IJTE, Vol. 8/ No.2/ (30) Autumn 2020, pp. 165-183

Research Paper

Origin-Destination Matrix Estimation Using Socio-Economic Information and Traffic Counts on Uncongested Networks

Hadi Karimi [1,*], Seyed-Nader Shetab-Boushehri [2], Ramin Nasiri [3]

Received: 25.04.2019 Accepted: 27. 09. 2020

ABSTRACT

The travel demand matrix, also known as an origin-destination matrix (OD matrix), is essential in transportation planning. Given their nature and extent of operation, direct methods of estimating the matrix often impose unusually high costs in terms of both time and human resources. Thus, over the past three decades, numerous attempts have been made to propose indirect methods of estimating and updating the OD matrix. Using traffic counts to estimate the OD matrix is one of those indirect methods. However, because there are insufficient of traffic counts, indirect methods mostly lead to multiple OD matrices. One way to overcome this drawback is to use a previously estimated matrix from available data (called the old matrix) for new matrix estimation. Since uncongested networks rarely suffer from congestion, they have not been at the center of attention by researchers and transportation planners; thus, no old OD matrix is available for these networks. This study proposes a two-stage approach for estimating the OD matrix on uncongested networks. Firstly, an initial OD matrix is built using a travel distribution model (e.g., gravity model) together with local socio-economic information and available traffic counts across the network. Secondly, by considering budget constraints and using Bayesian inference, the optimum counting sensor locations are determined and by applying the collected information and the precision of the initial OD matrix is improved. To evaluate the proposed solution, the algorithm is then applied to the Sioux Falls network. The results prove the efficiency and precision of the approach.

Keywords: OD matrix estimation; sensor location problem; counting sensor; uncongested networks; Sioux Falls network

---

Corresponding author E-mail: h.karimidehnavi@alumni.iut.ac.ir

[1] Ph.D., Department of Industrial and Systems Engineering, Isfahan University of Technology, Isfahan, Iran

[2] Associate Professor, Department of Transportation Engineering, Isfahan University of Technology, Isfahan, Iran

[3] MSc. Department of Industrial and System Engineering, Isfahan University of Technology, Isfahan, Iran

1. Introduction

Due to increment in urban areas, especially in the cities in developing countries, the utilization of cars is increasing, resulting in many problems like congestion, pollution, noise cost, and more side effects (Hafezi et al. 2013). Also, the worldwide population growth and economic developments of the past decades have led to a substantially increased demand for transportation [Elyasi et al. 2018]. Identifying accurate OD matrices is one of the most important and challenging problems in transportation planning and traffic management [Kim et al. 2018]. The elements of this matrix contain the demand of traffic flow from origins indexed by rows, to destinations, indexed by columns [Michau et al. 2017]. This matrix is used as a critical source of input information to evaluate traffic management and policy measures [Antoniou et al. 2016]. OD matrix data can be collected through both direct and indirect approaches. Direct approaches are divided into three categories: i) observational studies; ii) questionnaire surveys; and iii) person interviews. Frequent data collection via direct methods is highly expensive [Hadavi and Shafahi, 2016]. Moreover, the interval between two consecutive surveys is extremely long. Therefore, it is quite difficult to derive reliable OD matrices for the intervening years [Ge and Fukuda, 2016]. This is why indirect approaches for updating and modifying the OD matrix have received more attention in recent years. Among the indirect methods, mathematic models enhanced with direct traffic counts have recently grown in popularity. Using traffic counts is an appropriate approach in terms of data availability, avoiding traffic interruption during data collection, low costs, and the accuracy of the collected data [Xie, Kockelman and Waller, 2011] making it a popular OD estimation choice for researchers during the last three decades. However, when applying indirect approaches, the number of unknown variables (i.e., the number of network OD pairs) often exceeds that of the known variables (i.e., traffic counts, even if observed on all network links), making the model a non-deterministic problem [Gentili and Mirchandani, 2011; Castillo et al. 2010] and resulting in multiple demand matrices. This means that the majority of OD flows will have infinite answers, all of which conform to traffic flow conservation law.

The issue has been studied extensively. To yield a reliable and unique OD matrix, some authors propose the application of auxiliary data such as an initial OD matrix estimated from previous studies. One can expect such matrices to be available for urban areas; however, the number of studies is limited across uncongested networks because of fewer congestion problems and limited research funds spent beyond urban boundaries. In most cases, the only available data for uncongested networks are those collected from a limited number of roads where permanent sensors are installed. Subsequently, there are only a few roads for which traffic counts can be collected. On the contrary, new technologies like cell phones and in-vehicle GPS have enabled accurate data collection and reliable information at low cost, while due to the low penetration rates of these devices among transportation networks' users, excessive costs and the probability of statistical bias for specific social groups (such as young people or business people), applying these data, face some restrictions [Bauer et al. 2018].

Since the accuracy of the estimated OD matrix depends on the quality and quantity of collected data, determining the number and the location of counting sensors is critically important [Larsson, Lundgren and Peterson, 2010]. Clearly, counting traffic volumes on all network links is the best option for reliable OD matrix estimation. However, given budget limitations, it may not be realistic to deploy sensors densely over the entire network for practical applications [Bao et al. 2016]. The problem for finding the best counting sensor location to achieve maximum possible data collection on a given network is known as the network sensor location problem [Viti et al. 2014].

In this study, first, an initial OD matrix is estimated by applying a travel distribution model whose coefficients are estimated using local socio-economic information and traffic counts obtained from existing counting sensors. Second, regarding budget limitations, a limited number of network links (other than those with existing counting sensors) are selected for traffic volume data collection. Finally, the data acquired in the second step are used to improve the accuracy of the initial matrix.

The remainder of this paper is organized as follows: Section 2 briefly surveys the literature dealing with the principal aspects of our model. Section 3 presents the methods for solving the OD matrix estimation problem by applying socio-economic information and traffic counts. Section 4 reports the results obtained from running the proposed models on the Sioux Falls network to analyze model behaviors. Finally, we conclude the paper with a discussion of avenues for future research in Section 5.

2. Literature Review

In this paper, firstly, an initial OD matrix is built using a mathematical travel distribution model and, secondly, the precision of this matrix is improved through traffic counts on the selected optimal links. In the following subsections, the relevant literature is reviewed.

2.1 Travel Demand Matrix Estimation Models

Conventional methods for OD matrix estimation have been categorized by [Doblas and Benitez, 2005] as follows:

2-1-1 Direct methods such as data surveys

Although direct methods have so far been extensively used across the world, a variety of limitations like high costs in terms of both time and human resources have challenged the use of these methods. This has encouraged researchers to try better solutions for OD matrix estimation. Similarly, [Stopher and Greaves, 2007] stated that questionnaire surveys might cause meaningful errors in matrix estimation because of small sample size, high rate of refusals, unreported travels, etc. Furthermore, [Cools, Moons and Wets, 2010] indicated that OD matrix estimation based on interview surveys is biased even when large samples are used and the resulting matrix estimation is often unreliable. Given these drawbacks, direct methods can be applied to make an initial OD matrix or to provide supplementary data for other OD matrix estimation methods.

