

Research Article

Origin-Destination Estimation Using Probe Vehicle Trajectory and Link Counts

Xianfeng Yang,¹ Yang Lu,² and Wei Hao³¹Department of Civil, Construction & Environmental Engineering, San Diego State University, San Diego, CA, USA²Baidu Online Network Technology Co., Ltd., Beijing, China³University Transportation Research Center, City College of New York, New York, NY, USA

Correspondence should be addressed to Xianfeng Yang; xyang@mail.sdsu.edu

Received 23 June 2016; Revised 13 October 2016; Accepted 14 November 2016; Published 23 January 2017

Academic Editor: Dongjoo Park

Copyright © 2017 Xianfeng Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents two origin-destination flow estimation models using sampled GPS positions of probe vehicles and link flow counts. The first model, named as SPP model (scaled probe OD as prior OD), uses scaled probe vehicle OD matrix as prior OD matrix and applies conventional generalized least squares (GLS) framework to conduct OD correction using link counts; the second model, PRA model (probe ratio assignment), is an extension of SPP in which the observed link probe ratios are also included as additional information in the OD estimation process. For both models, the study explored a new way to construct assignment matrices directly from sampled probe trajectories to avoid sophisticated traffic assignment process. Then, for performance evaluation, a comprehensive numerical experiment was conducted using simulation dataset. The results showed that when the distribution of probe vehicle ratios is homogeneous among different OD pairs, both proposed models achieved similar degree of improvement compared with the prior OD pattern. However, under the case that the distribution of probe vehicle ratios is heterogeneous across different OD pairs, PRA model achieved more significant reduction on OD flow estimations compared with SPP model. Grounded on both theoretical derivations and empirical tests, the study provided in-depth discussions regarding the strengths and challenges of probe vehicle based OD estimation models.

1. Introduction

The origin-destination flow matrix (referred to as OD matrix) is one essential input to many dynamic traffic assignment and traffic simulation systems. Conventionally, a prior OD matrix (or seed OD matrix) is estimated from a large-scale travel survey which is usually conducted every several years due to prohibitively high cost. As a result, survey of both sampling rate and update frequency of travel is constrained which sometimes leads to a biased estimation of OD matrix. To contend with such problem, researchers developed various OD correction models using link flow counts which are collected routinely by detectors. According to Cascetta and Nguyen [1], most popular OD estimation models utilizing link flow counts can be formulated as an optimization problem taking the following generalized form:

$$\begin{aligned} \min_x \quad & F_1(X, \hat{X}) + F_2(Y, \hat{Y}) \\ \text{s.t.} \quad & Y = A(X), \quad X \in \Omega, \end{aligned} \quad (1)$$

where X is the unknown OD flow matrix and \hat{X} is the prior OD matrix (or seed matrix) representing modeler's prior belief regarding the temporal and spatial distribution of the travel pattern. Y and \hat{Y} are the estimated and observed link flow vectors, respectively; F_1 (F_2) is distance metric function to measure the discrepancy between X and \hat{X} (Y and \hat{Y}). In the first constraint of (1), the abstract function A essentially represents the traffic assignment process through which OD flows X are mapped to estimated link flows Y . A is usually referred to as assignment matrix. Ω is the feasible domain of X .

Depending on the form of metric functions, previous studies can be categorized into three groups: (1) generalized least squares models (GLS) studied by Bell [2, 3], Cascetta [4], Cascetta et al. [5], and Cascetta et al. [6]; (2) maximum likelihood models (ML) such as Spiess [7] and Cascetta and Nguyen [1]; and (3) Bayesian inference models (BI) studied by Maher [8]. In the above methods, the reference time period is

divided into a sequence of uniform intervals and OD flows of all intervals are estimated simultaneously. This type of modeling approach is called simultaneous OD estimation method which is developed primarily for offline applications. Meanwhile, for online applications, sequential OD estimation technique based on Kalman filter technique received intensive research attention during the last several decades. Starting from Okutani and Stephanedes [9], subsequent studies along this direction include Nihan and Davis [10], Chang and Wu [11], Ashok and Ben-Akiva [12, 13], and Antoniou et al. [14]. The major difference between sequential and simultaneous OD estimation is that sequential estimation performs OD adjustment recursively at each interval based on measurements of current interval and estimated results from previous intervals. Consequently, sequential estimation methods are more suitable for online traffic state estimation applications due to higher computational efficiency.

Recently, the application of emerging surveillance technologies in the OD estimation field has drawn increasing concerns from the research community. Many researchers developed their own models to combine the information from conventional detectors and emerging sensing technologies. Representative studies along this line include AVI system (Dixon and Rilett [15], Zhou and Mahmassani [16], and Chen et al. [17]), vehicle plate scanning (Castillo et al. [18]), sporadic routing data (Parry and Hazelton [19]), GPS probe vehicles (Eisenman and List [20], Cao et al. [21]), floating car data (Ásmundsdóttir [22], Yang et al. [23]), and cell phone data (Sohn and Kim [24], Calabrese et al. [25], and Iqbal et al. [26]). The basic idea behind these studies is to improve OD estimation accuracy using extra information which was not available before. Conventional OD estimation model given by formulation (1) faces three major challenges: (1) the underspecified nature of OD estimation problem; (2) reliability of prior OD matrix; and (3) accuracy of assignment matrix estimation.

First of all, in most transportation networks, the number of OD pairs exceeds the number of sensors. Therefore, without specifying additional constraints, such as prior OD matrix, the OD estimation problem is an underspecified problem. And through both laboratory experiments and real world dataset, Marzano et al. [27] and Cascetta et al. [6] concluded that “a satisfactory updating, regardless of the quality of the prior estimate, can be obtained generally only when the ratio between the number of equations (independent observed link flows) and the number of unknowns (i.e., OD flows) is close to one.” Hence, increasing the amount of observation is a major motivation of exploiting those emerging data sources. A second challenge is the availability and accuracy of the prior OD matrix. Both Marzano et al. [27] and Cascetta et al. [6] demonstrated the importance of the prior OD matrix; also Frederix et al. [28] discussed the possibility of falling into local optimal solutions given an inappropriate seed OD matrix. Besides, depending on the area of analysis, the prior OD matrix with acceptable level of reliability may not even exist. The third challenge is the computation of the assignment which usually involves traffic assignment process. A benchmark study in this regard is the bilevel optimization

model proposed by Yang et al. [29, 30]. Although the method can to some extent capture the impact of congestion on drivers' route choice, some of the assumptions (such as perfect information) of user equilibrium condition may not be satisfied in reality. And to the best of our knowledge, no systematical way has been proposed to correct such potential bias with observed data due to the low observability of the assignment matrix.

In view of the above challenges and potential of GPS location data, this study discusses two GLS based estimation frameworks for OD flows. The basic idea is to take full advantage of the sampled probe vehicle trajectory data to tackle all three above-mentioned difficulties. First of all, by aggregating the origin and destination zones of each probe vehicle trace, one can obtain observed probe vehicle OD flows. Then by scaling up the probe OD matrix using certain set of penetration ratios (the proportion of probe vehicles among the entire vehicle population), a crude estimation of OD flows can be obtained. Such scaled probe OD matrix can serve as a perfect supplement or replacement of the target OD matrix. Eisenman and List [20] conducted an exploratory study in which the scaled probe OD matrix is used in conjunction with prespecified target OD matrix in the GLS formulation with externally computed assignment matrix. Van Aerde et al. [31] focused on the computation of probe penetration ratios by averaging the observed probe ratios at different sensor locations across the network. Although the model does not require target OD matrix as input, unsatisfactory estimation results were reported even under 20% probe penetration rate.

Another relevant study was conducted by Iqbal et al. [26] using mobile phone call records. The phone call records were first used to generate tower-to-tower transient OD matrix which is then scaled up using an optimization model in conjunction with microscopic simulation model. Also, Cao et al. [21] proposed a two-step framework to incorporate probe vehicle data: in the first phase, link flows without traffic sensors are estimated based on observed link speed (from probe vehicles) and precalibrated macroscopic speed-density relationship; in the second phase, a bilevel GLS estimator is formulated to estimate OD flows. Similarly, Tan et al. [32] developed a dynamic OD estimation model using Automatic Vehicle Location information, where DTA is used for obtaining traffic assignment matrix. An important issue that is overlooked by above studies is the heterogeneity of probe penetration ratios among different OD pairs. Such situation may occur especially when probe vehicles are certain type of commercial vehicles. In practice, both delivery vehicles and taxi vehicles could be used as probes. When using delivery vehicles as probes, their restricted OD distribution in a given network would inevitably be an issue and it may affect the accuracy of OD estimation. So taxis would be a better choice. However, the proportion of taxi trips between distant OD pairs may still significantly differ from the other OD pairs. Such probe ratio heterogeneity is considered explicitly in this study. Another key concept proposed in this study is to estimate the assignment matrix directly from sampled trajectory data instead of running some complex traffic assignment process. There are two benefits in doing so: first is by replacing traffic assignment with map matching and data processing,

one can avoid sophisticated traffic assignment computation and parameter calibration; moreover, the model does not depend on any theoretical assumption regarding drivers' behavior.

Based on above discussions, two models are presented in this paper. The first model, SPP model (stands for scaled probe OD matrix as prior OD), uses scaled probe vehicle OD matrix as prior OD matrix and applies conventional GLS framework to conduct OD correction using link counts; the second model, PRA model (stands for probe ratio assignment), is an extension of SPP in which the observed link probe ratios are included as additional information in order to explicitly account for heterogeneity of probe penetration ratios. The remaining part of this paper is organized as follows: Section 2 explains basic concepts and notations. Section 3 contains detailed model specifications and is organized into three subsections: the development of SPP and PRA model and computation of assignment matrices. Section 4 discusses the solution algorithm. Section 5 presents the numerical experiment. Finally, Section 6 summarizes the conclusions.

2. Some Definitions and Notations

Considering a road network represented by a direct graph $\langle N, L \rangle$ where N is the node set and L is the link set, the analysis period is divided into K uniform intervals. Each interval is called a demand interval. Let I denote the origin-destination pair set; then the travel demand pattern of the network during the analysis period is represented by the OD flow matrix $x_{k,i}$ $i = 1, 2, \dots, I$ and $k = 1, 2, \dots, K$. Let J denote a collection of links installed with sensors; J is a subset of L . $\hat{y}_{k,j}$ $k = 1, 2, \dots, K$, $j = 1, 2, \dots, J$ are observed traffic flow counts of all vehicles at sensor locations.

It is assumed that there are two types of vehicles travel in the network: probe vehicles and regular vehicles. Each probe vehicle is able to actively report its position in the form of GPS coordinates. And through map matching algorithm, those GPS coordinates are transformed to corresponding locations in the network. In reality, the actual locations of probe vehicles usually cannot be determined fully due to measurement error of GPS. Since this study emphasizes more on the theoretical aspects, it is assumed that the exact location information is available whenever a probe vehicle reports its GPS coordinate.

