeley, Calif.

Ashok, K., and Ben-Akiva, M. E. (2000). "Alternative approaches for real-time estimation and prediction of time-dependent origin-destination flows." *Transp. Sci.*, 34, 21–36.

Ashok, K., and Ben-Akiva, M. E. (2002). "Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows." *Transp. Sci.*, 36(2), 184–198.

Bell, M. G. H. (1984). "The estimation of junction turning volumes from traffic counts: The role of prior information." *Traffic Eng. Control*, 25(5), 279–283.

Bell, M. G. H. (1991). "The estimation of origin-destination matrices by constrained generalized least squares." *Transp. Res., Part B: Methodol.* 25B, 13–22.

Bell, M. G. H., and Grosso, S. (1998). "The path flow estimator as a network observer." *Traffic Eng. Control*, 39, 540–549.

Bell, M. G. H., and Iida, Y. (1997). *Transportation network analysis*, Wiley, New York.

Bell, M. G. H., and Shield, C. M. (1995). "A log-linear model for path flow estimation." *Proc., 4th International Conf. on the Applications of Advanced Technologies in Transportation Engineering*, Carpi, Italy, 695–699.

Bierlaire, M. (2002). "The total demand scale: A new measure of quality for static and dynamic origin-destination trip tables." *Transp. Res., Part B: Methodol.*, 36B, 755–851.

Bierlaire, M., and Crittin, F. (2004). "An efficient algorithm for real-time estimation and prediction of dynamic OD tables." *Oper. Res.*, 52(1), 116–127.

Cascetta, E. (1984). "Estimation of origin-destination matrices from traffic counts and survey data: A generalized least squares estimator." *Transp. Res., Part B: Methodol.* 18B, 289–299.

Fisk, C. (1980). "Some developments in equilibrium traffic assignment." *Transp. Res., Part B: Methodol.* 14B, 243–255.

Fisk, C. (1988). "On combining maximum entropy trip matrix estimation with user equilibrium." *Transp. Res., Part B: Methodol.*, 22B, 69–73.

Fisk, C. (1989). "Trip matrix estimation from link traffic counts: The congested case." *Transp. Res., Part B: Methodol.*, 23B, 245–250.

Hazelton, M. L. (2000). "Estimation of origin-destination matrices from link flows on uncongested networks." *Transp. Res., Part B: Methodol.* 34B, 549–566.

Madanat, S. M., Hu, S. R., and Krogmeier, J. (1996). "Dynamic estimation and prediction of freeway O-D matrices with route switching considerations and time-dependent model parameters." *Transp. Res. Rec.*, 1537, Transportation Research Board, Washington, D.C., 98–105.

Maher, M. (1983). "Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach." *Transp. Res., Part B: Methodol.* 17B, 435–447.

Maher, M., Zhang, X., and Van Vliet, D. (2001). "A bi-level programming approach for trip matrix estimation and traffic control problems with stochastic user equilibrium link flows." *Transp. Res., Part B: Methodol.* 35B, 23–40.

Nanne, J., Zijpp, V. D., and Hamerslag, R. (1994). "Improved Kalman filtering approach for estimating origin-destination matrices for freeway corridors." *Transportation Research Record, 1443*, Transportation Research Board, Washington, D.C., 54–63.

Nguyen, S. (1977). "Estimating an OD matrix from network data: A network equilibrium approach." *Transportation planning models*, M. Florian, ed., CRT, Univ. de Montréal, Montréal, 363–380.

O'Neil, W. A. (1987). "Origin–destination trip table estimation using traffic counts." PhD dissertation, Univ. of New York, Buffalo, N.Y.

Sherali, H. D., Arora, N., and Hobeika, A. G. (1997). "Parameter optimization methods for estimating dynamic origin-destination trip tables." *Transp. Res., Part B: Methodol.*, 31B, 141–147.

Sherali, H. D., and Park, T. (2001). "Estimation of dynamic origin-destination trip tables for a general network." *Transp. Res., Part B: Methodol.*, 35B, 217–235.

Sherali, H. D., Sivanandan, R., and Hobeika, A. G. (1994). "A linear programming approach for synthesizing origin-destination trip tables from link traffic volumes." *Transp. Res., Part B: Methodol.*, 28B, 213–234.

Spiess, H. (1987). "A maximum likelihood model for estimating origin-destination matrices." *Transp. Res., Part B: Methodol.*, 21B, 395–412.

Van Zuylen, H. J. (1979). "The estimation of turning flows on a junction." *Traffic Eng. Control*, 20(11), 539–541.

Van Zuylen, H. J., and Willumsen, L. G. (1980). "The most likely trip estimated from traffic counts." *Transp. Res., Part B: Methodol.* 14B, 281–293.

Yang, H., Iida, Y., and Sasaki, T. (1992). "Estimation of origin-destination matrices from traffic counts on congested networks." *Transp. Res., Part B: Methodol.* 26B, 417–434.

Yang, H., Iida, Y., and Sasaki, T. (1994). "The equilibrium-based origin-destination matrix estimation problem." *Transp. Res., Part B: Methodol.*, 28B, 23–33.