Journal of Al-Azhar University Engineering Sector

Vol. 18, No. 66, January 2023, 57 - 73

# ORIGIN DESTINATION MATRIX ESTIMATION BASED ON TRAFFIC COUNTS USING FMINCON FUNCTION IN MATLAB

**Georgette Syffen[1,*], Ahmed Khalil[1], Ibrahim Ramadan[1]**

[1] Civil Engineering Department, Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt.
* Correspondence: jorget.lamey18@feng.bu.edu.eg; georgettesyffen@gmail.com

**Received: 17 Oct. 2022** **Accepted: 12 Nov. 2022**

## ABSTRACT

Most methods of traffic planning study require origin-destination (OD) matrices, which describe the trip demands between the origin and destination nodes in the network. This work is about the OD matrix estimation problem, which means estimating OD matrices from observed link flows. This study aims to provide an easy-to-use and cost-effective alternative software that researchers and planners in developing countries can access without restriction. The Fmincon function in Matlab software was used to create specialized code for estimating the OD matrix in congested networks using three different optimization approaches at the upper level: Generalized least square (GLS), Ordinary least square (OLS), and Maximum Entropy (ME) aggregated with user equilibrium assignment in the lower level. To test the performance of the provided models, a corridor hypothetical network was created, and two cases of reliability factors were suggested, followed by a detailed statistical analysis. The three proposed approaches produce adequate results, the improvement rate in error between observed and estimated traffic counts in the two cases was (52.11% and 52.01%) according to ME approach, (75.39% and 75.53%) according to OLS approach, and (75.56% and 75.53%) according to GLS approach. Also, to guarantee performance, the results were compared to PTV Visum's TFlow fuzzy (TFF) and least square (LS) results. The results show that both the GLS and OLS methods can produce extremely accurate results, with the mean absolute error value approximating the results of the LS method by at least 98%, and the error rate is reduced by approximately 11% over the TFF method.

**Keywords**: OD matrix, Bi-level optimization, Fmincon function, Traffic counts, Equilibrium trip assignment.

# تقدير مصفوفة المصدر – الهدف بناءً على تعدادات المرور باستخدام اقتران Fmincon في الماتلاب

جورجيت سيفين[1,*]، أحمد خليل[1]، ابراهيم رمضان[1]

[1] قسم الهندسة المدنية، كلية الهندسة بشبرا، جامعة بنها، القاهرة، مصر.

*البريد الالكتروني: jorget.lamey18@feng.bu.edu.eg; georgettesyffen@gmail.com

**الملخص**

تتطلب معظم طرق دراسة تخطيط حركة المرور مصفوفات المصدر – الهدف ( م هـ) والتي تمثل الطلب على الرحلات بين نقاط المصدر و الهدف في الشبكة. يناقش هذا العمل مشكلة تقدير مصفوفة ( م هـ )، مما يعني تقدير مصفوفات ( م هـ ) باستخدام التدفقات المرورية المرصودة على الروابط. تهدف هذه الدراسة إلى توفير برنامج بديل سهل الاستخدام و مجدٍ اقتصادياً بحيث يمكن للباحثين و المخططين في الدول النامية استخدامه دون قيود. تم استخدام اقتران Fmincon في برمجية الماتلاب لانشاء كود متخصص لتقدير مصفوفة ( م هـ ) في الشبكات المزدحمة باستخدام ثلاثة أساليب تحسين مختلفة في المستوى العلوي: المربعات الصغرى المعممة (GLS) ، المربعات الصغرى العادية (OLS) ، الحد الأقصى للانتروبيا (ME) مقرونة بطريقة الاتزان في توزيع الرحلات في المستوى السفلي. لاختبار أداء النماذج المقترحة ، تم إنشاء شبكة ممر افتراضية ، وتم اقتراح حالتين من عوامل الموثوقية ، يليها تحليل إحصائي مفصل. وأسفرت الطرق الثلاثة المقترحة عن نتائج مرضية، حيث بلغ معدل التحسن في الخطأ بين تعدادات المرور المرصودة والمقدرة في الحالتين (52.11٪ و52.01٪) وفقا لطريقة ME، و(75.39٪ و75.53٪) وفقا لطريقة OLS، و(75.56٪ و75.53٪) وفقا لطريقة GLS. أيضا ، لضمان الأداء ، تمت مقارنة النتائج بنتائج طريقتي (TFF) والمربعات الصغرى (LS) المستخدمتين في برمجية PTV Visum. وقد أظهرت النتائج أن كلا من طريقتي GLS و OLS يمكن أن تسفر عن نتائج دقيقة للغاية ، حيث يقارب متوسط قيمة الخطأ المطلق نتائج طريقة LS بنسبة 98٪ على الأقل ، ويتم تقليل معدل الخطأ بنسبة 11٪ تقريبا مقارنة بطريقة TFF.

**الكلمات المفتاحية:** مصفوفة المصدر الهدف ، التحسين ثنائي المستوى ، اقتران Fmincon ، تعدادات المرور، الاتزان في تخصيص الرحلات.

## 1. INTRODUCTION

People travel for several reasons, to mention few, working, studying, shopping or leisure. To achieve these activities, many trips are done. However, transportation might be a challenge, that countries seek to overcome by creating a good and reliable transportation system. At the same time, many difficulties, such as pollution, congestion, and traffic accidents, have arisen, as a result of the growing number of vehicles, changes in land use, and socio-economic conditions, particularly in developing countries [1]. Therefore, obtaining an accurate matrix is an important and essential factor in transportation planning and using the OD matrix reflects the trip requirements placed on specific areas in the network when traveling from the origin (Oi) to the destination (Dj) [2].

The conventional way of estimating the OD matrix is through a field survey such as a household survey, roadside survey, and vehicle license plate survey [3]. However, as traffic has grown, this system has become time-consuming, financially costly, and inaccurate. OD estimation with link traffic counts has been created as an alternative to the field survey methods, to overcome its limitations. This technique was first devised in the early 1980s, and it became a relevant research area and a practical challenge [4].

The method's main goal was to generate an OD matrix from available data to reproduce observed link counts as accurately as possible. The purpose of link traffic counts was to compare assigned link traffic counts with observed data and to minimize the gap between them. This could be done by an iterative process, in which the OD demands were adjusted and assigned to the study network, allowing for the generation of an estimation OD table [2]. The usage of data produced from traffic counts is attractive since it provides inexpensive, simple, and immediate data [5].

