Transportation Research Part B 43 (2009) 852–872

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Norm approximation method for handling traffic count inconsistencies in path flow estimator

Anthony Chen [a,*], Piya Chootinan [b], Will Recker [c]

[a] Department of Civil and Environmental Engineering, Utah State University, Logan, UT 84322-4110, USA
[b] Bureau of Planning, Department of Highways, Bangkok 10400, Thailand
[c] Department of Civil and Environmental Engineering, University of California, Irvine, CA 92697-3600, USA

ARTICLE INFO

Article history:
Received 28 April 2007
Received in revised form 26 February 2009
Accepted 27 February 2009

Keywords:
Origin–destination estimation
Path flow estimator
Stochastic user equilibrium
Norm approximation
Partial linearization method

ABSTRACT

Path flow estimator (PFE) is a one-stage network observer proposed to estimate path flows and hence origin–destination (O–D) flows from traffic counts in a transportation network. Although PFE does not require traffic counts to be collected on all network links when inferring unmeasured traffic conditions, it does require all available counts to be reasonably consistent. This requirement is difficult to fulfill in practice due to errors inherited in data collection and processing. The original PFE model handles this issue by relaxing the requirement of perfect replication of traffic counts through the specification of error bounds. This method enhances the flexibility of PFE by allowing the incorporation of local knowledge, regarding the traffic conditions and the nature of traffic data, into the estimation process. However, specifying appropriate error bounds for all observed links in real networks turns out to be a difficult and time-consuming task. In addition, improper specification of the error bounds could lead to a biased estimation of total travel demand in the network. This paper therefore proposes the norm approximation method capable of internally handling inconsistent traffic counts in PFE. Specifically, three norm approximation criteria are adopted to formulate three L_p-PFE models for estimating consistent path flows and O–D flows that simultaneously minimize the deviation between the estimated and observed link volumes. A partial linearization algorithm embedded with an iterative balancing scheme and a column generation procedure is developed to solve the three L_p-PFE models. In addition, the proposed L_p-PFE models are illustrated with numerical examples and the characteristics of solutions obtained by these models are discussed.

1. Introduction

Path flow estimator (PFE), originally developed by Bell and Shield (1995), is one of the efficient methods for estimating path flows (hence origin–destination (O–D) flows) from traffic counts. The attractiveness of PFE lies on the fact that it is a single level mathematical program in which the interdependency between O–D demands and route choice behavior (congestion effect) is taken into account without the need to employ the bi-level mathematical program (one level estimates the O–D trip table while the other level represents the behavioral responses of network users). Network users are assumed to follow the stochastic user equilibrium (SUE) principle, which allows the selection of non-equal travel time paths due to the imperfect knowledge of network travel times and yields unique path flow estimates. Besides, PFE can perform the estimation using traffic counts collected only on a subset of network links. Nevertheless, these available counts must be

* Corresponding author. Tel.: +1 435 797 7109; fax: +1 435 797 1185.
E-mail address: achen@engineering.usu.edu (A. Chen).

0191-2615/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.trb.2009.02.007

reasonably consistent or constitute a consistent system of linear constraints. This requirement is difficult to fulfill in practice due to the errors involved in data collection and processing. If the system of linear constraints is inconsistent, there may not exist any feasible path flow solution that is able to reproduce all traffic counts exactly. This is one source of the inconsistency problem while the other is caused by the capacity constraints used to restrict the estimated link-flows on the unobserved links. It is often observed that the total observed flows entering node is greater than the capacity of all exiting links combined or vice versa. In such conditions, no feasible path flow solution is able to satisfy both observation and capacity constraints at the same time.

The original PFE model handles the inconsistency problem by allowing user-specified error bounds (i.e., confidence interval) on the traffic counts. A more reliable traffic count would constrain the estimated link-flow within a smaller tolerance, while a less reliable traffic count would allow for a larger deviation. This method enhances the flexibility of PFE by allowing the user (e.g., an experienced traffic engineer who is familiar with the network conditions) to incorporate local knowledge about the network conditions to the estimation process. However, specifying appropriate error bounds for every observed link in a network of realistic size could be very laborious. One could set a uniform error bound (e.g., a default value of 10% error) across all observations, but this setting might be too loose for the more reliable traffic counts and too tight for the less reliable traffic counts. In addition, setting a uniform error bound could lead to biased estimates of the O–D demands. Typically, the total travel demand of the study area is biased downward due to the minimization of the objective function used in the PFE model (Chootinan et al., 2005a).

Although several preprocessing procedures (Van Zuylen and Branston, 1982; Kikuchi et al., 2000) have been proposed to remove the inconsistency problem among traffic counts prior to the O–D estimation process, it might be more appealing to let the mathematical program handle this task by itself since the inconsistency of traffic counts is a natural part of the O–D estimation problem. Jornsten and Wallace (1993) formulated an unconstrained stochastic program to maximize the entropy objective function and to minimize the expected deviation between the observed and estimated link-flows simultaneously. The requirement of exact replication of the traffic counts was relaxed and explicitly incorporated into the entropy objective function as a penalty term. Similarly, Sherali et al. (1994) proposed a linear PFE model in which two sets of non-negative artificial variables are included into the observation constraints to account for the positive and negative deviations of the link-flow estimates from the traffic counts. These artificial variables are concurrently minimized while solving for the deterministic user equilibrium (DUE) path flow pattern. Instead of considering the deviations of link-flow estimates directly, Van Aerde et al. (2003) incorporated the first-order necessary conditions of the generalized least squares (GLS) model into the maximum likelihood framework. The first-order necessary conditions were included directly into the objective function as a penalty term to obtain an unconstrained maximum likelihood model, which determines the most likely trip table that produces link-flow estimates with the least deviations (similar to the GLS model) from traffic counts.

In this paper, we propose using the norm approximation method to internally handle the inconsistent traffic counts within the nonlinear PFE model proposed by Bell and Shield (1995). Three criteria (i.e., L_∞-norm, L_1-norm, and L_2-norm) for approximating a solution of unsolvable (inconsistent) system of linear equations (i.e., constraint set) are considered in formulating three L_p-PFE models. A partial linearization algorithm embedding an iterative balancing scheme and a column generation is developed to solve the L_p-PFE models.

The organization of this paper is as follows. Section 2 provides formulations of three L_p-PFE models. Section 3 describes a solution procedure for solving the proposed L_p-PFE models. Section 4 provides numerical results to demonstrate the applicability of the proposed models and solution algorithm. Finally, concluding remarks are provided in Section 5.

2. Model formulations

In this study, we propose the norm approximation method to internally handle the inconsistent traffic counts within the nonlinear PFE model proposed by Bell and Shield (1995). Due to measurement errors inherited in traffic counts, there may not exist a path flow solution that can reproduce all traffic counts exactly; however, if measurement errors are allowed in the estimation, a path flow solution may be found to match all traffic counts with different degrees of deviation between the estimated and observed link-flows. This path flow pattern is usually associated with some estimation errors given by:

$$\psi_a = \left| v_a - \sum_{rs} \sum_k f_k^{rs} \delta_{ka}^{rs} \right|, \quad \forall a \in M, \tag{1}$$

where M is the set of links with traffic counts; v_a is the observed flow on link a; f_k^{rs} is the estimated flow on path k connecting origin r and destination s; δ_{ka}^{rs} is the path-link indicator: 1 if link a is on path k between origin r and destination s, and 0 otherwise; ψ_a is the error associated with the selected path flow pattern fails to satisfy the observed flow on link a.

Intuitively, the best approximation of a path flow pattern is the solution that keeps the deviation between the estimated and observed link-flows as small as possible. However, there are several ways to define the approximate solutions as follows:

$$\|\boldsymbol{\Psi}\|_p = \left(\sum_{a \in \mathbf{M}} \psi_a^p \right)^{1/p}. \tag{2}$$

In practice, three different norms (i.e., different p values) – namely the L_∞ norm, L_1-norm, and L_2-norm – are considered for evaluating the approximate solutions. They constitute the criteria that aim to (i) minimize the maximum absolute error ($p \Rightarrow \infty$), (ii) minimize the average absolute error ($p = 1$), and (iii) minimize the average squared error ($p = 2$), respectively. The question as to which criterion should be adopted is not trivial and depends upon, for examples, the nature of errors causing the inconsistency problem, the required characteristics of the approximate solution, etc. As discussed by Chvatal (1983), minimizing the L_1-norm leads to the most robust approximate solution. Here, robustness is defined by the insensitivity to the outliers (e.g., flawed data). Minimizing the L_∞ norm, on the other hand, tends to minimize gross discrepancies between the observed and adjusted values to accommodate all the data points as much as possible, thus it is quite sensitive to the outliers (i.e., less robust). Lastly, minimizing the L_2-norm can be shown to be suitable for the applications in which the errors causing the inconsistency tend to be small and follow the normal distribution. Using the above norm approximations, we develop three L_p-PFE formulations to determine the stochastic user equilibrium (SUE) path flow pattern that minimizes the estimation errors.

2.1. L_∞ approximation

Let us consider the maximum absolute error defined as

$$\psi_o = \underset{\forall a \in M}{Max}\{\psi_a\}. \tag{3}$$

Since ψ_o is the maximum absolute error among all observations, the following condition must hold:

$$-\psi_o \le v_a - \sum_{rs}\sum_{k} f_k^{rs}\delta_{ka}^{rs} \le \psi_o, \quad \forall a \in M. \tag{4}$$

A feasible path flow solution can be defined by two inequalities in Eq. (4), which represents the lower and upper limits of the estimated link-flows. One can view the maximum absolute error (ψ_o) as a flow on the virtual path, which traverses through all measured links. It plays the role of setting the boundaries acceptable for the link-flow estimates. Since the virtual path does not really exist, intuitively it should not be used very often. In addition, flow on the virtual path should be small, but large enough to ensure the existence of a feasible path flow solution. Hence, the L_∞-PFE formulation is to search for a SUE path flow pattern that produces a link-flow pattern with the least maximum absolute error as follows:

$$\text{Minimize}: \quad Z_{L_\infty} = \sum_{a\in A}\int_0^{x_a} t_a(w)dw + \frac{1}{\theta}\sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs}(\ln f_k^{rs} - 1) + \frac{1}{\theta}\psi_o(\ln\psi_o - 1) + \rho_o\psi_o \tag{5a}$$

subject to:

$$x_a \ge v_a - \psi_o, \quad \forall a \in M, \tag{5b}$$

$$x_a \le v_a + \psi_o, \quad \forall a \in M, \tag{5c}$$

$$x_a \le C_a, \quad \forall a \in U, \tag{5d}$$

$$x_a = \sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs}\delta_{ka}^{rs}, \quad \forall a \in A, \tag{5e}$$

$$q_{rs} = \sum_{k\in K_{rs}} f_k^{rs}, \quad \forall rs \in \mathrm{RS}, \tag{5f}$$

$$f_k^{rs} \ge 0, \quad \forall k \in K_{rs},\ rs \in \mathrm{RS}, \tag{5g}$$

$$\psi_o \ge 0, \tag{5h}$$

where U is the set of links without traffic counts; A is the set of links in the network, $A = M \cup U$; RS is the set of O–D pairs; K_{rs} is the set of paths connecting O–D pair rs; θ is the dispersion parameter; ρ_o *is the penalty* cost; x_a is the estimated flow on link a; C_a is the Capacity of link a; $t_a(\cdot)$ is the travel time function of link a; q_{rs} is the estimated flow of O–D pair rs.

