

RESEARCH ARTICLE

A Two-Stage Algorithm for Origin-Destination Matrices Estimation Considering Dynamic Dispersion Parameter for Route Choice

Yong Wang^{1,2}, Xiaolei Ma^{3,4*}, Yong Liu¹, Ke Gong¹, Kristian C. Henricakson², Maozeng Xu¹, Yinhai Wang^{2*}

1 School of Management, Chongqing Jiaotong University, Chongqing, China, **2** Department of Civil and Environmental Engineering, University of Washington, Seattle, Washington, United States of America, **3** School of Transportation Science and Engineering, Beijing Key Laboratory for Cooperative Vehicle Infrastructure, Systems, and Safety Control, Beihang University, Beijing, China, **4** Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, SiPaiLou #2, Nanjing, Jiangsu, China

* xiaolei@buaa.edu.cn (XM); yinhai@uw.edu (YH)



OPEN ACCESS

Citation: Wang Y, Ma X, Liu Y, Gong K, Henricakson KC, Xu M, et al. (2016) A Two-Stage Algorithm for Origin-Destination Matrices Estimation Considering Dynamic Dispersion Parameter for Route Choice. PLoS ONE 11(1): e0146850. doi:10.1371/journal.pone.0146850

Editor: Zhong-Ke Gao, Tianjin University, CHINA

Received: August 11, 2015

Accepted: December 21, 2015

Published: January 13, 2016

Copyright: © 2016 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research is supported by National Natural Science Foundation of China (Project No. 71402011, 71471024, 51408019, 71301180, 51329801), National Social Science Foundation of Chongqing of China (No. 2013YBJJ035), and the Scientific and Technological Research Program of Chongqing Municipal Education Commission (No. KJ1400307), and the Natural Science Foundation of Chongqing of China (No. cstc2015jcyjA30012), National Key Technologies R&D Program of China (2014BAG01B03), Science and Technology Project

Abstract

This paper proposes a two-stage algorithm to simultaneously estimate origin-destination (OD) matrix, link choice proportion, and dispersion parameter using partial traffic counts in a congested network. A non-linear optimization model is developed which incorporates a dynamic dispersion parameter, followed by a two-stage algorithm in which Generalized Least Squares (GLS) estimation and a Stochastic User Equilibrium (SUE) assignment model are iteratively applied until the convergence is reached. To evaluate the performance of the algorithm, the proposed approach is implemented in a hypothetical network using input data with high error, and tested under a range of variation coefficients. The root mean squared error (RMSE) of the estimated OD demand and link flows are used to evaluate the model estimation results. The results indicate that the estimated dispersion parameter theta is insensitive to the choice of variation coefficients. The proposed approach is shown to outperform two established OD estimation methods and produce parameter estimates that are close to the ground truth. In addition, the proposed approach is applied to an empirical network in Seattle, WA to validate the robustness and practicality of this methodology. In summary, this study proposes and evaluates an innovative computational approach to accurately estimate OD matrices using link-level traffic flow data, and provides useful insight for optimal parameter selection in modeling travelers' route choice behavior.

Introduction

Urban sprawl and population growth have resulted in increasingly severe traffic congestion in major cities around the world. City planners and decision makers have recognized the need for comprehensive traffic management strategies to meet the challenges of rapidly evolving built environments and population demographics. Effective transportation polices and control

on Transportation Construction by the Ministry of Transport of China (2015318835200).

Competing Interests: The authors have declared that no competing interests exist.

measures can improve traffic safety and quality of service, as well as promoting economic development and reducing air pollution. Obtaining origin-destination (OD) traffic demand matrix in low-cost and high-accuracy manner not only becomes a problem transportation science, but also draws attentions from many scholars in various scientific fields. For example, researchers in statistical physics and complex systems recently proposed a number of novel methods to estimate OD matrix directly from population data [1, 2, 3, 4, 5, 6]. Reliable OD matrix estimation can provide critical insight for traffic management, operations, and urban planning efforts to mitigate congestion [7, 8]. Thus, a reliable OD matrix estimation method is indispensable for both transportation planners and traffic engineers.

A number of approaches have been developed for estimating OD matrices in the past several decades [9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. Compared with conventional survey-based method, data-driven OD estimation methods relying on link-level traffic flow measurements require less effort and offer significantly reduced time and cost for data acquisition and processing. For such methods, observed traffic flows at key points throughout the network should be known as prior information for OD matrix initialization.

Past research on this topic has considered range of different optimization methods, including entropy maximizing estimators [21, 22], maximum likelihood estimation [23], Bayesian inference estimation [24], generalized least squares (GLS) [9, 10, 25] to estimate OD demands. Entropy maximizing estimators are used to maximize the spread of trip distributions on all available paths (routes) where the observed traffic flows are used as the only information (i.e. without a target trip matrix). Maximum likelihood estimation aims to maximize the likelihood of the closeness between target OD matrix and estimated OD matrix. In the Bayesian inference approach, the target OD matrix is a prior probability function of the estimated OD matrix on a basis of observed traffic count data. The GLS estimator is a robust and efficient linear unbiased estimator, which can solve the estimation of OD matrix by minimizing the Weighted Euclidean Distances (WED) between the target data and the solution data.

User equilibrium (UE) assignment models are commonly used to obtain path choice behavior based on the estimated OD demand. Deterministic UE assignment models assume that all users have access perfect information about the generalized link travel costs, and select a route with the lowest perceived travel cost [26]. Beckman [27] formulated the UE assignment model by assuming that the OD demands are a function of level of service. A combined distribution and assignment model which relies on link-level traffic flow data was presented by Fisk and Boyce [28], and extended by Lam and Huang [29] to address multiclass-user transportation networks. Fisk [30, 31] proposed a combined entropy maximizing model with UE constraints. Yang et al. [11] integrated the GLS technique with a UE traffic assignment model for OD matrix estimation, presented in the form of a convex bi-level optimization problem. Summaries of the more recent contributions to UE-based traffic assignment are provided in Han [32], Lu et al. [33], Inoue and Maruyama [34], Kumar and Peeta [35].

The stochastic user equilibrium (SUE) principle allows the perceived cost to vary between individuals in a heterogeneous population, which can be seen as a more realistic approach than deterministic UE [15, 36], in which the perceived travel costs cannot vary between travelers. The probit SUE was first formulated as a generalization of user equilibrium by Daganzo and Sheffi [37], and developed by Sheffi and Powell [38] as a mathematical programming problem. Liu and Fricker [39] presented a two-stage SUE approach to estimate OD matrices and the probit dispersion parameter in an iterative manner. Yang et al. [15] improved on the methods described in Liu and Fricker by incorporating link traffic flows and travel cost obtained using logit-based SUE traffic assignment. Meng et al. [40] presented a linearly constrained model and solution algorithm for the probit SUE problem with fixed demand and separable link travel time functions. This modeling approach was extended in Meng et al. [41] using elastic demand and non-

separable link travel time functions. Time-dependent traffic assignment can be also formulated as a multinomial logit model [42, 43, 44], and this has become one of the most common methods for SUE-based traffic assignment [45, 46, 47]. In a fixed-point formulation, fixed target demands or link flows are used to establish model based on UE and SUE principles [13, 14, 19].

In the multinomial logit model formulation, the link choice probability is a function of a dispersion parameter θ [16], which describes road users' perception of travel costs. Though the dispersion parameter θ is predetermined in many previous studies [14, 36, 45, 46, 47, 48], here we assume that this value should be allowed to change with traffic conditions. In addition, Lo and Chan [16] proposed a maximum likelihood procedure for simultaneously estimating the OD matrix and the dispersion parameter θ , while the link choice proportions and link flows can be further calculated based on the maximum likelihood estimators of OD matrix and θ . Compared with the previous studies, the main contributions of this paper lie in: (1) A fixed-point model is formulated with a dynamic dispersion parameter θ , where the estimation of link choice proportions is integrated into the optimization procedure; (2) A GLS estimator is utilized to train this model, and the link choice proportions can be simultaneously calculated based on the OD matrix and dispersion parameter through a multinomial logit model; (3) A two-stage iterative algorithm is presented to refine the OD matrix and dispersion parameter estimates, and Sequential Quadratic Programming (SQP) from the extended quasi-Newton method is applied in the two-stage algorithm process [49].

The remainder of this paper is organized as follows: In Section 2, relevant notation, definitions, and model formulations are presented, followed by a link choice proportion approach to calculate the observed link flow using a true OD matrix. A two-stage algorithm is described in Section 3, along with model implementation details. The performance of the proposed approach is tested in a hypothetical network, and a sensitivity analysis is conducted using a range of variation coefficients. Results are presented and compared with those obtained through other established OD estimation methods. In section 4, results are presented for a real-world network using loop detector data in the city of Seattle, WA to demonstrate the practicality of the proposed approach. Finally, conclusions are summarized in Section 5.

Model Formulation

Related Notations and Definitions

The notation and parameter definitions used throughout the paper are as follows:

K the set of network links $k \in K$, where T denotes the total number of links

L the set of observed links $l \in L$, where Γ denotes the number of observed links

J the set of OD pairs $j \in J$, where τ indicates the total number of OD pairs

A the set of paths connecting the OD pair j , $a \in A$

c_k travel cost of link k

c_{rj} travel cost of path r connecting the OD pair j

t_k the free flow travel time of link k

C_k the capacity of link k

α_k the performance function parameter of link k

β_k the exponential value of link k 's performance function

M the set of nodes in the network

f the vector for estimated link flows, where f_k is the estimated flow for link k
 \hat{f} the vector for observed link flow, where \hat{f}_l is the observed link flow for link l
 d the estimated OD vector matrix, where d_j is the j th element of d for OD pair j
 \bar{d} the target OD vector matrix, where \bar{d}_j is the j th element of \bar{d} for OD pair j
 \tilde{d} the initial OD vector matrix for SQP algorithm optimization
 W the initial weight matrix in the group of all paths connecting each OD pair
 w'_{mn} the initial weight element of all paths connecting nodes m and n , $m, n \in M$
 W the weight matrix in the group of all paths connecting each OD pair
 w_{mn} the weight element of all paths connecting nodes m and n , $m, n \in M$
 E the identity matrix with the same dimension as the initial weight matrix
 G the vector matrix of observed link flows, where G_i is the i th element of G
 U the covariance matrix for the target OD vector and estimated OD vector
 V the covariance matrix for the observed link flows and estimated link flows
 P the matrix of link choice proportions, where p_{kj} is the k th element of P . This is equivalent to the proportion of OD pair j traveling on the observed link k
 P_{rj} the probability of path r that connects OD pair j being chosen for a trip
 x_k the observed traffic flow of link k
 $x_k^{(s)}$ the estimated traffic flow of link k at the s th iteration in the traffic assignment stage
 $y_k^{(s)}$ the auxiliary mean traffic flow of link k at the s th iteration in the traffic assignment stage
 θ the estimated dispersion parameter for OD estimation
 $\bar{\theta}$ the target dispersion parameter for OD estimation
 $\tilde{\theta}$ the initial dispersion parameter for SQP algorithm optimization
 Q the covariance matrix for θ and $\bar{\theta}$
 S_d the feasible solution set for OD matrix
 S_θ the feasible solution set for θ parameter
 $\sigma_{d_j}^2$ the variance for OD demands
 $\sigma_{x_k}^2$ the variance for link flows
 σ_θ^2 the variance for dispersion parameter
 λ_d random term for the target OD matrix
 λ_f the random term for observed link flows
 F_1 the “distance” between the estimated OD vector matrix d and target demand OD matrix \bar{d}
 F_2 the “distance” between the estimated link flow vector f and observed link flow vector \hat{f}