Previous studies have investigated many optimization techniques, including entropy maximizing estimators (ME) [6], [7], maximum likelihood estimation (ML) [8]–[10], generalized least square (GLS) [8], [11], [12], Neural networks (NNs) [13] and Bayesian inference estimation (BI) [14].

ME estimators are used to find the most likely OD matrix estimate based on road traffic counts. The ML technique optimizes the joint probability of the closeness between the target OD and the estimated OD matrix. The GLS approach, which is based on the well-known Gauss-Markov theory, is commonly used in regression models, particularly when correlation must be considered. It minimizes the sum of the squared residuals of the OD matrix and link flows. The target OD matrix in the (BI) approach is a prior probability function of the estimated OD matrix based on observed traffic count data. NNs have been offered as a feasible solution to a range of transportation problems due to their extremely dynamic, large-scale, complicated, and unexpected nature.

Although the above models provided strong mathematical foundations for the OD matrix estimation problem, they are not without drawbacks. The fact that these models were created under uncongested network conditions is a key restriction. Further, the assumption that route choice proportions are determined independently outside the OD demand estimating process is impractical because it implies that OD flow estimation must not account for congestion [4].

Traffic equilibrium models are commonly used to estimate traffic flows in congested transportation networks. Wardrop's first principle states that “ The journey times on all routes used are equal, and less than those which would be experienced by a single vehicle on any unused route” [15]. Several optimization models used to estimate OD demands were discussed based on this assumption, including combined distribution and assignment models which are based on link-level based data [16]–[18]. Another model is a Bi-level optimization system that integrated any estimation technique with the user equilibrium assignment model [19]–[24]. Moreover, there is a fuzzy-based approach that employed fuzzy weights to estimate the OD matrix by applying some "fuzziness" to link data. Finally, the TFF algorithm has been widely used in the last years because it is implemented by PTV Visum software [25]–[27].

The main drawback in most equilibrium models is that they assume the trip proportion map obtained from assigning the target OD matrix remains constant through successive iterations. Although certain software can estimate OD matrices, planners and researchers in developing countries face many restrictions to use them. As a result, we plan to develop a code that overcomes these drawbacks.

The key contributions of this paper, when compared to earlier studies, are: (1) Three different bi-level iterative optimization models are coded and implemented by using Fmincon Function in MATLAB software. (2) GLS, OLS, and ME are used in the upper level and the trip proportion map P(g) can be updated successively based on the OD matrix and paths flow. (3) The results obtained from the three above models are compared with those obtained from PTV Visum software.

## 2. MATERIAL AND METHODS

The primary goal of this study is to develop an inexpensive, accurate, and easy-to-use code for estimating the OD matrix in congested networks which can be available for researchers and planners in developing countries. For this reason, coding was conducted using the MATLAB 2021 software and the results were compared to those obtained using the PTV Visum 2022 software.

### 2.1. Network Characteristics Data

Although the formulations of OD matrix estimation models using traffic counts varies, they have something in common, which is these approaches frequently require a target OD matrix ($\hat{g}$) and a collection of observed link traffic counts ($\hat{v}$) [25]. The target OD matrix is usually an outdated OD matrix updated by a growth factor or a matrix generated by a trip distribution technique. When a target OD matrix is used, as an input, the OD matrix estimation is considered a problem, known as a calibration or adjustment problem.

Zones are treated as nodes, and connectors are treated as links, in this MATLAB code. The code automatically determines the outputs that are required in the optimization process after entering the below inputs. The inputs and outputs from this stage are described in **Tables (1 and 2)** below.

**Table 1:** Network characteristics inputs

| Input | Description | Remarks |
|---|---|---|
| Q | Nodes name vector matrix | |
| R | Zones logic vector-matrix | Being 1 if node Q is a zone and 0 otherwise |
| S | Links logic matrix | Being 1 if there is a link connected between the two nodes and 0 otherwise. |
| $\bar{g}$ | Target OD matrix | A matrix where rows specify origin zones and columns specify destination zones. |
| W | link performance function matrix | A matrix where each row specifies (the link's name, $\alpha_a$, $\beta_a$, $t_a$, $C_a$, $\widehat{v_a}$) |
| $\gamma_1$ & $\gamma_2$ | the relative weight (reliability factor) for each objective | Values range between 0 and 1 |

**Table 2**: Network characteristics outputs

| Output | Description | Remarks |
|---|---|---|
| A | Links matrix | A matrix where each row specifies (name of the link, start node, end node) |
| $\boldsymbol{\pi_{OD}}$ | Path set matrix | A matrix where each row specifies (the name of the path, origin zone, destination zone, and names of links contained in the path). |
| $\boldsymbol{\hat{g}}$ | Target OD vector-matrix | Target OD pairs calculating from reshaping of $\bar{g}$ to a column vector |

## 2.2. Bi-level Approach for OD Matrix Estimation

All estimation models have the same goal of minimizing the distance between observed and estimated link traffic counts, as well as between estimated and target O-D demand. Equation (1) represents the general form of these models [28] :

$$Min_g \ \gamma_1 F_1(\hat{g}, g) + \gamma_2 F_2(\hat{v}, v) \tag{1}$$

$$s.t: \quad g \geq 0 \tag{1a}$$
$$v = P(g).g \tag{1b}$$

Where $g$ represents the estimated OD vector-matrix, $v$ represents the vector for estimated link counts, ($\gamma_1$ and $\gamma_2$) represent the relative weight (reliability factor) for each objective, and P(g) represents the trip proportion map.

The bi-level procedure is a hierarchical optimization problem in which one problem's constraints are partially specified, by a second parametric optimization problem. Equation (1) is usually written as a bi-level program to handle the problem efficiently; the upper-level problem is the trip matrix estimation problem, while the lower-level problem is the network equilibrium assignment problem.

### 2.2.1. Lower-level formulation

Equation (2) computes a traffic network equilibrium as described in [29]

$$h_k > 0, \ t_k = t_{cur} \tag{2a}$$
$$h_k = 0, \ t_k > t_{cur} \tag{2b}$$

Where $h_k$ denotes the path flow and t denotes the path travel time. The value $t_{cur}$ represents the shortest possible travel time in each OD-Pair. The path travel time $t_k$ is computed as a sum of the link travel time that makes it up.