2-1-2 Using travel distribution models

By applying both current inter-zonal travel values and growth factors, growth models can predict future inter-zonal travel interchange. [Wilson, 1967] introduced a model based on Newton’s gravity law according to which the travel value between two traffic zones is directly proportional to the relative attractiveness of the zones while inversely proportioned to the general travel cost. In his approach, the general cost of a travel can be expressed as travel time, distance, or the cost of fuel or ticket. Eq. (1) which presents one of the simplest forms of gravity models, is used for OD matrix estimation.

$$T_{ij} = \kappa P_i P_j \big/ d_{ij}^2 \tag{1}$$

where P_i and P_j are the populations of zones i and j respectively; d_{ij} is the distance between zone i and zone j; and κ is the model calibration constant. Assigning this matrix to the network yields Eq. (2).

$$V_a = \sum_{ij} p_{ij}^a \kappa P_i P_j \big/ (d_{ij})^2 = \kappa \sum_{ij} p_{ij}^a P_i P_j \big/ (d_{ij})^2 \tag{2}$$

where V_a is the traffic volume on link a and p_{ij}^a is the proportion of travels between zone i to zone j which use link a. In Eq. (2), the only unknown parameter is κ, which can be estimated using zone population data and traffic counts.

2-1-3 Using traffic counts for updating an old OD matrix.

[De Grange, González, and Bekhor, 2017] identified two traditional modeling approaches to estimate the OD matrix by relying on traffic counts: maximum entropy or minimum information and statistical techniques.

The first approach includes maximum entropy or minimum information models. It involves mathematical models that use maximum entropy from the basic OD matrix as the objective function in order to estimate a matrix having maximum consistency with the collected information (traffic counts). As a downside, these methods fail to take into account the uncertainty associated with traffic counts and the information in the basic matrix, which may include erroneous data and therefore impact the results [Bera and Rao, 2011]. Some researchers have used maximum entropy methods to estimate the OD matrix (e.g. [Tang and Zhang, 2013] and [Ryu et al. 2014]).

The second approach (i.e., statistical models) uses techniques like maximum likelihood, generalized least square, and Bayesian inference to estimate the OD matrix. [Castillo, Menéndez and Jiménez, 2008] classified statistical methods into classic statistical models and Bayesian statistical models. The most important benefit of this approach is considering the variability of input data. Traditional statistical models assume traffic flows as multivariate random variables belonging to a probability distribution such as Poisson, Gamma, multivariate normal, etc. Thus, the OD matrix estimation problem can be converted to a usual statistical parameter estimation problem. On the other hand, in Bayesian inference methods, previous understandings are combined with new observations to generate new insights [Bera and Rao, 2011]. In the Bayesian inference method, an initial OD matrix is calculated as a prior probability distribution ($\Pr(T)$) of the OD matrix while traffic counts from selected links are considered as another source of data with probability $L(V\,|\,T)$. Bayes' theorem, then, combines two sources of information so that the posterior probability of observing the OD matrix T conditional on the observed counts of certain links in the network $f(T\,|\,\hat{V})$ is obtained as follows:

$$f(T\,|\,\hat{V}) \approx L(\hat{V}\,|\,T).\Pr(T) \tag{3}$$

Similar to traditional statistical models, Bayesian inference methods assume traffic flow as multivariate random variables with the only difference that in the latter, parameters are random variables as well. Bayesian inference methods have been used for OD matrix estimation by many researchers (e.g. [Wei and Asakura, 2013] and [Perrakis et al. 2015]). Since this study aims to improve an initial matrix using Bayesian inference, this method is discussed in depth in the third section.

2.2 Network Sensor Location Models

OD matrix estimation involves optimally locating sensors on a transportation network to measure traffic counts [Ye and Wen 2017]. The network sensor location problem is a bi-level iterative problem. In the higher level, the optimal location of counting sensors is determined while, in the lower level, the best estimation of link flows is made using collected data from the counting sensors in the previous level. Accordingly, the sensor location problem for matrix estimation involves two problems: sensor location and estimation.

The first attempts to determine the number and optimal locations of counting sensors were made by [lam and lo, 1990]. Their method was developed by [Yim and Lam, 1998]. The authors introduced a heuristic algorithm in which sampling methods are used to prioritize the best network links for sensor location. By assuming that traffic flows between each OD pair select the shortest path, [Hodgson, 1990], they tried to reduce the complexity of sensor location problem and used the heuristic approach to solve the problem for a network comprised of 25 nodes. Based on the same assumption, two other studies [Berman, Bertsimas and Larson, 1995] developed Hodgson's model to determine the number and optimal location of counting sensors. Their goal was to choose the links with the highest collectible information. This approach was premised on the assumption of available traffic flow volume data for all network paths and subsequently for all OD pairs, which means that its application is only restricted to small networks regarding the criticality of volume data availability.

Using reliability theory, [Yang, Iida, and Sasaki, 1991] evaluated the estimated OD matrix based on maximal possible relative error (MPRE). Having precisely analyzed the characteristics of the MPRE index with regards to the optimum counting sensor locations, they formulated the MPRE using a simple quadratic programming problem and indicated that the MPRE is in effect the upper boundary of the actual relative error for the actual OD matrix. One of the most prominent investigations on the traffic sensor location problem was performed by [Yang and Zhou, 1998] based on previous research conducted by [Yang, Iida and Sasaki, 1991]. [Yang and Zhou 1998] proposed a systematic methodology for determining the number and optimal locations of counting sensors which finally led to defining four applicable rules including OD covering rule, maximal flow fraction rule, maximal flow intercepting rule, and link independence rule. Future studies like [Larsson, Lundgren and Peterson, 2010] and [Gentili and Mirchandani, 2011] proposed methodologies in line with Yang and Zhou's approach. Regarding the impossibility of full coverage of the all OD pairs in medium to large networks, and considering the fact that some OD pairs contain more information, added average data concept and sensor installation costs to previous models.

Moreover, [Castillo, Menéndez and Jiménez, 2008] proposed procedures based on Bayesian inference in order to select optimum links for installing counting sensors. They used Bayesian networks together with traffic counts collected from selected links to estimate the OD matrix in Sociedad, Spain. Their approach is appropriate for small networks but it is time-consuming in medium to large networks.

3. Methods

The underlying assumption in this study is that no initial OD matrix exists for the studied area, but socio-economic information is available. Using the available data, first, a primary model is constructed in order to estimate an initial OD matrix. Second, the proposed algorithm for determining the number and location of additional counting sensors is explained. The data obtained from these extra counting sensors are used to improve the accuracy of the initial OD matrix.