F_3 the “distance” between the estimated dispersion parameter θ and the target dispersion parameter $\bar{\theta}$

a_{krj} decision variable, if the link k lies on path r connecting the OD pair j , and set $a_{krj} = 1$, and 0 otherwise

η the percentage of traffic flow traveling from each node to the most adjacent node

$RMSE(OD)$ the root mean squared error between estimated and true OD matrices

$RMSE(LF)$ the root mean squared error based on estimated and observed (true) link flows

Based on the above notations, the OD estimation model development and validation procedure can be described as follows:

1. A ground truth OD matrix is used as prior information to calculate the link flows based on the link choice proportion model. These link flows represent the measured traffic flows obtained through fixed mechanical sensors or other means;
2. The observed link flows are chosen from those calculated link flows at fixed points throughout the network;
3. The estimated OD matrix, link flows, and dispersion parameter are obtained via the fixed-point model and two-stage iterative algorithm using the partial observed link flows from step (2);
4. Results are evaluated and compared with the ground truth as established in step (1).

The Fixed Point Model with Dynamic Dispersion Parameter. As presented in the previous subsection, the estimated OD vector matrix is expressed as $d = [d_1, d_2, \dots, d_j, \dots, d_\tau]'$, where d_j denotes the mean traffic flow of the j th element of d for OD pair j . Consider an OD pair j connected by a link k which is associated with a link performance cost function $c_k(f_k)$ equal to the cost of using link k . The link performance cost function [50] is expressed during the traffic assignment procedure in Eq 1:

$$c_k(f_k) = t_k [1.0 + \alpha_k (\frac{f_k}{C_k})^{\beta_k}], \forall k \in K \tag{1}$$

The link flow vector is defined as $f = [f_1, f_2, \dots, f_\Gamma]'$, and the matrix of link choice proportions is denoted as $P = [p_{kj}]$, where $0 \leq p_{kj} \leq 1$. This represents the proportion of OD pair j connected by the link k . Thus, the mathematical expectation of link flow vector f can be calculated as $E[f] = [Pd]_{\Gamma \times 1}$, where Pd is the product of the observed mean link flow vector and the matrix of link choice proportions P . P can be adjusted by the link flows and the dispersion parameter θ .

The OD matrix can be estimated via a fixed point formulation by considering the target OD matrix and observed link flows as follows [9, 10, 13, 14, 15, 16, 18, 19, 51]:

$$\begin{aligned} d &= \arg \min_{d \in S_d} [F_1(d, \bar{d}) + F_2(f, \hat{f})] \\ &= \arg \min_{d \in S_d} [(d - \bar{d})^T U^{-1} (d - \bar{d}) + (P(d)d - \hat{f})^T V^{-1} (P(d)d - \hat{f})] \end{aligned} \tag{2}$$

Where:

$P(d) = \{p_{kj}(d_j)\}$ is the assignment matrix, which represents the proportion of OD pair j using the observed link k ;

$f = P(d)d$ is the estimated link flow vector. $f = \{f_k\}$, where $f_k = \sum_j p_{kj}(d_j)d_j$.

In this study, the dispersion parameter is integrated into the objective function (Eq 2) [16, 19] as follows:

$$(d, \theta) = \arg \min_{\substack{d \in S_d \\ \theta \in S_t}} [F_1(d, \bar{d}) + F_2(f, \hat{f}) + F_3(\theta, \bar{\theta})] \tag{3}$$

This model can be seen as a Stochastic User Equilibrium (SUE) problem [13]. The Generalized Least Square (GLS) estimator can be used to solve Eq 3 by minimizing the Weighted Euclidean Distances (WED) between the target data and the solution vector, and Eq 3 can be then reorganized as shown in Eq 4 [9, 10, 15, 48]:

$$(d, \theta)^{GLS} = \arg \min_{\substack{d \in S_d \\ \theta \in S_t}} [(d - \bar{d})^T U^{-1} (d - \bar{d}) + (P(d, \theta)d - \hat{f})^T V^{-1} (P(d, \theta)d - \hat{f}) + (\theta - \bar{\theta})^2 Q^{-1}] \tag{4}$$

Where:

$P(d, \theta) = \{p_{kj}(d_j, \theta)\}$ is the assignment matrix, and is a function of both OD matrix and dispersion parameter θ .

$f = P(d, \theta)d$ is the estimated link flow vector. $f = \{f_k\}$, where $f_k = \sum_j p_{kj}(d_j, \theta)d_j$.

The matrix for link choice proportions P can be generally assumed fixed during the optimization procedure [10, 13, 14, 19]. This procedure performs well for uncongested traffic conditions or an idealized traffic network with fixed link costs. However, when the network becomes congested, users' choices are increasingly influenced by adverse traffic condition. In this case, link flow and cost are not independent, and the assignment matrix P should be assumed to vary within each optimization step for link flow and OD estimation. Similarly, the GLS estimators of d and θ can be also obtained by solving Eq 4.

The Link Choice Proportion Calculation Using the Dispersion Parameter. As mentioned in notation and definitions subsection, the link flow and cost will be updated when a new set of values of d and θ is received. Drivers' link choice decisions are influenced by the network-wide traffic condition, and thus the link choice proportion matrix P should be allowed to vary as well. The method of successive average (MSA) is adopted to calculate equilibrium link flows in the traffic assignment procedure [7, 16, 45, 52].

The cost of path r connecting the OD pair j can be expressed as:

$$c_{rj} = \sum_k a_{krj} c_k, \forall k \in K \tag{5}$$

The probability P_{rj} can be then computed according to the path choice logit model [45]:

$$\begin{aligned} P_{rj} &= \frac{\exp(-c_{rj}\theta)}{\sum_a \exp(-c_{aj}\theta)} \\ &= \frac{1}{\sum_a \exp(-c_{aj}\theta)} \exp(-\theta \sum_k a_{krj} c_k) \\ &= \frac{1}{\sum_a \exp(-c_{aj}\theta)} [\exp(-\theta a_{1rj} c_1) \cdot \exp(-\theta a_{2rj} c_2) \cdot \dots \cdot \exp(-\theta a_{Trj} c_T)] \end{aligned} \tag{6}$$

For a driver traveling along the path r , the weight assigned to link k is equal to $\exp(-c_k\theta)$. It is worth noting that the sum of probabilities over all feasible paths for each OD pair is equal to one.

As previously noted, $W' = [w'_{mn}]$ is the initial weight matrix of all possible paths connecting each OD pair. With the initial weight is set to $w'_{mn} = \exp(-c_k\theta)$, then W' , W'_2 , and W'_3 represent the weight matrix in the group of paths with one link, two links and three links respectively. Therefore, the weight matrix for all possible paths can be formulated as:

$$W' + W'_2 + W'_3 + \dots = (E - W')^{-1} - E \tag{7}$$

Wong [53] and Lo and Chan [16] have proven that the right side of Eq 7 is convergent for any acyclic networks, and is equal to $W = (E - W')^{-1} - E$. Therefore, the probability of a trip from node m to node n (OD pair j) choosing link k can be calculated as follows:

$$p_{kj} = \frac{w_{mg} \exp(-\theta c_k) w_{vn}}{w_{mn}} \tag{8}$$

Where link k connects node g and node v , and w_{mn} expresses weight matrix of all possible paths connecting nodes m and n , $m, n \in M$. w_{mn} is set to 1 for all nodes in the network.

Following the previous definition, the auxiliary mean traffic flow $y_k^{(s)}$ of link k is defined for each incoming d and θ via the following equation:

$$\begin{aligned} y_k^{(s)} &= [Pd]_k \\ &= \sum_j p_{kj} d_j, \forall k \in K \end{aligned} \tag{9}$$

The equilibrium traffic link flows can be then obtained using the MSA method. Specifically, the flow of link k can be calculated at the $(s+1)$ th iteration with the following equation:

$$\begin{aligned} x_k^{(s+1)} &= x_k^{(s)} + \frac{1}{s} (y_k^{(s)} - x_k^{(s)}) \\ &= \frac{1}{s} \sum_{j=1}^s y_k^{(j)}, \forall k \in K \end{aligned} \tag{10}$$

As shown in Eq 10, the flow of link k at the $(s+1)$ th iteration is equal to the mean of the auxiliary traffic flow of link k in the previous s iterations.

When a new set of values of d and θ is received, the matrix P of link choice proportions is updated following the procedure described above, and is then integrated into the Eq 4 to update the values of d and θ . This optimization procedure continues until convergence of the OD matrix and dispersion parameter estimation is reached.

Model Solution Algorithm

To solve the Stochastic User Equilibrium (SUE) problem described above, a two-stage algorithm for GLS estimation and SUE traffic assignment is proposed: First, the OD matrix d and the dispersion parameter θ are simultaneously estimated under the condition of the fixed link flows, link costs, and weight matrix. Second, the link flows, link costs, and link choice proportions are updated according to the new values of d and θ in the SUE assignment process. The two-stage algorithm is executed iteratively until the convergence of values of d and θ is reached. Sequential quadratic programming (SQP) from the extended quasi-Newton method is chosen as the solution method [49].