The solution method used in the procedure alternates contributes at solving an equilibrium problem on a restricted set of paths, and the generation of new paths. The equilibrium problem with the restricted set of paths is calculated by using Beckmann's mathematical formulation in Equation (3) [30]:

$$v = \min \sum_{a \in A} \int_0^{Va} t_a . \left(1 + \alpha_a \left(\frac{x_a}{C_a}\right)^{\beta_a}\right) . dx \quad s.t: \tag{3}$$

$$va = \sum_{O \in Q} \sum_{D \in Q} \sum_{k \in \pi_{OD}} \delta_{ak}^{OD} . h_k^{OD} \qquad \forall a \in A \tag{3a}$$

$$\sum_{k \in \pi_{OD}} h_k^{OD} = g_{OD} \quad , \forall O, D \in Q \tag{3b}$$

$$h_k^{OD} \geq 0, \forall k \in \pi_{OD}, \forall O, D \in Q \tag{3c}$$

Where $v_a$: is the flow on the link a with the trip matrix g, $h_k$: is the flow on the $k^{th}$ path for the OD pair, OD: is the set of all Origin-Destination pairs in the network, $\pi_{OD}$: is the set of paths connecting the OD pair, $\delta_{ak}$: is the link route incidence variable being 1 if link a is included in route k and 0 otherwise, $C_a$: is the capacity of link a, $t_a$: free-flow travel time of link a, and ($\alpha_a$, $\beta_a$): the performance functions parameter of link a.

ALLPATHS function was used to determine ( $\pi_{OD}$) which represents the potential paths between any pair of origin-destination. From there, the travel time for each path could be determined. This non-linear optimization problem approximates the optimal value of the traffic equilibrium problem by determining the value of (Va) where the paths' time is equal for all potential paths between each OD pair.

### 2.2.2. Upper-level formulation

The upper level aims to minimize the distance measures between observed and estimated variables. As mentioned in [24], [28], [31], three different distance measures based on GLS, OLS, and ME concepts (given in **Table 3**) are substituted in Equation(1):

**Table 3:** Upper-level formulation

| Model Name | $\boldsymbol{F_1(\hat{g}, g)}$ | $\boldsymbol{F_2(\hat{v}, v)}$ |
|---|---|---|
| GLS | $((\hat{g} - g)^T U^{-1} (\hat{g} - g))$ | $\left((\hat{v} - P(g).g)^T Z^{-1} (\hat{v} - P(g).g)\right)$ |
| OLS | $((\hat{g} - g)^T (\hat{g} - g))$ | $\left((\hat{v} - P(g).g)^T (\hat{v} - P(g).g)\right)$ |
| ME | $\sum g \left(\log(\frac{g}{\hat{g}}) - 1\right)$ | $\sum P(g).g \left(\log(\frac{P(g).g}{\hat{v}}) - 1\right)$ |

where U & Z represent the variance-covariance matrix for the error in the target OD matrix and observed traffic counts respectively.

## 2.3. Trip Proportion Map

The fraction of OD demand on each link is represented by P(g), which is a (a X g) matrix. When travel time fluctuates with traffic flows in congested networks, the trip proportion map p(g) is a function of the estimated traffic flow vector (v).

In conventional models, the value of P(g) calculated from the target OD matrix remains constant over iterations, whereas our code updates P(g) after each iteration. P(g) is calculated by using Equation (4) as described in [8].

$$P\,(g) = \delta_{ak}^{OD} \,.\, P_a^{OD} \tag{4}$$

$$P_a^{OD} = \frac{h_k^{OD}}{g^{OD}} \quad \text{and} \begin{Bmatrix} \sum_k h_k^{OD} = g^{OD} \\ h_k^{OD} \geq 0 \end{Bmatrix} \tag{4a}$$

### 2.4. Convergence Criteria

The user can choose whether to use Equation (5) as the convergence criteria or to determine the number of iterations.

$$\frac{|V_a^{n+1} - V_a^n|}{V_a^{n+1}} \leq 0.0001 \tag{5}$$

### 2.5. Model Solution Algorithm

Since the generated mathematical model in this optimization study is multi-variable and non-linear, the MATLAB function FMINCON Function is used to optimize both the upper and lower levels. This function is based on gradients [32]. Gradient-based approaches use first and second derivatives, gradients, and Hessians to locate local minima, as the name implies. Fmincon uses one of five implemented optimization algorithms to iterate from a given starting point to a local minimum. The algorithms are trust-region reflective, SQP, SQP-legacy, active set, and interior-point [33], the generated code enables the user to select the best algorithm for his needs. The interior-point algorithm was used to calculate the results in the next section. **Fig.1** describes the flow of the complete procedure.

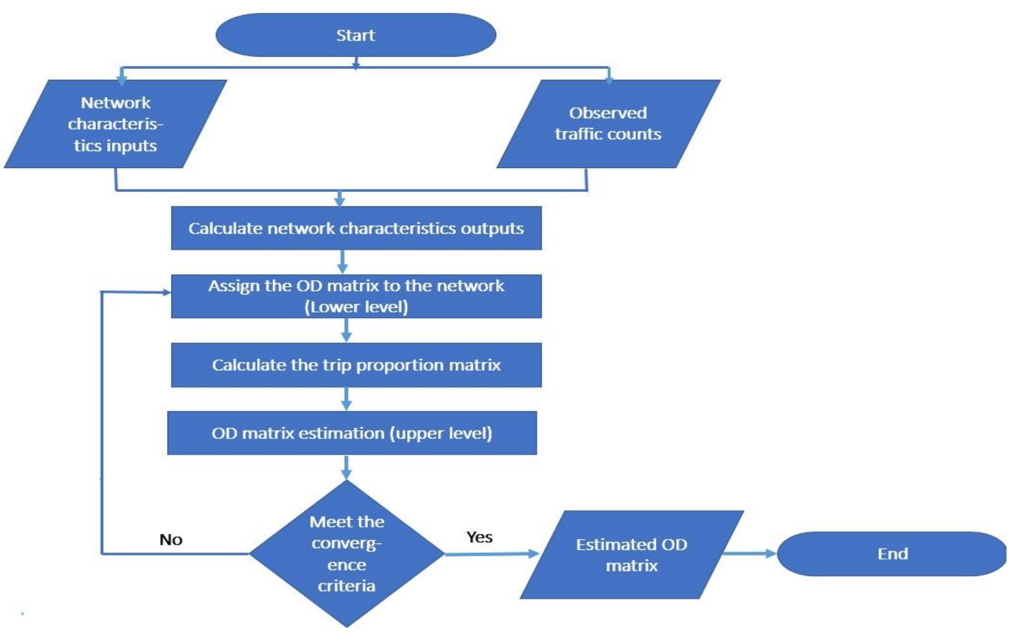


**Fig.1:** Solution Algorithm.

## 3. MODEL VALIDATION

PTV Visum software was used to validate the proposed models and guarantee the code performance. The purpose of the matrix estimation technique in Visum is to adjust a demand matrix so that the assignment results for a supply match the actual supply observation. PTV Visum has a variety of adjustment techniques, but here we employed both the Least Squares (LS) and TFlow Fuzzy (TFF) procedures, to compare our results.