3.1 Initial OD Matrix Estimation using Socio-Economic Information

It is possible to estimate an initial OD matrix by constructing an appropriate mathematical model for inter-zonal travel distribution across a given uncongested network. A number of mathematical models have been developed to sketch travel distribution across the zones of a given area, among which the gravity model is one of the most well-known ones. The general form of the gravity model is as follows:

$$T_{ij} = A_i B_j O_i D_j e^{-\alpha C_{ij}} + \varepsilon_{ij} \qquad i \in I, j \in J, (i,j) \in W \tag{4}$$

where T_{ij} is the number of travels from zone i to zone j; O_i is the number of travels produced in zone i; D_j is the number of travels attracted at zone j; C_{ij} is travel cost from zone i to zone j; I denotes origins set, J denotes destinations set; W denotes OD pairs set; α is the model parameter; ε_{ij} is structural model error; and A_i and B_j are scale parameters for zone i and j, respectively, which act as moderator and are calculated through the Eq. (5).

$$\begin{aligned} A_i &= \Big(\sum_{j \in J} B_j D_j e^{-\alpha C_{ij}}\Big)^{-1} \\ B_j &= \Big(\sum_{i \in I} A_i O_i e^{-\alpha C_{ij}}\Big)^{-1} \end{aligned} \tag{5}$$

Considering p_i and e_i as population and employment of zone i respectively, and assuming simple mathematical models for travel production and attraction of zones i and j are as Eq. (6).

$$\begin{aligned} O_i &= \alpha_1 p_i + \alpha_2 e_i \\ D_j &= \alpha_3 p_j + \alpha_4 e_j \end{aligned} \tag{6}$$

The inter-zonal travel generation calculated from gravity models could be linked to the regional socio-economic information through Eq. (7).

$$\begin{aligned} O_i &= \alpha_1 p_i + \alpha_2 e_i \\ D_j &= \alpha_3 p_j + \alpha_4 e_j \\ A_i &= \Big(\sum_{j \in J} B_j D_j e^{-\alpha_5 C_{ij}}\Big)^{-1} \\ B_j &= \Big(\sum_{i \in I} A_i O_i e^{-\alpha_5 C_{ij}}\Big)^{-1} \\ T_{ij} &= A_i B_j O_i D_j e^{-\alpha_5 C_{ij}} + \varepsilon_{ij} \end{aligned} \tag{7}$$

If the values of α_1 to α_5 can be estimated by using auxiliary data, an initial OD matrix for the study area can be estimated using Eq. (7) along with applying socio-economic and employment information associated to traffic zones.

With regards to the gravity type travel distribution model, there are two important issues. Considering the theoretical base of the gravity model, it can be proved that the OD matrix obtained using this model is most likely to occur [Hong and Jung, 2016]. Second, solving Eq. (4) seems to be difficult since it is nonlinear. However, using an iterative algorithm (called AB) [Wilson, 1967; Kanafany, 1983], it can be solved. Algorithm AB to determine the values of A_i and B_j is as follows:

Step 0: set $B_j = 1 \quad j \in J$

Step 1: calculate A_i values using equation $A_i = (\sum_{j \in J} B_j D_j e^{-\alpha_5 C_{ij}})^{-1}, \quad i \in I$

Step 2: calculate B_j values using equation $B_j = (\sum_{i \in I} A_i O_i e^{-\alpha_5 C_{ij}})^{-1}, \quad j \in J$

Step 3: stop if the difference between each A_i and B_j in the two consecutive iterations is less than the threshold value; otherwise, go back to step 1.

Determining the travel cost between zones is also essential and is proportionate to the path traveled by the passenger. Assuming the passenger travels across the shortest possible path, C_{ij} in the Eq. (7) will be the cost of the shortest path from zone i to zone j. If all passengers in the network are assumed to travel through reasonable paths to get to their destinations, C_{ij} can be defined as the average travel cost from zone i to zone j and can be calculated as follows:

$$C_{ij} = \sum_{r \in R_{ij}} C_{ij,r} P_{ij,r} \quad (i,j) \in W \tag{8}$$

where $C_{ij,r}$ is the travel cost from zone i to zone j via the rational path r; $P_{ij,r}$ is the passengers proportion travel from zone i to zone j through rational path r; and R_{ij} is the sum of all rational paths between zone i to zone j.

In this study, passengers are assumed to travel through rational paths that are obtained from the Dial's traffic assignment [Ortuzar and Willumsen, 2011]. In this assignment, $P_{ij,r}$ value is calculated as follows:

$$P_{ij,r} = \frac{\exp(-\theta C_{ij,r})}{\sum_{r \in R_{ij}} (-\theta C_{ij,r})} \tag{9}$$

where θ is determined based on network characteristics. In this study, θ and other model parameters are estimated simultaneously.

The majority of previous research on the estimation of the OD matrix using traffic counts has applied the least squares approach, in which a bi-level model structure is used; the OD matrix is calculated in the higher level, and in the lower one, the estimated matrix is assigned to the transportation network.

The general structure of the bi-level approach is formulated as follows [Abrahamson, 1998]:

$$\begin{aligned} &\min F(T,V) = \gamma_1 F_1(T,\hat{T}) + \gamma_2 F_2(V,\hat{V}) \\ &s.t. \quad V = assign(T) \end{aligned} \tag{10}$$

In this model, the OD pair's flow is estimated by minimizing the two corresponding distance values: the distance between the estimated matrix T and initial matrix $\hat{T}$ (using distance function $F_1(T,\hat{T})$) and the distance between estimated volume V and links' observed volume $\hat{V}$ (by distance function $F_2(V,\hat{V})$). In the objective function, the weight coefficients γ_1 and γ_2 are assigned to distance functions whose values depend on the reliability of the available data. The constraint defined in Eq. (10) shows the overall relationship between the estimated links' volumes and the estimated OD matrix.

To estimate travel distribution model parameters in uncongested networks in this study, Eq. (11) is proposed based on Eq. (10) with two assumptions as follows:

1. There is no access to an initial OD matrix for the study area, thus γ_1 = 0 and γ_2 = 1;

2. Network users are assumed to travel through rational paths to get to their destinations.

$$
\begin{aligned}
&\min \ \sum_{a\in\bar{A}}(V_a-\hat{V}_a)^2 \\
&s.t. \quad P_{ij,r}=\frac{\exp(-\theta C_{ij,r})}{\sum_{r\in R_{ij}}(-\theta C_{ij,r})} && r\in R_{ij},(i,j)\in W \\
&C_{ij}=\sum_{r\in R_{ij}}C_{ij,r}P_{ij,r} && (i,j)\in W \\
&O_i=\alpha_1 p_i+\alpha_2 e_i && i\in I \\
&D_j=\alpha_3 p_j+\alpha_4 e_j && j\in J \qquad (11) \\
&A_i=(\sum_{j\in J}B_j D_j e^{-\alpha_5 C_{ij}})^{-1} && i\in I \\
&B_j=(\sum_{i\in I}A_i O_i e^{-\alpha_5 C_{ij}})^{-1} && j\in J \\
&T_{ij}=A_i B_j O_i D_j e^{-\alpha_5 C_{ij}} && i\in I, j\in J,(i,j)\in W \\
&V_a=\sum_{(i,j)\in W}\sum_{r\in R_{ij}}\delta^a_{ij,r}P_{ij,r}T_{ij} && a\in\bar{A} \\
&\alpha_1,\alpha_1,\alpha_1,\alpha_1,\alpha_1\geq 0
\end{aligned}
$$

where V_a is volume of link a resulting from traffic assignment; $\hat{V}_a$ volume of link a obtained from counting sensors ($a\in\bar{A}$), $\delta^a_{ij,r}$ is binary variable which take 1 if the link a located in path r from zone i to zone j otherwise takes 0, $\bar{A}$ is the set of network links with sensors, $\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5$ are model parameters, and θ is the Dial's traffic assignment model's parameter obtained by solving Eq. (11). Other parameters are defined as before.

3.2 Improving the Accuracy of the Initial Matrix

In section 3.1, we explained the procedure of estimating the initial OD matrix ($\hat{T}$) using only socio-economic information and traffic counts (using existing traffic sensors). In this section, the approach for improving the quality of the initial OD matrix is explained. Among many approaches for matrix quality improvement, one is selecting extra optimum links to use their traffic count data (i.e., extra links other than those with existing sensors). In the current study, a two-step methodology is applied to improve the initial OD matrix ($\hat{T}$) using traffic counts. In the first step, the optimal links are selected and, then, using their traffic counts the initial OD matrix ($\hat{T}$) will be updated and improved.