Two-Stage Algorithm

The initialization procedure of the two-stage algorithm can be described as follows:

1. Initialize the counter $t = 0$, set the initial OD vector matrix $d^{(0)} = \bar{d}$, the initial dispersion parameter $\theta^{(0)} = \bar{\theta}$, and the initial link flow $x_k^{(0)} = 0, k \in K$.
2. Calculate the initial link costs for all links in the network using [Eq 1](#), and calculate the weight matrix W for all paths based on the initial link costs and $\theta^{(0)}$.
3. Calculate the link choice proportion matrix P using the weight matrix W and $\theta^{(0)}$.
4. Calculate the initial mean auxiliary traffic flow for all the observed links with [Eq 9](#), and update $t = t + 1$.

The first stage of the algorithm is described as follows:

Step 1. The objective function ([Eq 4](#)) can be updated with the new mean auxiliary observed link flows as follows:

$$(d^{(t)}, \theta^{(t)})^{GLS} = \arg \min_{\substack{d \geq 0 \\ \theta > 0}} [(d^{(t)} - \bar{d})^T U^{-1} (d^{(t)} - \bar{d}) + (P^{(t)} d^{(t)} - \hat{f})^T V^{-1} (P^{(t)} d^{(t)} - \hat{f}) + (\theta^{(t)} - \bar{\theta})^2 Q^{-1}] \quad (11)$$

Where:

U^{-1} , $P^{(t)}$, V^{-1} , and Q^{-1} can be updated using the new mean auxiliary observed link flows, estimated OD vector matrix, dispersion parameter, and link flow vector respectively; The feasible set for d and θ should meet the requirements $d \geq 0, \theta > 0$. When the value of θ approaches zero, the path choice probabilities for all paths tend to be equal. As the value of θ increases, the path choice probabilities tend to be deterministic.

Step 2. Use the SQP algorithm to obtain a new set of values of $d^{(t)}$ and $\theta^{(t)}$ that minimizes the objective function. The starting point for optimizing the OD vector $\tilde{d}^{(t)}$ and dispersion parameter $\tilde{\theta}^{(t)}$ should be fixed in advance. During the iterative process of the SQP algorithm, whenever a new value θ is received, the link choice proportion matrix P will be updated by changing the value of $\exp(-\theta c_k)$ in [Eq 8](#), while the link cost and weight matrix should remain unchanged. The second stage of the algorithm can be described as follows:

Step 3. Initialize the counter $s = 1$.

Step 4. Calculate the weight matrix W with the new dispersion parameter $\theta^{(t)}$.

Step 5. Calculate the link choice proportion matrix $P^{(t)}$ using the weight matrix W and dispersion parameter $\theta^{(t)}$.

Step 6. Calculate the mean auxiliary traffic flow for all observed links as follows:

$$y_l^{(s)} = [P^{(t)} d^{(t)}]_l = \sum_j p_{lj} d_j, l \in L$$

Step 7. Calculate the equilibrium traffic link flow of link k via the MSA method:

$$x_l^{(s+1)} = x_l^{(s)} + \frac{1}{s} (y_l^{(s)} - x_l^{(s)}), l \in L$$

Step 8. The maximum relative difference between current and previous mean link flows should satisfy the following requirement:

$$\max_{\forall l \in L} \left\{ \frac{|x_l^{(s+1)} - x_l^{(s)}|}{x_l^{(s+1)}} \right\} \leq \varepsilon_1 \tag{12}$$

If the above requirement is met, the algorithm proceeds directly to step 11, otherwise proceed to step 9.

Step 9. Calculate the new link costs according to $x_l^{(s+1)}, l \in L$.

Step 10. Calculate the weight matrix using the updated link costs, set $s = s + 1$, and return to step 5.

Step 11. The maximum relative difference between the current and previous OD matrix estimates should satisfy the following requirement:

$$\max_{\forall j \in J} \left\{ \frac{|d_j^{(t)} - d_j^{(t-1)}|}{d_j^{(t)}} \right\} \leq \varepsilon_2 \tag{13}$$

If the above requirement is met, terminate the procedure and output the current estimates of OD vector matrix d and dispersion parameter θ as $d^{(t)}$ and $\theta^{(t)}$. Otherwise, set $t = t + 1$, and proceed to step 12.

Step 12. Calculate the new starting points as follows: $\tilde{d}^{(t+1)} = \frac{1}{t}d^{(t)} + \frac{t-1}{t}d^{(t-1)}$, $\tilde{\theta}^{(t+1)} = \frac{1}{t}\tilde{\theta}^{(t)} + \frac{t-1}{t}\tilde{\theta}^{(t-1)}$, and return to step 1.

Model Evaluation

To evaluate the performance of the proposed method, the root mean squared errors (RMSE) for OD matrix and link flows after convergence are defined as follows:

(1) The root mean squared error (RMSE) of the estimated link flows $x_l^{(s+1)}$ relative to the true link flow x_l is computed as follows:

$$RMSE(LF) = \sqrt{\frac{1}{\Gamma} \sum_{l=1}^{\Gamma} (x_l^{(s+1)} - x_l)^2} \tag{14}$$

Similarly, the RMSE of the observed (target) link flows \hat{f}_l relative to the true link flows x_l can be defined as $RMSE(L\bar{F})$, where $x_l^{(s+1)}$ is replaced by \hat{f}_l in Eq 14.

(2) The RMSE of the estimated OD matrix $d^{(t)}$ relative to the true OD matrix d can be defined as $RMSE(OD)$:

$$RMSE(OD) = \sqrt{\frac{1}{\tau} \sum_{j=1}^{\tau} (d_j^{(t)} - d_j)^2} \tag{15}$$

Likewise, the RMSE of target OD matrix \bar{d} relative to the true OD matrix d is defined as RMSE ($O\bar{D}$), where $d^{(i)}$ is replaced by \bar{d}_j in Eq 15.

Numerical Experiment and Result Analysis

A Hypothetical Network Test

In this section, the performance of the proposed approach is tested in a hypothetical network. The network and data proposed by Yang et al. [15] and Caggiani et al. [19] are adopted as the test bed with some slight modifications. The network (presented in Fig 1), is composed of 9 nodes (3 origin centroids and 3 destination centroids), and 14 links. The true and initial OD vector matrices d and \tilde{d} for the SQP algorithm are shown in Table 1. The initial dispersion parameter $\tilde{\theta}$ is assumed to be 4, and the true dispersion parameter $\hat{\theta}$ is fixed to 1.5. Note that the initial OD matrix \tilde{d} and dispersion parameter $\tilde{\theta}$ are quite dissimilar from those of the ground truth data.

The following parameters in the Bureau of Public Roads (BPR) [50] link performance function are used: $\alpha_k = 0.15$ and $\beta_k = 4, \forall k \in K$. In addition, the free flow travel time (t_k) and capacity (C_k) for each link are predetermined as shown in Table 2.

The ground truth link flows can be generated by allocating the true OD matrix to the traffic network using SUE-Logit assignment method presented in Section 2.3. The true dispersion parameter is $\theta = 1.5$, resulting in the link flows shown in Table 3. The set of links {5, 6, 7, 11, 13} is selected as the observed links.

In this example, we assume that the OD vector and link flow vector follow the Poisson distribution. The covariance matrices U (for OD demands) and V (for link flows) in Eq 4 can be assumed to be diagonal matrices [9, 14, 54]. The diagonal element for U , V and Q can be computed respectively through the following equations:

$$\sigma_{d_j}^2 = (cv_d \cdot d_j)^2, \sigma_{x_k}^2 = (cv_x \cdot x_k^{(s+1)})^2, \sigma_{\theta}^2 = (cv_{\theta} \cdot \theta)^2$$

Where cv_d , cv_x and cv_{θ} represent the variation coefficients for OD demands, link flows, and dispersion parameter respectively. Specifically, these parameters are set as $cv_d = 0.3$, $cv_x = 0.05$, and $cv_{\theta} = 0.1$.

The target OD matrix \bar{d} , observed link flow vectors \hat{f} , and target dispersion parameter $\bar{\theta}$ can be generated separately by adding random terms into the corresponding true values. The random terms are sampled from independent normal variables with zero means. For instance, the target OD matrix can be calculated by adding a random term with $\lambda_d = 0.3$ to the values of the true OD matrix divided by two, the observed link flow vectors can be generated by adding a random term with $\lambda_f = 0.1$, and the target parameter can be set as $\bar{\theta} = 4$. In addition, the error tolerance threshold used in the optimization is set to $\epsilon_1 = \epsilon_2 = 10^{-3}$. The convergence for theta is plotted in Fig 2, which shows the estimate slowly falling in the first 120 iterations before rapidly converging to the true value at 1.5099. This is a very slight deviation with the true value of 1.5. In addition, the convergence of the objective function is presented in Fig 3, where the value of the objective function sharply falls at the first iteration and then gradually decreases and levels off at a lower value. Poor initial choices of OD input vector and dispersion parameter may lead to the slower convergence.

In order to further evaluate the effectiveness of the proposed approach, a sensitivity analysis is conducted with parameter cv_{θ} (CVT) varying from 0.1 to 0.5 and cv_d (CVD) changing from

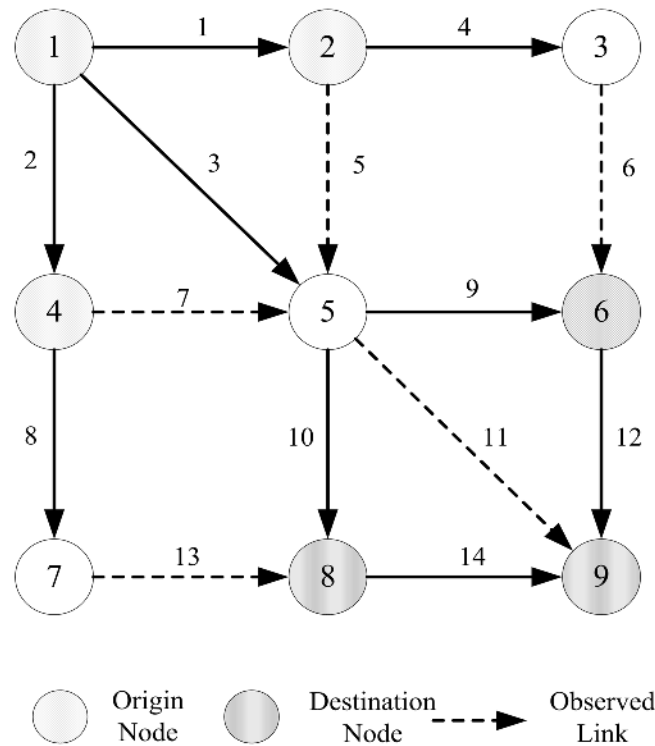


Fig 1. The test network used in the numerical example.

doi:10.1371/journal.pone.0146850.g001

Table 1. The true and initial OD vector matrices.