The numbers of probe vehicles traveling between different OD pairs within each interval are called probe vehicle OD flows (simplified as probe OD flows) and are denoted by $\hat{x}_{k,i}$, $i = 1, 2, \dots, I$ and $k = 1, 2, \dots, K$; for each OD pair, the proportion of probe vehicle in the total vehicle population within the same interval is called OD probe vehicle penetration ratio (simplified as OD probe ratio) and is denoted by $\gamma_{k,i}$, $i = 1, 2, \dots, I$ and $k = 1, 2, \dots, K$. Note that OD probe ratios are both time- and OD-dependent.

The observed numbers of probe vehicles passing each sensor location are called probe vehicle link flows (simplified as probe link flows) and are represented by $\hat{h}_{k,j}$, $j = 1, 2, \dots, J$ and $k = 1, 2, \dots, K$. The ratio of probe link flow to

corresponding link traffic flow during each interval is called link probe vehicle penetration ratio (simplified as link probe ratio) and is represented by $\hat{\theta}_{k,j}$, $j = 1, 2, \dots, J$ and $k = 1, 2, \dots, K$.

3. Model Specifications

This section introduced two different models for OD estimations. It is noticeable that traffic congestion may bring great challenge to the estimation process. For simplicity, this study does not account for the dynamic routings issue which may have happened in practice.

3.1. Scaled Probe OD Matrix as Prior Matrix (SPP) Method. SPP method consists of two steps: firstly, the prior OD matrix $\hat{x}_{k,i}$ is estimated by scaling up the probe OD flows $\hat{z}_{k,i}$ with corresponding OD probe ratios $\gamma_{k,i}$ estimated by averaging the link probe ratios across the network; then OD flows $x_{k,i}$ are solved correcting the prior OD matrix with a GLS formulation.

To make the paper self-contained, the "direct scaling" method proposed by Van Aerde et al. [31] is summarized briefly in this section. To estimate OD probe ratios, the average link probe ratios across the entire network within each interval is computed by the following expression:

$$\gamma_k = \frac{\sum_{j \in J} \hat{h}_{k,j}}{\sum_{j \in J} \hat{y}_{k,j}} \quad k = 1, 2, \dots, K, \quad (2)$$

where $\hat{h}_{k,j}$ are the observed link probe flows, $\hat{y}_{k,j}$ are the observed link flows, and γ_k is the common value of $\gamma_{k,i}$. Equation (2) computes the average ratio of the total number of probe vehicles to the total number of vehicles observed across the entire network during one interval. Therefore expression (2) implicitly assumes that OD probe ratios are homogeneous among all OD pairs. Then γ_k are used to compute the prior OD flows $\hat{x}_{k,i}$ using the following equation:

$$\hat{x}_{k,i} = \frac{\hat{z}_{k,i}}{\gamma_k} \quad i = 1, 2, \dots, I, \quad k = 1, 2, \dots, K, \quad (3)$$

where $\hat{z}_{k,i}$ are the observed probe OD flows.

It is noticeable that the prior OD flows $\hat{x}_{k,i}$ themselves are an estimator of OD flows. Such method is referred to as direct scaling model (DS) in this study. Since the prior OD matrix is only a crude estimation, it is then adjusted using the following GLS formulation:

SPP model:

$$\min_{x_{k,i}} \sum_{k=1}^K \sum_{i=1}^I \frac{(x_{k,i} - \hat{x}_{k,i})^2}{w_{k,i}^2} + \sum_{k=1}^K \sum_{j=1}^J \frac{(y_{k,j} - \hat{y}_{k,j})^2}{q_{k,j}^2} \quad (4a)$$

$$y_{k,j} = \sum_{t=1}^T \sum_{i=1}^I a_{i,j}^t x_{k-t,i} \quad \forall k, j \quad (4b)$$

$$-\beta x_{k,i} \leq x_{k+1,i} - x_{k,i} \leq \beta x_{k,i} \quad \forall k, i \quad (4c)$$

$$x_{k,i} \geq 0 \quad \forall k, i, \quad (4d)$$

where $x_{k,i}$, $\hat{x}_{k,i}$ are unknown and prior OD flows, respectively; $y_{k,j}$, $\hat{y}_{k,j}$ are estimated and observed link flows, respectively; $w_{k,i}^2$, $q_{k,j}^2$ are, respectively, variances of $\hat{x}_{k,i}$ and $\hat{y}_{k,j}$; and β is the maximum percentage change of OD flows between two consecutive intervals.

Constrain (4b) essentially represents the flow assignment process. $a_{i,j}^t$ are the proportion (or probability) of vehicles departed during k th interval traveling between OD pair i to pass link j during $(k + t)$ th interval; $a_{i,j}^t$ is referred to as flow assignment fractions in this study. The estimated link flows, $y_{k,j}$, are then expressed as the weighted sum of all OD flows departed before interval k . T is the maximum travel time among all journeys converted to the number of demand intervals. In this study, the flow assignment fractions are computed directly from probe vehicle trajectories.

The basic idea of SPP model is to reduce the estimation bias of the direct scaling method using sensor count information. Let d and n be the numbers of OD pairs and traffic sensors, respectively; then the total number of unknowns to independent observations ratio of SPP model is $d/(d + n)$. Note that the target OD flows are obtained from GPS probe vehicle data; no additional target OD matrix from other sources is required by model.

Finally, formulations (4a)~(4d) constitute a nonlinear optimization problem with convex objective function and linear constraints. Therefore the global optimal solution exists and any solution satisfying local optimality condition is also the global optimal solution of the problem. A gradient based searching algorithm as discussed in Appendix A is adopted to solve SPP model.

3.2. Probe Ratios Assignment Model (PRA). The second formulation proposed in this study is called the probe ratio assignment model (referred to as PRA). The underlying idea of PRA is to explicitly consider the correlation between OD probe ratios and observed link probe ratios.

Thus, there exist some function P that links OD probe ratios $\gamma_{k,i}$ and estimated link probe ratios $\theta_{k,j}$:

$$\theta_{k,j} = P(\gamma_{k,i}) \quad k = 1, 2, \dots, K, \quad j = 1, 2, \dots, J. \quad (5)$$

The structure of P is similar to that of flow assignment matrix A ; the difference is that P represents the assignment of probe vehicle ratios instead of flows. In this study, P is called probe ratio assignment matrix. Based on the discussion, we can now define the following equation:

$$\theta_{k,j} = \sum_{t=1}^T \sum_{i=1}^I \rho_{i,j}^t \gamma_{k,i} = \sum_{t=1}^T \sum_{i=1}^I \rho_{i,j}^t \left(\frac{\hat{z}_{k-t,i}}{x_{k-t,i}} \right) \quad (6)$$

$$k = 1, 2, \dots, K, \quad j = 1, 2, \dots, J.$$

Here $\rho_{i,j}^t$ are probe ratio assignment fractions. $\rho_{i,j}^t$ represents the contribution of the probe ratio during k th interval between OD pairs i on the probe ratio of link j during $(k + t)$ th interval. Note that in above equation, the OD probe ratios are estimated by taking the ratio between $\hat{z}_{k,i}$, the observed OD flows, and $x_{k,i}$, the OD flows we want to estimate. Let $\hat{\theta}_{k,j}$ be

the observed link probe ratios at all sensor locations. Then PRA model can be obtained by extending SPP formulation by considering $\hat{\theta}_{k,j}$ and incorporating equation (6):

$$\min_{x_{k,i}} \sum_{k=1}^K \sum_{i=1}^I \frac{(x_{k,i} - \hat{x}_{k,i})^2}{w_{k,i}^2} + \sum_{k=1}^K \sum_{j=1}^J \frac{(y_{k,j} - \hat{y}_{k,j})^2}{q_{k,j}^2} + \sum_{k=1}^K \sum_{j=1}^J \frac{(\theta_{k,j} - \hat{\theta}_{k,j})^2}{v_{k,j}^2} \quad (7a)$$

$$y_{k,j} = \sum_{t=1}^T \sum_{i=1}^I a_{i,j}^t x_{k-t,i} \quad \forall k, j \quad (7b)$$

$$\theta_{k,j} = \sum_{t=1}^T \sum_{i=1}^I \rho_{i,j}^t \left(\frac{\hat{z}_{k-t,i}}{x_{k-t,i}} \right) \quad \forall k, j \quad (7c)$$

$$-\beta x_{k,i} \leq x_{k+1,i} - x_{k,i} \leq \beta x_{k,i} \quad \forall k, i \quad (7d)$$

$$x_{k,i} \geq 0 \quad \forall k, i. \quad (7e)$$

In the above formulations, please note that $v_{k,j}^2$ is the variance of $\hat{\theta}_{k,j}$. Other notations are introduced previously.

The objective function of PRA model (7a) adds a third term to that of SPP (4a) which is the sum of weighted distances between estimated and observed link probe ratios. And the first and second terms of (7a) are identical to those of (4a). Constraints (7b) and (7c) represent the assignment of OD flows and OD probe ratios.

The primary feature of PRA as given by (7a)~(7e) is the utilization of a new set of field observations: the observed link probe ratios which is the combined information of flow counts and probe vehicle trajectories. Also the OD probe ratio assignment matrix is a new concept introduced in this study. Let d and n be the number of OD flows and traffic sensors, respectively; then the total number of unknowns to independent observation ratio of PRA model is $d/(d + 2 * n)$ which is lower than that of SPP model. Essentially, each sensor provides two observations instead of only one after considering the probe vehicle trajectories: the first is the flow count of the entire vehicle population and the second is the proportion of probe vehicles passing the sensor location. The computation of $\rho_{i,j}^t$ is discussed in the next subsection.

Note that the PRA formulation is no longer a convex optimization problem due to the existence of (7c); therefore solving the problem using gradient based searching algorithm faces the possibility of being trapped in local optimums. The solution algorithm of PRA model is discussed in Section 4.

3.3. Computation of Assignment Matrices Using Probe Vehicle Trajectories. In this study, both flow assignment fractions $a_{i,j}^t$ and probe ratio assignment fractions $\rho_{i,j}^t$ are estimated through analyzing the GPS trajectories of probe vehicles. Compared with conventional methods (i.e., traffic assignment model), the main feature of the proposed approach is

that it replaced dynamic traffic assignment process with map matching procedure of GPS coordinates.

The underlying concept is relatively straightforward. The entire vehicle population is divided into two groups: probe vehicles and regular vehicles. And all the assignment fractions of probe vehicle population can be obtained from their time-dependent location information. Then one can use those to approximate the assignment fractions of the entire vehicle population assuming that probe vehicles are randomly sampled.