The objective function (5a) of the L_∞-PFE formulation is to minimize the travel costs and path entropies for both physical and virtual paths. The entropy of the virtual path is treated in the same manner as those of the physical paths while the travel cost of the virtual path is treated as a penalty term. The penalty cost (ρ_o) should be chosen judiciously such that the maximum deviation is minimized. Eqs. (5b) and (5c) define the lower and upper limits of the estimated link-flows, respectively. These two constraints restrict the estimated link-flows (derived from the physical path flow estimates) to be within the boundaries defined by the traffic counts and the minimized maximum absolute error (ψ_o). Note that there is no need to specify the error bound in advance. For the unobserved links, the estimated link-flows are constrained not to exceed their capacities defined by Eq. (5d). Eqs. (5e) and (5f) are definitional constraints to obtain link-flows and O–D flows from the path flow solution. Eqs. (5g) and (5h) ensure the non-negativity of both physical and virtual path flows.

The Lagrangian function of the L_∞-PFE formulation and its first partial derivatives with respect to the path flow variables can be expressed as:

$$L(\mathbf{f},\Psi,\ell,\mathbf{u},\mathbf{d}) = Z_{L_\infty} + \sum_{a\in M}\ell_a\cdot\left(v_a - \sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs}\delta_{ka}^{rs} - \psi_o\right) + \sum_{a\in M}u_a\cdot\left(v_a - \sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs}\delta_{ka}^{rs} + \psi_o\right) + \sum_{a\in U}d_a\cdot\left(C_a - \sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs}\delta_{ka}^{rs}\right), \tag{6}$$

$$\frac{\partial L}{\partial f_k^{rs}} = 0 \Rightarrow \frac{1}{\theta}\ln f_k^{rs} + \sum_{a\in A} t_a(x_a)\delta_{ka}^{rs} - \sum_{a\in M}\ell_a\delta_{ka}^{rs} - \sum_{a\in M}u_a\delta_{ka}^{rs} - \sum_{a\in U}d_a\delta_{ka}^{rs} = 0, \quad \forall k\in K_{rs},\ rs\in \mathrm{RS}, \tag{7}$$

$$\frac{\partial L}{\partial \psi_o} = 0 \Rightarrow \frac{1}{\theta}\ln\psi_o + \rho_o - \sum_{a\in M}\ell_a + \sum_{a\in M}u_a = 0. \tag{8}$$

The optimality conditions lead to the analytical expressions of flows on both physical paths and virtual path as follows:

$$f_k^{rs} = \exp\left(\theta\left(-\sum_{a\in A} t_a(x_a)\delta_{ka}^{rs} + \sum_{a\in M}(u_a^- + u_a^+)\delta_{ka}^{rs} + \sum_{a\in U}d_a\delta_{ka}^{rs}\right)\right), \quad \forall k\in K_{rs},\ rs\in \mathrm{RS}, \tag{9}$$

$$\psi_o = \exp\left(\theta\left(-\rho_o + \sum_{a\in M}(\ell_a - u_a)\right)\right), \tag{10}$$

where ℓ_a, u_a, and d_a are the dual variables of constraints (5b)–(5d), respectively. The values of u_a and d_a are restricted to be non-positive while the value of ℓ_a must be non-negative; ℓ_a and u_a can be viewed as the corrections in the link cost function, which bring the estimated path flows into agreement with the observed link volumes; and d_a is related to the link delay when the estimated link-flow reaches its capacity.

2.2. L_1 approximation

In the L_1 approximation, instead of using only one virtual path to absorb the residuals of all link-flow estimates, there are as many virtual paths as the number of observed links. In other words, there is one virtual path for each observation. Likewise, there is one penalty cost for each estimation error (ρ_a), which reflects the reliability (or confidence) of each observation. It should be noted that the L_1-norm is associated with the mean absolute error (MAE) commonly used as a statistical measure to evaluate the closeness of link-flow replication. Hence, the L_1-PFE formulation is to search for a SUE path flow pattern that produces a link-flow pattern with the minimum MAE as follows:

$$\text{Minimize}:\quad Z_{L_1} = \sum_{a\in A}\int_0^{x_a} t_a(w)dw + \frac{1}{\theta}\sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs}(\ln f_k^{rs} - 1) + \frac{1}{\theta}\sum_{a\in M}\psi_a(\ln\psi_a - 1) + \sum_{a\in M}\rho_a\psi_a \tag{11a}$$

subject to:

$$x_a \ge v_a - \psi_a, \quad \forall a\in M, \tag{11b}$$
$$x_a \le v_a + \psi_a, \quad \forall a\in M, \tag{11c}$$
$$x_a \le C_a, \quad \forall a\in U, \tag{11d}$$
$$x_a = \sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs}\delta_{ka}^{rs}, \quad \forall a\in A, \tag{11e}$$
$$q_{rs} = \sum_{k\in K_{rs}} f_k^{rs}, \quad \forall rs\in \mathrm{RS}, \tag{11f}$$
$$f_k^{rs} \ge 0, \quad \forall k\in K_{rs},\ rs\in \mathrm{RS}, \tag{11g}$$
$$\psi_a \ge 0, \quad \forall a\in M. \tag{11h}$$

Following the same derivation as the L_∞-PFE formulation above, the solution to the L_1-PFE formulation is given by:

$$f_k^{rs} = \exp\left(\theta\left(-\sum_{a\in A} t_a(x_a)\delta_{ka}^{rs} + \sum_{a\in M}(\ell_a + u_a)\delta_{ka}^{rs} + \sum_{a\in U}d_a\delta_{ka}^{rs}\right)\right), \quad \forall k\in K_{rs},\ rs\in \mathrm{RS}, \tag{12}$$

$$\psi_a = \exp(\theta(-\rho_a + \ell_a - u_a)), \quad \forall a\in M. \tag{13}$$

As can be seen above, the analytical expression of path flow estimates of the L_1-PFE formulation, Eq. (12), and of the L_∞-norm formulation, Eq. (9), are exactly the same, although the analytical expressions of their virtual paths are different. Besides the difference in the number of virtual paths, the single virtual path flow in the L_∞-norm formulation is controlled by the magnitude of all dual variables for all observation constraints, while each virtual path flow in the L_1-norm formulation is controlled by the magnitude of two dual variables of each observed link: the lower and upper limits (ℓ_a and u_a). These dual variables represent the difficulty in replicating each observed link volume, reflecting the magnitude of flow on each virtual path (i.e., residual).

2.3. L_2 approximation

Similar to the L_1 approximation, the L_2-norm is also a statistical measure known as the root mean squared error (RMSE). The L_2-PFE formulation is to search for a SUE path flow pattern that produces a link-flow pattern with the minimum RMSE as follows:

$$\text{Minimize: } Z_{L_2} = \sum_{a\in A}\int_0^{x_a} t_a(w)dw + \frac{1}{\theta}\sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs}(\ln f_k^{rs}-1) + \frac{1}{\theta}\sum_{a\in M}\psi_a(\ln\psi_a - 1) + \sum_{a\in M}\rho_a\psi_a^2 \tag{14a}$$

subject to:

$$x_a \geq v_a - \psi_a, \quad \forall a\in M, \tag{14b}$$
$$x_a \leq v_a + \psi_a, \quad \forall a\in M, \tag{14c}$$
$$x_a \leq C_a, \quad \forall a\in U, \tag{14d}$$
$$x_a = \sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs}\delta_{ka}^{rs}, \quad \forall a\in A, \tag{14e}$$
$$q_{rs} = \sum_{k\in K_{rs}} f_k^{rs}, \quad \forall rs\in \text{RS}, \tag{14f}$$
$$f_k^{rs} \geq 0, \quad \forall k\in K_{rs},\ rs\in \text{RS}, \tag{14g}$$
$$\psi_a \geq 0, \quad \forall a\in M. \tag{14h}$$

The first partial derivatives of the Lagrangian function of the L_2-PFE formulation with respect to the path flow variables can be expressed as:

$$\frac{\partial L}{\partial f_k^{rs}} = 0 \Rightarrow \frac{1}{\theta}\ln f_k^{rs} + \sum_{a\in A} t_a(x_a)\delta_{ka}^{rs} - \sum_{a\in M}\ell_a\delta_{ka}^{rs} - \sum_{a\in M}u_a\delta_{ka}^{rs} - \sum_{a\in U}d_a\delta_{ka}^{rs} = 0, \quad \forall k\in K_{rs},\ rs\in \text{RS}, \tag{15}$$

$$\frac{\partial L}{\partial \psi_a} = 0 \Rightarrow \frac{1}{\theta}\ln\psi_a + 2\rho_a\psi_a - \ell_a + u_a = 0, \quad \forall a\in M. \tag{16}$$

These optimality conditions lead to the following analytical expressions of flows on the physical paths and the virtual paths:

$$f_k^{rs} = \exp\left(\theta\left(-\sum_{a\in A} t_a(x_a)\delta_{ka}^{rs} + \sum_{a\in M}(\ell_a + u_a)\delta_{ka}^{rs} + \sum_{a\in U} d_a\delta_{ka}^{rs}\right)\right), \quad \forall k\in K_{rs},\ rs\in \text{RS}, \tag{17}$$

$$\psi_a = \exp(\theta(-2\rho_a\psi_a + \ell_a - u_a)), \quad \forall a\in M. \tag{18}$$

Similarly, the analytical expression of path flow estimates of the L_2-PFE formulation, Eq. (17), is the same as those of the L_1-PFE formulation, Eq. (12), and the L_∞-norm formulation, Eq. (9). In terms of the expression of the virtual paths in Eq. (18), it is quite different. In the L_2-PFE formulation, ψ_a appears on both sides of the equation due to the quadratic penalty term in the objective function.

It should be noted that the three L_p-PFE formulations proposed above can also include additional constraints to increase the observability of the O–D estimation problem from traffic counts (see Chen et al. (2005), Chootinan et al. (2005a) for examples of adding the target of O–D demands of selected O–D pairs as constraints to improve the reasonableness of path flow and O–D flow estimates). The next section provides the partial linearization procedure for solving these formulations.

3. Solution procedure

Let us first consider the L_p-PFE formulations presented in Section 2 without the definitional constraints in vector form as follows:

$$\underset{\mathbf{f}}{\text{Minimize}}: \quad Z(\mathbf{f}) = P(\mathbf{f}) + G(\mathbf{f}) \tag{19a}$$

subject to:

$$\mathbf{\Delta}_1\mathbf{f} \geq \mathbf{v}, \quad \mathbf{\Delta}_2\mathbf{f} \leq \mathbf{v}, \quad \mathbf{\Delta}_3\mathbf{f} \leq \mathbf{C} \quad \text{and} \quad \mathbf{f} \geq \mathbf{0}, \tag{19b}$$

where $\mathbf{f}$ is a solution vector to the problem, $\mathbf{f} = (\ldots, f_k^{rs}, \ldots, \psi_a, \ldots)$, (i.e., flows on both physical and virtual paths); $\mathbf{v}$ is a vector of observed link volumes, $\mathbf{C}$ is a vector of link capacities; $\mathbf{0}$ is a vector of zeros; $\mathbf{\Delta}_1$, $\mathbf{\Delta}_2$, and $\mathbf{\Delta}_3$ represent the coefficient matrices of the left-hand side of the constraint set; $P(\mathbf{f})$ (i.e., travel cost and penalty terms) is the part of the objective function to be linearized while $G(\mathbf{f})$ (i.e., entropy terms for both physical and virtual paths) is the remaining part of the objective function.