As mentioned in the literature review section, a variety of models and methodologies have been used for optimal link selection for sensor location and OD matrix improvement; one of the most widely used methods is the Bayesian inference method. Using the methodology proposed by [Castillo, Menéndez and Jiménez, 2008], extra links are selected and their traffic counts are used. In this methodology, with some assumptions, the relationships between variance-covariance matrices and the OD flows are calculated by Eqs. (12)-(15) [Castillo, Menéndez, and Jiménez, 2008]:

$$\Sigma_{TT}=\sigma_U^2 KK'+D_\eta \qquad (12)$$

$$\Sigma_{TV}=\Sigma_{TT}\beta^t \qquad (13)$$

$$\Sigma_{VT}=\Sigma_{TV} \qquad (14)$$

$$\Sigma_{VV}=\beta\Sigma_{TT}\beta^t+D_\varepsilon \qquad (15)$$

where U is the mean flow of the entire network; σ_U^2 is the variance of variable U; K is the relative weights of the OD pairs; $\boldsymbol{\beta}$ is the assignment matrix with elements β_a^w and denotes the portion of travels between OD pair w which pass through the counted link a; D_η and D_ε are the variance-covariance matrices of independent random variables η and ε respectively, which are assumed to be diagonal, $\sum_{TT}$ shows the variance-covariance matrix of OD pair flows; $\sum_{TV}$ denotes the variance-covariance matrix of OD pair flows and link flows; and $\sum_{VV}$ shows the variance-covariance matrix of link flows.

Additionally, provided that some of the variables are observed, the Bayesian approach can be used to update the mean and variance-covariance matrix of OD flows in the network through Eqs. (16)- (17).

$$\mu_{T|\hat{V}=\hat{v}} = \mu_T + \Sigma_{T\hat{V}}\Sigma_{\hat{V}\hat{V}}^{-1}(\hat{v} - \hat{V}(T)) \quad (16)$$

$$\Sigma_{T|\hat{V}=\hat{v}} = \Sigma_{TT} - \Sigma_{T\hat{V}}\Sigma_{\hat{V}\hat{V}}^{-1}\Sigma_{\hat{V}T} \quad (17)$$

where, $\hat{v}$ is the vector of the traffic counts passing through observed network links and $\hat{V}(T)$ is the vector of volumes obtained through an assignment for the same links.

Using the above equations, [Castillo, Menéndez and Jiménez, 2008] introduced algorithms for locating optimal links and matrix estimation. [Karimi, Ebrahimi, and Shetab Bousshehri, 2017] put forward suggestions for improving the efficiency of these algorithms particularly for large scale networks.

3-2-1 Selecting extra optimum links

If the links with existing counting sensors are named as set $\bar{A}$, new links must be selected and added to set $\bar{A}$ to be used for improving the initial OD matrix using their observed traffic counts. Regarding budget constraints, only h sensors could be added to existing ones. For network sensor location, the algorithm proposed by [Castillo, Menéndez and Jiménez, 2008] is used with some modifications as follows:

In the original algorithm, first, the correlation coefficient matrix (ρ) with components ρ_{TV} is calculated using the variance-covariance matrix formulated as Eqs. (12)- (15). Second, the component corresponding to the biggest correlation coefficient in the correlation coefficient matrix (ρ) is selected as the optimal link for counting. Third, the selected link is added to the existing links and the correlation coefficient matrix (ρ) is updated using Eq. (17). If the number of selected links is less than h, the process is repeated to choose the next link.

In Castillo's approach for sensor location, the correlation OD pair flow matrix and link volumes are used and the link with the largest matrix component is picked as the candidate link for sensor location. This link provides substantial amounts of information for only one OD pair. While at the present study, a new index called "sum of the differences of the corresponding OD pair variances in the network within the running step and the previous step" is proposed as an appropriate indicator for choosing the optimal link in a given step. Accordingly, in the Castillo's first improvement step, instead of the link with the largest matrix component, the link that yields the highest value for the proposed index (i.e. sum of the differences of the corresponding OD pair variances within the running step and the previous step) is selected as the optimum link for sensor location.

3-2-2 Improvement step

Assuming to put the set of links $\bar{A}$ together with the selected links obtained from the first step in a new set called $\bar{A}'$, considering the initial OD matrix ($\hat{T}$) as the base matrix, volume counts from link set $\bar{A}'$ as $\hat{V}'$, then the initial OD matrix will by improved using Castillo's algorithm (base algorithm 1) through Eqs. (12)- (17). Figure 1 shows a flowchart of the proposed algorithm.

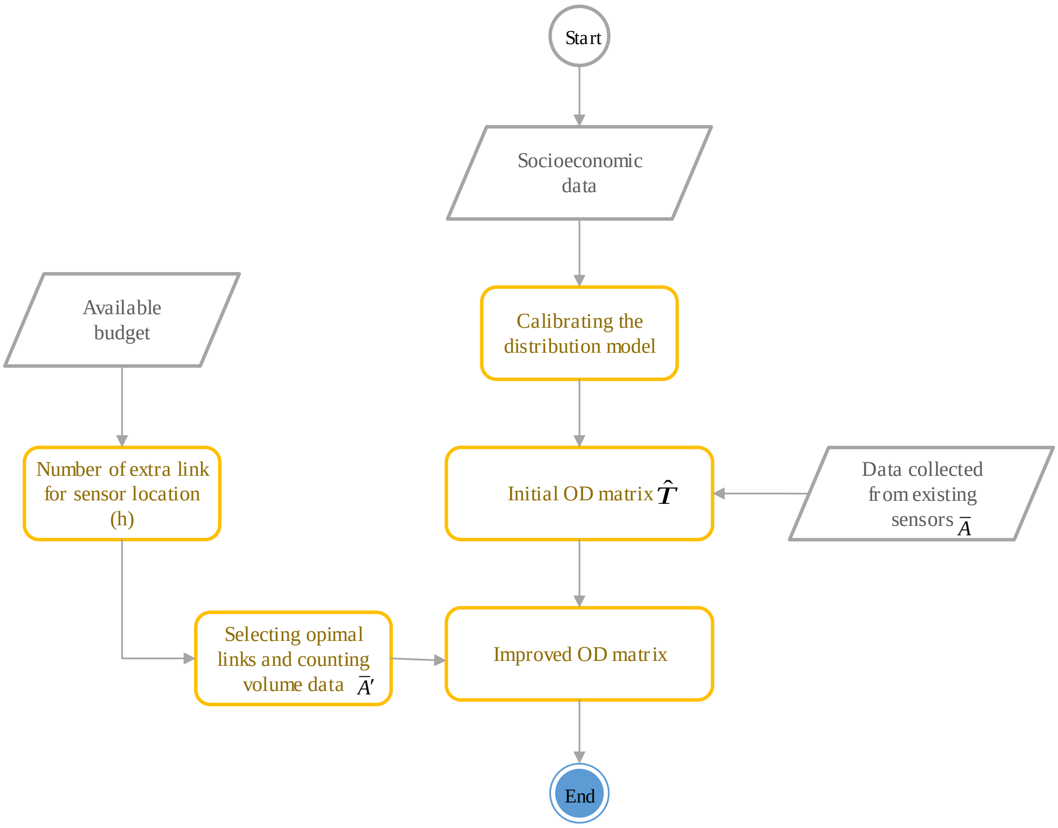


Figure 1. The proposed algorithm for estimating and improving the OD matrix on uncongested networks

4. Results

In this section, in order to analyze and evaluate the proposed algorithm, it is run on the Sioux Falls network and the results are reported.