Defining $\hat{z}_{i,j}^{k,k+t}$ as the observed number of probe vehicles traveling between OD pair i and departed during interval k passed sensor j during interval $k+t$, then the flow assignment fractions of probe vehicles can be computed as

$$a_{i,j}^t = \frac{1}{K} \sum_k \left(\frac{\hat{z}_{i,j}^{k,k+t}}{\hat{z}_{k,i}} \right) \quad t = 1, 2, \dots, T. \quad (8)$$

In the above equation, the denominator is essentially the total number of probe vehicles departed during interval k between OD pairs i . Therefore the fraction inside the parenthesis is the proportion of vehicles passing link j after t intervals counted from their departure time interval. Note that probe vehicles departed at different time intervals are aggregated together to obtain a single estimation of the assignment matrix. The underlying assumption is that there exists some time period during which drivers' route choice behavior and network traffic condition remain approximately stable, and one can use a single flow assignment matrix (or OD probe ratio assignment matrix) to represent the correlation between OD flows and link flows (or OD probe ratios and link probe ratios). The purpose is to increase the estimation accuracy of assignment matrix by aggregating probe vehicles from multiple demand intervals. Extending (8) into a time-dependent form is also straightforward. Essentially probe vehicles are grouped according to their departure time. Consider the fact that the entire period of analysis is divided into S assignment intervals. Each assignment interval contains K demand intervals. Then let $a_{i,j}^{s,t}$, $s = 1, 2, \dots, S$ be the assignment matrices; then $a_{i,j}^{s,t}$ can be estimated based on probe data collected during assignment interval s using

$$a_{i,j}^{s,t} = \frac{1}{K} \sum_{k_s} \left(\frac{\hat{z}_{i,j}^{k_s, k_s+t}}{\hat{z}_{k_s, i}} \right) \quad (9)$$

$k = 1, 2, \dots, K, \quad s = 1, 2, \dots, S, \quad t = 1, 2, \dots, T,$

where k_s is the k th demand interval during s th assignment interval. It is noticeable that traffic congestion level on the network would directly affect the value of T .

The probe ratio assignment fractions can be computed using the following equation:

$$\rho_{i,j}^{s,t} \approx \frac{1}{K} \sum_{k_s} \left(\frac{\hat{z}_{i,j}^{k_s-t, k_s}}{\gamma_{k_s, j}} \right) \quad (10)$$

$k = 1, 2, \dots, K, \quad s = 1, 2, \dots, S, \quad t = 1, 2, \dots, T.$

TABLE 1: Summary of sensor locations passed by probe vehicles.

Interval	Sensor locations passed											
	OD pair 1				OD pair 2				OD pair 3			
	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	—	2	2	4	—	3	3	—	5
2	4	3	4	1	5	—	—	4	5	5	5	—
3	—	5	—	4	—	5	—	—	—	—	—	—

Similar to (9), the estimation of probe ratio assignment fractions in (10) can be approximated by using the observed number of probe vehicles departed during interval k_s - t passed sensor j during interval k_s and sensor detected link flows.

This section provides an illustrative example to show how (9) and (10) work. Consider a hypothetical network consisting of four nodes and five links (shown in Figure 1(a)), among all the nodes, nodes 1, 2, and 3 are demand generation nodes and node 4 is a demand absorption node. Therefore there are three OD pairs, i.e., 1-4, 2-4, and 3-4. It is assumed that traffic sensors are installed at the middle of each link. A total of 12 probe vehicles departed during one demand interval (therefore $K = S = 1$); the distribution of the three OD pairs is 6, 4, and 2 vehicles, respectively.

Note that the initial locations of all probe vehicles are plotted in Figure 1(b).

Suppose that all vehicles finished their trip within the following four time intervals and each vehicle is observed at least once during each interval. The positions of all vehicles at the end of each subsequent interval are visualized by Figures 2(a)~2(d). Note that vehicles traveling between different OD pairs are painted with different patterns.

According to observed vehicle positions at the end of each interval, one can identify, from individual vehicle's viewpoint, the sensor location passed by each probe vehicle during each interval which is summarized in Table 1.

According to Table 1, $\hat{z}_{i,j}^{k,k+t}$ are computed and summarized in Table 7. Then the flow assignment fractions and probe ratio assignment fractions are computed according to (9) and (10). The results are summarized in Table 8.

4. Computation Procedures

This section presents the numerical solution algorithm for the proposed models. Our objective is a nonlinear optimization problem with equality and inequality constraints. Mathematically, such optimization problem takes the following general form:

$$\begin{aligned} \min_{x_1, x_2, \dots, x_n} \quad & f(x_1, x_2, \dots, x_n) \\ \text{s.t.} \quad & \sum_{r=1}^n a_{m,r} x_r = b_m \quad m = 1, 2, \dots, u \\ & \sum_{r=1}^n a_{m,r} x_r \leq b_m \quad m = u+1, 2, \dots, M. \end{aligned} \quad (11)$$

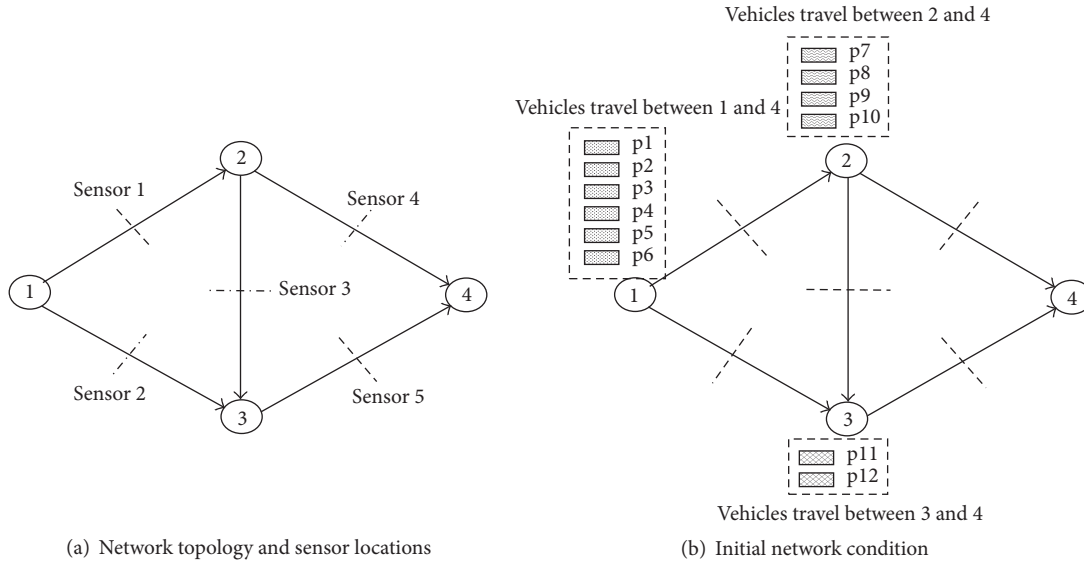


FIGURE 1: Network topology and initial probe vehicle locations.

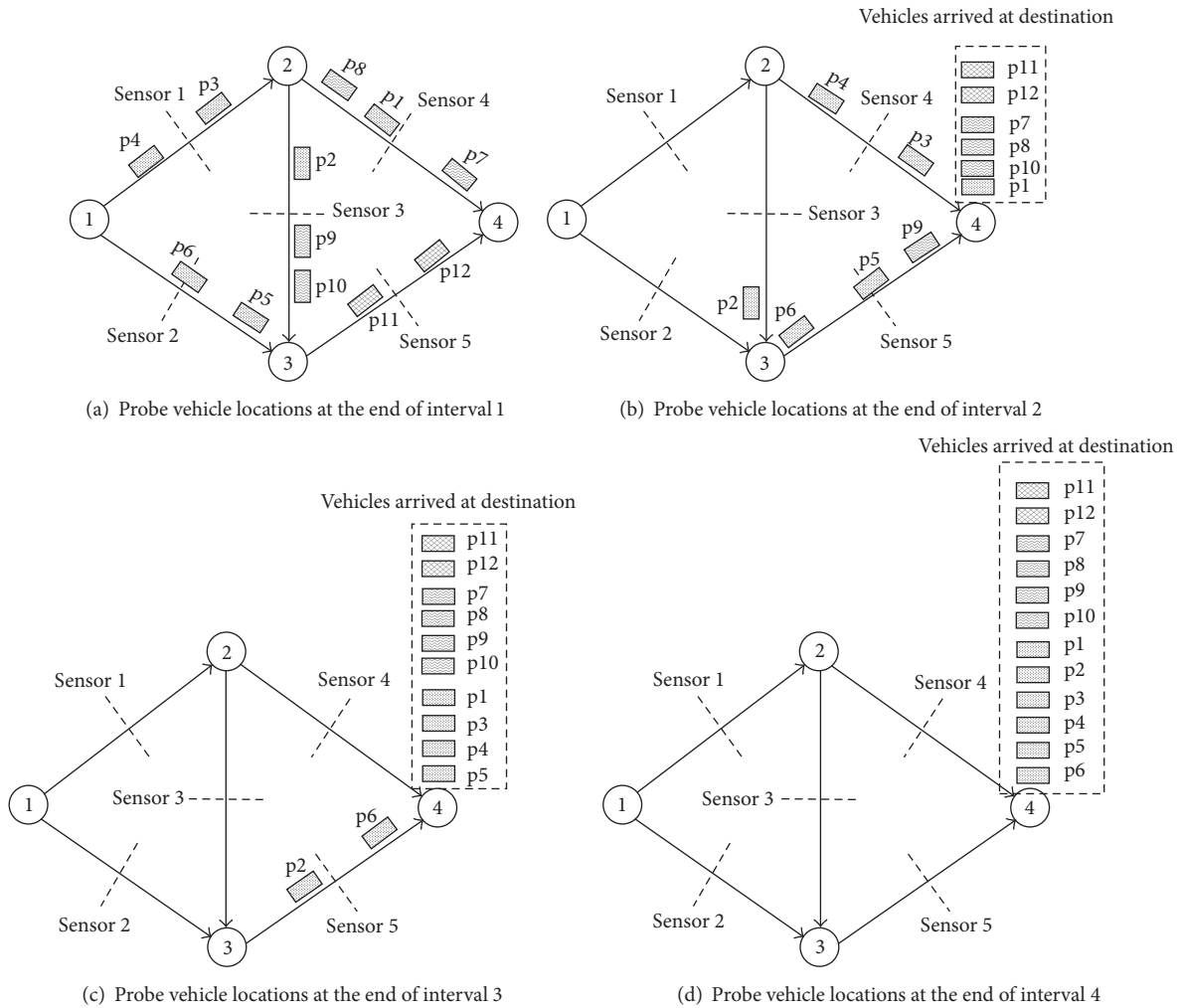


FIGURE 2: Illustration of probe vehicle locations captured at the end of each interval.

In formulation (11), the objective function f is an n -dimensional scalar function. Since f is continuous and differentiable in this study, the gradient of f can be obtained analytically. Also there are u linear equality constraints and $M-u$ linear inequality constraints. The gradient vector of f is denoted by ∇f :

$$\nabla f(x_1, x_2, \dots, x_n) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right]^T. \quad (12)$$

The solution algorithm is summarized as follows.

Step 1 (initialization). Determine an initial feasible solution that satisfies all constraints. Let X_0 be the initial feasible solution and set the current iteration $h = 0$; then enter the main optimization loop consisting of Steps 2~6.

Step 2 (gradient computation). Based on X_h , compute the gradient of objective function $\nabla f(X_h)$. The specific formulas used for gradient computation of SPP and PRA model are given by equations (A.3) and (B.3) in Appendices A and B, respectively.