The L_p-PFE formulations above can be solved by the partial linearization method (Evans, 1976; Patriksson, 1994). The method is based on an iterative solution of subproblems that are generated through a partial linearization of the objective function. Suppose at iteration $n-1$, a feasible solution vector is given. $P(\mathbf{f})$ is linearized, which amounts to assuming that the travel costs are fixed at their current values. The resulting subproblem defined the search direction is given by

$$\underset{\mathbf{h}}{\text{Minimize}}: \quad \nabla_{\mathbf{f}}P(\mathbf{f}^{n-1})^T\mathbf{h} + G(\mathbf{h}) \tag{20a}$$

subject to:

$$\mathbf{\Delta}_1\mathbf{h} \geq \mathbf{v}, \quad \mathbf{\Delta}_2\mathbf{h} \leq \mathbf{v}, \quad \mathbf{\Delta}_3\mathbf{h} \leq \mathbf{C} \quad \text{and} \quad \mathbf{h} \geq \mathbf{0}, \tag{20b}$$

where $\nabla_{\mathbf{f}}P(\mathbf{f}^{n-1})$ is the derivative of P with respect to $\mathbf{f}$ evaluated at iteration $n-1$. This subproblem is a nonlinear program with linear inequality constraints and can be solved by the iterative balancing scheme used in the original PFE model (Bell and Shield, 1995; Bell et al., 1997). The line search step determines how far the current solution should move in the search direction. The new solution is found as a convex combination of the solution of the above subproblem and the current solution. A column generation can also be implemented to avoid path enumeration for a general transportation network. Given the above descriptions, the solution procedure can be summarized into the following steps:

Step 0 (*Initialization*): Generate an initial feasible solution vector, $\mathbf{f}^1$.
- Set $x_a^0 = \ell_a^0 = u_a^0 = d_a^0 = 0,\ \forall a \in A,\ K_{rs}^0 = \emptyset, \forall r, s.$
- Set iteration counter: $n = 1$.
- Determine the shortest path for all O–D pairs based on the free-flow travel times: $\bar{k}_{rs}^n,\ \forall r, s$.
- Update the path set: $K_{rs}^n = K_{rs}^{n-1} \cup \bar{k}_{rs}^n,\ \forall r, s.$
- Solve the following partial linearized subproblem, (20a) and (20b), for $\mathbf{h} = (\ldots, h_k^{rs}, \ldots, \tau_a, \ldots)$:

$$\underset{\mathbf{h}}{\text{Minimize}}: \quad \nabla_{\mathbf{f}}P(0)^T\mathbf{h} + G(\mathbf{h})$$

$$\text{subject}: \quad \mathbf{\Delta}_1\mathbf{h} \geq \mathbf{v}, \quad \mathbf{\Delta}_2\mathbf{h} \leq \mathbf{v}, \quad \mathbf{\Delta}_3\mathbf{h} \leq \mathbf{C} \quad \text{and} \quad \mathbf{h} \geq \mathbf{0}.$$

The above partial linearized subproblem is solved using the iterative balancing scheme to be discussed later. Besides the primal variables ($\mathbf{h}$), the dual variables $(\ldots,\ell_a,\ldots,u_a,\ldots,d_a,\ldots)$ are also available. Set $\ell_a^n = \ell_a, u_a^n = u_a, d_a^n = d_a$.
- Set $\mathbf{f}^1 = \mathbf{h}$, and update the link-flows: $x_a^n = \sum_{rs\in RS}\sum_{k\in K_{rs}^n} f_k^{rs}(n)\delta_{ka}^{rs},\ \forall a \in A.$

Step 1 (*Column generation*):
- Set iteration counter: $n = n + 1$.
- Update the link costs: $\tilde{t}_a^n = t_a(x_a^{n-1}) - \ell_a^{n-1} - u_a^{n-1} - d_a^{n-1},\ \forall a \in A.$
- Determine the shortest path for all O–D pairs based on $\tilde{t}_a^n$: $\bar{k}_{rs}^n,\ \forall r, s.$
- Update the path set: $K_{rs}^n = K_{rs}^{n-1} \cup \bar{k}_{rs}^n,\ \forall r, s.$

Step 2 (*Direction finding*): Solve the following partial linearized subproblem for $\mathbf{h}$:

$$\underset{\mathbf{h}}{\text{Minimize}}: \quad \nabla_{\mathbf{f}}P(\mathbf{f}^{n-1})^T\mathbf{h} + G(\mathbf{h})$$

$$\text{subject}: \quad \mathbf{\Delta}_1\mathbf{h} \geq \mathbf{v}, \quad \mathbf{\Delta}_2\mathbf{h} \leq \mathbf{v}, \quad \mathbf{\Delta}_3\mathbf{h} \leq \mathbf{C} \quad \text{and} \quad \mathbf{h} \geq \mathbf{0}.$$

Step 3 (*Line search*): Solve the following one-dimension optimization problem for an optimal step size (ϕ):

$$\underset{\phi\in[0,1]}{\text{Minimize}} \quad Z(\mathbf{f}^{n-1} + \phi \cdot (\mathbf{h} - \mathbf{f}^{n-1}))$$

Step 4 (*Solution update*):
- Update the solution vector: $\mathbf{f}^n = \mathbf{f}^{n-1} + \phi \cdot (\mathbf{h} - \mathbf{f}^{n-1})$
- Update the link-flows: $x_a^n = \sum_{rs\in RS}\sum_{k\in K_{rs}^n} f_k^{rs}(n)\delta_{ka}^{rs},\ \forall a \in A.$

Step 5 (*Convergence test*): If a convergence criterion (e.g., maximum change of the path flow solution between two consecutive iterations is less than a predetermined threshold) is met, stop; otherwise, go to Step 1.

It should be noted that the updated link costs in Step 1 could be negative in the early iterations because some observed links were not on the paths generated by the path set. In such a case, the dual variables (i.e., for the lower limit) are adjusted to a large positive number in order to induce a virtual flow that satisfies the flow on the measured links to maintain feasibility, see Eqs. (10), (13), and (18). Accordingly, these measured links become very attractive (i.e., link with a negative cost) so as to be included in the shortest paths of the next iteration.

In our implementation, we used Pape's algorithm (1974), which is a label correcting method that can work with negative link costs as long as the network does not contain a negative cost loop. However, in the event that the algorithm detects a path with a negative cost loop, which may occur when working with realistic networks, all negative link costs are set to a very small positive number and the algorithm is then repeated. This approach although may bias the paths generated, it is believed to be simple and have minor impact on the solution quality since sufficient number of paths would eventually be generated in the later iterations. This is very common in many traffic assignment algorithms with a column generation scheme in which some paths generated in the early iterations may not be included in the final path set (e.g., column dropping) or may carry negligible flows.

Nonetheless, more appropriate modification to Pape's algorithm could be implemented to generate the shortest "simple path" with the presence of negative cycle (e.g., the labeling and scanning method). The basic idea is to include a scan operation to the shortest path algorithm to eliminate negative cost cycles, see Tarjan (1983) for additional information. Moreover, a path set generated by a choice set generation scheme based on behavioral survey (Bekhor et al., 2001) could also be used as an initial path set in the initialization (Step 0) to minimize the occurrence of generating paths with a negative cost loop.

3.1. Direction finding: iterative balancing scheme

As mentioned above, the partial linearized subproblem is a nonlinear program with linear inequality constraints. One efficient approach for solving this subproblem is the iterative balancing scheme adopted by Bell and Shield (1995) to solve the original PFE formulation with right-angle cost function (i.e., flow-independent link cost function). The basic idea is to scale the path flows to fulfill one constraint at a time by adjusting the dual variables (e.g., ℓ_a, u_a, and d_a). The adjustment of the dual variable is to bring the violated constraint satisfactory. Here we just provide the adjustment equations for different types of constraint (e.g., lower limit, upper limit, and link-capacity constraints). The full derivations of the adjustment equations are provided in Appendix A. The adjustment of d_a is provided first since the capacity constraint is common in all three L_p-PFE models and followed by the lower and upper limits for each of the three L_p-PFE formulations.

3.1.1. Handling link-capacity constraint

The adjustment factor for the dual variable (d_a) of the link-capacity constraint (5d), (11d), (14d) is:

$$\lambda_a = \frac{1}{\theta}\ln\left(\frac{C_a}{x_a}\right), \quad \forall a \in U. \tag{21}$$

3.1.2. Handling lower limit of link-flow estimate constraint

The adjustment factors for the dual variables (β_a) of the lower limit of link-flow estimate constraints for each of the three L_p-PFE formulations are given as follows.

Constraint (5b) of the L_∞-norm model

$$\beta_a = \frac{1}{\theta}\ln\left(\frac{v_a}{x_a + \psi_o}\right), \quad \forall a \in M. \tag{22}$$

Constraint (11b) of the L_1-norm model and constraint (14b) of the L_2-norm model

$$\beta_a = \frac{1}{\theta}\ln\left(\frac{v_a}{x_a + \psi_a}\right), \quad \forall a \in M. \tag{23}$$

3.1.3. Handling upper limit of link-flow estimate constraint

The adjustment factors for the dual variables (π_a) of the upper limit of link-flow estimate constraints for each of the three L_p-PFE formulations are given as follows.

Constraint (5c) of the L_∞-norm model

$$\pi_a = \frac{1}{\theta}\ln\left(\frac{v_a + \sqrt{v_a^2 + 4x_a\psi_o}}{2x_a}\right), \quad \forall a \in M. \tag{24}$$

Constraint (11c) of the L_1-norm model and constraint (14c) of the L_2-norm model

$$\pi_a = \frac{1}{\theta}\ln\left(\frac{v_a + \sqrt{v_a^2 + 4x_a\psi_a}}{2x_a}\right), \quad \forall a \in M. \tag{25}$$

For a given path set fixed at (outer) iteration n (the master problem), the iterative balancing scheme is performed using the adjustment equations developed above until the algorithm converges (e.g., insignificant change of the primal and dual variables). The steps are summarized as follows.