OD pair	1–6	1–8	1–9	2–6	2–8	2–9	4–6	4–8	4–9
j	1	2	3	4	5	6	7	8	9
d	120	150	100	130	200	90	80	180	110
\tilde{d}	30	20	10	30	30	30	30	40	20

doi:10.1371/journal.pone.0146850.t001

Table 2. Free flow travel time and capacity for each link.

link	1	2	3	4	5	6	7	8	9	10	11	12	13	14
t_k	2.0	1.5	3.0	1.0	1.0	2.0	2.0	1.0	1.0	1.0	2.0	1.0	1.0	1.0
C_k	280	290	280	280	600	300	500	400	500	700	250	300	350	520

doi:10.1371/journal.pone.0146850.t002

Table 3. True link flows in the hypothetical network.

link	1	2	3	4	5	6	7	8	9	10	11	12	13	14
x_k	125	143	103	172	474	172	201	313	307	393	279	148	313	475

doi:10.1371/journal.pone.0146850.t003

0.1 to 1. This generates 50 different estimates for RMSE (OD), RMSE (LF) and Theta as presented in Figs 4–6.

As shown in Fig 4, RMSE (OD) increases with the variation coefficient cv_θ when cv_d falls between 0.2 and 0.8. With cv_θ fixed between 0.3 and 0.5, RMSE(OD) can be seen as a convex

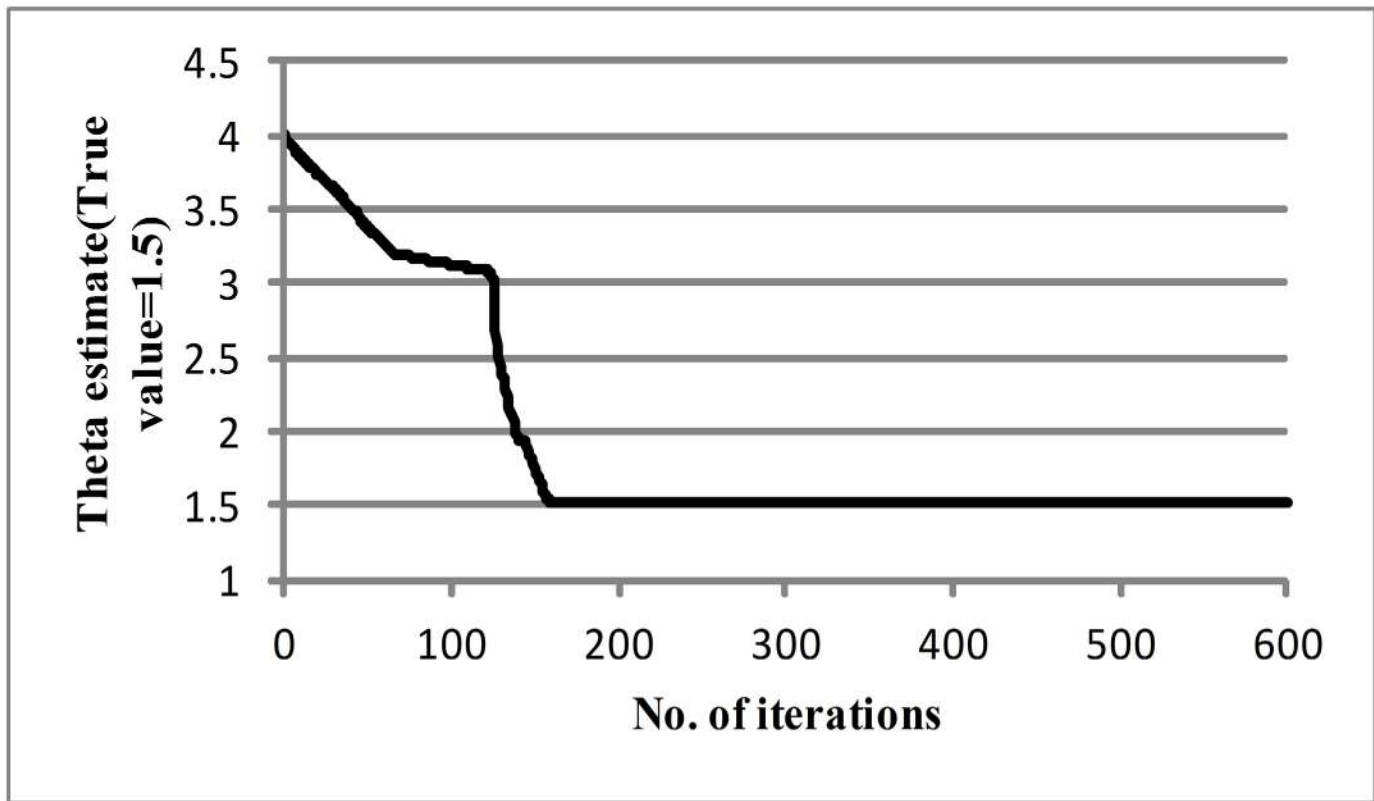


Fig 2. Convergence of the theta estimate.

doi:10.1371/journal.pone.0146850.g002

function of cv_d . Alternatively, when cv_θ is between 0.1 and 0.2, cv_d has a negligible impact on RMSE (OD). Thus, the maximum value (19.9022) of RMSE (OD) can be found at $cv_d = 0.5$ and $cv_\theta = 0.5$, and the minimum value (4.6859) is obtained at $cv_d = 0.3$ and $cv_\theta = 0.1$. Compared with the initial RMSE (OD) of 93.8971 calculated from Table 1, a 78.8% reduction is achieved at the maximum RMSE (OD), and a 95% reduction is obtained at the minimum RMSE (OD).

Fig 5 shows the impact of cv_d and cv_θ on RMSE (LF). With the value of cv_d fixed, RMSE (LF) increases with the variation coefficient cv_θ . For a fixed value of cv_θ , the RMSE (LF) decreases with an increase in cv_d . Thus, we can conclude that maximum RMSE (LF) value of 19.6597 is located at $cv_d = 0.1$ and $cv_\theta = 0.5$, and the minimum value of 8.6696 can be found at $cv_d = 0.6$ and $cv_\theta = 0.1$. Compared with the initial RMSE (LF) value of 30.8347, 36.2% and 71.9% reductions can be achieved for the maximum value of RMSE (LF) and minimum value of RMSE (LF) respectively.

As shown in Fig 6, the value of theta varies negligibly with the choice of cv_d and cv_θ . In other words, the estimated value of theta always converges to approximately the true value. As shown in Fig 6, for a fixed value of cv_d , the estimated θ is close to the true value for any given cv_θ . For example, the value of θ fluctuates between 1.37 and 1.51 when $cv_d = 0.3$. Likewise, for any fixed cv_θ , the estimated θ varies minimally about the true value of θ using the proposed method. For example, the estimated θ is between 1.35 and 1.52 for $cv_\theta = 0.1$.

The above discussion reveals a fact that the initial value of \bar{d} , $\bar{\theta}$, and observed link flow vectors \hat{f} do not affect the theta estimation performance. This is equivalent to a convex optimization problem, where the optimal results tend to converge near the true dispersion parameter value. This implies that the estimate of θ is insensitive to the variation coefficients, and can be used as a stable and accurate parameter to determine travelers' route decisions.

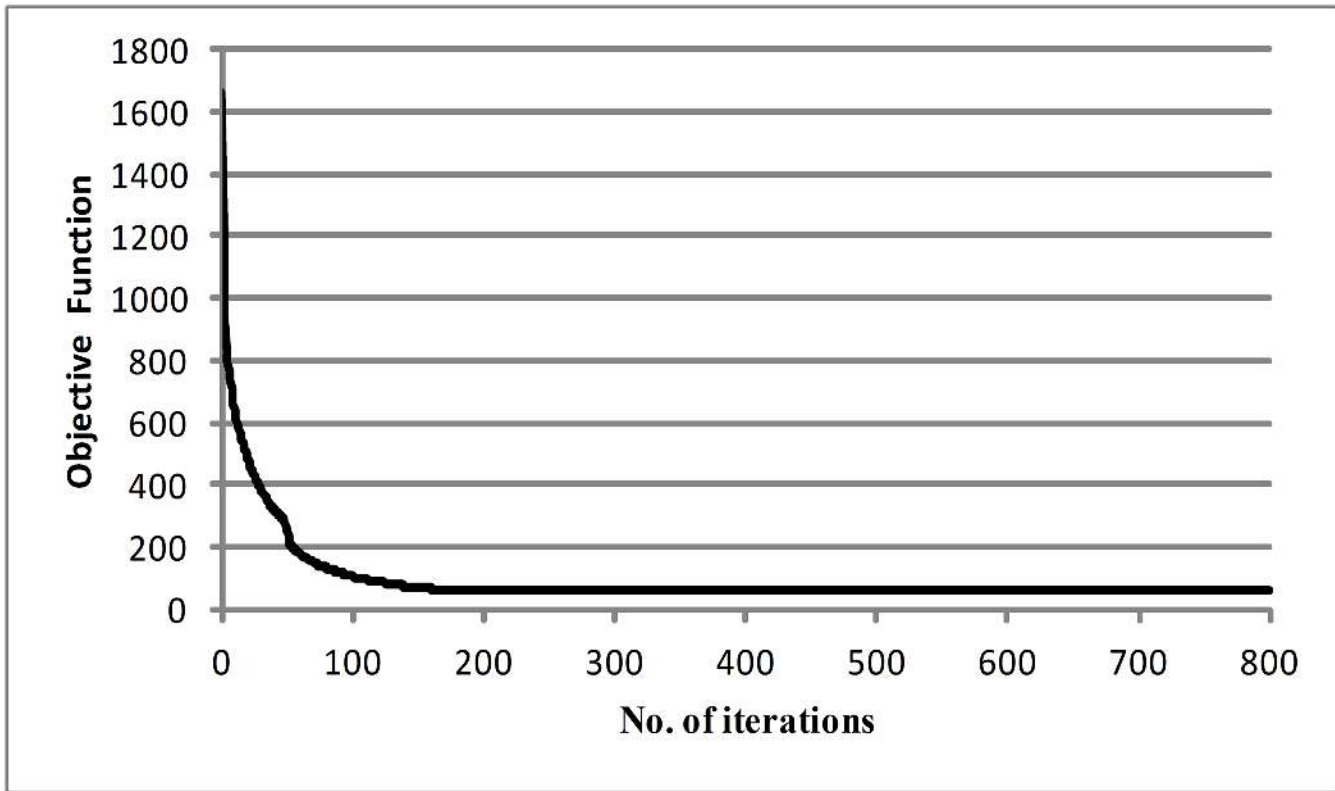


Fig 3. Convergence of the objective function.