4.1 Sioux Falls Transportation Network

In this research, a medium-size Sioux Falls network with 24 nodes, 76 links, and 576 OD pairs is applied which is a typical network widely used in transportation studies. The required information for this network, including link attributes and OD matrix, was extracted from data reported by [Leblank, 1975]. It is assumed that nodes 1, 2, 4, 5, 10, 11, 13, 14, 15, 19, 20, 21, 22, 24 are origin destination nodes. Figure 2 shows the Sioux Falls street network of the largest city in South Dakota with highlighted nodes representing origin destination ones.

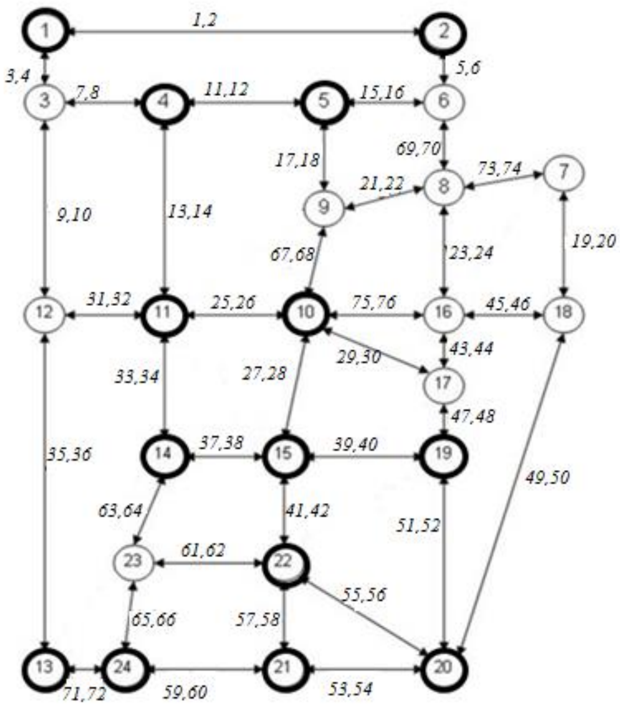


Figure 2. Sioux Falls transportation network

4.2 Transportation Simulation Program

The results of the proposed model are validated using a simulator program whose inputs include the OD matrix built upon population data and zonal employment vectors within the study area as well as traffic counts resulting from assigning this matrix on Sioux Falls network. In each run, the simulator randomly creates one population and one employment vector and then, generates the target OD matrix using a mathematical model together with population as well as employment data and, finally, adds a random value as error component to the matrix. Afterwards, by assuming this matrix as the actual OD matrix and assigning it to Sioux Falls network, the simulator calculates volumes on the links and represents them as if they have been collected from network sensors. Thus, in each run, the simulator outputs are population vector, employment vector, actual OD matrix, and traffic counts.

4.3 Estimating the Initial OD Matrix

In this section, we describe how the parameters of Eq. (11) are estimated and subsequently evaluate the results.

Eq. (11), which is a methodical programming model is solved with the aim of determining parameters $\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5$ and θ in the travel distribution gravity model while assuming passengers travel across rational paths to reach their destinations. The model is solved using the Quasi Newton module in Gauss.

By determining model parameters, the travel distribution model is calibrated and the OD matrix can be estimated (regenerated) using available socio-economic information (p and e vectors are obtained from the simulator program).

Using the OD matrix generated by the simulator program which is assumed as actual OD (T) the estimated OD matrix calculated by Eq. (11) ($\hat{T}$) can be evaluated. To evaluate $\hat{T}$ and determine the accuracy of the proposed model, the regression equation $\hat{T}_{ij} = a + bT_{ij}$ is estimated using the OLS method and the correlation of two the matrices is then determined based on the statistical results summarized in Figure 3.

As can be seen, R^2 is approaching 1 which demonstrates the high accuracy of the model.

Figure 4 depicts the actual and estimated OD flows. The results indicate that the proposed method enjoys high accuracy in estimating the real OD matrix in the Sioux Falls network. However, with simply one iteration, model reliability cannot be approved. Therefore, the simulator was run on the studied network 50 times and the results were evaluated in order to obtain consistent results. Figure 5 depicts the values of T and $\hat{T}$ for these 50 times iterations. Moreover, R^2 values are shown in Figure 6.

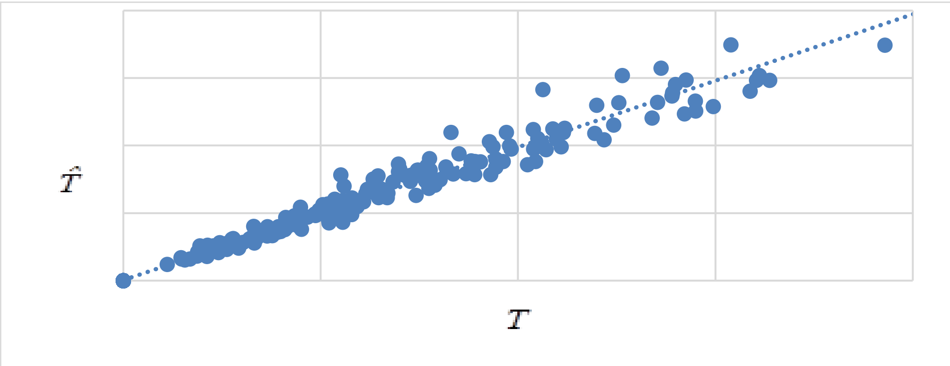


Figure 3. Actual OD matrix versus Estimated OD matrix for the Sioux Falls Transportation Network