Step 3 (optimal search direction calculation). Based on the current gradient, compute an optimal search direction considering all constraints. Let $D_h = [d_{h1}, d_{h2}, \dots, d_{hn}]$ be the search direction of current iteration; D_h is computed by solving the following LP model:

$$\begin{aligned} \min_{D_h} \quad & \nabla f(X_h)^T D_h \\ \text{s.t.} \quad & \alpha_m^T D_h = 0 \quad m \in E \\ & \alpha_m^T D_h \leq 0 \quad m \in I_B \\ & -1 \leq d_{hr} \leq 1 \quad r = 1, 2, \dots, n. \end{aligned} \quad (13)$$

In the above LP problem, α_m is the coefficient vector of m th constraint, $\alpha_m = [a_{m,1}, a_{m,2}, \dots, a_{m,n}]$; E and I_B are the equality constraint set and bounded inequality constraint set.

Step 4 (optimal search step length calculation). According to the optimal search direction D_h , perform the following line search to determine optimal step length τ_h :

$$\begin{aligned} \min_{0 \leq \tau_h \leq \tau_{\max}} \quad & f(X_h + \tau_h D_h) \\ \tau_{\max} = \min \quad & \left\{ \frac{b_m - \alpha_m^T X_h}{\alpha_m^T D_h} \mid \alpha_m^T D_h > 0, m \notin I_B \right\}. \end{aligned} \quad (14)$$

This algorithm follows an iterative procedure and it is quite similar to traditional Frank-Wolfe algorithm.

Step 5 (update X_h). $X_{h+1} = X_h + \tau_h D_h$.

Step 6 (check convergence criteria). If $|f(X_{h+1}) - f(X_h)| < \varepsilon$, then terminate the computation; otherwise repeat the process from Step 2 to Step 6.

5. Numerical Examples

5.1. Simulation Setup and Results. To evaluate effectiveness of the proposed models, numerical experiments are conducted using VISSIM as a laboratory experiment tool. The use of synthetic dataset is due to the lack of real world dataset. A road network in the northern part of Maryland State (near I-495 beltway) is selected as the test site. The network consists of 28 nodes and 74 links. The bird view map of the target area and the network topology constructed in VISSIM are shown by Figures 3 and 4.

The simulation period is set to 3 hours which is divided into 18 demand intervals (each interval is 10 minutes). To simplify the simulation process, 39 major OD pairs are selected in the simulation. As shown in Figure 4, 10 out of 38 links are installed with traffic sensors for traffic data collection. Table 10 summarizes the origin and destination nodes of each OD pair along with its time-dependent demand volumes; Table 11 summarizes the route choice probability of all paths between all OD pairs. For convenience of study, between each OD pair shown in Table 11, we only selected those paths whose lengths are obviously shorter than the others. Given the OD flows and route choice in the simulation network, it shall be noted that congestion v/c ratios on all links are below 0.4. By running the simulation network in VISSIM, we collected the GPS trajectory of each vehicle and traffic flow rate on those links with sensors. Notably, the GPS trajectory of each vehicle can directly yield the ground truth of OD flows. For model evaluations and comparisons, this study uses a part of trajectory dataset (based on the preset probe ratio in each scenario) as model inputs which represent the probe vehicles.

To reveal the model property under different network conditions, two scenarios are simulated. The first scenario (referred to as scenario A) represents the situation in which the probe vehicle penetration ratios are approximately homogeneous among different OD pairs; and the second scenario (referred to as scenario B) represents the heterogeneous probe ratios across multiple OD case. For both scenarios, the average probe vehicle penetration ratio is set as 15%. The probe ratios across different OD flows in scenarios B ranges from 5% to 30%.

For each model (DS, SPP, and PRA), the estimation accuracy of four sets of parameters is examined: (1) OD flows; (2) OD probe ratios; (3) link flow counts at sensor locations; and (4) link probe vehicle ratios at sensor locations. The estimated values are compared with ground truth values extracted from the simulator. The estimated qualities are then quantified by the following five performance indicators: MSE (Mean Square Error), RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), MSPE (Mean Square Percentage Error), and RMSPE (Root Mean Square Percentage Error). The definitions of performance indicators are summarized by the following equation.

Performance Indicators

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2$$

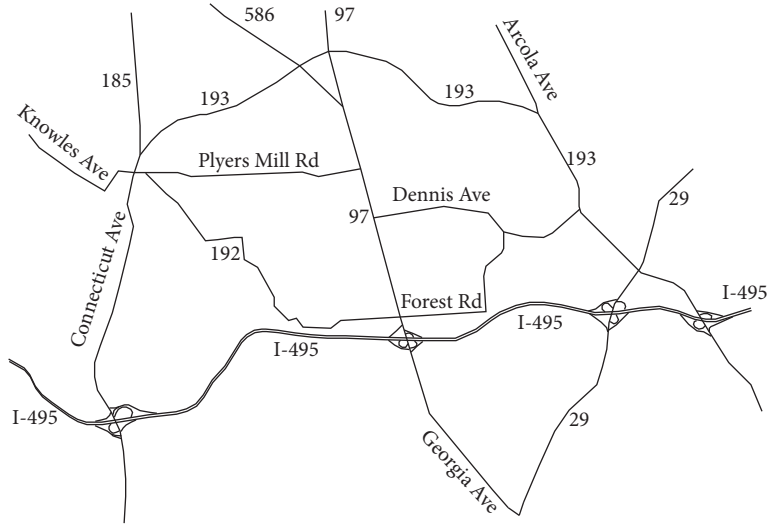
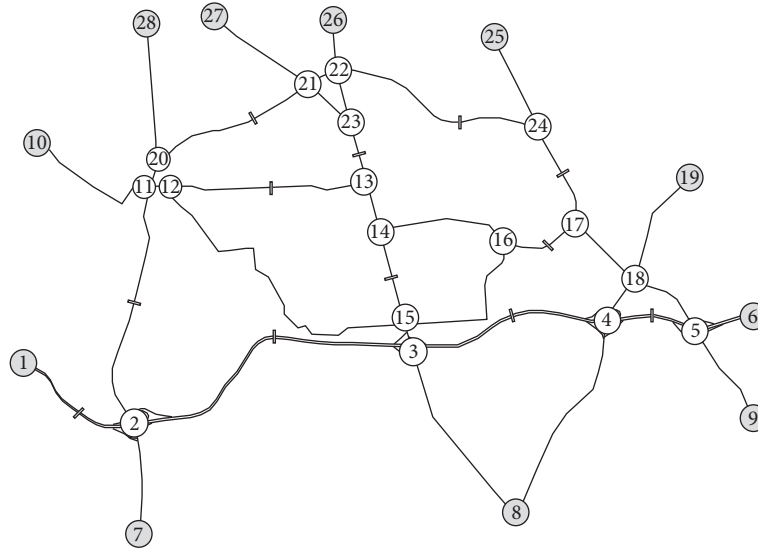


FIGURE 3: The bird view map of the test network.



! : Detectors for both directions

FIGURE 4: Topology of the test road network in VISSIM.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2}$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{x_i - \hat{x}_i}{\hat{x}_i} \right|$$

$$\text{MSPE} = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \hat{x}_i}{\hat{x}_i} \right)^2$$

$$\text{RMSPE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \hat{x}_i}{\hat{x}_i} \right)^2}.$$

(15)

x_i and \hat{x}_i are estimated and ground truth values. N is the number of estimates.

Tables 2~5 summarize all the performance indicators for OD flows, OD probe ratios, link flows, and link probe ratios given by DS, SPP, and PRA models under scenarios A and B. Also using SPP model as the benchmark, the improvement of PRA in estimations is also presented. By using the proposed solution algorithm, the estimations with all models can be completed within 3 minutes.

5.2. Results Interpretation. Figures 5, 6, and 7 show the comparison between estimated and ground truth values of different parameters displayed in the form of scatter plots.

TABLE 2: Estimation accuracy of OD flows.

Scenario	Models	Performance indicators				
		MSE	RMSE	MAPE	MSPE	RMSPE
A	DS	1679.4	41.0	55.7%	79.1%	88.9%
	SPP	416.6	20.4	29.90%	24.00%	49.00%
	PRA	343.1	18.5	29.5%	22.2%	47.1%
	PRA improvement	-17.6%	-9.3%	-1.3%	-7.5%	-3.9%
B	DS	2295.9	47.9	81.9%	156.5%	125.1%
	SPP	805.7	28.4	50.1%	58.4%	76.4%
	PRA	612.6	24.8	40.2%	37.1%	60.9%
	PRA improvement	-24.0%	-12.7%	-19.8%	-36.5%	-20.3%

TABLE 3: Estimation accuracy of OD probe ratios.

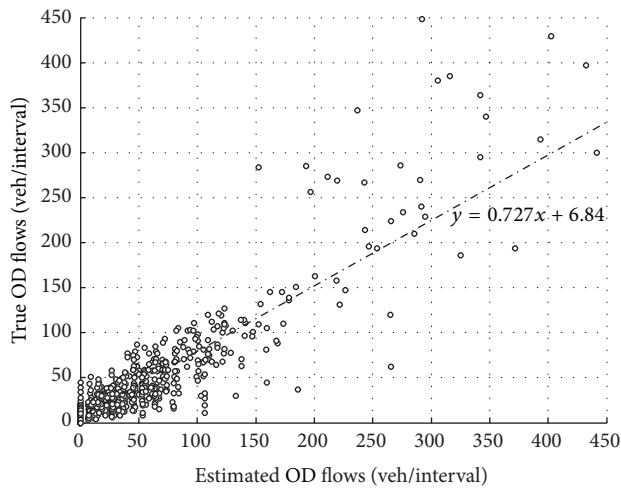
Scenario	Models	Performance indicators				
		MSE	RMSE	MAPE	MSPE	RMSPE
A	DS	0.00598	0.07735	34.9%	19.2%	43.8%
	SPP	0.00353	0.05945	25.6%	11.6%	34.0%
	PRA	0.00351	0.05926	24.7%	10.0%	31.6%
	PRA improvement	-0.6%	-0.3%	-3.5%	-13.8%	-7.1%
B	DS	0.02176	0.14750	45.4%	29.8%	54.6%
	SPP	0.01527	0.12357	35.6%	20.4%	45.2%
	PRA	0.01041	0.10204	29.8%	14.5%	38.1%
	PRA improvement	-31.8%	-17.4%	-16.3%	-28.9%	-15.7%

TABLE 4: Estimation accuracy of link flow counts.