doi:10.1371/journal.pone.0146850.g003

Comparison and Analysis

To further demonstrate the advantages of the proposed methodology, two OD matrix estimation methods are implemented and compared with the proposed approach. To make this comparison, we first implement the algorithm described in Yang et al. [15], which presents an optimization model for OD matrix estimation in congested networks using the logit-based

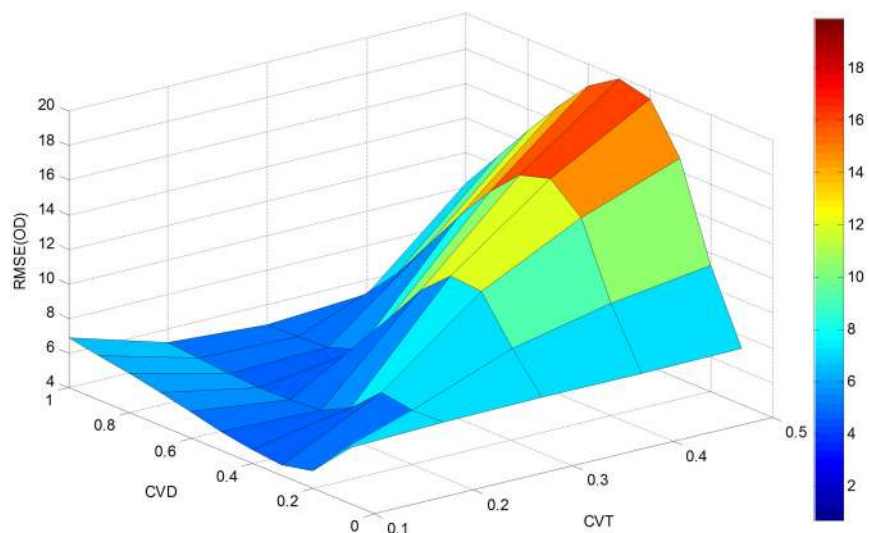


Fig 4. RMSE(OD) versus cv_d and cv_θ .

doi:10.1371/journal.pone.0146850.g004

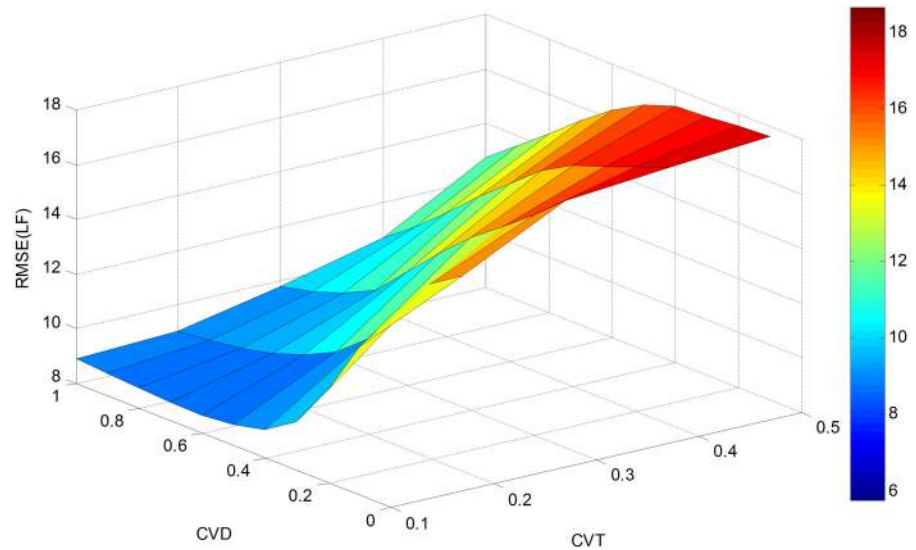


Fig 5. RMSE(LF) versus cv_d and cv_θ .

doi:10.1371/journal.pone.0146850.g005

SUE. The method described in Lo and Chan [16] is implemented for the second comparison. This method applies both statistical estimation and traffic assignment to simultaneously calculate the OD matrix and link choice proportions based on OD survey data and traffic counts. To maintain a fair comparison, the same test network and data set are applied in all cases.

The OD matrix estimation method proposed by Yang et al. [15] is given in section 2.2. The objective function is shown in Eq 16.

$$\begin{aligned}
 (d, \theta) &= \arg \min_{d \in S_d} [F_1(d, \bar{d}) + F_2(f, \hat{f})] \\
 &= \arg \min_{d \geq 0} \left[\frac{1}{2} (d - \bar{d})^2 + \frac{1}{2} (Pd - \hat{f})^2 \right] \quad (16) \\
 &\theta > 0
 \end{aligned}$$

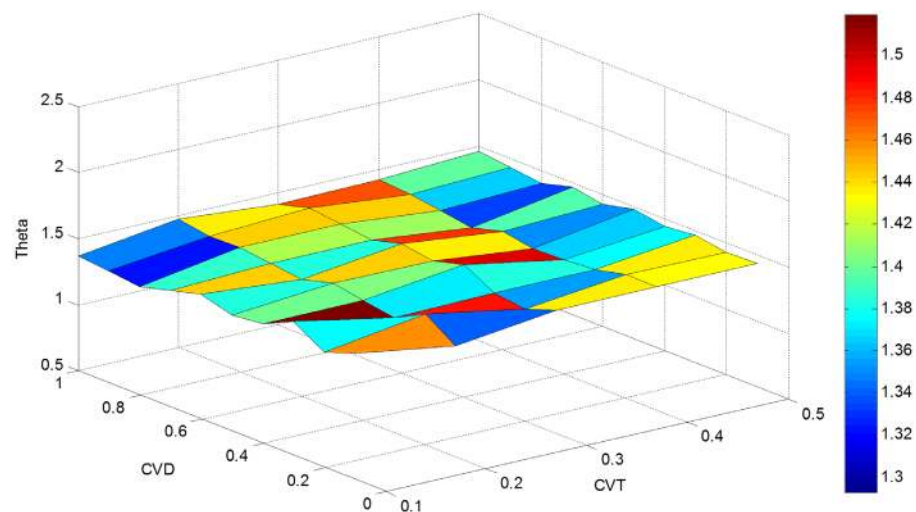


Fig 6. Estimated Theta versus cv_d and cv_θ .

doi:10.1371/journal.pone.0146850.g006

Table 4. Comparison between Yang et al.’s approach and the proposed approach with OD matrix and link flow estimation.

Approach	Estimated OD matrix		Estimated Link flows	
	RMSE (OD)	RMSE (OD)	RMSE (LF)	RMSE (LF)
Yang et al.’s approach	24.27	20.85	26.65	19.02
Proposed approach	24.27	18.79	26.65	17.39

doi:10.1371/journal.pone.0146850.t004

In Yang et al.’s work, the weighted Euclidean distance function is used to develop a unit weighting matrix and the value of theta is set to 1.5. The RMSE(OD), RMSE(LF), RMSE (OD) and RMSE(LF) for Yang et al.’s approach are calculated and compared with the proposed approach in [Table 4](#):

As shown in [Table 4](#), the proposed method yields significantly lower RMSE (OD) and RMSE (LF) relative to Yang et al.’s approach. Compared with the initial RMSE values, a 22.6% reduction in RMSE (OD) is achieved using the proposed approach, while only a 14.1% reduction is achieved using the method described in Yang et al. Similarly, the proposed approach resulted in a 34.7% reduction in RMSE(LF), while only a 28.6% reduction was achieved using Yang et al.’s approach. One reason that the dispersion parameter is estimated and integrated into the [Eq 3](#) by $F_3(\theta, \bar{\theta})$ in the proposed method, and it is able to yield a better estimate of the dispersion parameter than previous approaches. The other reason is that the covariance matrices U (for OD demands), V (for link flows) and Q (for dispersion parameter) are not a fixed variable during the calculation. These improvements can help the method enhance the estimation performance for the OD matrix and link flow vectors.

Lo and Chan [[16](#)] present the following maximum likelihood objective function:

$$(d, \theta) = \arg \max_{d \geq 0, \theta > 0} \ln L(\theta, d | \hat{f}, \bar{d}) \tag{17}$$

In Lo and Chan [[16](#)], it is assumed that the observed flows are equal to the true flows in the test network. For Lo and Chan’s algorithm, we set the target dispersion parameter to $\bar{\theta} = 4$ (This is also equal to the initial dispersion parameter value used in Lo and Chan [[16](#)]’s work), and the variation coefficients as follows: $cv_{\theta} = 0.1$, $cv_x = 0.05$, and $cv_d = 0.3$. In order to evaluate the performance of the proposed approach relative to that of Lo and Chan [[16](#)]’s method, RMSE (OD), RMSE (LF), and the estimated Theta are selected for comparison and shown in [Table 5](#).

Unlike Lo and Chan’s method, random terms are added to the observed link flows in the proposed approach, thus introducing additional challenges for estimation. However, the results presented in [Table 5](#) demonstrate that the method proposed in this paper outperforms Lo and Chan’s approach in terms of OD matrix, link flow, and Theta estimation accuracy.

Table 5. Comparison between Lo and Chan’s approach and the proposed approach with OD matrix, link flow and Theta estimation.