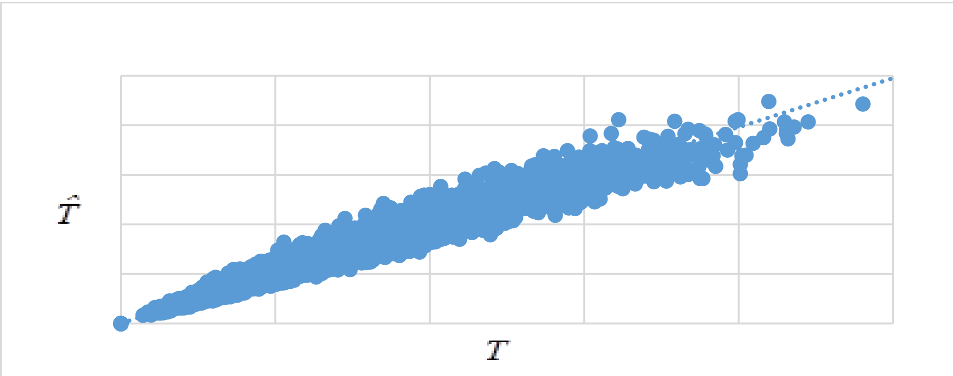


Figure 4. Actual OD matrix versus Estimated OD matrix for 50 iterations

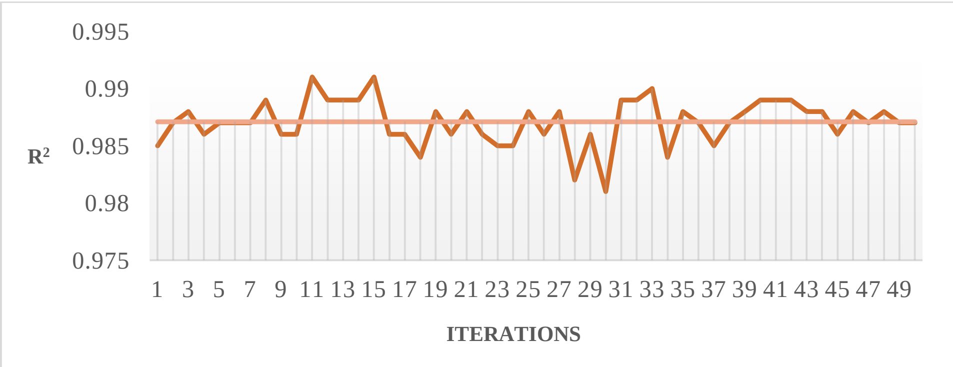
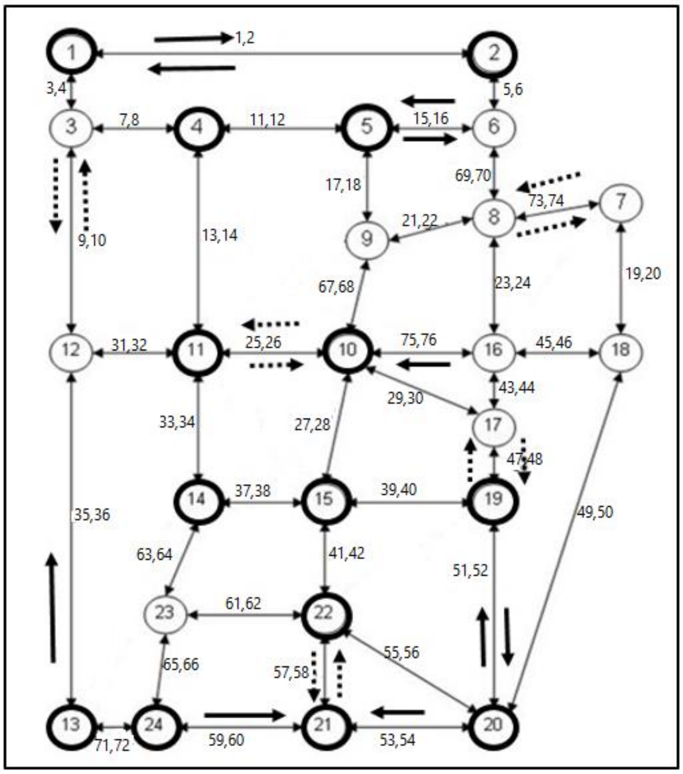


Figure 5. R^2 **values after 50 iterations of Eq. (11)**

4.4 Initial OD Matrix Improvement

There are two assumptions here: (1) the available sensors are installed on links 9, 10, 25, 26, 47, 48, 57, 58, 73, and 74 (figure 7); and (2) due to budget constraints only 10 more links can be selected for sensor installation. The actual OD matrix (T), and traffic counts on observed links ($\hat{V}$) can be determined using the simulation program, population vector (P) and employment vector (e). Table 1 summarizes the population and employment vector data for each corresponding zone. The actual OD matrix is also listed in Table 2. Subsequently, using the proposed model, the initial demand matrix ($\hat{T}$) is estimated whose results are summarized in Table 3.

To identify the difference between the actual OD matrix and the initial OD matrix after adding the optimal links for matrix improvement, we used $e^2 = \sum_{(i,j)\in W} (T_{ij} - \hat{T}_{ij})^2$ index. Before adding new links, this index was 88.73 for the example being investigated. Using the [Castillo, Menéndez and Jiménez, 2008] methodology and making the suggested improvements, 10 additional links are selected, which are displayed in Figure 7. As a result, the final improved demand matrix ($\hat{T}'$) was estimated based on the initial matrix ($\hat{T}$), which is summarized in Table 4. Calculating the e^2 index, the difference between the actual OD matrix (T) and the improved matrix ($\hat{T}'$), after incorporating the data for 10 links yielded 58.872 which shows a decrease compared to 88.73 before adding the new links, indicating that the information acquired from the selected links have led to a considerable improvement in the accuracy of the initial matrix.



The location of existing counting sensors Extra optimum links

Figure 6. Sioux Falls network with preinstalled sensors and optimal links selected in the first improvement step.

Table 1. Zonal population and employment data for the study area

Zone number	P	e
1	27.01	5.3
2	29.1	9.2
4	15.6	6.5
5	25.3	6.2
10	10.9	9.3
11	24.5	9.1
12	17.0	9.4
14	18.4	7.9
15	28.9	5.9
19	20.9	5.4
20	29.8	8.9
21	19.5	7.0
22	11.2	6.8
24	28.9	5.4

Table 2. The actual travel **demand matrix**

OD	1	2	4	5	10	11	13	14	15	19	20	21	22	24
1	0	11/84	3/07	2/93	0/57	0/76	1/46	0/36	0/46	0/31	1/17	1/23	0/78	3/07
2	8/25	0	2/20	6/28	0/52	2/09	0/62	0/78	2/97	1/44	0/74	0/75	1/01	2/49
4	1/42	1/98	0	2/84	1/07	2/49	1/07	0/71	0/58	0/15	0/77	0/34	0/45	1/29
5	1/51	5/53	4/14	0	2/39	4/61	0/99	1/41	0/91	0/88	0/77	0/58	0/58	1/57
10	0/47	0/95	0/53	2/00	0	2/43	0/68	1/56	1/65	0/97	0/97	0/85	0/22	0/44
11	1/12	0/59	3/38	2/79	1/61	0	2/82	1/68	1/01	0/56	1/57	1/12	1/30	3/44
13	0/43	0/39	0/46	1/21	2/01	4/36	0	0/97	1/91	1/23	1/78	1/30	0/90	1/64
14	0/57	0/78	0/54	1/12	1/34	2/14	3/03	0	2/82	1/15	3/17	2/30	0/58	1/73
15	0/54	3/43	0/53	1/68	1/72	2/23	3/68	3/7	0	1/84	0/7	1/00	1/85	2/29
19	0/70	1/23	0/25	0/96	0/85	0/69	0/52	1/90	3/85	0	4/21	1/24	1/02	0/90
20	1/49	1/35	0/77	0/98	1/50	0/56	1/75	2/87	4/66	2/02	0	4/83	4/32	2/28
21	0/66	0/40	0/17	0/69	1/06	1/50	1/50	1/73	2/93	2/32	4/09	0	0/94	0/96
22	0/78	0/70	0/48	0/47	0/33	0/73	1/06	1/08	0/99	0/62	1/60	1/50	0	1/29
24	4/39	2/17	2/08	2/38	0/99	2/42	2/28	1/45	0/97	1/31	5/03	1/50	2/87	0

Table 3. The initial demand matrix

OD	1	2	4	5	10	11	13	14	15	19	20	21	22	24
1	0	7/81	2/42	3/11	0/73	2/15	0/98	0/49	0/83	0/62	1/72	0/87	0/98	2/98
2	7/24	0	2/17	4/58	1/25	1/75	0/72	0/95	2/65	1/66	1/25	0/82	0/75	2/27
4	2/03	1/96	0	3/09	0/83	2/12	1/02	0/58	0/68	0/38	0/62	0/33	0/39	1/18