### 3.1. Comparison Criteria

The estimation accuracy is compared using the errors between observed and estimated values. GEH, Coefficient of determination ($R^2$), and Relative Absolute Error (RAE) are used to estimate accuracy in this paper. GEH is not a legitimate statistical test, although its mathematical shape is like the chi-squared test. It was proposed by Geoffrey E. Havers and is used to compare test results. According to the DMRB [34], GEH values of less than 5% are acceptable and show a satisfactory match between real and simulated data. Furthermore, the GEH should be less than 5% at 85% of the data.

GEH and relative mean absolute errors (RMAE) for OD matrix and link flow after convergence are defined in Equations (6-9) as follows:

- **For OD matrices:**

$$GEH = \sqrt{\frac{2.(\hat{g}-g)^2}{(\hat{g}+g)}} \tag{6}$$

$$RMAE\% = \frac{1}{\mu}\sum_{1}^{\mu}\left|\frac{(\hat{g}-g)}{\hat{g}}\right| * 100 \tag{7}$$

Where μ represents the number of non-zero OD pairs.

- **For links traffic counts**

$$GEH = \sqrt{\frac{2.(\hat{v}-v)^2}{(\hat{v}+v)}} \tag{8}$$

$$RMAE\% = \frac{1}{\psi}\sum_{1}^{\mu}\left|\frac{(\hat{v}-v)}{\hat{v}}\right| * 100 \tag{9}$$

Where ψ represents the number of observed links.

## 4. RESULTS AND DISCUSSION

The computational results of the proposed algorithm are shown and compared to the PTV Visum estimation results (TFF and LS). The essential disadvantage of the PTV Visum software is that it assumes P(g) remains constant. As a result, the corridor network (shown in **Fig.2**) is used to demonstrate the algorithms' efficiency.

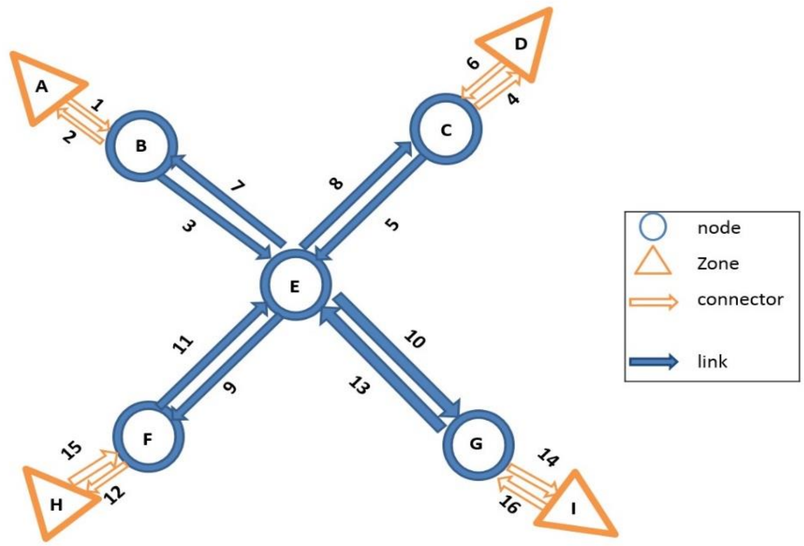


**Fig.2**: Corridor Network

### 4.1. Inputs

The first input is the network characteristics. For this reason, the current nodes and zones should be plotted first, followed by the creation of links and connectors, to transfer demand to the network. Each link is made up of two connected nodes, and each connector is made up of one zone connected to one node. The corridor network in **Fig.2** includes 4 zones, 5 intersection nodes, 8 links, and 8 connectors. **Tables (4 and 5)** list the important characteristics of this network.

The value of the ($\alpha_a$ and $\beta_a$) is presumed to be 0.15 and 4 respectively. The free flow time and capacity of connectors are assumed to be Zero and infinity respectively assumed to match the "Absolute" option in the zone to connectors share in PTV Visum.

Two cases of reliability factors were discussed because, in theory, most contributions consider both F1 and F2 in the upper-level problem, even though in some of them only F2 is present in the objective function. The reliability factors $\gamma_1$ & $\gamma_2$ were assumed equally in the first case, which means both the target OD matrix and observed link counts have the same reliability level, but in the second case, we assumed that the target OD matrix was unreliable.

In this bi-level optimization, both equilibrium assignment and matrix estimation should be carried out iteratively, with each algorithm running for 30 iterations.

**Table 4:** Links characteristics

| Link index | From | To | Free flow time (min) | observed link counts (pcu/h) |
|---|---|---|---|---|
| 3 | B | E | 42.85 | 4,950 |
| 5 | C | E | 34.28 | 5,130 |
| 7 | E | B | 42.85 | 4,725 |
| 8 | E | C | 34.28 | 4,845 |
| 9 | E | F | 38.57 | 4,686 |
| 10 | E | G | 69 | 5,088 |
| 11 | F | E | 38.57 | 4,600 |
| 13 | G | E | 69 | 5,136 |

**Table 5:** Target OD matrix

| O/D | A | D | H | I |
|---|---|---|---|---|
| A | 0 | 900 | 1,600 | 2,000 |
| D | 1,000 | 0 | 1,950 | 1,800 |
| H | 1,500 | 2,050 | 0 | 1,000 |
| I | 2,000 | 1,800 | 1,000 | 0 |

## 4.2. Estimation Results

The estimation results from the MATLAB code and the PTV Visum software are shown below. Then, the estimation accuracy is compared, using the errors between observed and estimated values.