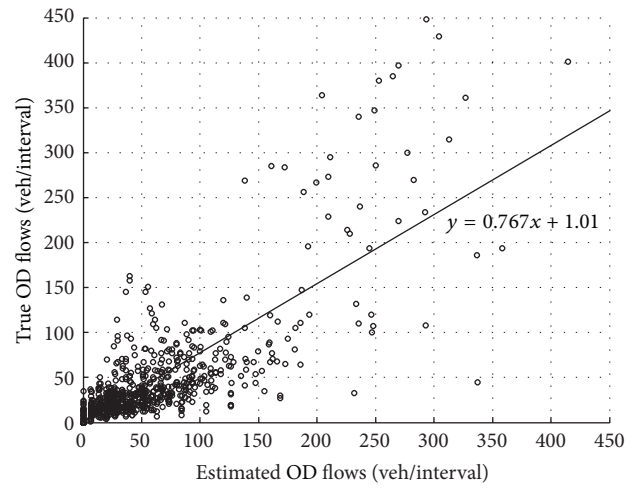
Scenario	Models	Performance indicators				
		MSE	RMSE	MAPE	MSPE	RMSPE
A	DS	9124.4	95.5	31.76%	18.93%	43.51%
	SPP	7356.5	85.8	22.29%	10.48%	32.37%
	PRA	7040.4	83.9	20.89%	9.66%	31.08%
	PRA improvement	-4.3%	-2.2%	-6.3%	-7.8%	-4.0%
B	DS	12889.3	113.5	51.0%	55.9%	74.8%
	SPP	8133.4	90.2	27.6%	16.3%	40.4%
	PRA	7273.0	85.3	24.4%	12.2%	34.9%
	PRA improvement	-10.6%	-5.4%	-11.6%	-25.2%	-13.6%

TABLE 5: Estimation accuracy of link probe ratios.

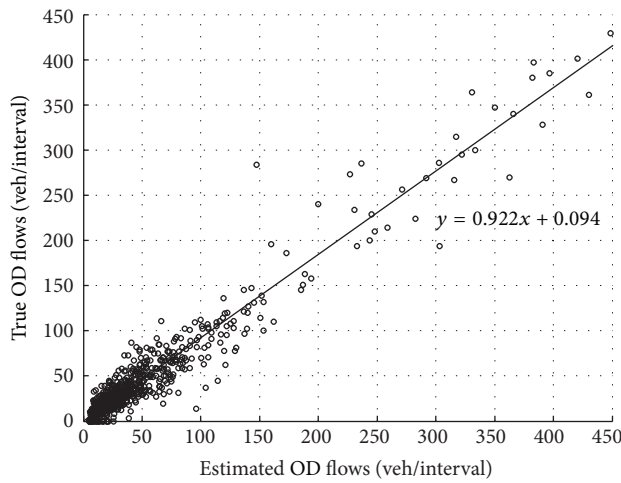
Scenario	Models	Performance indicators				
		MSE	RMSE	MAPE	MSPE	RMSPE
A	DS	0.00205	0.04523	25.4%	10.5%	32.4%
	SPP	0.00160	0.03999	20.4%	7.7%	27.8%
	PRA	0.00166	0.04072	18.8%	6.3%	25.1%
	PRA improvement	3.8%	1.8%	-7.8%	-18.2%	-9.7%
B	DS	0.00604	0.07770	30.7%	14.3%	37.8%
	SPP	0.00290	0.05382	20.2%	7.8%	27.9%
	PRA	0.00290	0.05383	17.3%	5.6%	23.6%
	PRA improvement	0.0%	0.0%	-14.4%	-28.2%	-15.4%



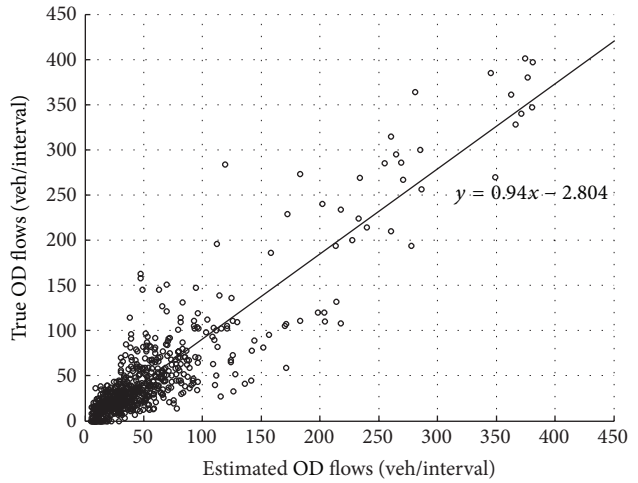
(a) DS, scenario A



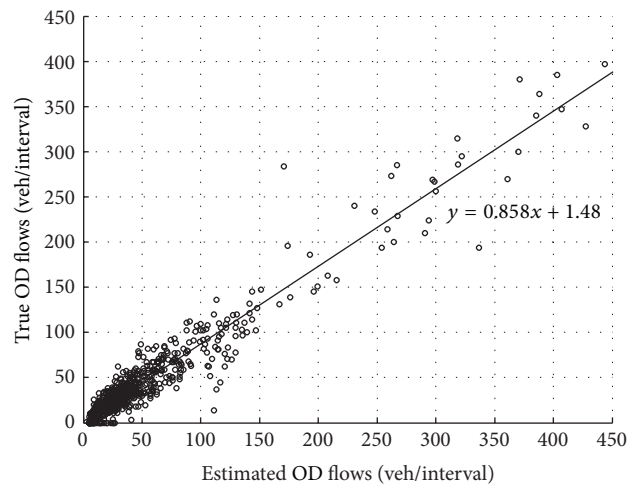
(d) DS, scenario B



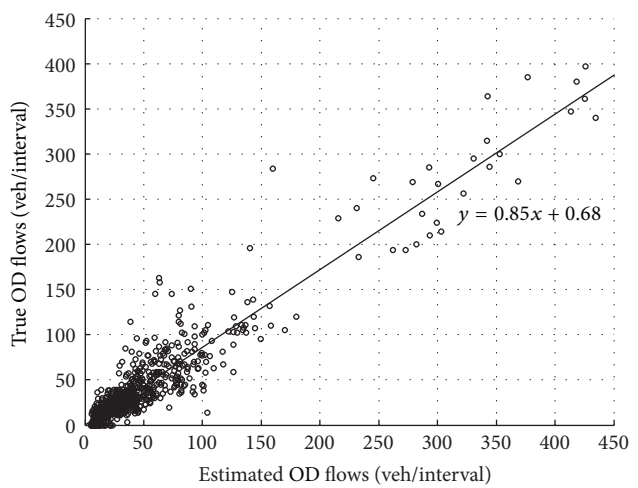
(b) SPP, scenario A



(e) SPP, scenario B

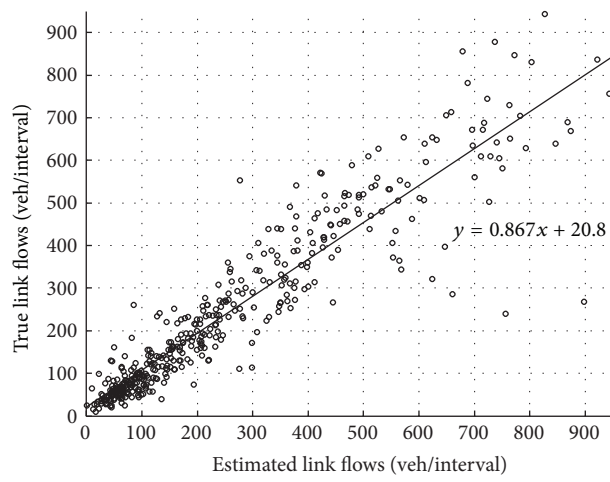


(c) PRA, scenario A

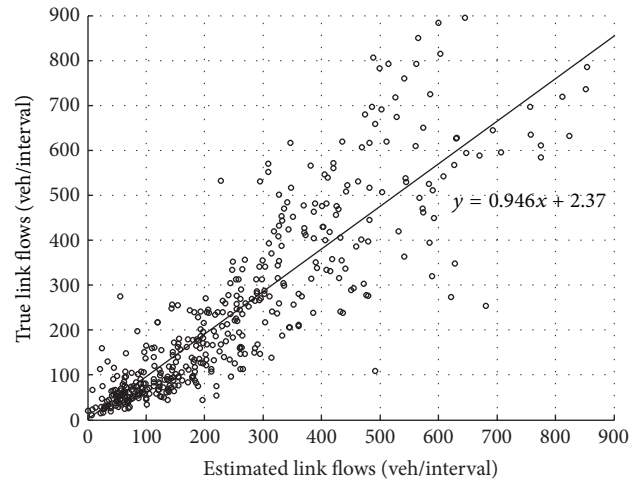


(f) PRA, scenario B

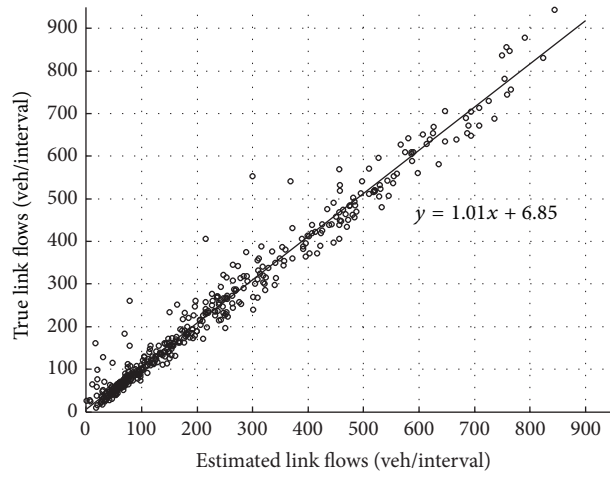
FIGURE 5: Scatter plots of estimated and ground truth OD flows.