Step 0 *Initialization*. Set $m = 0$, $u_a^m = d_a^m = 0, \ell_a^m = \eta_o$ for all links. Note that η_o is the threshold for termination (e.g., the maximum adjustment among all dual variables – 10^{-6}). Set $m = m + 1$ (m is the counter of the iterative balancing scheme).
Step 1 *Update dual variables*.
Set $\ell_a^m = \ell_a^{m-1},\ u_a^m = u_a^{m-1},$ and $d_a^m = d_a^{m-1}$.
Compute primal variables (path flows).
Compute flow on the physical paths:

$$h_k^{rs} = \exp\left(\theta\left(-\sum_{a\in A} t_a(x_a^{n-1})\delta_{ka}^{rs} + \sum_{a\in M}(\ell_a^m + u_a^m)\delta_{ka}^{rs} + \sum_{a\in U} d_a^m\delta_{ka}^{rs}\right)\right), \quad \forall k \in K_{rs},\ rs \in \mathrm{RS}.$$

Compute flow on the virtual paths:
For the L_∞-norm model, $\tau_o = \exp\left(\theta(-\rho_o + \sum_{a\in M}(\ell_a^m - u_a^m))\right)$
For the L_1-norm model, $\tau_a = \exp(\theta(-\rho_a + \ell_a^m - u_a^m)),\ \forall a \in M$

For the L_2-norm model, $\tau_a = \exp(\theta(-2\rho_a\psi_a^{n-1} + \ell_a^m - u_a^m)),\ \forall a \in M$

Compute the link-flows: $y_a = \sum_{rs\in RS}\sum_{k\in K_{rs}} h_k^{rs}\delta_{ka}^{rs},\ \forall a \in A$.
For each unmeasured links $a \in$ U, update the dual variables using the adjustment factor (λ_a) computed by Eq. (21)

$$d_a^m = \mathrm{Min}\{0, d_a^m + \lambda_a\}. \tag{26}$$

For each measured links $a \in$ M, compute the adjustment factors (β_a and π_a).
For the L_∞-norm model, use Eqs. (22) and (24).
For the L_1-norm and L_2-norm models, use Eqs. (23) and (25).
Update the dual variables of the lower and upper limits of link-flow estimates as follows:

$$\ell_a^m = \mathrm{Max}\{0, \ell_a^m + \beta_a\}, \tag{27}$$
$$u_a^m = \mathrm{Min}\{0, u_a^m + \pi_a\}. \tag{28}$$

Step 3 *Convergence testing*. Determine the maximum adjustment of all dual variables (ξ) as shown below.

$$\xi = \underset{\forall a}{\mathrm{Max}}\left\{|\ell_a^m - \ell_a^{m-1}|, |u_a^m - u_a^{m-1}|, |d_a^m - d_a^{m-1}|\right\}. \tag{29}$$

If $\eta_o < \xi < \eta$, set $m = m + 1$ and go to Step 1. If $\xi \geqslant \eta$ or $\xi \leqslant \eta_o$, stop and set $\ell_a^n = \ell_a^m,\ u_a^n = u_a^m,$ and $d_a^n = d_a^m$. Note that η is the upper limit of the adjustment allowed (e.g., 10^6 for detecting divergence).

3.2. Column generation

A column generation is also embedded into the solution procedure to avoid path enumeration for a general transportation network. Link costs are updated based on the new estimated flows and the dual variables. The shortest path is then generated for each O–D pair accordingly and added into the current path set in order to reduce the discrepancies between estimated and observed link-flows as much as possible. The idea is similar to that of the original PFE model, which is to generate new paths in order to ease the replication of link-flows. The difference is that the L_p-PFE models take into account the necessity of replicating individual observations explicitly (or directly) in the objective function (i.e., the penalty term). Therefore, the values of dual variables associated with individual constraints represent not only the corrections of mis-specified link costs, but also the penalty when the estimation fails to replicate the observed link-flows.

4. Numerical examples

4.1. Problem descriptions

To illustrate the application of the L_p-PFE formulations in handling inconsistency among traffic counts, one grid network and one real network are used. Due to its simplicity in topology, the grid network as depicted in Fig. 1 is used to study the solution properties as well as the characteristics of three L_p-PFE models. The grid network consists of 9 nodes, 14 uni-directional links, and 9 O–D pairs. Nodes 1, 2, and 4 are origin nodes while nodes 6, 8, and 9 are destination nodes (all shaded

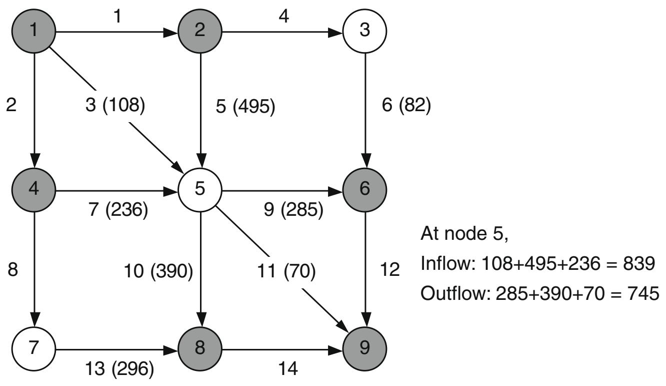


Fig. 1. Grid network and observed link volumes.

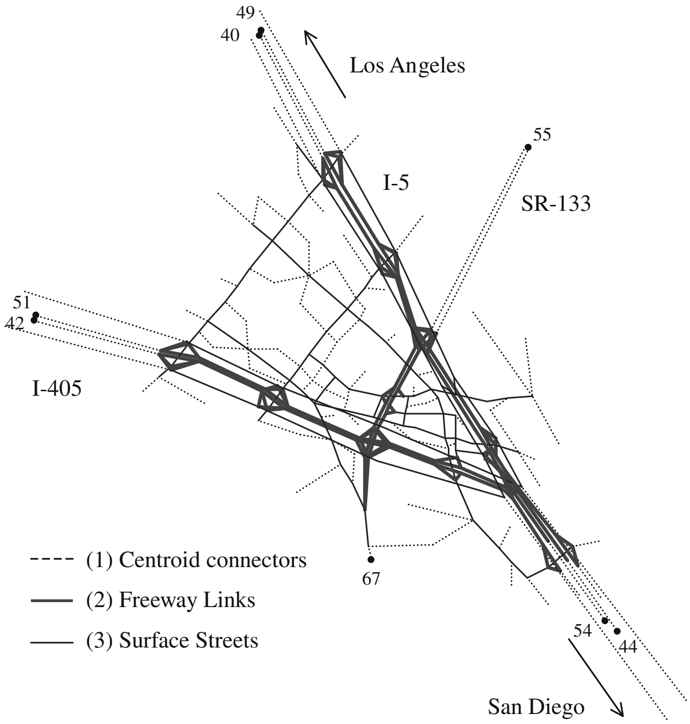


Fig. 2. Irvine network, City of Irvine, California.

nodes in Fig. 1). For this network, there are a total of 33 paths, all of which will be included in the estimation. The real road network in the City of Irvine (Orange County, California, USA) depicted in Fig. 2 is used to illustrate the practicality of the proposed L_p-PFE models. It consists of three major freeways (I-5, I-405, and SR-133) and several arterials within the study area. The Irvine network and its associated demand data were extracted from the Orange County Transportation Analysis (OCTA) model. The extracted network is composed of 162 nodes, 496 links, 39 traffic analysis zones (TAZ), 28 external stations, and 1,547 O–D pairs.

In this study, two statistical measures for the accuracy of link-flow estimates, which are root mean squared error (RMSE) and mean absolute error (MAE) defined below, are considered.

$$\mathrm{RMSE} = \sqrt{\frac{1}{|M|}\sum\nolimits_{a\in M}(x_a - v_a)^2}, \tag{30}$$

$$\mathrm{MAE} = \frac{1}{|M|}\sum\nolimits_{a\in M}|x_a - v_a|, \tag{31}$$

where $|M|$ is the number of link observations, x_a and v_a are the estimated and observed flows on link a, respectively. Further note that this study adopts the standard Bureau of Public Road (BPR) function as the link travel time function:

$$t_a = t_a^f \cdot (1 + 0.15 \cdot (x_a/C_a)^4), \tag{32}$$

where t_a^f is the free-flow travel time on link a.

4.2. Grid network

The characteristics of the grid network at the link level are summarized in Table 1 (see also Yang et al., 2001). For demonstration purpose, it is assumed that the true O–D trip table is available as shown in Table 2. In addition, this trip table is used to synthesize the observed traffic volumes according to the logit-based SUE model with a dispersion parameter of 1.50. Since the first set of traffic counts reported in Table 1 is a direct result of the logit-based SUE model, they are consistent (i.e., satisfy the conservation of flows at all intermediate nodes). To create inconsistency in traffic counts, it is assumed that the observed traffic volumes are independent Poisson variates with means and variances equal to the link volumes in Set 1. The second set of traffic counts (also see Table 1) is one instance (sample) generated according to this assumption. Observed traffic volumes are assumed available only on links 3, 5, 6, 7, 9, 10, 11, and 13 (8 out of 14 links). Links 6, 9, 10, 11, and 13 were selected to intercept the total demand from all origins to all destinations (see Yang et al. (1991,2005), Yang and Zhou (1998), Bierlaire (2002), Chen et al. (2005,2007), Chootinan et al. (2005a,b) for a discussion on selecting traffic counts to intercept the total demand). Three additional links (links 3, 5 and 7) are included to create nodal-inconsistency at node 5 (see Fig. 1).

4.2.1. Setting of experiments

To examine how the original PFE model and the three L_p-PFE models resolve the inconsistency problem of traffic counts, six experiments using the observed volumes in Set 2 are considered. Experiments A, B, and C are designed to examine the effects of user-specified error bounds in the original PFE model. Experiments D, E, and F are for the three proposed L_p-PFE models, which do not require the user-specified error bounds for the measured links.

In experiment A, the user is assumed to know the exact errors for all measured links. In other words, the user knows a set of consistent link volumes with the least deviation from the observed values. Several criteria discussed earlier (e.g., L_1-norm or L_2-norm) could be used to define such a deviation when preprocessing these inconsistent data. The L_1-norm criterion is used to determine the error bounds to obtain consistent link volumes for experiment A1 and the L_2-norm criterion for experiment A2 (see Table 3). An optimal determination of a consistent set of link volumes (i.e., preprocessing method) involves solving a mathematical program that minimizes the deviation between observed and adjusted volumes subject to conservation of flows at intermediate nodes (Kikuchi et al., 2000). It should be mentioned that the preprocessing method for PFE has to take the link-capacity constraints (Eq. (3)) into account as well. The reader may refer to Chootinan (2006) for the detailed descriptions of the preprocessing method employed in this study. In experiment B, the user is assumed to have a rough idea

Table 1
Link characteristics of grid network and observed link volumes.

Link	Node		Capacity	Free-flow travel time	SUE link-flow (Set 1)	Observed link-flow (Set 2)
	From	To				
1	1	2	280.00	2.00	124	–
2	1	4	290.00	1.50	137	–
3	1	5	280.00	3.00	109	108
4	2	3	280.00	1.00	77	–
5	2	5	600.00	1.00	467	495
6	3	6	300.00	2.00	77	82
7	4	5	500.00	2.00	212	236
8	4	7	400.00	1.00	295	–
9	5	6	500.00	1.50	303	285
10	5	8	700.00	1.00	400	390
11	5	9	250.00	2.00	85	70
12	6	9	300.00	1.00	50	–
13	7	8	350.00	1.00	295	296
14	8	9	220.00	1.00	165	–

Table 2
True O–D trip table of grid network.

From/to	6	8	9
1	120	150	100
2	130	200	90
4	80	180	110

Table 3
Consistent (preprocessed) observed link volumes for grid network.

Link	Observation	L_1-norm criterion			L_2-norm criterion		
		Consistent flow[a]	Absolute error	% Error	Consistent flow[a]	Absolute error	% Error
3	108	108.00	0.00	0.00	92.33	15.67	14.51
5	495	491.80	3.20	0.65	479.33	15.67	3.17
6	82	82.00	0.00	0.00	82.00	0.00	0.01
7	236	184.78	51.22	21.71	220.33	15.67	6.64
9	285	285.00	0.00	0.00	300.67	15.67	5.50
10	390	394.00	4.00	1.03	405.67	15.67	4.02
11	70	105.58	35.58	50.84	85.67	15.67	22.39
13	296	296.00	0.00	0.00	296.00	0.00	0.00
			Average	9.28		Average	7.03

[a] One pattern among many possibilities (non-unique).

on the quality of traffic counts and uses the average error determined by the L_2-norm criterion as the default error bound for all measured links (uniform error bound). In fact, the minimum uniform error bound, which still results in a solvable system of equations, could be determined by the L_∞-norm criterion (also see Chootinan (2006)). The minimum uniform error bound used for this data set is 5.39%. Experiment C examines the effect of mis-specified (i.e., unnecessarily large) error bound on the estimation results.