### 4.2.1. Case one: assume ($\gamma_1 = \gamma_2$):

**Table (6)** shows the estimated traffic counts, GEH, and relative absolute error (RAE%) before adjustments (equilibrium case) and after using the five proposed adjustment models. The first estimated link counts on network links (Equilibrium) are produced by assigning the target OD matrix ($\hat{g}$) to the network, whereas the other estimated link counts are produced by assigning the estimated OD matrix (g) to the network.

**Table 6:** Case one estimation results

| Link index | Equilibrium | | | GLS | | | OLS | | |
|---|---|---|---|---|---|---|---|---|---|
| | V | GEH | RAE | V | GEH | RAE | V | GEH | RAE |
| 3 | 4,500 | 6.546537 | 9.090909 | 4,884.874 | 0.928714 | 1.315667 | 4,814.529 | 1.938818 | 2.736796 |
| 5 | 4,750 | 5.406549 | 7.407407 | 5,067.458 | 0.875869 | 1.219138 | 5,021.862 | 1.517822 | 2.107955 |
| 7 | 4,500 | 3.312946 | 4.761905 | 4,779.407 | 0.789242 | 1.151479 | 4,728.829 | 0.055686 | 0.081028 |
| 8 | 4,750 | 1.371564 | 1.960784 | 4,902.92 | 0.829632 | 1.195454 | 4,888.162 | 0.618712 | 0.890855 |
| 9 | 4,550 | 2.008602 | 2.912621 | 4,745.942 | 0.865556 | 1.268363 | 4,740.595 | 0.787925 | 1.154278 |
| 10 | 4,800 | 4.095937 | 5.660377 | 5,141.815 | 0.752461 | 1.057684 | 5,082.629 | 0.075323 | 0.10557 |
| 11 | 4,550 | 0.739222 | 1.086957 | 4,544.73 | 0.817374 | 1.201526 | 4,577.095 | 0.338134 | 0.49793 |
| 13 | 4,800 | 4.767037 | 6.542056 | 5,073.021 | 0.881487 | 1.226219 | 5,026.729 | 1.532909 | 2.127559 |
| Link index | ME | | | LS | | | TFF | | |
| | V | GEH | RAE | V | GEH | RAE | V | GEH | RAE |
| 3 | 4,725.394 | 3.229254 | 4.537496 | 4,835.631 | 1.635047 | 2.310492 | 4,834.199 | 1.655642 | 2.339423 |
| 5 | 4,987.916 | 1.997631 | 2.769672 | 5,036.35 | 1.313532 | 1.825538 | 5,045.243 | 1.188277 | 1.652181 |
| 7 | 4,725.394 | 0.005731 | 0.008337 | 4,743.94 | 0.275268 | 0.400858 | 4,753.251 | 0.410386 | 0.597915 |
| 8 | 4,987.916 | 2.038234 | 2.949759 | 4,893.204 | 0.690806 | 0.994916 | 4,918.607 | 1.053492 | 1.519244 |
| 9 | 4,777.898 | 1.328638 | 1.950247 | 4,743.837 | 0.835002 | 1.223453 | 4,777.945 | 1.329314 | 1.951244 |
| 10 | 5,040.42 | 0.668601 | 0.935138 | 5,100.643 | 0.177135 | 0.248486 | 5,126.51 | 0.538865 | 0.756879 |
| 11 | 4,777.898 | 2.597969 | 3.867354 | 4,568.994 | 0.457927 | 0.674037 | 4,630.232 | 0.445014 | 0.657215 |
| 13 | 5,040.42 | 1.339935 | 1.860978 | 5,040.649 | 1.336706 | 1.856514 | 5,066.64 | 0.971104 | 1.35046 |

Before matrix adjustment, the MRAE for links was (4.927877%), $R^2 = 0.5574$, and 75% of links had a GEH value < 5, but the MRAE, GEH, and $R^2$ for estimated links produced by MATLAB and PTV Visum are better, indicating the high accuracy of the matrix estimation method. **Fig. 3** illustrates the dispersal diagrams of the observed and estimated link volumes

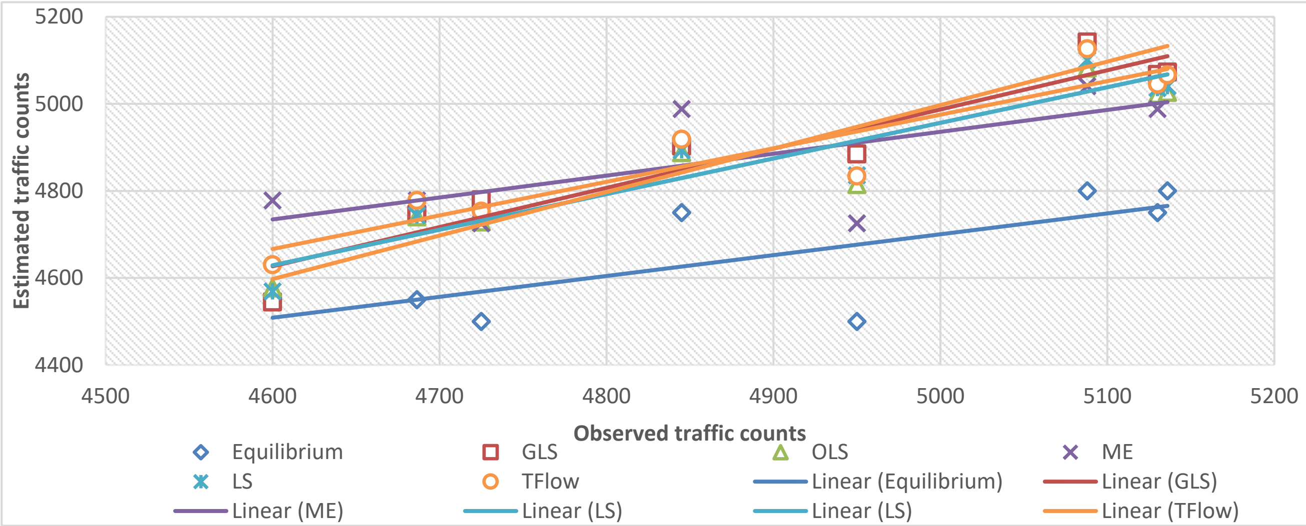


**Fig.3:** The dispersal diagram of the observed and estimated link volumes – case one

### 4.2.2. Case two: assume ($\gamma_1 < \gamma_2$)

In this case, we assumed that the only reliable and accurate input is observed link counts. Hence, we suggest that $\gamma_1 = 0$ and $\gamma_2 = 1$ consider F2 only in the upper level. Although variations between ($\widehat{V}$ and V) are more important, larger differences between (g and $\hat{g}$) might be acceptable in this situation. TFlow fuzzy fails to accurately reflect this situation because it expects the same degree of reliability for both distance measurements. The outcomes stayed the same as a result.