OD	1	2	4	5	10	11	13	14	15	19	20	21	22	24
5	2/92	4/64	3/46	0	1/87	2/64	1/27	1/26	1/49	0/85	0/95	0/70	0/45	1/70
10	0/47	0/86	0/63	1/27	0	1/64	0/75	0/98	1/17	0/62	1/04	0/74	0/30	0/55
11	1/81	1/59	2/13	2/36	2/15	0	2/84	1/60	1/83	0/93	1/72	1/23	0/97	2/58
13	0/74	0/58	0/91	1/01	0/88	2/54	0	1/63	1/94	0/95	1/73	1/21	1/01	1/58
14	0/40	0/83	0/56	1/09	1/24	1/55	1/76	0	2/85	1/42	2/54	1/80	0/74	1/16
15	0/79	2/74	0/78	1/51	1/75	2/08	2/47	3/36	0	3/84	3/31	2/36	1/01	1/57
19	0/58	1/67	0/42	0/85	0/9	1/03	1/19	1/64	3/75	0	3/24	2/22	0/99	1/55
20	1/54	1/21	0/66	0/90	1/46	1/83	2/07	2/81	3/09	3/10	0	3/93	2/37	3/71
21	0/7	0/76	0/34	0/64	0/99	1/25	1/38	1/90	2/11	2/04	3/76	0	1/15	1/79
22	0/72	0/59	0/34	0/35	0/35	0/85	0/99	0/67	0/78	0/78	1/96	0/99	0	1/74

Table 4. The improved travel **demand matrix**

OD	1	2	4	5	10	11	13	14	15	19	20	21	22	4
1	0	10/39	2/42	3/11	0/72	2/09	0/97	0/49	0/78	0/56	1/63	0/86	0/96	2/76
2	8/62	0	2/38	5/83	1/28	1/97	0/75	0/95	2/29	1/06	1/05	0/74	0/75	2/28
4	2/03	2/02	0	3/09	0/81	2/12	1/02	0/57	0/66	0/35	0/61	0/32	0/38	1/16
5	2/92	4/61	3/46	0	1/87	3/08	1/37	1/26	1/40	0/70	0/85	0/63	0/46	1/66
10	0/49	0/86	0/66	1/27	0	2/07	0/80	0/98	1/13	0/57	1/04	0/74	0/30	0/56
11	1/83	1/60	2/13	2/21	1/82	0	2/84	1/52	1/58	0/81	1/68	1/19	0/97	2/58
13	0/75	0/58	0/91	0/98	0/85	2/54	0	1/63	1/94	0/90	1/73	1/21	1/01	1/58
14	0/41	0/83	0/58	1/09	1/24	1/72	1/76	0	2/85	1/26	2/54	1/80	0/74	1/16
15	0/8	2/77	0/81	1/56	1/76	2/58	2/47	3/36	0	1/53	2/21	1/96	0/93	1/38
19	0/58	1/27	0/43	0/82	0/92	1/15	1/21	1/72	4/53	0	3/22	2/22	0/99	1/55
20	1/67	1/08	0/68	0/88	1/52	2/02	2/15	3/04	3/21	2/59	0	4/38	2/37	3/71
21	0/78	0/72	0/34	0/62	0/99	1/32	1/38	1/90	2/22	2/04	3/76	0	1/15	1/79
22	0/75	0/63	0/34	0/35	0/34	0/87	0/99	0/68	0/78	0/76	1/93	1/00	0	1/74
24	3/39	2/96	1/48	1/94	0/84	3/40	2/07	1/42	1/65	1/53	3/98	2/11	2/32	0

5. Conclusion

Throughout this paper we, like the vast majority of authors on the subject, have considered O-D matrix estimation for uncongested networks. The most common way to estimate this matrix is to use an initial (or old) OD matrix. But in uncongested networks, usually, no initial OD matrix is available. In this study, to tackle this problem, firstly, by applying local socio-economic information and available traffic counts an initial OD matrix is made. Secondly, by applying Bayesian inference and considering budget constraints, the optimum counting sensor locations are determined and by applying the collected information, the precision of the initial OD matrix is improved.

In this way, a two-step methodology was proposed. In the first step, a mathematical travel distribution model was calibrated using both zonal socio-economic and existing traffic counts to obtain an initial OD matrix. In the second step, the estimated initial matrix was improved by using traffic counts collected from additional links on the network. For this purpose, after determining the number of additional links, regarding budget limitations, these links are identified through a statistical approach based on the Bayesian inference method and finally using new links' traffic count data the initial matrix is improved. The results on the Sioux

Falls network show that the proposed methodology is low-cost and yields precise and appropriate OD matrix.

Future research steps should include research along with three main directions. The first one is OD matrix estimation by combining information collected from various types of sensors (both active and passive). The second direction of further research could focus on the estimation of OD pairs flow and considering sensor costs (device, installation, and maintenance). The third direction for further research involves examining large-scale networks to understand the efficiency of the proposed methodology better.

6. References

- Abrahamsson, T. (1998) "Estimation of origin-destination matrices using traffic counts-a literature survey".

- Antoniou, C., Barceló, J., Breen, M., Bullejos, M., Casas, J., Cipriani, E., ... & Montero, L. (2016) "Towards a generic benchmarking platform for origin–destination flows estimation/updating algorithms: Design, demonstration and validation", Transportation Research Part C: Emerging Technologies, No. 66, pp. 79-98. https://doi.org/10.1016/j.trc.2015.08.009

- Bao, X., Li, H., Qin, L., Xu, D., Ran, B., & Rong, J. (2016) "Sensor location problem optimization for traffic network with different spatial distributions of traffic information", Sensors, Vol. 16, No. 11, pp. 1790. https://doi.org/10.3390/s16111790

- Bauer, D., Richter, G., Asamer, J., Heilmann, B., Lenz, G. and Kölbl, R. (2018) "Quasi-Dynamic Estimation of OD Flows From Traffic Counts Without Prior OD Matrix", IEEE Transactions on Intelligent Transportation Systems, Vol. 19, No. 6, pp. 2025-2034. https://doi.org/10.1109/TITS.2017.2741528

- Bera, S. and Rao, K.V. (2011) "Estimation of origin-destination matrix from traffic counts: the state of the art".

- Berman, O., Bertsimas, D. and Larson, R.C. (1995) "Locating discretionary service facilities, II: maximizing market size, minimizing inconvenience", Operations Research, Vol. 43, No. 4, pp. 623-63. https://doi.org/10.1287/opre.43.4.623

- Castillo, E., Gallego, I., Menéndez, J.M. and Rivas, A. (2010) "Optimal use of plate-scanning resources for route flow estimation in traffic networks", IEEE Transactions on Intelligent Transportation Systems, Vol. 11, No. 2, pp. 380-391. https://doi.org/10.1109/TITS.2010.2042958

- Castillo, E., Menéndez, J.M. and Jiménez, P. (2008) "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. https://doi.org/10.1016/j.trb.2007.09.004

- Cools, M., Moons, E. and Wets, G. (2010) "Assessing the quality of origin-destination matrices derived from activity travel surveys: Results from a Monte Carlo experiment", Transportation Research Record: Journal of the Transportation Research Board, No. 2183, pp. 49-59. https://doi.org/10.3141/2183-06