4.2.2. Handling traffic inconsistency with different PFE models

Table 4 summarizes the estimation results of the experiments discussed in Section 4.2.1 by solving different PFE models. Several numerical indices, for instance, two statistical measures (MAE and RMSE) of link-flow estimates, PFE objective value, and computational requirements, are provided for comparison purpose (see also Fig. 3 for a graphical comparison of these measures). Table 5 also provides the estimation of individual O–D demands and total demand for each experiment. With a proper setting of the error bounds, the original PFE model can perform the estimation quite well. A well performance here is

Table 4
Estimation results for the grid network.

Experiment	Maximum error	MAE	RMSE	PFE objective	Norm objective[a]	Penalty[b]	No. of iterations		
							Total	Main (n)	Average inner (m)
A1	51.22	11.75	22.12	5967.29	–	–	1207	9	134
A2	15.67	11.75	13.57	6007.24	–	–	480	10	48
B (average error)	34.80	15.07	18.69	5792.24	–	–	500	10	50
C1 (5.39%)	29.38	14.55	16.85	5873.17	–	–	564	12	47
C2 (10.0%)	49.50	23.60	27.22	5577.17	–	–	296	10	30
C3 (12.5%)	61.88	26.56	31.63	5401.02	–	–	210	10	21
D (L_∞-norm)	15.67	15.67	15.67	5828.77	2370.43	150.10	4467	15	298
E (L_1-norm)	45.49	11.75	20.38	5711.35	1216.68	11.27	5441	19	286
F (L_2-norm)	21.60	13.73	14.84	5820.61	604.11	0.27	110,428	474	233

[a] Virtual path entropy term in the modified L_p-PFE objective function.
[b] Penalty parameter.

Table 5
O–D trip tables estimated by various PFE models.

O–D pair	Ref. demand	Experiment								
		A1	A2	B	C1 (5.39%)	C2 (10.0%)	C3 (12.5%)	D (L_∞)	E (L_1)	F (L_2)
(1,6)	120.00	47.42	45.42	46.12	48.30	41.35	40.20	44.81	35.94	43.11
(1,8)	150.00	84.56	80.40	84.42	84.03	84.36	81.74	79.14	68.16	77.37
(1,9)	100.00	49.07	42.36	42.45	42.71	41.76	40.88	41.99	32.73	39.93
(2,6)	130.00	203.01	205.60	191.58	198.82	175.34	170.47	193.40	206.00	198.29
(2,8)	200.00	193.12	192.03	192.59	190.61	194.50	188.29	191.97	195.25	191.61
(2,9)	90.00	150.45	137.07	126.84	127.59	124.83	122.17	134.42	131.26	132.99
(4,6)	80.00	49.65	61.87	59.86	62.82	53.37	51.89	61.87	58.15	60.51
(4,8)	180.00	288.73	303.71	291.58	293.09	285.90	277.51	291.97	299.68	296.41
(4,9)	110.00	96.57	101.53	96.01	96.81	93.89	91.73	99.09	95.85	98.38
Total	1160.00	1162.58	1170.00	1131.44	1144.78	1095.30	1064.87	1138.67	1123.01	1138.60

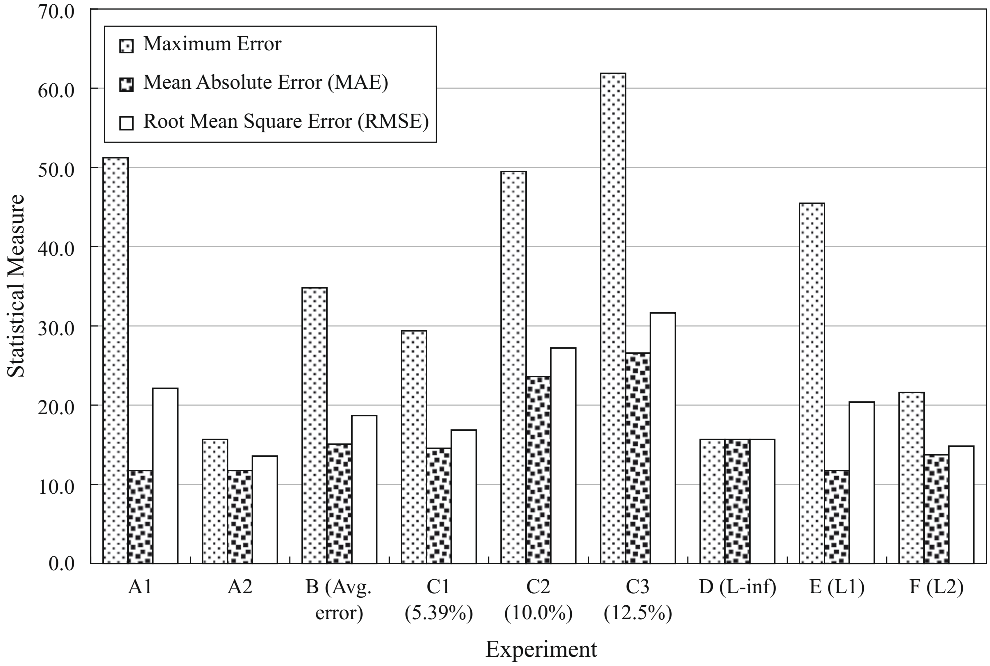


Fig. 3. Statistical measures of link-flow estimates by various PFE models.

indicated by the ability to produce a close estimation of observed link volumes (quantified by MAE and RMSE) and the ability to capture the total demand of the network. As expected, in experiment A1 when the error bounds were determined by the L_1-norm criterion, the MAE of link-flow estimates is minimized (compared to all other cases being discussed later). On the other hand, when the L_2-norm criterion is used as in experiment A2, the RMSE of link-flow estimates is minimized instead. In both experiments (A1 and A2), the total demand estimate do not deviate much from the true value (1,160 units) despite that the estimated demands of individual O–D pairs are quite different from the known values (see Table 5). This characteristic (i.e., the under-determinate nature) is common for the O–D estimation problem when only traffic counts (even on all network links) are used. One way to resolve this issue is to incorporate a target or reference O–D trip table into the O–D estimation method (see Yang (1995) for details). In the PFE model, this can be accomplished by adding selected O–D pairs of the

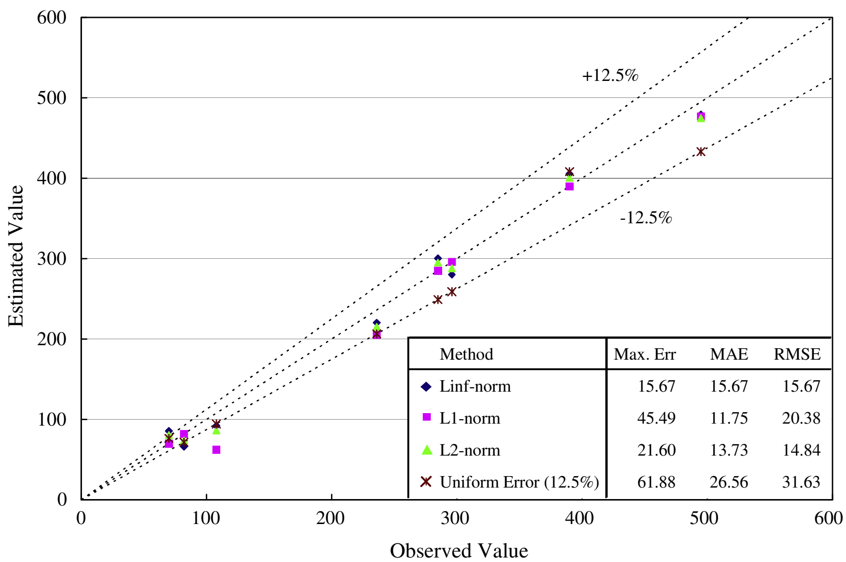


Fig. 4. Comparison of link-flow estimates for the grid network.

target trip table as constraints with appropriate bounds to improve the observability of the O–D estimation problem from traffic counts (see Chen et al. (2005) and Chootinan et al. (2005a) for examples).

In experiment B when the average error is specified for all measured link-flows (i.e., uniform error bound), the performance of the original PFE model deteriorates as indicated by the increased value of both statistical measures (MAE and RMSE) – link-flow estimates are farther away from the observed values. Experiment C examines the effects of specifying different uniform error bounds for all measured links. As mentioned earlier, the smallest uniform error bound required to resolve the inconsistency of this data set is 5.39% (experiment C1). If a smaller error bound (e.g., <5.39%) is used, there is no feasible solution. When the uniform error bound (in the original PFE model) is set larger than 5.39% (see experiments C2 – 10% and C3 – 12.5%), there could be several feasible (consistent) link-flow patterns within the specified error bounds. The original PFE model will select a path flow pattern that gives the lowest objective value despite that the solution has a higher MAE/RMSE value (a lower quality of link-flow replication). This is reasonable because neither MAE nor RMSE is a quantity to be minimized in the original PFE objective function. This phenomenon, as discussed earlier, will lead to the underestimation