**Table (7)** below shows the estimation results for case two. the MARE% was increased to (1.206013%) for the GLS approach, (2.364683%) for ME approach, (1.205956%) for LS approach, and decreased only to (1.20601028%) for the OLS approach, and remain (1.35307%) for TFlow fuzzy approach.

**Table 7:** Case two estimation results

| Link index | Equilibrium | | | GLS | | | OLS | | |
|---|---|---|---|---|---|---|---|---|---|
| | V | GEH | RAE | V | GEH | RAE | V | GEH | RAE |
| 3 | 4,500 | 6.546537 | 9.090909 | 4,891.062 | 0.840215 | 1.19067 | 4,891.052 | 0.84035 | 1.19086 |
| 5 | 4,750 | 5.406549 | 7.407407 | 5,071.062 | 0.825252 | 1.148887 | 5,071.057 | 0.825328 | 1.148992 |
| 7 | 4,500 | 3.312946 | 4.761905 | 4,783.937 | 0.854746 | 1.247345 | 4,783.93 | 0.854647 | 1.247199 |
| 8 | 4,750 | 1.371564 | 1.960784 | 4,903.937 | 0.844165 | 1.216458 | 4,903.936 | 0.844143 | 1.216427 |
| 9 | 4,550 | 2.008602 | 2.912621 | 4,745.438 | 0.858236 | 1.257603 | 4,745.438 | 0.858249 | 1.257622 |
| 10 | 4,800 | 4.095937 | 5.660377 | 5,146.937 | 0.823873 | 1.158352 | 5,146.929 | 0.823765 | 1.1582 |
| 11 | 4,550 | 0.739222 | 1.086957 | 4,541.063 | 0.871777 | 1.281242 | 4,541.069 | 0.871693 | 1.281118 |
| 13 | 4,800 | 4.767037 | 6.542056 | 5,077.062 | 0.824768 | 1.147545 | 5,077.056 | 0.824854 | 1.147665 |
| Link index | ME | | | LS | | | TFF | | |
| | V | GEH | RAE | V | GEH | RAE | V | GEH | RAE |
| 3 | 4,735.025 | 3.089239 | 4.342919 | 4,890.992 | 0.841211 | 1.192077 | 4,834.199 | 1.655642 | 2.339423 |
| 5 | 4,998.082 | 1.853761 | 2.571491 | 5,070.995 | 0.826194 | 1.150194 | 5,045.243 | 1.188277 | 1.652181 |
| 7 | 4,735.025 | 0.145772 | 0.21218 | 4,783.994 | 0.855572 | 1.248553 | 4,753.251 | 0.410386 | 0.597915 |
| 8 | 4,998.082 | 2.182101 | 3.159597 | 4,903.998 | 0.845029 | 1.217707 | 4,918.607 | 1.053492 | 1.519244 |
| 9 | 4,787.637 | 1.46945 | 2.158048 | 4,745.000 | 0.851878 | 1.248258 | 4,777.945 | 1.329314 | 1.951244 |
| 10 | 5,050.694 | 0.523967 | 0.733218 | 5,146.993 | 0.824658 | 1.15946 | 5,126.51 | 0.538865 | 0.756879 |
| 11 | 4787.637 | 2.738767 | 4.079062 | 4541.003 | 0.872663 | 1.282539 | 4630.232 | 0.445014 | 0.657215 |
| 13 | 5050.694 | 1.195304 | 1.660945 | 5076.994 | 0.825716 | 1.148861 | 5066.64 | 0.971104 | 1.35046 |

**Fig. 4** illustrates the dispersal diagrams of the observed and estimated link volumes and **Table 8** shows the value of the estimated OD matrices obtained from the five models in the two cases.

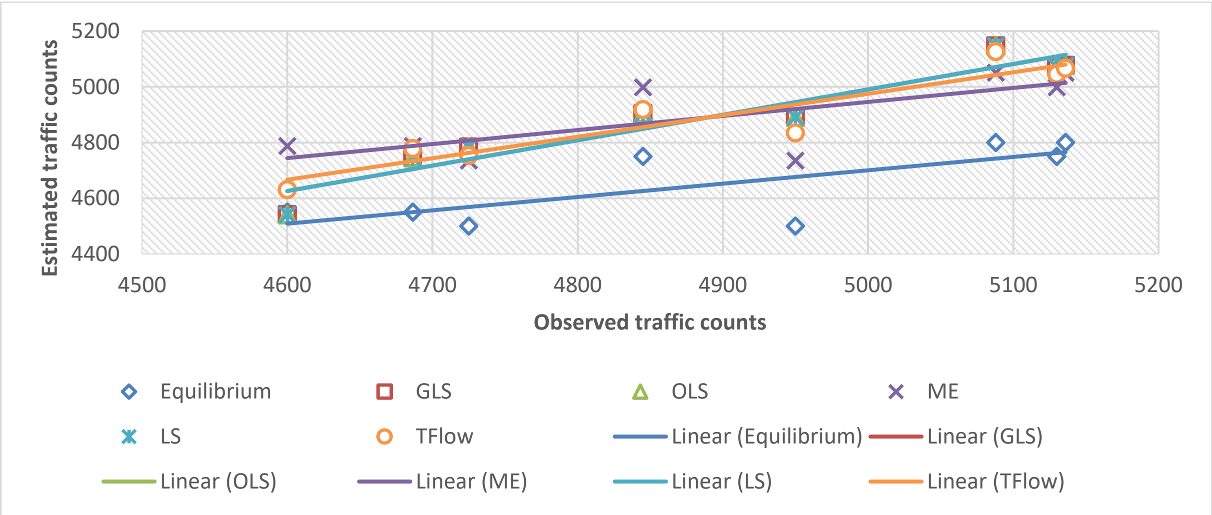


**Fig. 4:** The dispersal diagram of the observed and estimated link volumes - case two

Furthermore, statistical measures of relative errors (differences between $(\hat{g}\ , g)$ and $(\hat{v}, v)$ for the two cases are reported in **Table 9**. As can be observed from the data, the MARE (OD) has been increased in case two. The negligently of F1 from the upper level only improve the error rate of OLS approach by 0.14% but they make the optimization problem under-specified in theory which leads to several solutions that satisfy the observed link flow constraints as shown in **Table 8**.