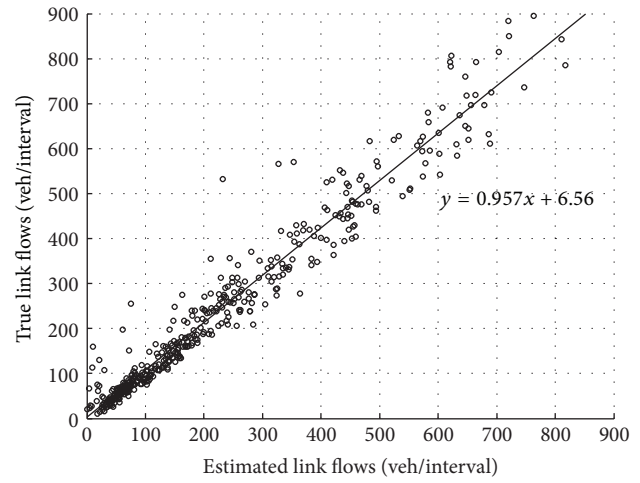
(a) DS, scenario A



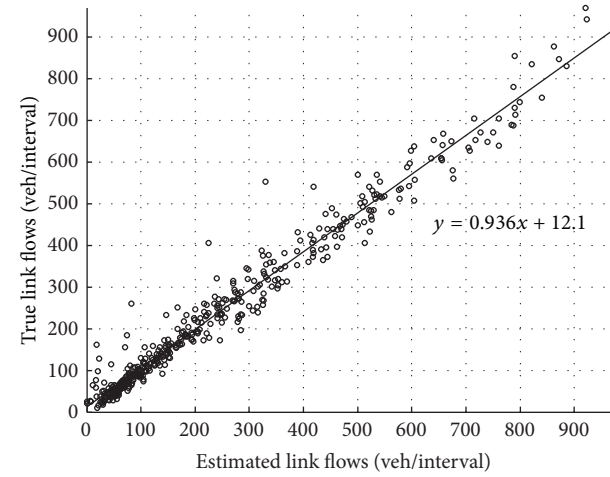
(d) DS, scenario B



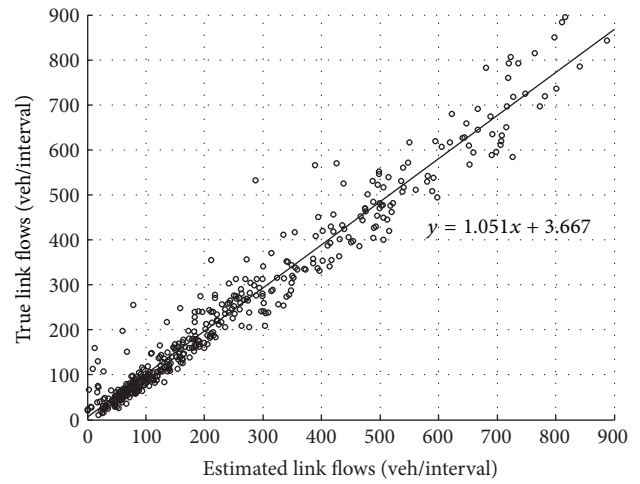
(b) SPP, scenario A



(e) SPP, scenario B

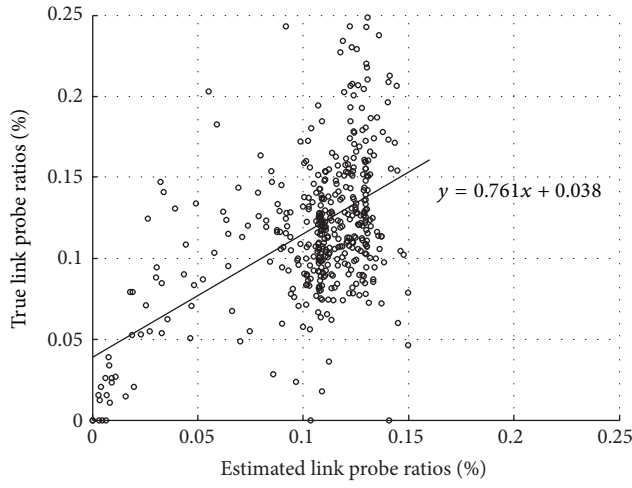


(c) PRA, scenario A

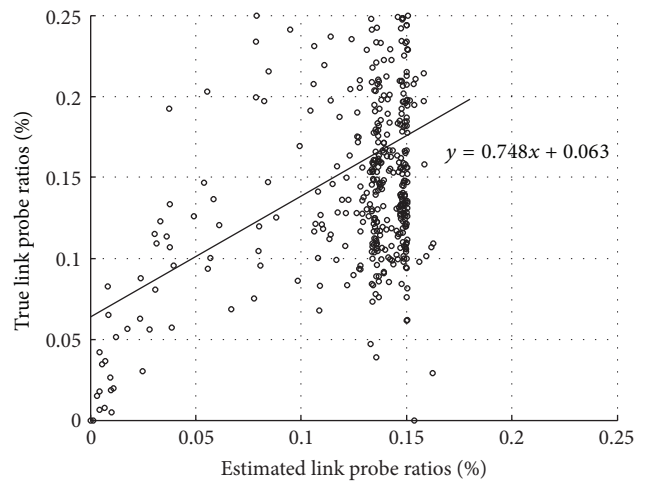


(f) PRA, scenario B

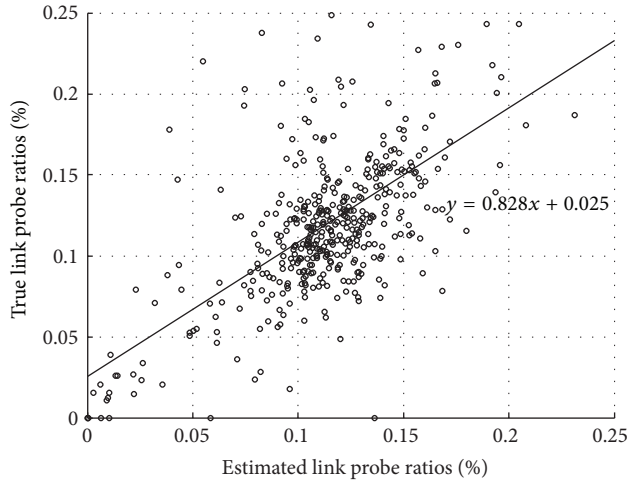
FIGURE 6: Scatter plots of estimated and observed link flow counts.



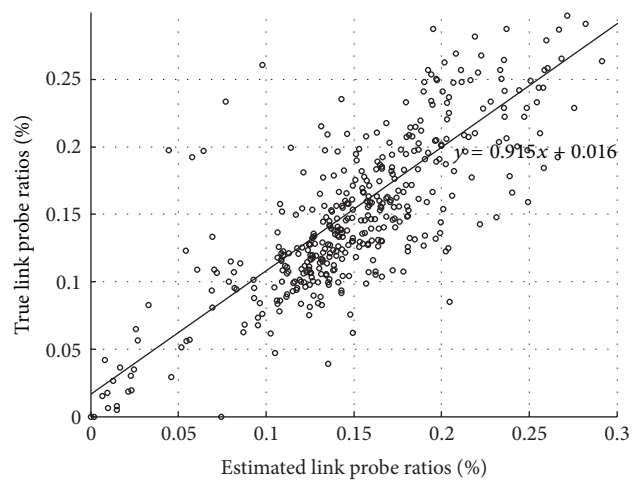
(a) DS, scenario A



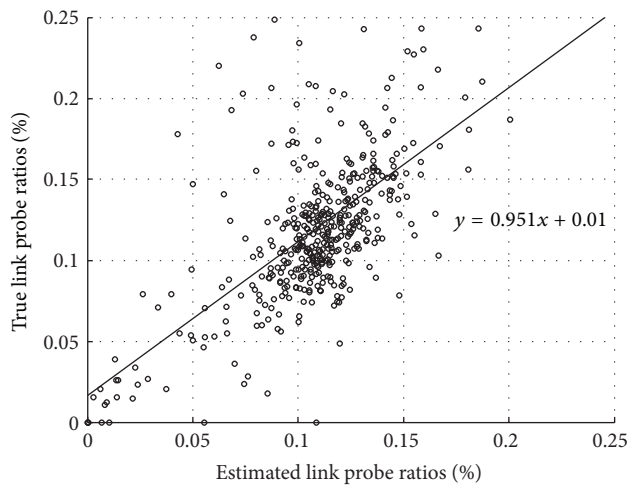
(d) DS, scenario B



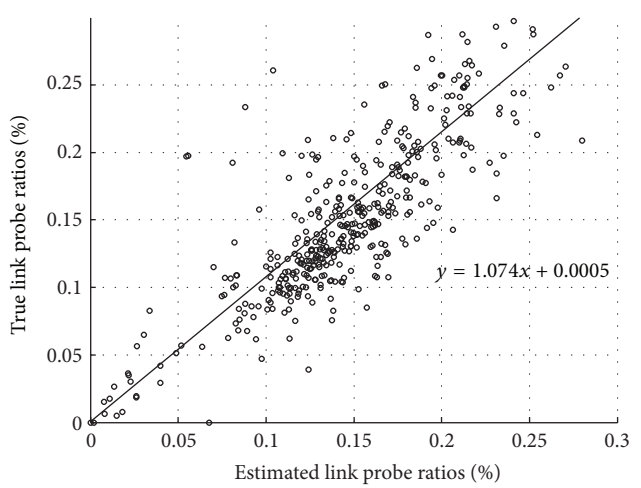
(b) SPP, scenario A



(e) SPP, scenario B



(c) PRA, scenario A



(f) PRA, scenario B

FIGURE 7: Scatter plots of estimated and observed link probe ratios.

TABLE 6: Estimation accuracy of OD flows with different probe ratios.

Probe penetration ratio	Performance indicators				
	MSE	RMSE	MAPE	MSPE	RMSPE
5%	1545.3	41.6	75.3%	113.6%	99.5%
10%	754.4	27.4	48.1%	53.2%	71.3%
15%	612.6	24.8	40.2%	37.1%	60.9%
20%	554.4	21.4	38.2%	35.4%	53.1%
30%	415.7	18.2	34.7%	28.7%	50.1%

TABLE 7: Summary of $\hat{z}_{i,j}^{k,k+t}$ in the hypothetical example.

Travel time	OD $i/\text{link } j$	1	2	3	4	5
$t = 1$	1	3	2	0	0	0
	2	0	0	2	1	0
	3	0	0	0	0	2
$t = 2$	1	1	0	1	2	1
	2	0	0	0	1	2
	3	0	0	0	0	0
$t = 3$	1	0	0	0	1	2
	2	0	0	0	0	0
	3	0	0	0	0	0

Several important observations made from model outputs are discussed in this section. First of all, DS model tends to produce biased OD estimation in both scenarios which can be observed from Table 2 and Figures 5(a) and 5(d). Particularly when the probe vehicle ratios are not homogeneous across different OD pairs, the percentage error of DS model reached as high as 82% in MAPE and 125% in RMSPE. Therefore, directly scaled OD flows based solely on observed link probe ratios are not reliable estimator of OD flows and require additional adjustment.

Secondly, in scenario A, both SPP and PRA models offered significantly higher estimation quality comparing with the DS model. Using the SPP as the benchmark, in scenario A, the PRA model reduced the MAPE and MSPE of the OD matrix, respectively, by 1.3% and 7.5%. Based on the performance indicators reported by the tables, one can argue that when the probe ratios are homogeneous across OD pairs, the two models have similar degree of accuracy.

As for scenario B where the probe ratios are not homogeneous, results showed that PRA model is more effective compared with SPP model. According to Table 2, PRA model reduced the MAPE and MSPE of SPP model by 19.8% and 36.5%, respectively. Therefore it indicates that PRA model offers additional correction capability compared with SPP model. Such conclusion can be also inspected visually through Figures 5(e) and 5(f). Similar results can also be found in Tables 3~4. In Table 5, both SPP and PRA can outperform DS in estimating link probe ratios while the MES and RMSE showed that the two models are almost identical. However, in terms of MAPE, MSPE, and RMSPE, the PRA model can

yield significant improvement which validates the model's effectiveness.

Another important finding of the experiment is that it is generally more challenging to correct link probe ratios comparing with flow counts. From Figures 6(b), 6(c), 6(e), and 6(f), one can observe that the regression line between estimated and observed link flow counts is very close to 45-degree line for both models, while, in Figures 7(b), 7(c), 7(e), and 7(f), the dispersions of link probe ratio scatter plots are obviously larger. One viable explanation is that the computation of the probe ratio assignment matrix is more difficult than flow assignment matrix and a better correction of link probe ratios can be achieved only when the probe ratio assignment matrix is more precisely specified.