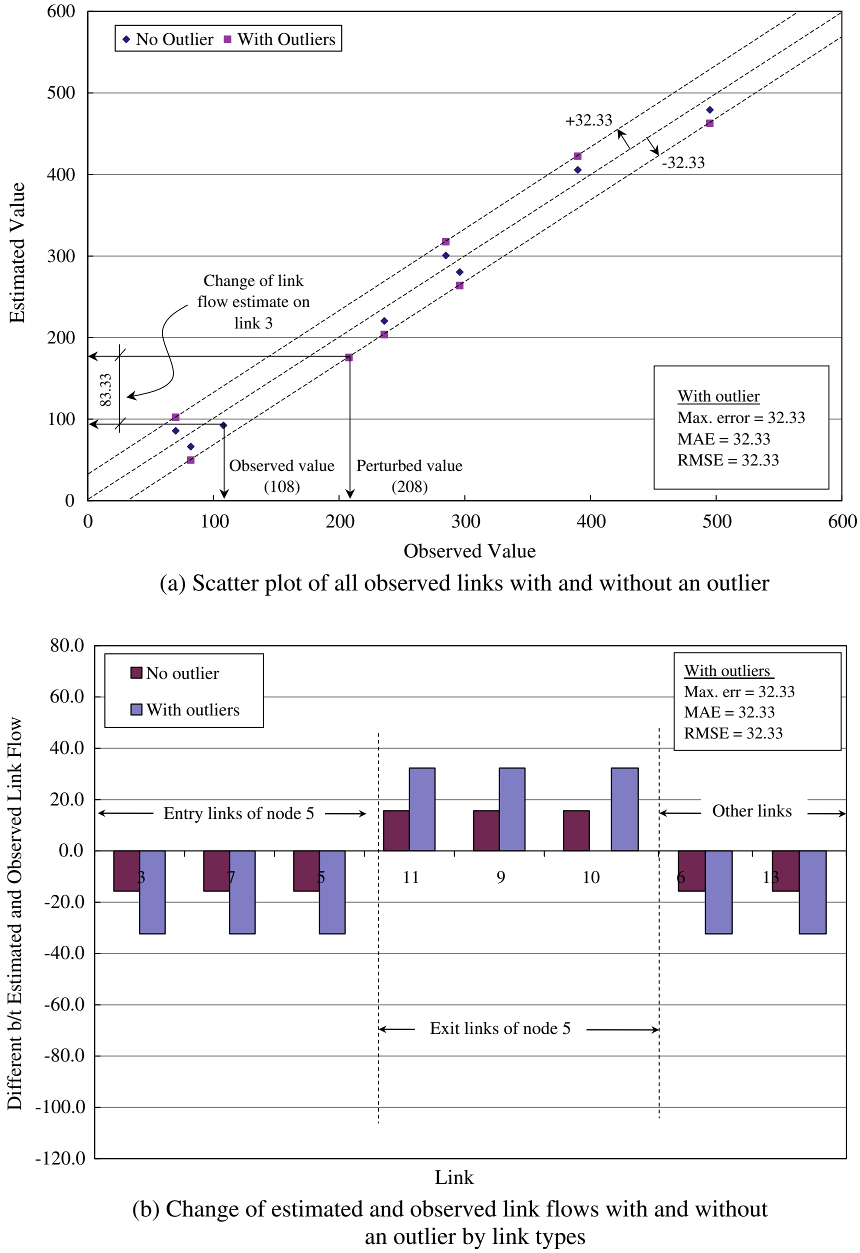
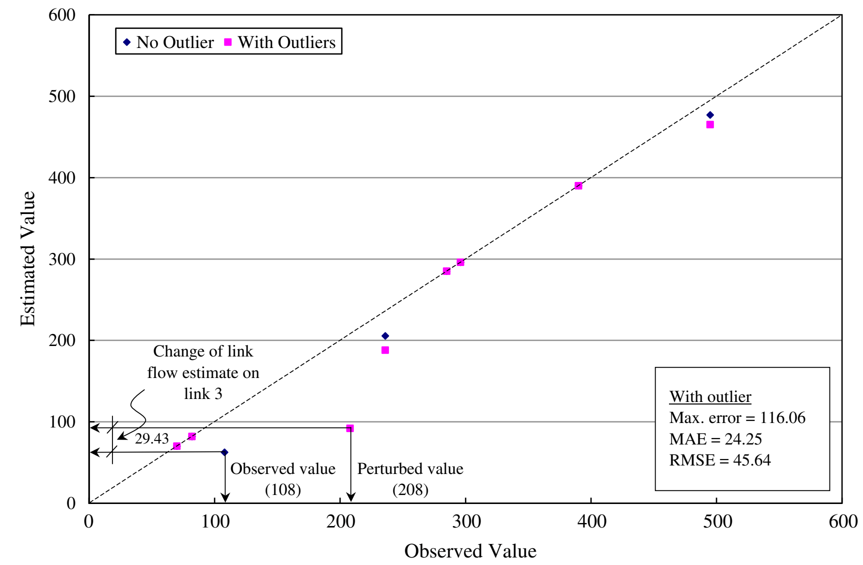


Fig. 5. Effect of an outlier on the link-flow estimates obtained by the L_∞-PFE model.

of the total travel demand. That is, if the error bounds are specified too loose, the original PFE model will select a solution with a lower objective value (i.e., lower travel cost and path entropy), thus a lower total demand. From Table 4 (among the three C experiments), it is also observed that PFE is likely to find a solution faster (e.g., less iterations) when the error bounds are large.

When the three L_p-PFE models are applied to this data set, they are expected to resolve the inconsistency problem differently according to the criterion incorporated into the model. As expected, in experiment D, the L_∞-norm model minimizes the maximum absolute error among all observations, the L_1-norm model minimizes the mean absolute error (MAE) as in experiment E, and the L_2-norm model minimizes the root mean squared error (RMSE) as in experiment F. However, by comparing the results of experiments F and A2, they are quite different (i.e., RMSE in case A2 is lower than that in case F) even though they are based on the same criterion for handling data inconsistency (L_2-norm). This is due to the difficulty in setting the penalty cost (ρ) in the L_p-PFE models. Basically, the penalty cost in the L_2-norm model could not be increased to the level at which the RMSE will be minimal without causing any numerical difficulty (i.e., ill-conditioned problem). Based on the limited results, this issue only occurs with the L_2-norm model because of the quadratic cost function (penalty term) in the



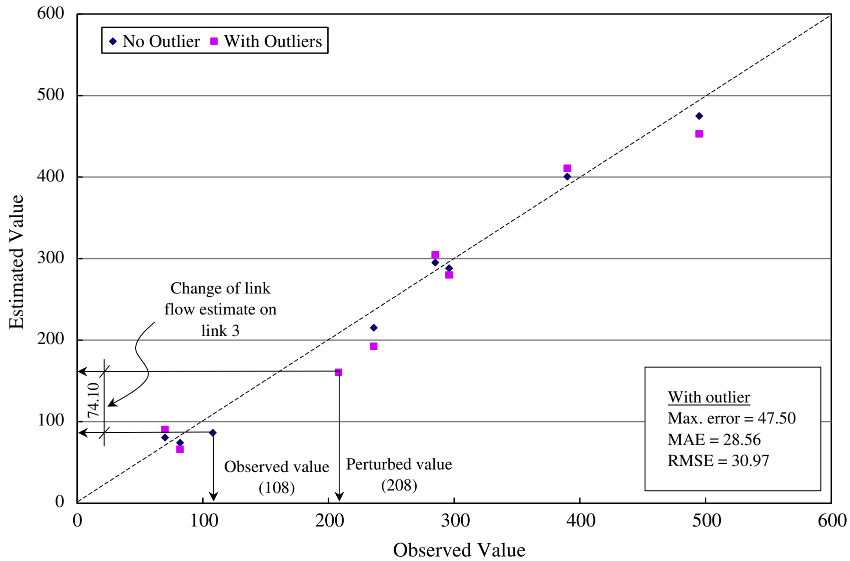
(a) Scatter plot of all observed links with and without an outlier

(b) Change of estimated and observed link flows with and without an outlier by link type

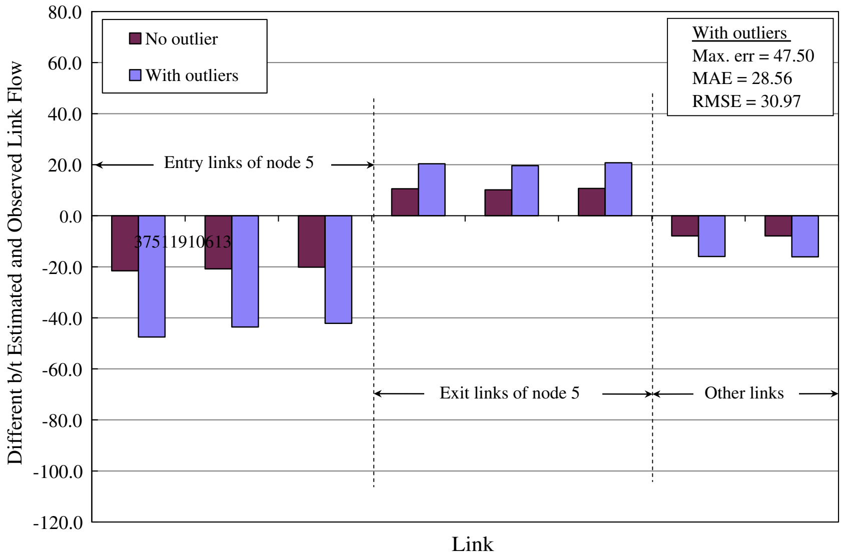
Fig. 6. Effect of an outlier on the link-flow estimates obtained by the L_1-PFE model.

modified objective function. This explanation is supported by the computation required to solve the L_2-norm model, which is generally higher than those required by the other models. Among the three proposed L_p-PFE models and the original PFE model with uniform error bounds, the L_2-PFE model, however, gives the lowest RMSE (14.84), which indicates that such a modification still inherits the property of L_2-norm criterion in handling the inconsistent traffic counts.

Another observation is that experiments A1 and E obtain the estimations with the lowest MAE (11.75). In both experiments, the L_1-norm criterion is used either for preprocessing inconsistent data or for modifying the PFE formulation. However, the detailed solutions (e.g., individual link-flow estimates) obtained from both experiments are different (see also other indices in Table 4). This observation is due to the non-uniqueness property of the mathematical formulation (linear program) for preprocessing traffic counts (Kikuchi et al., 2000). In other words, the predetermined error bound provided in Table 3 is one solution among many possibilities. In general, the L_p-PFE models require a higher computational time to internally resolve the inconsistency problem of traffic counts compared to the PFE model with pre-specified error bounds as shown in experiments A, B, and C of Table 4. It is also observed that the original PFE is likely to find a solution faster with less number



(a) Scatter plot of all observed links with and without an outlier



(b) Change of estimated and observed link flows with and without an outlier by link type

Fig. 7. Effect of an outlier on the link-flow estimates obtained by the L_2-PFE model.

of iterations when the error bounds are large (however, the results will be biased toward a lower total demand). As for the three L_p-PFE models proposed in this paper, the L_∞-norm model takes the least number of iterations to converge, followed by the L_1-norm model, and the L_2-norm model takes the most number of iterations to converge due to the extra computational efforts required to solve the quadratic penalty term as shown in experiments D, E, and F of Table 4.

Fig. 4 compares the link volumes estimated by different models. The data points along the 45° line represent an accurate estimate while the data points under (above) this line represent an underestimate (overestimate) of the observed link volume. As can be seen, most of the link-flow estimates obtained by the L_1-norm model lie almost exactly on the 45-degree line with the exception of a few data points while those obtained by the L_∞-norm and L_2-norm models cluster around the 45° line. When unnecessarily large uniform error bounds (12.5%) are used in the original PFE model, majority of the estimated volumes (6 out of 8 data points) lie on the lower limit (−12.5%), which results in a lower PFE objective value. The distributions of link-flow estimates obtained by the L_p-PFE models are consistent with the characteristics of different norms discussed earlier. Namely, the L_1-norm model aims to reproduce most of the data points by disregarding a few points, which are sometimes believed to be outliers. On the other hand, the L_∞-norm and L_2-norm models distribute the amount of underestimated and overestimated flows (i.e., the number of data points below and above the 45° line) such that the maximum absolute error and the RMSE are, respectively, minimized.

4.2.3. Effects of an outlier in the observed link volumes

This section is provided to investigate the effects of an outlier in the observed link volumes on the performance of the L_p-PFE models. The observed traffic volume on link 3 is intentionally perturbed from 108 units to 208 units to create an outlier in the observed link volumes. The estimations using the proposed L_p-PFE models are repeated on this perturbed data set in which a high inconsistency of traffic counts is expected. Figs. 5–7 compare the estimated link volumes obtained using the original and the perturbed data sets. Figs. 5a, 6a, and 7a, respectively, show the scatter plots of the observed and estimated link volumes of both data sets for each of the three L_p-PFE models. Figs. 5b, 6b, and 7b further show how each model adjusts the estimated link volumes by link types surrounding node 5 (e.g., entry links, exiting links, and other links). As can be seen, each model has to redistribute the link-volume estimates in order to minimize the corresponding objective value (e.g., Max.