GLS, OLS, LS approaches are the best in terms of links MARE, GEH, and $R^2$. However, as ME is not a quadratic problem, it is very sensitive to the target OD matrix values. Therefore, the outcomes were always focused on reducing the distance between $(\hat{g}\ , g)$ rather than $(\hat{v}, v)$ For this reason, it is to be considered the worst of the five methods.

**Table 8:** Estimated OD matrix (g)

| OD Pair | Case One | | | | | Case Two | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | GLS | OLS | ME | LS | TFF | GLS | OLS | ME | LS | TFF |
| AA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| AD | 1,015 | 992 | 945 | 999 | 963 | 1,017 | 1,017 | 947 | 986 | 963 |
| AH | 1,690 | 1,681 | 1,680 | 1,685 | 1,697 | 1,691 | 1,691 | 1,684 | 1,699 | 1,697 |
| AI | 2,180 | 2,141 | 2,100 | 2,152 | 2,174 | 2,183 | 2,183 | 2,104 | 2,207 | 2,174 |
| DA | 1,129 | 1,104 | 1,050 | 1,111 | 1,075 | 1,132 | 1,132 | 1,052 | 1,101 | 1,075 |
| DD | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| DH | 1,999 | 2,004 | 2,048 | 2,004 | 2,040 | 1,998 | 1,998 | 2,052 | 2,026 | 2,040 |
| DI | 1,939 | 1,914 | 1,890 | 1,921 | 1,929 | 1,941 | 1,941 | 1,894 | 1,944 | 1,929 |
| HA | 1,514 | 1,519 | 1,575 | 1,518 | 1,540 | 1,513 | 1,513 | 1,578 | 1,522 | 1,540 |
| HD | 2,008 | 2,030 | 2,153 | 2,024 | 2,067 | 2,006 | 2,006 | 2,157 | 2,024 | 2,067 |
| HH | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| HI | 1,023 | 1,028 | 1,050 | 1,027 | 1,024 | 1,022 | 1,022 | 1,052 | 996 | 1,024 |
| IA | 2,136 | 2,105 | 2,100 | 2,114 | 2,138 | 2,139 | 2,139 | 2,104 | 2,162 | 2,138 |
| ID | 1,880 | 1,866 | 1,890 | 1,871 | 1,889 | 1,882 | 1,882 | 1,894 | 1,895 | 1,889 |
| IH | 1,056 | 1,055 | 1,050 | 1,056 | 1,040 | 1,056 | 1,056 | 1,052 | 1,020 | 1,040 |
| II | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Table 9**: Statistical measures for the two cases

| Statistical Measure | Case One | | | | | Case Two | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | GLS | OLS | ME | LS | TFF | GLS | OLS | ME | LS | TFF |
| MARE (OD) | 6.062 | 5.120 | 5.009 | 5.404 | 5.236 | 6.141 | 6.141 | 5.223 | 5.545 | 5.236 |
| MARE (links) | 1.204 | 1.213 | 2.359 | 1.192 | 1.353 | 1.206 | 1.206 | 2.365 | 1.206 | 1.353 |
| $R^2$ (OD) | 0.981 | 0.990 | 1.000 | 0.988 | 0.993 | 0.980 | 0.980 | 1.000 | 0.982 | 0.993 |
| $R^2$ (links) | 0.912 | 0.903 | 0.557 | 0.903 | 0.879 | 0.912 | 0.912 | 0.557 | 0.912 | 0.879 |
| % Of links have (GEH <5) | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| %Of OD pairs have (GEH <5) | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |

## 5. CONCLUSION AND RECOMMENDATIONS

Surveys and interviews have long been used in developed countries to collect traffic data. Since developing countries lack the financial resources to collect data regularly for traffic planning, they usually multiply the target OD matrix by the growth factor, which is inefficient. As a result, using the MATLAB code presented in this paper will provide an easy, accurate, and inexpensive way to estimate the OD matrix that reflects current traffic flows.

As the target OD matrix, observed traffic counts, and route choice proportion have a substantial impact on OD matrix estimation accuracy, we included reliability factors and trip proportion map in our code to account for all possible scenarios.

The three proposed models (GLS, OLS, and ME) increase the percentage of links that have a GEH value less than 5 from 75% to 100% after matrix adjustment and decrease the MARE from (4.927877%) to (1.204441,1.212746, 2.359873) % respectively in case one and to (1.206013,1.20601,2.364683) % respectively in case two. So, the improvement rate in error between observed and estimated traffic counts in the two cases was (52.11% and 52.01%) according to the ME approach, (75.39% and 75.53%) according to the OLS approach, and (75.56% and 75.53%) according to the GLS approach.

Both GLS and OLS approaches can produce extremely accurate results, with the MRAE value approximating the results of the LS method by at least 98%, and the error rate is reduced by approximately 11% over the TFlow fuzzy method. The ME approach is the worst of the five models, for estimating the link flow values, because it is more sensitive to the target OD matrix values.

More accurate outcomes are explicitly estimated when OD matrix variables (F1) are used. The target OD matrix appears to increase goodness-of-fit when it is included in GLS, ME, and LS. Using (F1) prevents the possibility that there are multiple OD demands that identically match the observed link flows.

The generated code does traffic assignment in each iteration because it estimates the value of P(g) after each iteration. It takes a longer run time than PTV Visum, which assumes a constant value of P(g), and we hope to enhance this aspect in future research. Additionally, more research should be done to see how alternative traffic assignment methods, such as stochastic and dynamic user equilibrium approaches, affect the precision of this adjustment. It will also be necessary to investigate models for assessing public urban transit trips.

## REFERENCES

[1] Karimi, H., Boushehri, S.N., and Nasiri, R.. 2020. “Origin-Destination Matrix Estimation Using Socio-Economic Information and Traffic Counts on Uncongested Networks.” International Journal of Transportation Engineering 8(2): 165–83

[2] 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 41(3): 265–85.

[3] Gangeraj, E.N, Behzadi, G.A., and Behzad, R. 2017. “Estimation of Origin-Destination Matrix from Traffic Counts Based On Fuzzy Logic.” Civil Engineering Journal 3(11): 1166–75.

[4] Behara, K.N.S. 2019. “Origin-Destination Matrix Estimation Using Big Traffic Data: A Structural Perspective.”