Approach	Estimated OD matrix		Estimated Theta	
	RMSE (OD)	RMSE (LF)	Theta target	Theta estimated
Lo and Chan’s approach	5.34	12.08	4	1.572
Proposed approach	4.69	9.77	4	1.509

doi:10.1371/journal.pone.0146850.t005

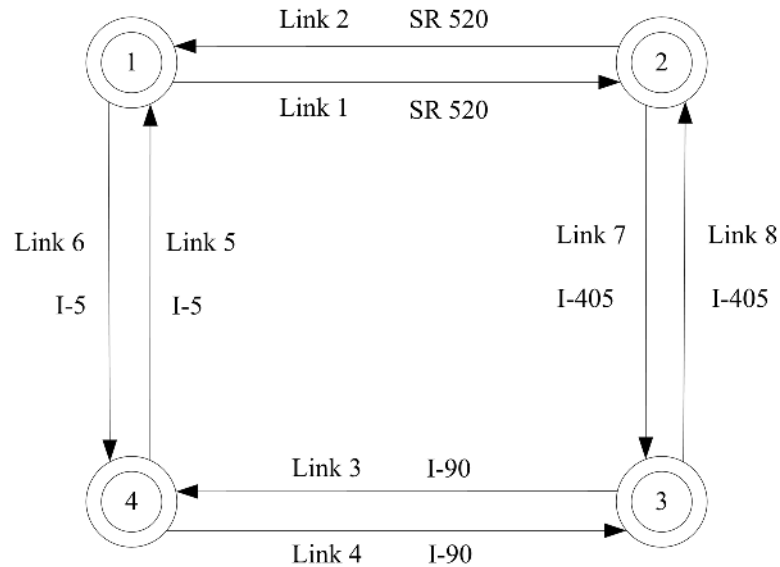


Fig 7. Square network in Seattle. Double circle nodes represent zone centroids (origins and destinations).

doi:10.1371/journal.pone.0146850.g007

Application to A Square Network in Seattle

A square network in Seattle is used as a congested network case study to demonstrate the applicability and transferability of the proposed approach in a real-world traffic network (Shown in Fig 7). Empirical data was collected from loop detectors located along one freeway section in Seattle area, and obtained for this research through the Strategic Highway Research Program 2 (SHRP 2 program) supported by Washington State Department of Transportation (WSDOT) [55].

The square test network used in this case study consists of 4 nodes and 8 links, where all nodes are centroids (origins and destinations). The topology of the test network is outlined in Fig 7. We assume that the study network is acyclic, such that the traffic flow starting from one node will leave the network before returning to the original node. Specifically, Links 1 and 2 represent the SR 520 Bridge connecting I-5 in Seattle and SR 202 in Redmond. Interstate 90 (I-90) is represented by Links 3 and 4, and Interstate 5 (I-5) is represented by Links 5 and 6. Links 7 and 8 represent Interstate 405 (I-405), which intersects I-90 in the south and SR 520 in the north.

Traffic flows were obtained from loop detectors installed at nodes 1, 2, 3 and 4, illustrated in Fig 8. The parameters for the BPR link performance cost function (Eq 18) were estimated based on the empirical data and are presented in Table 6.

$$c_k(f_k) = t_k [1.0 + \alpha_k (\frac{f_k}{C_k})^{\beta_k}], \forall k \in K \tag{18}$$

Table 7 indicates the external traffic flow recorded for each node during peak hour, where 1-Link 1 represents the external traffic flow on Link 1 from node 1, and 2-Link 7 represents the external traffic flow on Link 7 from node 2, and so forth. To convert true link flows into a ground truth OD matrix, the flow proportion for each node $\eta = 0.6$ is assumed based on extensive video records and filed surveys. This implies that, for the traffic leaving each node, 60% exits the network from an adjacent node while 40% exits from the other nodes. In order to avoid circular flow in the OD calculation process, it is assumed that the final remaining traffic

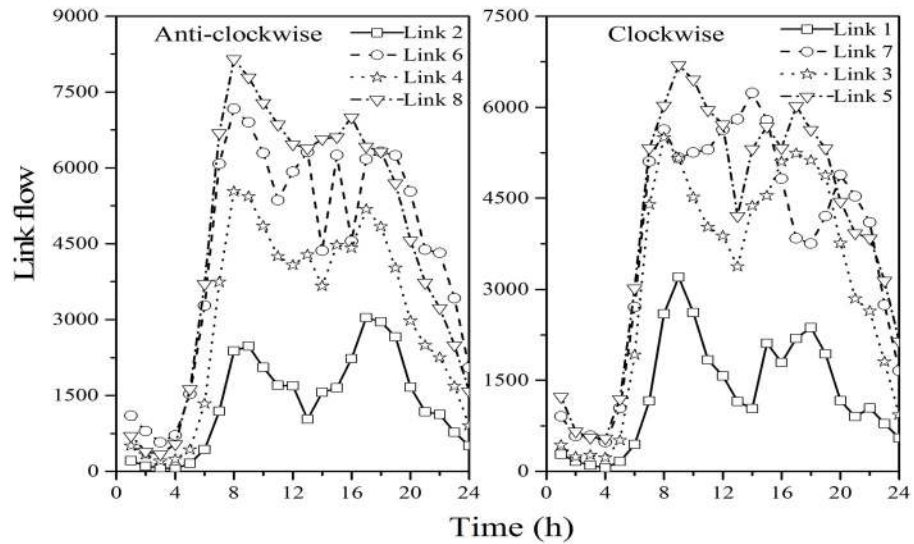


Fig 8. Traffic flow fluctuation by time of day.

doi:10.1371/journal.pone.0146850.g008

flow will leave from the last node before returning to the original node. Based on these assumptions, the ground truth OD matrix is calculated and shown in Table 8. In addition, the initial OD matrix \tilde{d} can be computed by rounding the last digit of the true OD matrix as shown in Table 8.

The true OD matrix in Table 8 is then used to assign the corresponding traffic flow into each link according to Eq 5 through Eq 10. The calculated traffic flows can be assumed to represent the true link flows, where link 1, 3, 5, 6, and 8 are selected as the observed links to estimate OD matrix shown in Table 9.

Similar to the hypothetical network, we assume that the OD demands and observed link flows follow the Poisson distribution, and the covariance matrices U and V can be assumed to be diagonal. The initial value of the dispersion parameter $\tilde{\theta}$ is set to 40.5. The remaining input parameters are set identically to the hypothetical network. In addition, a sensitivity analysis with 50 different combinations of variation coefficients cv_d and cv_θ was conducted to investigate the optimal parameter initialization for the proposed approach. The results of this sensitivity analysis are shown in Figs 9–11.

Table 6. BPR link performance cost function parameter calibration.

Links	Road Name	Length(km)	t_k (h)	C_k (pcu/h)	α_k	β_k
1,2	SR 520	7.5	0.1162	4149	0.1450	3.5
3,4	I-90	3.6	0.0667	8685	0.1035	2.7
5,6	I-5	3.5	0.1016	9683	0.0988	2.7
7,8	I405	7.8	0.1332	7961	0.1242	3.5

doi:10.1371/journal.pone.0146850.t006

Table 7. The external true traffic flow for each node at peak hour.

Link direction	1-Link 1	2-Link 2	3-Link 3	4-Link 4	4-Link 5	1-Link 6	2-Link 7	3-Link 8
Traffic flow	3199	2480	5499	5535	6018	7169	5628	8153

doi:10.1371/journal.pone.0146850.t007

Table 8. True OD matrix and initial OD matrix at peak hour for each OD pair.

OD pair	1–2	1–3	1–4	2–1	2–3	2–4	3–1	3–2	3–4	4–1	4–2	4–3
j	1	2	3	4	5	6	7	8	9	10	11	12
d	3067	2489	4814	2389	3774	1946	3277	5772	4604	4497	2773	4284
\tilde{d}	307	249	482	239	378	195	328	578	461	450	278	429

doi:10.1371/journal.pone.0146850.t008

Table 9. Observed link flows at peak hour.

Link No.	1	2	3	4	5	6	7	8
x_k	3711	-	8282	-	8439	7857	-	7591

doi:10.1371/journal.pone.0146850.t009

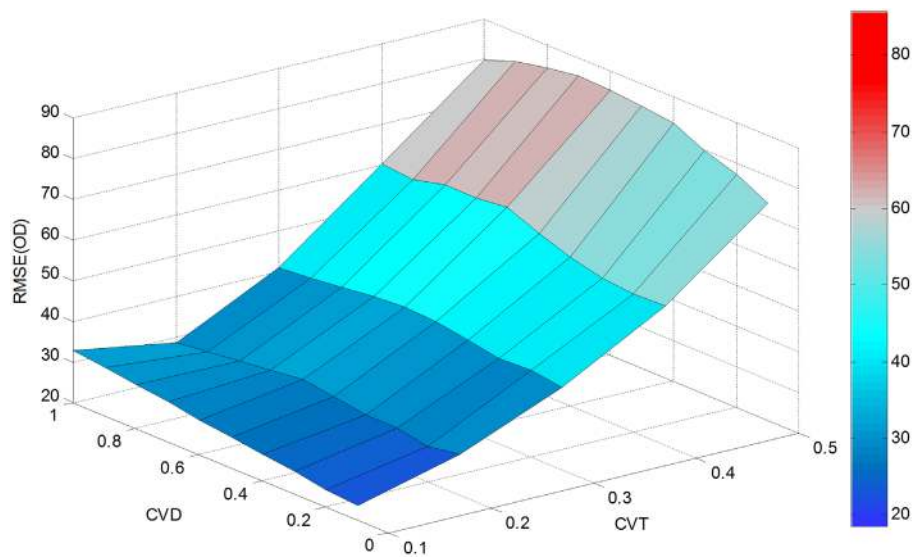


Fig 9. RMSE(OD) versus cv_d and cv_θ in the actual network.

doi:10.1371/journal.pone.0146850.g009

Fig 9 shows the value of RSME (OD) versus cv_d and cv_θ . For a fixed value of cv_d between 0.1 and 0.7, RMSE(OD) increases with cv_θ . RMSE(OD) is a concave function of cv_θ when cv_d is fixed between 0.8 and 1.0, and a convex function of cv_d for a fixed value of cv_θ between 0.1 and 0.5. Thus, the maximum RMSE (OD) of 85.6123 can be obtained at $cv_d = 0.7$ and $cv_\theta = 0.5$, and the minimum value of 23.5917 can be obtained at $cv_d = 0.1$ and $cv_\theta = 0.1$.

As shown in Fig 10, the value of RMSE (LF) increases with cv_θ for a fixed value cv_d . For a fixed value of cv_θ , the value of RMSE (LF) decreases with an increase of cv_d . The maximum RMSE (LF) of 82.5113 is found at $cv_d = 0.1$ and $cv_\theta = 0.5$, and the minimum value of 7.3277 at $cv_d = 1.0$ and $cv_\theta = 0.1$.

As noted in the hypothetical case, the choice of cv_d and cv_θ has very little impact on the estimation of Theta. As shown in Fig 11, the estimated dispersion parameter θ is between 20.8327 ($cv_d = 0.5$ and $cv_\theta = 0.5$) and 22.7165 ($cv_d = 0.7$ and $cv_\theta = 0.2$) in all cases. The best estimate of dispersion parameter θ can be found between 20.8327 and 22.7165.

Finally, using the BPR link performance cost function parameters described in Table 6, different combinations of variation coefficient $cv_d = 0.3$ and $cv_\theta = 0.1$; $cv_d = 0.5$ and $cv_\theta = 0.5$; $cv_d = 0.7$ and $cv_\theta = 0.2$ are used to estimate theta for the actual network.