To summarize, as a direct enhancement of the DS method, SPP is very effective when the distribution of OD probe ratios is homogeneous; on the other hand, the PRA model, by considering additional link observations, generally outperforms SPP model when the OD probe ratios vary significantly among different OD pairs. However, compared with SPP and DS models, the PRA method also introduces several additional complexities. To maximize the benefit of PRA, one needs to correctly specify the probe ratio assignment matrix and link probe ratio variances and also take some care on the solution algorithm in order to avoid being trapped in local optimal points.

To further evaluate the effectiveness of PRA models in estimating OD flows with different probe penetration ratios, this study conducted a sensitivity analysis, which ranges the ratio from 5% to 30%. As shown in Table 6, the estimation

TABLE 8: Assignment matrix estimated.

Travel time	OD $i/\text{link } j$	1	2	3	4	5
$t = 1$	1	1/2	1/3	0	0	0
	2	0	0	1/2	1/4	0
	3	0	0	0	0	1/2
$t = 2$	1	1/6	0	1/6	1/3	1/6
	2	0	0	0	1/4	1/2
	3	0	0	0	0	1/2
$t = 3$	1	0	0	0	1/6	1/3
	2	0	0	0	0	0
	3	0	0	0	0	0

accuracy of PRA is quite sensitive to the average probe ratios. Higher probe ratios will result in more accurate OD flow estimation. In addition, it is noticeable that when the probe ratio drops to 5%, the estimation results from PRA are not sufficiently reliable. Under such condition, increasing the number of link sensors would help improve the model's performance.

6. Conclusions and Future Research

This paper presented the development of two offline OD estimation models using probe vehicle data: the SPP and PRA models. Both mathematical formulations and solution algorithm are discussed in detail. Also the study successfully explored the possibility of computing assignment matrices directly from GPS trajectories to avoid complex traffic assignment process.

Then, through a comprehensive numerical experiment, the performances of proposed models are analyzed. It is shown that the distribution of OD probe ratios can affect the correction power of different models when the probe vehicle data is used. When the OD probe ratios are approximately homogeneous across different OD pairs, both SPP and PRA performed equally well by reducing about half of the relative error of DS method; however when the OD probe ratios are nonhomogeneous, then PRA model outperformed SPP model to some extent. The results also implied that when the OD probe ratios are heterogeneous, incorporating observed link probe ratios into the objective function can improve the overall estimation accuracy. However, unlike link flows, the correction of link probe ratios turns out to be much more challenging and one needs to carefully specify the correlation between OD and link probe ratios which is the probe ratio assignment matrix P in this study.

As for future research, the proposed models can be integrated with the quasidynamic approach proposed by Cascetta et al. [6] to further reduce the unknown-to-observation ratio; also sensitivity analysis can be performed to investigate the impact of different input parameters on the final outcomes; moreover numerical experiment using real world dataset is another important future work in order to better assess the model performance. In addition, one can note that the probe OD ratios are used as the approximates of the flow assignment fractions, which may not be true in practice. Hence, it is also

critical to develop an advanced model to overcome this limitation and apply some state-of-the-art method to estimate the traffic assignment matrix.

Appendix

A. Gradient Computation of SPP Model

According to the objective function given by (4a), define the following functions:

$$f(x_{k,i}, y_{k,j}) = \sum_{i \in I} \sum_{k \in K} \frac{(x_{k,i} - \hat{x}_{k,i})^2}{w_{k,i}^2} + \sum_{j \in J} \sum_{k \in K} \frac{(y_{k,j} - \hat{y}_{k,j})^2}{q_{k,j}^2} \quad (\text{A.1})$$

$$f_1(x_{k,i}) = \sum_{i \in I} \sum_{k \in K} \frac{(x_{k,i} - \hat{x}_{k,i})^2}{w_{k,i}^2},$$

$$f_2(y_{k,j}(x_{k,i})) = \sum_{j \in J} \sum_{k \in K} \frac{(y_{k,j} - \hat{y}_{k,j})^2}{q_{k,j}^2}.$$

Essentially, original objective function f is divided into two parts; therefore $f = f_1 + f_2$.

Taking derivatives of f_1 and f_2 with respect to $x_{k,i}$, one can obtain the following expressions:

$$\begin{aligned} \frac{\partial f_1(x_{k,i})}{\partial x_{m,r}} &= \frac{2(x_{m,r} - \hat{x}_{m,r})}{w_{m,r}^2}, \\ \frac{\partial f_2(y_{k,j}(x_{k,i}))}{\partial x_{m,r}} &= \sum_{k \in K} \sum_{j \in J} \frac{\partial f_2}{\partial y_{k,j}} \frac{\partial y_{k,j}}{\partial x_{m,r}} \\ &= \sum_{k \in K} \sum_{j \in J} \frac{2a_{r,j}^{k-m}(y_{k,j} - \hat{y}_{k,j})}{q_{k,j}^2} \\ y_{k,j} &= \sum_{i \in I} \sum_{t \in T} a_{i,j}^t x_{k-t,i} \quad \forall k, j. \end{aligned} \quad (\text{A.2})$$

TABLE 9: Probe ratio assignment matrix estimated.

Travel time	OD $i/\text{link } j$	1	2	3	4	5
$t = 1$	1	1	1	0	0	0
	2	0	0	1	1	0
	3	0	0	0	0	1
$t = 2$	1	1	0	1	2/3	1/4
	2	0	0	0	1/3	1/2
	3	0	0	0	0	1/4
$t = 3$	1	0	0	0	1	1
	2	0	0	0	0	0
	3	0	0	0	0	0

Note that the second derivative is computed based on the correlation between $y_{k,j}$ and $x_{k,i}$ given by constraint (4b). To summarize,

$$\begin{aligned} \frac{\partial f(x_{k,i}, y_{k,j})}{\partial x_{t,r}} &= \frac{\partial f_1(x_{k,i})}{\partial x_{t,r}} + \frac{\partial f_2(y_{k,j}(x_{k,i}))}{\partial x_{m,r}} \\ &= \frac{2(x_{m,r} - \hat{x}_{m,r})}{w_{m,r}^2} \\ &\quad + \sum_{k \in K} \sum_{j \in J} \frac{2a_{r,j}^{k-m}(y_{k,j} - \hat{y}_{k,j})}{q_{k,j}^2}. \end{aligned} \quad (\text{A.3})$$

B. Gradient Computation of PRA Model

According to the objective function given by (6), define the following three functions:

$$\begin{aligned} f(x_{k,i}, y_{k,j}, \theta_{k,j}) &= \min_{x_{k,i}} \sum_{i \in I} \sum_{k \in K} \frac{(x_{k,i} - \hat{x}_{k,i})^2}{w_{k,i}^2} \\ &\quad + \sum_{j \in J} \sum_{k \in K} \frac{(y_{k,j} - \hat{y}_{k,j})^2}{q_{k,j}^2} \\ &\quad + \sum_{j \in J} \sum_{k \in K} \frac{(\theta_{k,j} - \hat{\theta}_{k,j})^2}{v_{k,j}^2} \\ f_1 &= \sum_{i \in I} \sum_{k \in K} \frac{(x_{k,i} - \hat{x}_{k,i})^2}{w_{k,i}^2} \\ f_2 &= \sum_{j \in J} \sum_{k \in K} \frac{(y_{k,j} - \hat{y}_{k,j})^2}{q_{k,j}^2} \\ f_3 &= \sum_{j \in J} \sum_{k \in K} \frac{(\theta_{k,j} - \hat{\theta}_{k,j})^2}{v_{k,j}^2}. \end{aligned} \quad (\text{B.1})$$

Essentially, the original objective function f is divided into three parts; therefore $f = f_1 + f_2 + f_3$. Taking derivatives of f_1 , f_2 , and f_3 with respect to $\gamma_{k,i}$, one can obtain the following expressions:

$$\begin{aligned} \frac{\partial f_1(x_{k,i})}{\partial x_{m,r}} &= \frac{2(x_{m,r} - \hat{x}_{m,r})}{w_{m,r}^2} \\ \frac{\partial f_2(y_{k,j}(x_{k,i}))}{\partial x_{m,r}} &= \sum_{k \in K} \sum_{j \in J} \frac{\partial f_2}{\partial y_{k,j}} \frac{\partial y_{k,j}}{\partial x_{m,r}} \\ &= \sum_{k \in K} \sum_{j \in J} \frac{2a_{r,j}^{k-m}(y_{k,j} - \hat{y}_{k,j})}{q_{k,j}^2} \\ \frac{\partial f_2(\theta_{k,j}(x_{k,i}))}{\partial x_{m,r}} &= \sum_{j \in J} \sum_{k \in K} \frac{\partial f_2}{\partial \theta_{k,j}} \frac{\partial \theta_{k,j}}{\partial x_{m,r}} \\ &= \sum_{j \in J} \sum_{k \in K} \frac{2(\theta_{k,j} - \hat{\theta}_{k,j})}{v_{k,j}^2} \frac{-a_{r,j}^{k-m} \hat{z}_{m,r}}{x_{m,r}^2}. \end{aligned} \quad (\text{B.2})$$

Therefore, to summarize,

$$\begin{aligned} \frac{\partial f(x_{k,i}, y_{k,j}, \theta_{k,j})}{\partial x_{m,r}} &= \frac{2(x_{m,r} - \hat{x}_{m,r})}{w_{m,r}^2} + \sum_{k \in K} \sum_{j \in J} \frac{2a_{r,j}^{k-m}(y_{k,j} - \hat{y}_{k,j})}{q_{k,j}^2} \\ &\quad + \sum_{j \in J} \sum_{k \in K} \frac{2(\theta_{k,j} - \hat{\theta}_{k,j})}{v_{k,j}^2} \frac{-a_{r,j}^{k-m} \hat{z}_{m,r}}{x_{m,r}^2}. \end{aligned} \quad (\text{B.3})$$

C. Assignment Fractions of Example

See Tables 7, 8, and 9.

TABLE 10: Time-dependent OD flows (veh/hr).

Node pair	Time period (hour)					
	0-0.5	0.5-1	1-1.5	1.5-2	2-2.5	2.5-3
1→6	1440	1800	2400	2800	2400	1800
1→26	240	480	600	720	600	480
1→8	360	600	840	1080	840	600
1→25	120	180	240	360	240	180
1→28	120	240	360	480	360	240
10→9	60	72	96	144	96	72
10→8	72	96	120	180	120	96
10→26	60	72	84	132	84	60
6→1	1320	1700	2300	2800	2300	1700
6→26	360	480	600	960	600	480
6→10	60	96	120	180	120	96
6→8	360	480	720	960	720	480
6→28	120	240	480	720	480	240
19→1	120	240	360	480	360	240
19→8	120	180	240	360	240	180
19→27	72	96	120	180	120	96
9→28	120	180	240	360	240	180
9→25	96	120	180	240	180	120
9→8	72	96	120	180	120	96
9→19	120	156	180	216	180	156
9→1	72	96	120	180	120	96
9→26	60	96	144	180	144	96
9→27	12	24	36	48	36	24
28→8	180	240	300	360	300	240
28→7	240	360	480	600	480	240
27→9	120	144	216	240	216	144
27→1	180	240	300	420	144	120
26→10	96	120	144	180	144	120
26→8	120	180	240	360	240	180
26→19	96	120	180	240	180	120
7→28	240	300	350	400	350	300
7→25	96	120	144	180	144	120
7→9	120	156	180	216	180	156
8→25	96	120	144	180	144	120
8→26	240	300	360	480	360	300
8→1	120	180	240	360	180	120
8→28	60	96	120	180	120	96
8→9	120	180	360	480	360	180
8→19	240	240	300	480	300	240
25→9	240	480	600	720	600	480

TABLE 11: Turning ratios via different paths.