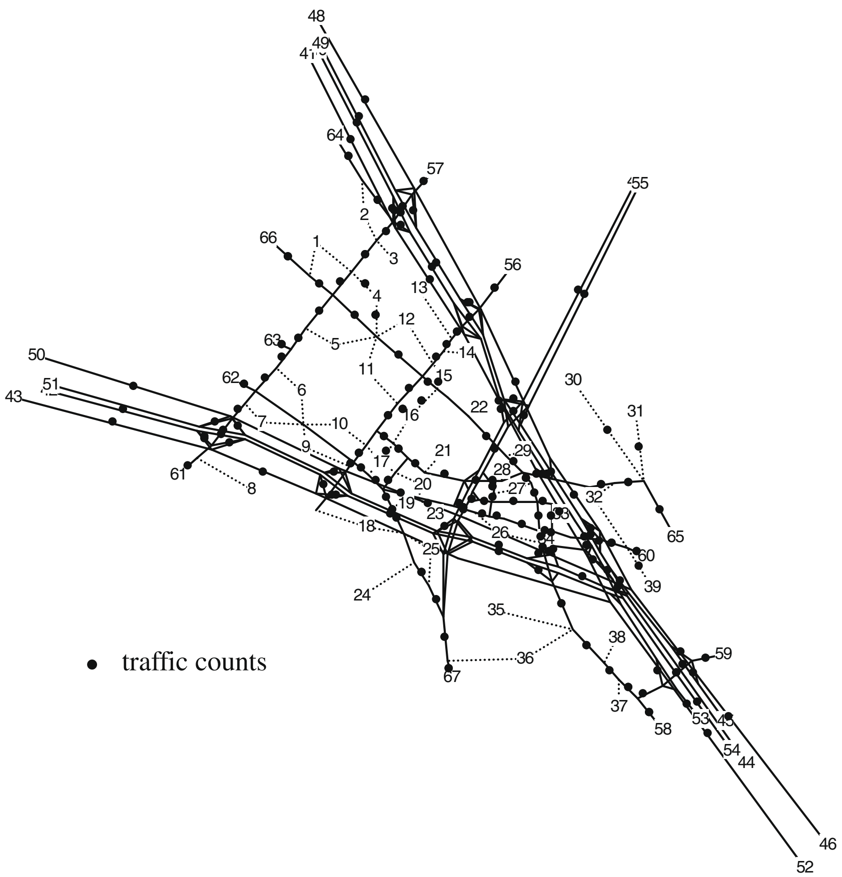


Fig. 8. Locations of traffic counts.

Table 6
Estimation results for the Irvine network.

Method	Maximum error	MAE	RMSE	Total demand[a]	No. of paths	Total no. of iterations
Uniform error (5%)	538.18	61.48	96.68	45269.34	2053	1743
Heuristic	405.14	22.77	49.96	47117.50	2071	48,712
L_∞-norm (maximum error)	134.49	93.19	104.75	44036.25	2097	35,082
L_1-norm (average error)	340.34	12.41	46.63	48176.97	1,939	154,989
L_2-norm (average square error)	204.56	15.80	32.81	47082.83	1,959	313,231

[a] The total demand from the OCTA model is 47,522.

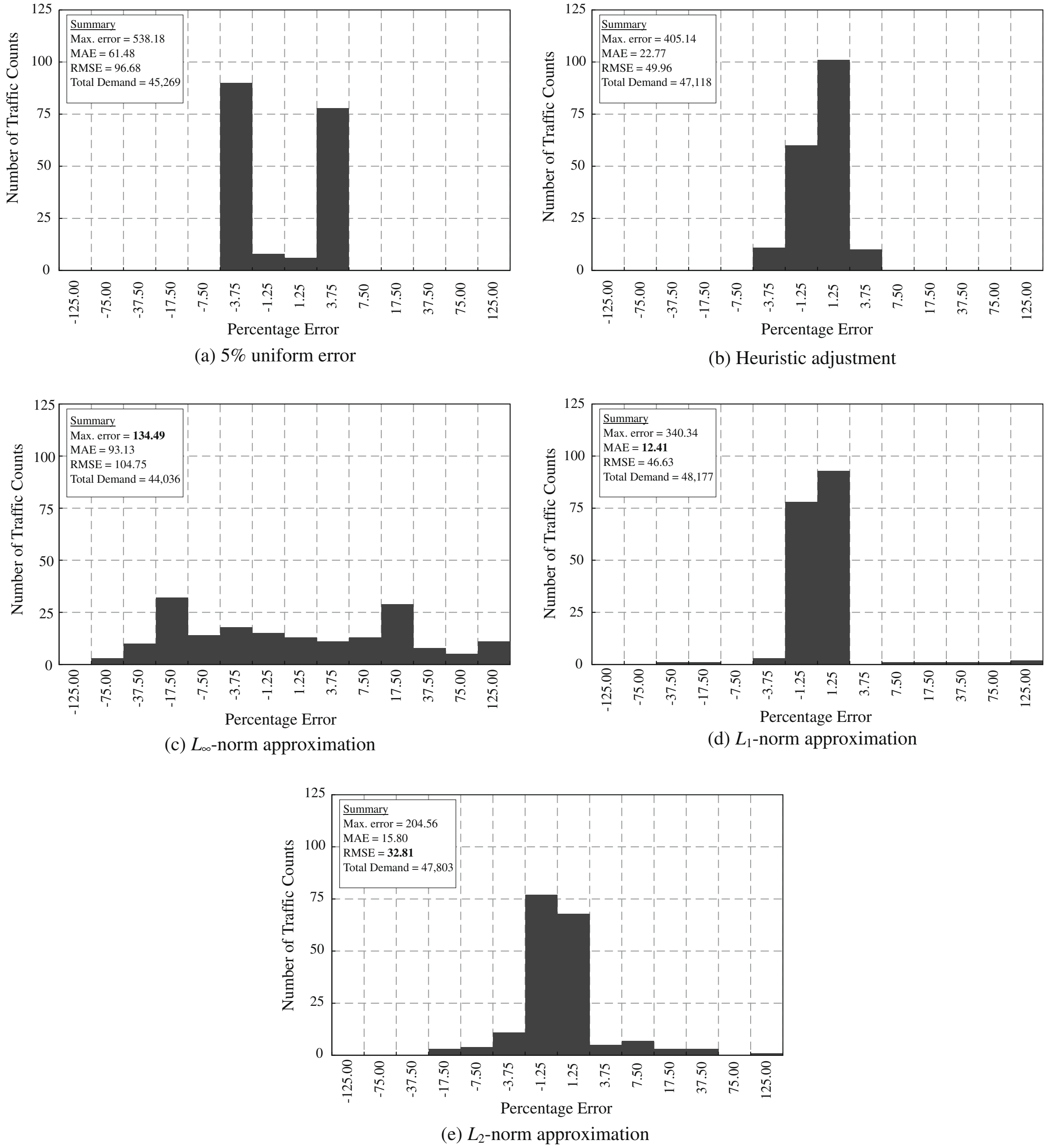


Fig. 9. Error distributions of link-flow estimates for different PFE models.

error, MAE, or RMSE) in the presence of an outlier in the observed link volumes. However, such a re-distribution is more pronounced in the L_∞-norm and L_2-norm models, but not in the L_1-norm model. As can be seen in Fig. 6a and b, majority of the link volumes estimated by the L_1-norm model remains unchanged. The perturbed data point is regarded as an outlier and does not affect the overall estimation of the L_1-norm model. As mentioned before, the L_1-norm model is insensitive to outliers. On the other hand, both L_∞-norm and L_2-norm models have to adjust the estimated flows on all observed links to accommodate the outlier in order to minimize the maximum absolute error and RMSE, respectively. It should also be noted that even with a proper re-estimation, the indices of link-flow deviation (e.g., Max. error, MAE, or RMSE) become higher in all cases due to a higher degree of data inconsistency as shown in the figures.

4.3. Irvine network

In this section, various PFE models are applied to estimate the O–D trip table for the Irvine network depicted in Fig. 2. Traffic counts are generated by assigning the O–D trip table extracted from the OCTA model onto the network according to the SUE principle. Only a subset of links is assumed to have traffic counts. Their locations are selected using the O–D separation rule (Yang et al., 2005) to completely observe the total trips traversing the network (see also Chen et al. (2005,2007), Chootinan et al. (2005a,b) for additional discussions on the selection of traffic counts for O–D trip table estimation). Accordingly, there are 186 observations (38% of network links) available for the estimation as depicted in Fig. 8. To create inconsistency in traffic counts, it is assumed that the measurement error of each traffic count follows a normal distribution with a zero mean and a standard deviation of one-third of its mean (note that other types of distribution could also be used to generate the traffic count data). Then 100 samples of traffic counts are generated based on the above assumption and the average volumes of individual links are used as the actual traffic counts. The results obtained by the original PFE model, the three L_p-PFE models, and the heuristic proposed by Chootinan et al. (2005a) are presented in Table 6.

It should be mentioned that the heuristic adjustment is based on repeatedly solving the original PFE model with a progressive adjustment of individual error bounds (see Chootinan et al. (2005a) for details of the heuristic procedure). The adjustment is guided by the difficulty in replicating individual observations as indicated by the magnitude of the dual variables of each observation constraint obtained by solving the subproblem of the original PFE model using the iterative balancing scheme.

The results of the Irvine network follow a similar trend as those of the grid network. In general, when compared to the original PFE model and the heuristic procedure, the three L_p-PFE models deliver trip tables that can better reproduce the observed link volumes. However, the three L_p-PFE models generally required a higher computational time to resolve the inconsistency problem of traffic counts compared to the original PFE model and the heuristic procedure as shown in Table 6. The L_∞-PFE model takes the least number of iterations to converge among the three L_p-PFE models. This is because the L_∞-PFE model only requires one additional variable (i.e., one virtual path for all observed links), while the L_1-PFE and the L_2-PFE models require as many additional variables as the number of observed links. However, the L_2-PFE model takes a significantly more iterations to converge compared to the L_1-PFE model due to the extra computational efforts required to solve the quadratic penalty term.

Fig. 9a–e depict the error distributions of link-flow estimates obtained by the various PFE models. From Fig. 9a, it shows that, with a uniform error bound, majority of the link-flow estimates are either at the upper limit or at the lower limit (±5%). Those at the upper limit are links typically associated with low flow values, while those at the lower limit are links with high flow values. The reason for such flow allocation pattern is that it can effectively minimize the objective value of the original PFE model while accommodating the inconsistent traffic counts. For the L_∞-PFE model, link-flow deviations are almost uniformly distributed across the boundary defined by the maximal error (±134.49 units of flow). For the L_1-PFE model, a majority of the link volumes (94% of traffic counts) has a small estimation error within ±2.5% (see Fig. 9d). A similar finding is also observed in the L_2-PFE model; approximately 80% of link volumes are within ±2.5% of their observed values (see Fig. 9e). From Fig. 9e, it can also be seen that the distribution of estimation errors obtained by the L_2-PFE model resembles the normal probability distribution; the amounts of underestimated flows (52%) and overestimated flows (48%) are approximately equal.

5. Concluding remarks

In this study, the PFE model was reformulated to internally resolve the inconsistency problem of traffic counts. Three different criteria for defining the estimation errors, namely L_∞-norm, L_1-norm, and L_2-norm, were considered and incorporated into the PFE model. A partial linearization algorithm embedded with an iterative balancing scheme and a column generation procedure was developed to solve the three L_p-PFE models. Numerical results indicate that the three L_p-PFE models are capable of estimating trip tables that better reproduce the observed traffic volumes. The advantage of the three L_p-PFE models is that they do not require a consistent (preprocessed) data or user-specified error bounds for each individual observation. However, it does require the specification of penalty parameters to internally handle the inconsistent traffic counts in the estimation. Ideally, the penalty cost has to be set as high as possible to minimize the estimation error. However, due to numerical difficulty (i.e., ill-conditioned problem), this penalty cost could not always be set at the level at which the estimation error will be minimized especially when the L_2-norm criterion is considered. In spite of this difficulty and a higher computational cost, the proposed L_p-PFE models generally perform better in terms of replicating the observed link volumes and eliminating the bias of underestimating the total demand in the original PFE model.