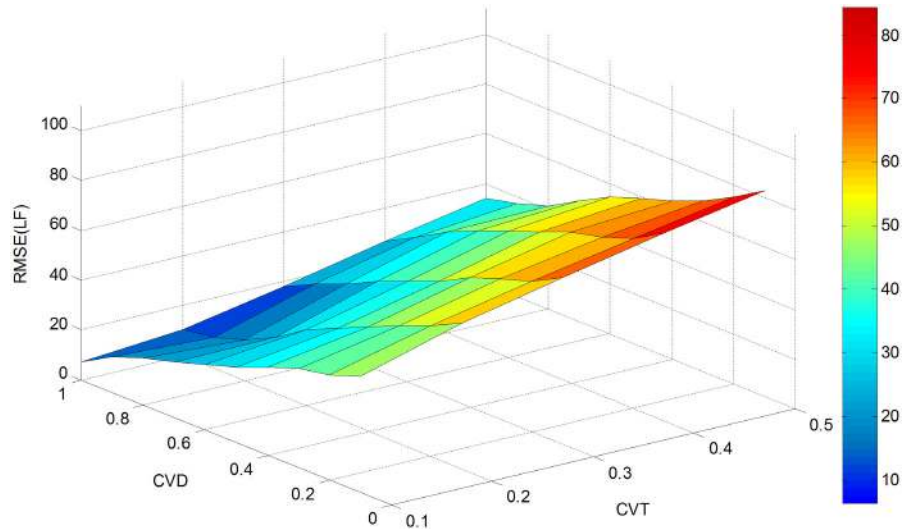


Fig 10. RMSE(LF) versus cv_d and cv_θ in the actual network.

doi:10.1371/journal.pone.0146850.g010

It is interesting to observe that the estimated RMSE(OD), RMSE(LF), and Theta for both hypothetical and actual networks exhibit a similar trend yet have obvious differences. Two primary reasons may explain these differences: First, the network topology is quite different for the two scenarios. The hypothetical network is unidirectional, where each node can be either origin or destination. In contrast, the actual network is bidirectional, where each node is both origin and destination, and thus multiple paths may exist between each OD pair. For example, the traffic flows on both 1-Link 1 and 1-Link 6 contribute to the OD demands from node 1 to node 2. Second, compared with the hypothetical network with equal cost parameters for all links, a more realistic BPR link performance cost function is adopted for the actual network. In the real-world network, the parameters (e.g. free-flow travel time and link capacity) are

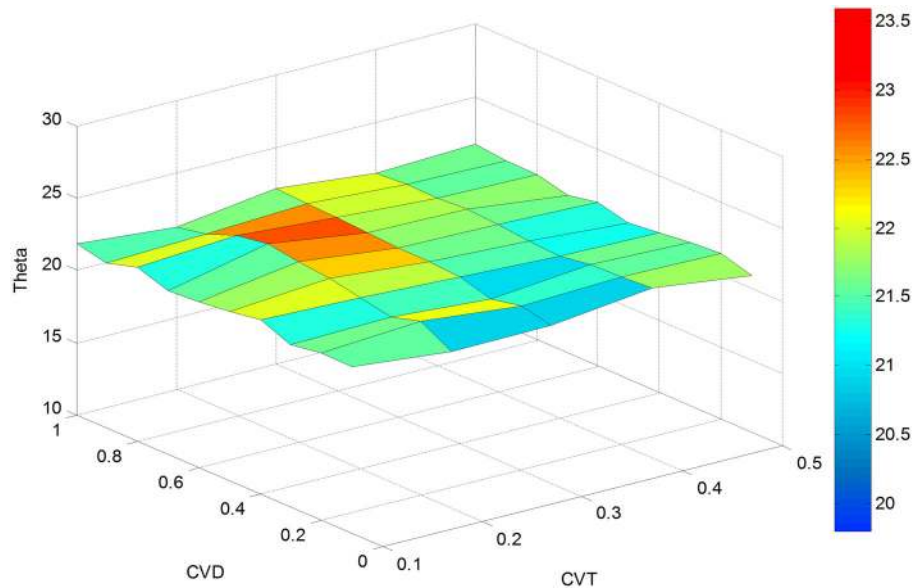


Fig 11. Theta estimated versus cv_d and cv_θ in the actual network.

doi:10.1371/journal.pone.0146850.g011

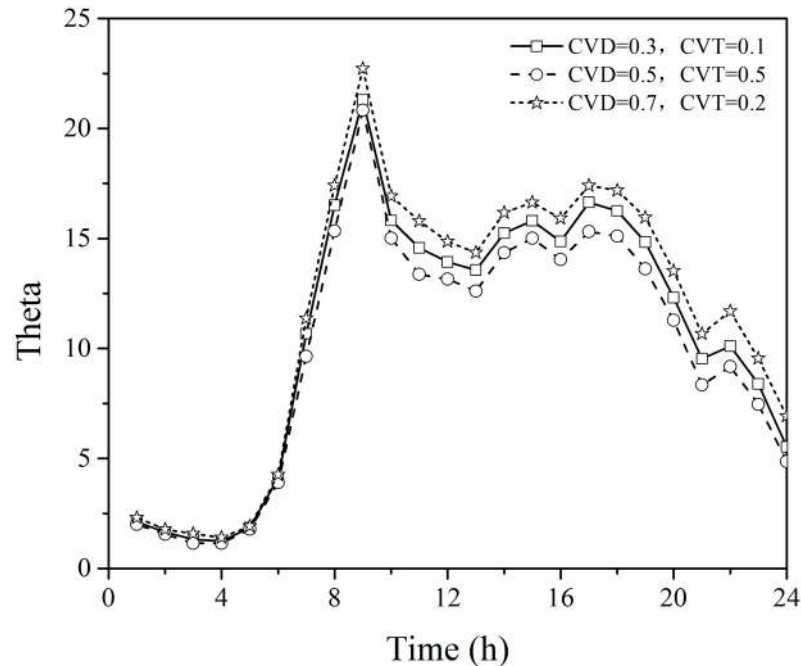


Fig 12. Estimated Theta values in the actual network by time of day.

doi:10.1371/journal.pone.0146850.g012

calibrated for each link based on empirical data. That said, the sensitivity analysis for Theta produced similar results for both the hypothetical and actual networks, indicating that this parameter is not sensitive to the choice of variation coefficients. In addition, the theta estimates obtained using a range of different parameter settings exhibits a similar and regular trend over time of day as shown in Fig 12. These findings provide guidance for initial parameter selection, and offer useful insight for interpreting modeling results.

Conclusions

This paper proposes a two-stage algorithm to simultaneously estimate origin-destination matrices and link choice proportions by incorporating a dynamic dispersion parameter into the route choice model. The dispersion parameter θ is of practical significance in describing travelers' route choice decisions, but has typically been assumed constant in previous studies. Finding the optimal dispersion parameter is not a straightforward task. To address this issue, this paper presents a model calibration procedure to simultaneously estimate the dispersion parameter θ , link choice proportions, and OD matrix. In order to obtain the Generalized Least Square (GLS) estimators of the above listed parameters, a two-stage algorithm is proposed which integrates GLS estimation into the SUE traffic assignment procedure. The first and second stages of the algorithm are applied iteratively until the maximum relative difference presented in Step 11 is achieved, after which the estimated OD matrix, link choice proportion, and dispersion parameter θ can be obtained. The SQP approach based on the extended quasi-Newton method is used to search for the optimal solution in the first stage of the algorithm. The SUE traffic assignment procedure is applied to incorporate both OD matrix and link choice proportion estimation into the second stage of the algorithm, and MSA is used to obtain the equilibrium link flows.

A hypothetical network was constructed to test the performance of the proposed approach, followed by a comprehensive sensitivity analysis with 50 combinations of variation coefficient

combinations cv_d (CVD) and cv_θ (CVT) to investigate the stability of the estimated OD matrix, link flows, and Theta. A comparison with two different methods described in Yang et al. [15] and Lo and Chan [16] suggests that the proposed approach can achieve superior performance in terms of RMSE (OD), RMSE (LF), and accuracy of the estimated Theta parameter. Moreover, a case study is presented using a real-world congested square network in Seattle, WA to demonstrate the practicality of the proposed approach, in which the true OD matrix and observed link flows are calculated via ground-truth traffic count data collected by loop detectors. The proposed method is shown to be robust under a range of initial parameter values. The RMSE (OD) can be reduced from 3426.9 to 23.6 at $cv_d = 0.1$ and $cv_\theta = 0.1$ when traffic flows are observed on five out of eight links. In addition, the estimated dispersion parameter exhibits a consistent and regular trend by time of day for all combinations of initial parameters. For future research, the proposed approach should be tested on a network of greater complexity and size, and the impact of input data inaccuracy should be considered. Additionally, further work is needed to determine the number and location of observed links required for accurate OD estimation using the proposed approach.

Supporting Information

S1 Dataset. The dataset includes the Link Speed Data and Link Volume data, and the data were collected from loop detectors located along the freeway section (I-5, I-90, I-405 and SR 520) in Seattle area, and are retrieved via the Strategic Highway Research Program 2 (SHRP 2 program). The file named as “S1 Link Speed Data” records the average speed for all links every 20-second time interval, and the other file named as “S1 Link Volume data” records volume for all links every 20-second time interval.
(RAR)

Acknowledgments

The helpful comments from the two anonymous reviewers are gratefully acknowledged. This research is supported by National Natural Science Foundation of China (Project No. 71402011, 71471024, 51408019, 71301180, 51329801), National Social Science Foundation of Chongqing of China (No. 2013YBJJ035), and the Scientific and Technological Research Program of Chongqing Municipal Education Commission (No. KJ1400307), and the Natural Science Foundation of Chongqing of China (No. cstc2015jcyjA30012), National Key Technologies R&D Program of China (2014BAG01B03), Science and Technology Project on Transportation Construction by the Ministry of Transport of China (2015318835200).

Author Contributions

Conceived and designed the experiments: Yong W. XM Yinhai W. Performed the experiments: Yong W. XM YL KG. Analyzed the data: Yong W. XM YL KG MX. Contributed reagents/materials/analysis tools: KCH Yinhai W. Wrote the paper: Yong W.