OD	Path	Ratio
1→6	1→2→3→15→16→17→18→5→6	0.05
	1→2→3→15→14→16→17→18→5→6	0.05
	1→2→3→4→5→6	0.9
1→26	1→2→11→12→13→23→22→26	0.3
	1→2→11→20→21→22→26	0.3
	1→2→3→15→14→13→23→22→26	0.4
1→8	1→2→3→8	1.0
	1→2→11→12→13→23→22→24→25	0.3
1→25	1→2→3→15→16→17→24→25	0.3
	1→2→11→20→21→22→24→25	0.4
1→28	1→2→11→20→28	1.0
	10→11→2→3→4→5→9	0.4
10→9	10→11→12→13→14→15→3→4→5→9	0.3
	10→11→12→13→14→16→17→18→5→9	0.3
10→8	10→11→2→3→8	0.5
	10→11→12→15→3→8	0.5
10→26	10→11→12→13→23→22→26	0.5
	10→11→20→21→22→26	0.5
6→1	6→5→4→3→15→14→13→12→11→2→1	0.1
	6→5→4→3→2→1	0.9
6→26	6→5→18→17→16→14→13→23→22→26	0.3
	6→5→18→17→24→22→26	0.3
6→10	6→5→4→3→15→14→13→23→22→26	0.4
	6→5→4→3→15→12→11→10	0.5
6→8	6→5→18→17→16→14→13→12→11→10	0.5
	6→5→4→8	1.0
6→28	6→5→4→3→2→11→20→28	0.4
	6→5→4→3→15→12→11→20→28	0.3
6→1	6→5→18→17→16→14→13→12→11→20→28	0.3
	19→18→17→16→15→3→2→1	0.5
19→8	19→18→4→3→2→1	0.5
	19→18→4→8	1.0
19→27	19→18→17→16→14→13→23→21→27	0.4
9→25	9→5→18→17→24→25	1.0
9→8	9→5→4→8	0.5
	9→5→18→4→8	0.5
9→19	9→5→18→19	1.0
9→1	9→5→4→3→2→1	1.0
	9→5→4→3→15→14→13→23→22→26	0.4
9→26	9→5→18→17→16→14→13→23→22→26	0.3
	9→5→18→17→24→22→26	0.3
9→27	9→5→18→17→24→22→21→27	0.5
	9→5→4→3→15→14→13→23→21→27	0.5
28→8	28→20→11→2→3→8	0.6
	28→20→11→12→13→14→3→8	0.2
28→7	28→20→11→12→15→3→8	0.2
	28→20→11→2→7	1.0
27→9	27→21→23→13→14→15→3→4→5→9	0.6
	27→21→22→24→17→18→5→9	0.4
27→1	27→21→20→11→2→1	1.0

D. OD Flows and Turning Ratios in the Network

See Tables 10 and 11.

TABLE II: Continued.

OD	Path	Ratio
26→10	26→22→23→13→12→11→10	0.3
	26→22→21→20→11→10	0.7
26→8	26→22→23→13→14→15→3→8	1.0
26→19	26→22→23→13→14→15→3→4→18→19	0.3
	26→22→24→17→18→19	0.7
7→28	7→2→11→20→28	1.0
7→25	7→2→3→4→18→17→24→25	0.4
	7→2→11→20→21→22→24→25	0.3
	7→2→3→15→14→16→17→24→25	0.3
7→9	7→2→3→4→5→9	1.0
8→25	8→3→15→16→17→24→25	0.5
	8→4→18→17→24→25	0.5
8→26	8→3→15→14→13→23→22→26	1.0
8→1	8→3→2→1	1.0
8→28	8→3→2→11→20→28	0.5
	8→3→15→14→13→12→11→20→28	0.5

Competing Interests

The authors declare that they have no competing interests.

References

- [1] E. Cascetta and S. Nguyen, "A unified framework for estimating or updating origin/destination matrices from traffic counts," *Transportation Research Part B: Methodological*, vol. 22, no. 6, pp. 437–455, 1988.
- [2] M. G. Bell, "Estimation of an origin-destination matrix from traffic counts," *Transportation Science*, vol. 17, no. 2, pp. 198–217, 1983.
- [3] M. G. H. Bell, "The real time estimation of origin-destination flows in the presence of platoon dispersion," *Transportation Research Part B: Methodological*, vol. 25, no. 2-3, pp. 115–125, 1991.
- [4] E. Cascetta, "Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator," *Transportation Research Part B: Methodological*, vol. 18, no. 4-5, pp. 289–299, 1984.
- [5] E. Cascetta, D. Inaudi, and G. Marquis, "Dynamic estimators of origin-destination matrices using traffic counts," *Transportation Science*, vol. 27, no. 4, pp. 363–373, 1993.
- [6] E. Cascetta, A. Papola, V. Marzano, F. Simonelli, and I. Vitiello, "Quasi-dynamic estimation of o-d flows from traffic counts: formulation, statistical validation and performance analysis on real data," *Transportation Research Part B: Methodological*, vol. 55, pp. 171–187, 2013.
- [7] H. Spiess, "A maximum likelihood model for estimating origin-destination matrices," *Transportation Research Part B*, vol. 21, no. 5, pp. 395–412, 1987.
- [8] M. J. Maher, "Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach," *Transportation Research. Part B: Methodological*, vol. 17, no. 6, pp. 435–447, 1983.
- [9] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transportation Research Part B*, vol. 18, no. 1, pp. 1–11, 1984.
- [10] N. L. Nihan and G. A. Davis, "Recursive estimation of origin-destination matrices from input/output counts," *Transportation Research Part B*, vol. 21, no. 2, pp. 149–163, 1987.
- [11] G.-L. Chang and J. Wu, "Recursive estimation of time-varying origin-destination flows from traffic counts in freeway corridors," *Transportation Research Part B: Methodological*, vol. 28, no. 2, pp. 141–160, 1994.
- [12] K. Ashok and M. E. Ben-Akiva, "Alternative approaches for real-time estimation and prediction of time-dependent origin-destination flows," *Transportation Science*, vol. 34, no. 1, pp. 21–36, 2000.
- [13] K. Ashok and M. E. Ben-Akiva, "Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows," *Transportation Science*, vol. 36, no. 2, pp. 184–198, 2002.
- [14] C. Antoniou, M. Ben-Akiva, and H. N. Koutsopoulos, "Dynamic traffic demand prediction using conventional and emerging data sources," *IEEE Proceedings-Intelligent Transport Systems*, vol. 153, no. 1, pp. 97–104, 2006.
- [15] M. P. Dixon and L. R. Rilett, "Real-time OD estimation using automatic vehicle identification and traffic count data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 17, no. 1, pp. 7–21, 2002.
- [16] X. Zhou and H. S. Mahmassani, "Dynamic origin-destination demand estimation using automatic vehicle identification data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 105–114, 2006.
- [17] R. Chen, J. Sun, and Y. Feng, "A novel OD estimation method based on automatic vehicle identification data," in *Intelligent Computing and Information Science*, vol. 135, pp. 461–470, Springer, Berlin, Germany, 2011.
- [18] E. Castillo, J. M. Menéndez, and P. Jiménez, "Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations," *Transportation Research Part B: Methodological*, vol. 42, no. 5, pp. 455–481, 2008.
- [19] K. Parry and M. L. Hazelton, "Estimation of origin-destination matrices from link counts and sporadic routing data," *Transportation Research Part B: Methodological*, vol. 46, no. 1, pp. 175–188, 2012.
- [20] S. M. Eisenman and G. F. List, "Using probe data to estimate OD matrices," in *Proceedings of the 7th International IEEE Conference on Intelligent Transportation Systems (ITSC '04)*, pp. 291–296, Washington, DC, USA, October 2004.
- [21] P. Cao, T. Miwa, T. Yamamoto, and T. Morikawa, "Bilevel generalized least squares estimation of dynamic origin-destination matrix for urban network with probe vehicle data," *Transportation Research Record*, vol. 2333, pp. 66–73, 2013.
- [22] R. Ásmundsdóttir, *Dynamic OD matrix estimation using floating car data [Ph.D. dissertation]*, Delft University of Technology, 2008.
- [23] Y. Yang, H. P. Lu, and Q. Hu, "A Bi-level programming model for origin-destination estimation based on FCD," in *Proceedings of the 10th International Conference of Chinese Transportation Professionals (ICCTP '10)*, pp. 117–124, American Society of Civil Engineers, Beijing, China, August 2010.
- [24] K. Sohn and D. Kim, "Dynamic origin-destination flow estimation using cellular communication system," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 5, pp. 2703–2713, 2008.
- [25] F. Calabrese, G. Di Lorenzo, L. Liu, and C. Ratti, "Estimating origin-destination flows using mobile phone location data," *IEEE Pervasive Computing*, vol. 10, no. 4, pp. 36–44, 2011.

- [26] M. S. Iqbal, C. F. Choudhury, P. Wang, and M. C. González, "Development of origin-destination matrices using mobile phone call data," *Transportation Research Part C: Emerging Technologies*, vol. 40, pp. 63–74, 2014.
- [27] V. Marzano, A. Papola, and F. Simonelli, "Limits and perspectives of effective O–D matrix correction using traffic counts," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 2, pp. 120–132, 2009.
- [28] R. Frederix, F. Viti, R. Corthout, and C. M. J. Tampère, "New gradient approximation method for dynamic origin-destination matrix estimation on congested networks," *Transportation Research Record*, vol. 2263, pp. 19–25, 2011.
- [29] H. Yang, T. Sasaki, Y. Iida, and Y. Asakura, "Estimation of origin-destination matrices from link traffic counts on congested networks," *Transportation Research Part B: Methodological*, vol. 26, no. 6, pp. 417–434, 1992.
- [30] H. Yang, Q. Meng, and M. G. H. Bell, "Simultaneous estimation of the origin-destination matrices and travel-cost coefficient for congested networks in a stochastic user equilibrium," *Transportation Science*, vol. 35, no. 2, pp. 107–123, 2001.
- [31] M. Van Aerde, B. Hellinga, L. Yu, and H. Rakha, *Vehicle Probes as Real-Time ATMS Sources of Dynamic OD and Travel Time Data*, Queen's University, Department Of Civil Engineering, 1993.
- [32] G. Tan, L. Liu, F. Wang, and Y. Wang, "Dynamic OD estimation using automatic vehicle location information," in *Proceedings of the 2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference (ITAIC '11)*, vol. 1, pp. 352–355, Chongqing, China, August 2011.