Queensland University of Technology.

[5] Ramli, M. I., Runtulalo, D., Yatmar, H., and Mangessi, A. 2020. "An Estimation of Origin-Destination Matrices for a Public Transport Network in Makassar Using Macrosimulation Visum." in IOP Conference Series: Materials Science and Engineering. Vol. 875.

[6] Hellinga, B. R. 1994. "ESTIMATING DYNAMIC ORIGIN- DESTINATION DEMANDS FROM LINK AND PROBE COUNTS By." Queen's University.

[7] Van Zuylen, H. J., and Willumsen, L.G. 1980. "The Most Likely Trip Matrix Estimated from Traffic Counts." Transportation Research Part B: Methodological 14(3):281–93.

[8] Cascetta, E., and Nguyen, S. 1988. "A Unified Framework for Estimating or Updating Origin/Destination Matrices from Traffic Counts." Transportation Research Part B 22(6):437–55.

[9] Hazelton, M. L. 2000. "Estimation of Origin-Destination Matrices from Link Flows on Uncongested Networks." Transportation Research Part B: Methodological 34(7):549–66.

[10] Spiess, H. 1987. "A Maximum Likelihood Model for Estimating Origin-Destination Matrices." Transportation Research Part B 21(5):395–412.

[11] Bell, M.G. H. 1991. "The Estimation of Origin-Destination Matrices by Constrained Generalised Least Squares." Transportation Research Part B 25B(1):13–22.

[12] Cascetta, E. 1984. "Estimation of Trip Matrices from Traffic Counts and Survey Data: A Generalized Least Squares Estimator." Transportation Research Part B 18(4–5):289–99.

[13] Gong, Z. 1998. "Estimating the Urban OD Matrix: A Neural Network Approach." European Journal of Operational Research 106(1):108–15.

[14] Maher, M. J. 1983. "Inferences on Trip Matrices from Observations on Link Volumes: A Bayesian Statistical Approach." Transportation Research Part B 17B(6):435–47.

[15] Bar-gera, H., Hellman,F., and Patriksson, M. 2013. "Computational Precision of Traffic Equilibria Sensitivities in Automatic Network Design and Road Pricing." Procedia - Social and Behavioral Sciences 80(2011):41–60.

[16] Fisk, C. S., and Boyce, D. E. 1983. "A Note on Trip Matrix Estimation from Link Traffic Count Data." Transportation Research Part B 17(3):245–50.

[17] Fisk, C. S. 1989. "Trip Matrix Estimation from Link Traffic Counts: The Congested Network Case." Transportation Research Part B 23(5):331–36.

[18] Kawakami, S., Lu, H., and Hirobata, Y. 1992. "ESTIMATION OF ORIGIN-DESTINATION MATRICES FROM LINK TRAFFIC COUNTS CONSIDERING THE INTERACTION OF THE TRAFFIC MODES." Regional Science Association International 71(2):139–51.

[19] 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 39(7):565–91.

[20] Florian, M., and Chen, Y. 1995. "A Coordinate Descent Method for the Bi-Level O-D Matrix Adjustment Problem." International Transactions in Operational Research 2(2):165–79.

[21] García-Ródenas, R., and Verastegui-Rayo,D. 2008. "A Column Generation Algorithm for the Estimation of Origin-Destination Matrices in Congested Traffic Networks." European Journal of Operational Research 184(3):860–78.

[22] Kim, H., Baek, S.,and Lim,Y. 2001. “Origin-Destination Matrices Estimated with a Genetic Algorithm from Link Traffic Counts.” Transportation Research Record: Journal of the Transportation Research Board 1771(1):156–63.

[23] Lundgren, J.T., and Peterson, A. 2008. “A Heuristic for the Bilevel Origin-Destination-Matrix Estimation Problem.” Transportation Research Part B: Methodological 42(4):339–54.

[24] Yang, H., Sasaki, T., Iida, Y., and Asakura, Y. 1992. “Estimation of Origin-Destination Matrices from Link Traffic Counts on Congested Networks.” Transportation Research Part B 26(6):417–34.

[25] Duraku, R., Atanasova, V., and Krstanoski, N. 2019. “Building and Calibration Transport Demand Model in Anamorava Region.” Tehnicki Vjesnik - Technical Gazette 26(6):1784–93.

[26] NOVAČKO, L., ŠIMUNOVIĆ, L., and KRASIĆ, D. 2014. “Estimation of origin-destination trip matrices for small cities". Promet Traffic & Transportation 26(5):419–28.

[27] Yousefikia, M., Mamdoohi, A.R., Moridpour, S., Noruzoliaee, M.H., and Mahpour, A. 2013. “A Study on the Generalized TFlowFuzzy O-D Estimation.” Pp. 1–9 in Australasian Transport Research Forum, ATRF 2013 - Proceedings. Brisbane, Australia.

[28] Wang, Y., Ma, X., Liu, Y., Gong, K., Henricakson, K.C., Xu, M., and Wang, Y. 2016. “A Two-Stage Algorithm for Origin-Destination Matrices Estimation Considering Dynamic Dispersion Parameter for Route Choice.” PLoS ONE 11(1):1–24.

[29] Josefsson, M., and Patriksson, M. 2007. “Sensitivity Analysis of Separable Traffic Equilibrium Equilibria with Application to Bilevel Optimization in Network Design.” Transportation Research Part B: Methodological 41(1):4–31.

[30] SHEFFI, Y. 1984. Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods. New Jersey, USA: PRENTICE-HALL, INC.

[31] Brenninger-Göthe, M., Jörnsten, K.O. ,and Lundgren, J.T.. 1989. “Estimation of Origin-Destination Matrices from Traffic Counts Using Multiobjective Programming Formulations.” Transportation Research Part B 23(4):257–69.

[32] Albaghdadi, A.M, Bin Baharom, M., and bin Sulaiman, S.A. 2021. “Parameter Design Optimization of the Crank-Rocker Engine Using the FMINCON Function in MATLAB.” IOP Conference Series: Materials Science and Engineering 1088(1):012072.

[33] Agnarsson, J., Sunde, M., and Ermilova, I. 2013. Parallel Optimization in Matlab.

[34] Villa, A. R.D., Casas, J., Breen, M., and Perarnau, J. 2014.“Static OD Estimation minimizing the Relative Error and the GEH Index,” Procedia - Soc. Behav. Sci., vol. 111, pp. 810–818.