Acknowledgements

The authors would like to thank the reviewer and the editor for their constructive comments. This research was supported in part by the California Partners for Advanced Transit and Highways (PATH) Program through a grant (TO 5502). The contents of this paper reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein and do not necessarily reflect the views of our sponsors.

Appendix A

This appendix provides the full derivations of the adjustment equations for the link-capacity constraints, the lower limit and upper limit of link-flow constraints for each of the three L_p-norm models.

A.1. Handling link-capacity constraint

Consider the link-capacity constraint (5d), (11d), (14d). If flow on link a exceeds its link-capacity (i.e., $x_a > C_a$), there exists an adjustment factor (λ_a) associated with d_a to reduce the flow on link a back to its link-capacity. The adjustment factor is the root of the following equation, which is obtained by replacing the analytical expressions of flows on the physical and virtual paths into Eqs. (5d), (11d), (14d):

$$\sum_{rs\in RS}\sum_{k\in K_{rs}} \exp\left(\theta\left(-c_k^{rs}+\sum_{b\in M}(\ell_b+u_b)\delta_{kb}^{rs}+\sum_{b\in U} d_b\delta_{kb}^{rs}+\lambda_a\right)\right)\delta_{ka}^{rs}=C_a, \tag{a.1}$$

which is equivalent to:

$$\sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs}\delta_{ka}^{rs}\exp(\theta\cdot\lambda_a\delta_{ka}^{rs})=C_a \Rightarrow \quad \exp(\theta\cdot\lambda_a)\cdot x_a=C_a. \tag{a.2}$$

In Eq. (a.2), it should be noted that only the paths passing through link a are involved in the computation of link-flow (e.g., $\delta_{ka}^{rs}=1$). The exponential term is common to all the relevant paths and therefore can be moved outside the summation. After re-arranging Eq. (a.2), the adjustment factor for the dual variable of the link-capacity constraint (5d), (11d), (14d) can be written as follows:

$$\lambda_a=\frac{1}{\theta}\ln\left(\frac{C_a}{x_a}\right),\quad \forall a\in U. \tag{a.3}$$

A.2. Handling lower limit of link-flow estimates

The adjustments of ℓ_a (β_a) in all three L_p-PFE models are very similar. Here the derivation is provided only for the L_∞-PFE model while that for the L_1-PFE and L_2-PFE models can be derived along the same line. Consider constraint (5b) for the L_∞-PFE model (constraint (11b) for the L_1-PFE model and constraint (14b) for the L_2-PFE model) and replace the analytical expressions of path flows (both physical and virtual paths) into Eq. (5b). The idea is to increase flow on link a if the estimated flow is less than the lower limit by adjusting the associated dual variable ℓ_a by β_a

$$\sum_{rs\in RS}\sum_{k\in K_{rs}} \exp\left(\theta\left(-c_k^{rs}+\sum_{b\in M}(\ell_b+u_b)\delta_{kb}^{rs}+\sum_{b\in U} d_b\delta_{kb}^{rs}+\beta_a\right)\right)\delta_{ka}^{rs}=v_a-\exp\left(\theta\left(-\rho_o+\sum_{b\in M}(\ell_b-u_b)+\beta_a\right)\right), \tag{a.4}$$

which is equivalent to:

$$\sum_{rs\in RS}\sum_{k\in K_{rs}} f_k^{rs}\delta_{ka}^{rs}\exp(\theta\beta_a\delta_{ka}^{rs})=v_a-\psi_o\exp(\theta\beta_a), \tag{a.5}$$

$$x_a\cdot\exp(\theta\beta_a)=v_a-\psi_o\cdot\exp(\theta\beta_a). \tag{a.6}$$

This leads to the following adjustment factor for the dual variable of the lower limit constraint of the L_∞-PFE model:

$$\beta_a=\frac{1}{\theta}\ln\left(\frac{v_a}{x_a+\psi_o}\right),\quad \forall a\in M. \tag{a.7}$$

The adjustment factor for the L_1-PFE and the L_2-PFE models can be similarly derived as follows:

$$\beta_a = \frac{1}{\theta}\ln\left(\frac{v_a}{x_a+\psi_a}\right), \quad \forall a \in M. \tag{a.8}$$

A.3. Handling upper limit of link-flow estimates

Similar to the lower limit constraint, the adjustments of u_a (π_a) for all three L_p-PFE models are also similar. Consider constraint (5c) for the L_∞-PFE model (or constraint (11c) and (14c) for the L_1-PFE and L_2-PFE models, respectively) and replace the analytical expression of path flows (both physical and virtual paths) into Eq. (5c). The idea is to reduce flow on link a if the estimated flow is greater than the upper limit by adjusting the associated dual variable u_a by π_a

$$\sum_{rs\in RS}\sum_{k\in K_{rs}}\exp\left(\theta\left(-c_k^{rs}+\sum_{b\in M}(\ell_b+u_b)\delta_{kb}^{rs}+\sum_{b\in U}d_b\delta_{kb}^{rs}+\pi_a\right)\right)\delta_{ka}^{rs} = v_a+\exp\left(\theta\left(-\rho_o+\sum_{b\in M}(\ell_b-u_b)-\pi_a\right)\right), \tag{a.9}$$

which is equivalent to:

$$\sum_{rs\in RS}\sum_{k\in K_{rs}}f_k^{rs}\delta_{ka}^{rs}\exp(\theta\pi_a\delta_{ka}^{rs}) = v_a+\frac{\psi_o}{\exp(\theta\pi_a)}, \tag{a.10}$$

$$x_a\cdot\exp(\theta\pi_a) = v_a+\frac{\psi_o}{\exp(\theta\pi_a)}. \tag{a.11}$$

Let $\sigma_a = \exp(\theta\pi_a)$. Eq. (a.11) can be rewritten as follows:

$$x_a\sigma_a^2 - v_a\sigma_a - f_o = 0. \tag{a.12}$$

Since $\exp(\theta\pi_a)$ can only be non-negative, the non-negative root of the above polynomial equation determines the value of π_a.

$$\sigma_a = \frac{v_a \pm\sqrt{v_a^2+4x_a\psi_o}}{2x_a}. \tag{a.13}$$

In Eq. (a.13), $v_a^2+4x_a\psi_o$ is clearly positive and greater than or equal to v_a^2. In addition, this adjustment is required only when the upper limit is activated (i.e., link-flow estimate exceeds the upper bound), thus x_a is always positive and Eq. (a.13) is always valid (i.e., never divide by zero). The adjustment for the dual variables of the upper limit constraints of the L_∞-PFE model is given in Eq. (a.14) while those of the L_1-PFE and L_2-PFE models are given in (a.15).

$$\pi_a = \frac{1}{\theta}\ln\left(\frac{v_a+\sqrt{v_a^2+4x_a\psi_o}}{2x_a}\right), \quad \forall a\in M. \tag{a.14}$$

$$\pi_a = \frac{1}{\theta}\ln\left(\frac{v_a+\sqrt{v_a^2+4x_a\psi_a}}{2x_a}\right), \quad \forall a\in M. \tag{a.15}$$

References

Bekhor, S., Ben-Akiva, M., Ramming, S., 2001. Route choice: choice set generation and route choice models. In: Proceedings of the IV Tristan Conference, Azores, Portugal.

Bell, M.G.H., Shield, C.M., 1995. A log-linear model for path flow estimation. In: Proceedings of the 4th International Conference on the Applications of Advanced Technologies in Transportation Engineering, Carpi, pp. 695–699.

Bell, M.G.H., Shield, C.M., Busch, F., Kruse, G., 1997. A stochastic user equilibrium path flow estimator. Transportation Research Part C 5 (3/4), 197–210.

Bierlaire, M., 2002. The total demand scale: a new measure of quality for static and dynamic origin–destination trip tables. Transportation Research Part B 36 (9), 837–851.

Chen, A., Chootinan, P., Recker, W., 2005. Examining the quality of synthetic origin–destination trip table estimated by path flow estimator. Journal of Transportation Engineering 131 (7), 506–513.

Chen, A., Pravinvongvuth, S., Chootinan, P., Lee, M., Recker, W., 2007. Strategies for selecting additional traffic counts for improving O–D trip table estimation. Transportmetrica 3 (3), 191–211.

Chvatal, V., 1983. Linear Programming. W.H. Freeman and Company, New York.

Chootinan, P., 2006. A framework for estimating origin–destination trip tables from observed traffic data: a quality assessment methodology. Ph.D. Dissertation, Department of Civil and Environmental Engineering, Utah State University.

Chootinan, P., Chen, A., Recker, W., 2005a. Improved path flow estimator for estimating origin–destination trip tables. Transportation Research Record 1923, 9–17.

Chootinan, P., Chen, A., Yang, H., 2005b. A bi-objective traffic counting location problem for origin–destination trip table estimation. Transportmetrica 1 (1), 65–80.

Evans, S., 1976. Derivation and analysis of some models for combining trip distribution and assignment. Transportation Research 9 (12), 241–246.

Jornsten, K., Wallace, S.W., 1993. Overcoming the (apparent) problem of inconsistency in origin–destination matrix estimations. Transportation Science 27 (4), 374–380.

Kikuchi, S., Miljkovic, D., Van Zuylen, H.J., 2000. Examination of methods that adjust observed traffic volumes on a network. Transportation Research Record 1717, 109–119.

Pape, U., 1974. Implementation and efficiency of Moore algorithms for the shortest route problem. Mathematical Programming 7, 212–222.

Patriksson, M., 1994. The Traffic Assignment Problem: Models and Methods. VSP, Utrecht, The Netherlands.

Sherali, H.D., Sivanandan, R., Hobeika, A.G., 1994. A linear programming approach for synthesizing origin–destination trip tables from link traffic volumes. Transportation Research Part B 28 (3), 213–233.

Tarjan, R.E., 1983. Data Structures and Network Algorithms. CBMS 44. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Van Aerde, M., Rakha, H., Paramahamsan, H., 2003. Estimation of O–D matrices: the relationship between practical and theoretical considerations. Transportation Research Record 1831, 122–130.

Van Zuylen, H.J., Branston, D., 1982. Consistent link flow estimation from counts. Transportation Research Part B 16 (6), 473–476.

Yang, H., 1995. Heuristic algorithms for the bilevel origin-destination matrix estimation problem. Transportation Research Part B 29 (4), 231–242.

Yang, H., Iida, Y., Sasaki, T., 1991. An analysis of the reliability of an origin–destination trip matrix estimated from traffic counts. Transportation Research Part B 25 (5), 351–363.

Yang, H., Meng, Q., Bell, M.G.H., 2001. Simultaneous estimation of the origin–destination matrices and travel–cost coefficient for congested networks in a stochastic equilibrium. Transportation Science 35 (2), 107–123.

Yang, H., Yang, C., Gan, L., 2005. Models and algorithms for the screen-line-based traffic counting location problem. Computers and Operations Research 33 (3), 836–858.

Yang, H., Zhou, J., 1998. Optimal traffic counting locations for origin–destination matrix estimation. Transportation Research Part B 32 (2), 109–126.