References

1. Simini F, González M C, Maritan A, Barabási A-L. A universal model for mobility and migration patterns. *Nature*, 2012; 484: 96–100. doi: [10.1038/nature10856](https://doi.org/10.1038/nature10856) PMID: [22367540](https://pubmed.ncbi.nlm.nih.gov/22367540/)
2. Simini F, Maritan A, Neda Z. Human mobility in a continuum approach. *PLoS one*, 2013; 8(3): e60069. doi: [10.1371/journal.pone.0060069](https://doi.org/10.1371/journal.pone.0060069) PMID: [23555885](https://pubmed.ncbi.nlm.nih.gov/23555885/)
3. Yan X-Y, Zhao C, Fan Y, Di Z, Wang W-X. Universal predictability of mobility patterns in cities. *Journal of the Royal Society Interface*, 2014; 11: 0834.

4. Gao Z K, Jin N D. A directed weighted complex network for characterizing chaotic dynamics from time series. *Nonlinear Analysis: Real World Applications*, 2012; 13(2): 947–952.
5. Tang J J, Wang Y H, Wang H, Zhang S, Liu F. Dynamic analysis of traffic time series at different temporal scales: A complex networks approach. *Physica A: Statistical Mechanics and its Applications*, 2014; 405, 303–315.
6. Gao Z K, Yang Y X, Fang P C, Zou Y, Xia C Y, Du M. Multiscale complex network for analyzing experimental multivariate time series. *Europhysics Letters*, 2015; 109(3): 30005.
7. Cheung WM, Wong SC, Tong CO. Estimation of a time-dependent origin-destination matrix for congested highway networks. *Journal of Advanced Transportation*, 2010; 40(1): 95–117.
8. Ma XL, Yu HY, Wang YP, Wang YH. Large-scale Transportation Network Congestion Evolution Prediction Using Deep Learning Theory. *PloS one*, 2015; 10(3): e0119044. doi: [10.1371/journal.pone.0119044](https://doi.org/10.1371/journal.pone.0119044) PMID: [25780910](https://pubmed.ncbi.nlm.nih.gov/25780910/)
9. Cascetta E. Estimation of trip matrices from traffic counts and survey data: a generalized least squares approach. *Transportation Research Part B: Methodological*, 1984; 18(4–5): 289–299.
10. Cascetta E, Nguyen S. A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B: Methodological*, 1988; 22(6): 437–455.
11. Yang H, Sasaki T, Iida Y, Asakura Y. Estimation of origin-destination matrices from link traffic counts on congested networks. *Transportation Research Part B: Methodological*, 1992; 26(6), 417–434.
12. Hazelton ML. Some comments on origin-destination matrix estimation. *Transportation Research Part A: Policy and Practice*, 2003; 37(10), 811–822.
13. Cantarella GE. A general fixed-point approach to multimode multi-user equilibrium assignment with elastic demand. *Transportation Science*, 1997; 31(2), 107–128.
14. Cascetta E, Postorino MN. Fixed point approaches to the estimation of O/D matrices using traffic counts on congested networks. *Transportation Science*, 2001; 35(2): 134–147.
15. Yang H, Meng Q, Bell MGH. Simultaneous estimation of the origin-destination matrices and travel-cost coefficient for congested networks in a stochastic user equilibrium. *Transportation Science*, 2001; 35(2): 107–123.
16. Lo HP, Chan CP. Simultaneous estimation of an origin-destination matrix and link choice proportions using traffic counts. *Transportation Research Part A: Policy and Practice*, 2003; 37(9): 771–788.
17. Manley E. Estimating urban traffic patterns through probabilistic interconnectivity of road network junctions. *PLOS ONE*, 2015; 10(5): e0127095. doi: [10.1371/journal.pone.0127095](https://doi.org/10.1371/journal.pone.0127095) PMID: [26009884](https://pubmed.ncbi.nlm.nih.gov/26009884/)
18. Ródenas RG, Marín Á. Simultaneous estimation of the origin–destination matrices and the parameters of a nested logit model in a combined network equilibrium model. *European Journal of Operational Research*, 2009; 197(1): 320–331.
19. Caggiani L, Ottomanelli M, Sassanelli D. A fixed point approach to origin-destination matrices estimation using uncertain data and fuzzy programming on congested networks. *Transportation Research Part C: Emerging Technologies*, 2013; 28: 130–141.
20. Shao H, Lam WHK, Sumalee A, Chen A, Hazelton ML. Estimation of mean and covariance of peak hour origin-destination demands from day-to-day traffic counts. *Transportation Research Part B: Methodological*, 2014; 68: 52–75.
21. Van Zuylen HJ, Willumsen LG. The most likely trip matrix estimated from traffic counts. *Transportation Research Part B: Methodological*, 1980; 14(3): 281–293.
22. Bell MGH. The estimation of an origin destination matrix from traffic counts. *Transportation Science*, 1983; 17(2): 198–217.
23. Spiess H. A maximum likelihood model for estimating origin-destination matrices. *Transportation Research Part B: Methodological*, 1987; 21(5): 395–412.
24. Maher MJ. Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. *Transportation Research Part B: Methodological*, 1983; 17(6): 435–447.
25. Bell MGH. The estimation of origin-destination matrices by constrained generalized least squares. *Transportation Research Part B: Methodological*, 1991; 25(1): 13–22.
26. Sheu JB. A composite traffic flow modeling approach for incident-responsive network traffic assignment. *Physica A: Statistical Mechanics and its Applications*, 2006; 367: 461–478.
27. Beckmann M, McGuire CB, Winsten CB. *Studies in the Economics of Transportation*. New Haven: Yale University Press; 1956.
28. Fisk CS, Boyce DE. A note on trip matrix estimation from link traffic count data. *Transportation Research Part B: Methodological*, 1983; 17(3): 245–250.

29. Lam WHK, Huang HJ. A combined trip distribution and assignment model for multiple user classes. *Transportation Research Part B: Methodological*, 1992; 26(4): 275–287.
30. Fisk CS. On combining maximum entropy trip matrix with user optimal assignment. *Transportation Research Part B: Methodological*, 1988; 22(1): 69–73.
31. Fisk CS. Trip matrix estimation from link traffic counts: The congested network case. *Transportation Research Part B: Methodological*, 1989; 23(5): 331–336.
32. Han SJ. A route-based solution algorithm for dynamic user equilibrium assignments. *Transportation Research Part B: Methodological*, 2007; 41(10): 1094–1113.
33. Lu CC, Mahmassani HS, Zhou XS. A bi-criterion dynamic user equilibrium traffic assignment model and solution algorithm for evaluating dynamic road pricing strategies. *Transportation Research Part C: Emerging Technologies*, 2008; 16(4): 371–389.
34. Inoue SI, Maruyama T. Computational Experience on Advanced Algorithms for User Equilibrium Traffic Assignment Problem and Its Convergence Error. *Procedia-Social and Behavioral Sciences*, 2012; 43: 445–456.
35. Kumar A, Peeta S. Entropy weighted average method for the determination of a single representative path flow solution for the static user equilibrium traffic assignment problem. *Transportation Research Part B: Methodological*, 2015; 71: 213–229.
36. Zhang HL, Mahmassani HS, Lu CC. Dynamic pricing, heterogeneous users and perception error: Probit-based bi-criterion dynamic stochastic user equilibrium assignment. *Transportation Research Part C: Emerging Technologies*, 2013; 27: 189–204.
37. Daganzo CF, Sheffi Y. On stochastic models of traffic assignment. *Transportation Science*, 1977; 11(3): 253–274.
38. Sheffi Y, Powell WB. An algorithm for the equilibrium assignment problem with random link times. *Networks*, 1982; 12(2): 191–207.
39. Liu S, Fricker JD. Estimation of a trip table and the θ parameter in a stochastic network. *Transportation Research Part A: Policy and Practice*, 1996; 30(4): 287–305.
40. Meng Q, Lam WH, Yang L. General stochastic user equilibrium traffic assignment problem with link capacity constraints. *Journal of Advanced Transportation*, 2008; 42(4): 429–465.
41. Meng Q, Liu Z. Mathematical models and computational algorithms for probit-based asymmetric stochastic user equilibrium problem with elastic demand. *Transportmetrica*, 2012; 8(4): 261–290.
42. Lam WHK, Yin YF. An activity-based time-dependent traffic assignment model. *Transportation Research Part B: Methodological*, 2001; 35(6): 549–574.
43. Lam WHK, Li ZC, Huang HJ, Wong SC. Modeling time-dependent travel choice problems in road networks with multiple user classes and multiple parking facilities. *Transportation Research Part B: Methodological*, 2006; 40(5): 368–395.
44. Londono G, Lozano A. Dissuasive queues in the time dependent traffic assignment problem. *Procedia-Social and Behavioral Sciences*, 2014; 162: 378–387.
45. Bell MGH. Alternatives to dial's logit assignment algorithm. *Transportation Research Part B: Methodological*, 1995; 29(4): 287–295.
46. Conti PL, Giovanni LD, Naldi M. Blind maximum likelihood estimation of traffic matrices under long-range dependent traffic. *Computer Networks*, 2010; 54(15): 2626–2639.
47. Guo XL, Yang H, Liu TL. Bounding the inefficiency of logit-based stochastic user equilibrium. *European Journal of Operational Research*, 2010; 201(2): 463–469.
48. Akamatsu T. A dynamic traffic equilibrium assignment paradox. *Transportation Research Part B: Methodological*, 2000; 34(6): 515–531.
49. Boggs PT, Tolle JW. Sequential quadratic programming for large-scale nonlinear optimization. *Journal of Computational and Applied Mathematics*, 2000; 124(1–2): 123–137.
50. Bureau of Public Roads. Traffic assignment manual. U.S. Department of Commerce, Urban Planning Division, Washington, D. C., 1964.
51. Lu ZB, Rao WM, Wu YJ, Guo L, Xia JX. A Kalman filter approach to dynamic OD flow estimation for urban road networks using multi-sensor data. *Journal of Advanced Transportation*, 2015; 49(2): 210–227.
52. Liu HX, He XZ, He BS. Method of successive weighted averages (MSWA) and self-regulated averaging schemes for solving stochastic user equilibrium problem. *Networks and Spatial Economics*, 2009; 9(4): 485–503.
53. Wong SC. On the convergence of Bell's logit assignment formulation. *Transportation Research Part B: Methodological*, 1999; 33(8): 609–616.

54. Cascetta E, Russo F. Calibrating aggregate travel demand model with traffic counts: estimators and statistical performance. *Transportation*, 1997; 24(3): 271–293.
55. Ma X, Wu Y, and Wang Y. DRIVE Net: An E-Science of Transportation Platform for Data Sharing, Visualization, Modeling, and Analysis. *Transportation Research Record: Journal of the Transportation Research Board*. 2011; 2215: 37–49.