- De Grange, L., González, F. and Bekhor, S. (2017) "Path flow and trip matrix estimation using link flow density. Networks and Spatial Economics", Vol. 17, No. 1, pp. 173-195. https://doi.org/10.1007/s11067-016-9322-1

- Doblas, J. and Benitez, F.G. (2005) "An approach to estimating and updating origin–destination matrices based upon traffic counts preserving the prior structure of a survey matrix", Transportation Research Part B: Methodological, Vol. 39, No. 7, pp. 565-591. https://doi.org/10.1016/j.trb.2004.06.006

- Elyasi, M., Faezi, S. F., Haghsheno-Sabet, M., & Mazaheri, M. (2018) "An ANP-based Model for Location of Fixed Speed Cameras", International Journal of Transportation Engineering, Vol. 6, No. 1, pp. 17-34. https://dx.doi.org/10.22119/ijte.2017.52975

- Ge, Q., & Fukuda, D. (2016) "Updating origin–destination matrices with aggregated data of GPS traces", Transportation Research Part C: Emerging Technologies, No. 69, pp. 291-312. https://doi.org/10.1016/j.trc.2016.06.002

- Gentili, M. and Mirchandani, P. (2011) "Survey of models to locate sensors to estimate traffic flows", Transportation Research Record: Journal of the Transportation Research Board, No. 2243, pp. 108-116. https://doi.org/10.3141/2243-13

- Hadavi, M., & Shafahi, Y. (2016) "Vehicle identification sensor models for origin–destination estimation", Transportation Research Part B: Methodological, No. 89, pp. 82-106. https://doi.org/10.1016/j.trb.2016.03.011

- Hafezi, Mohammad Hesam, et al. "A Novel Method for Travel System Patterns." International Journal of Transportation Engineering 1.2 (2013): 93-100. https://dx.doi.org/10.22119/ijte.2013.3235

- Hodgson, M. J. (1990) "A Flow-capturing location-allocation model", Geographical Analysis, Vol. 22, No. 3, pp. 270-279. https://doi.org/10.1111/j.1538-4632.1990.tb00210.x

- Hong, I. and Jung, W.S. (2016) "Application of gravity model on the Korean urban bus network", Physica A: Statistical Mechanics and its Applications, No. 462, pp. 48-55. https://doi.org/10.1016/j.physa.2016.06.055

- Kanafany, A. (1983). Transportation Demand Analysis. New York, NY: Mc Graw-Hil

- Karimi, H., Ebrahimi, A., Shetab Bousshehri, S. (2017) "Optimum Counting Location to update Origin-Destination Matrix Using Bayesian inference (Case Study: City of Isfahan) ", Quarterly Journal of Transportation Engineering, Vol. 8, No. 3, pp. 451-471, In Persian.

- Kim, H., Nam, D., Suh, W. and Cheon, S.H. (2018) "Origin-destination trip table estimation based on subarea network OD flow and vehicle trajectory data", Transportation Planning and Technology, Vol. 41, No. 3, pp. 265-285. https://doi.org/10.1080/03081060.2018.1435437

- Lam, W.H.K. and Lo, H.P. (1990) "Accuracy of OD estimates from traffic counts", Traffic engineering & control, Vol. 31, No. 6.

- Larsson, T., Lundgren, J.T. and Peterson, A. (2010) "Allocation of Link Flow Detectors for Origin-Destination Matrix Estimation—A Comparative Study", Computer-Aided Civil and Infrastructure Engineering, Vol. 52, No. 2, pp. 116-131. https://doi.org/10.1111/j.1467-8667.2009.00625.x

- Leblanc, L.J. (1975) "An algorithm for the discrete network design problem", Transportation Science, Vol. 9, No. 3, pp. 183-199. https://doi.org/10.1287/trsc.9.3.183

- Michau, G., Pustelnik, N., Borgnat, P., Abry, P., Nantes, A., Bhaskar, A., & Chung, E. (2017) "A primal-dual algorithm for link dependent origin destination matrix estimation", IEEE Transactions on Signal and Information Processing over Networks, Vol. 3, No. 1, pp. 104-113. https://doi.org/10.1109/TSIPN.2016.2623094

- Ortúzar, J., & Willumsen, L. G. (2011). Trip distribution modelling. Modelling transport (4th ed., pp. 175–206).

- Perrakis, K., Karlis, D., Cools, M. and Janssens, D. (2015) "Bayesian inference for transportation origin–destination matrices: the Poisson–inverse Gaussian and other Poisson mixtures", Journal of the Royal Statistical Society: Series A (Statistics in Society), Vol. 178, No. 1, pp. 271-296. https://doi.org/10.1111/rssa.12057

- Ryu, S., Chen, A., Zhang, H.M. and Recker, W. (2014) "Path flow estimator for planning applications in small communities", Transportation Research Part A: Policy and Practice, No. 69, pp. 212-242. https://doi.org/10.1016/j.tra.2014.08.019

- Stopher, P.R. and Greaves, S.P. (2007) "Household travel surveys: Where are we going? ", Transportation Research Part A: Policy and Practice, Vol. 41, No. 5, pp. 367-381. https://doi.org/10.1016/j.tra.2006.09.005

- Tang, S. and Zhang, H. (2013) "Primal-dual heuristic for path flow estimation in medium to large networks", Transportation Research Record: Journal of the Transportation Research Board, No. 2333, pp. 91-99. https://doi.org/10.3141/2333-11

- Viti, F., Rinaldi, M., Corman, F. and Tampère, C.M. (2014) "Assessing partial observability in network sensor location problems", Transportation research part B: methodological, Vol. 70, pp. 65-89. https://doi.org/10.1016/j.trb.2014.08.002

- Wei, C. and Asakura, Y. (2013) "A Bayesian approach to traffic estimation in stochastic user equilibrium etworks", Transportation Research Part C: Emerging Technologies, No. 36, pp. 446-459. https://doi.org/10.1016/j.sbspro.2013.05.032

- Wilson, A. G. (1967) "A statistical theory of spatial distribution models", Transportation research, Vol. 1, No. 3, pp. 253-269.

- Xie, C., Kockelman, K.M. and Waller, S.T. (2011) "A maximum entropy-least squares estimator for elastic origin–destination trip matrix estimation", Transportation Research Part B: Methodological, Vol. 45, No. 9, pp. 1465-1482. https://doi.org/10.1016/j.sbspro.2011.04.514

- Yang, H., Iida, Y. and Sasaki, T. (1991) "An analysis of the reliability of an origin-destination trip matrix estimated from traffic counts", Transportation Research Part B: Methodological, Vol. 25, No. 5, pp. 351-363. https://doi.org/10.1016/0191-2615(91)90028-H

- Yang, H. and Zhou, J. (1998) "Optimal traffic counting locations for origin–destination matrix estimation", Transportation Research Part B: Methodological, Vol. 32, No. 2, pp. 109-126. https://doi.org/10.1016/S0191-2615(97)00016-7

- Ye, P., & Wen, D. (2017) "Optimal traffic sensor location for origin–destination estimation using a compressed sensing framework", IEEE Transactions on Intelligent Transportation Systems, Vol. 18, No. 7, pp. 1857-1866. https://doi.org/10.1109/TITS.2016.2614828

- Yim, P.K. and Lam, W.H. (1998) "Evaluation of count location selection methods for estimation of OD matrices", Journal of transportation engineering, Vol. 124, No. 4, pp. 376-383. https://doi.org/10.1061/(ASCE)0733-947X(1998)124:4(376)