Australasian Transport Research Forum 2021 Proceedings
8-10 December, Brisbane, Australia
Publication website: http://www.atrf.info

Can partial structural information of travel demand improve the quality of OD matrix estimates?

Krishna N.S. Behara[1], Ashish Bhaskar[1]

[1]School of Civil and Environmental Engineering, Queensland University of Technology, Brisbane, Australia

Email for correspondence: ashish.bhaskar@qut.edu.au

Abstract

The traditional upper-level formulation in the bi-level OD matrix estimation process includes link flows and is mathematically an under-determinate problem. Various methods exist in literature to improve the quality of OD estimates by using additional traffic information including target OD, traffic speeds, travel times, turning proportions, trajectories, and partial OD flows. Irrespective of the type of traffic information, this study along with a few limited studies are of the opinion that structural information of OD needs to be accounted in the upper-level formulation to improve the quality of final OD estimates. To this end, this study investigates if additional structural knowledge available through partial OD flows (referred as sub-OD flows) can improve the OD estimation quality. This study used two methods to capture the OD structure at two levels and refer to them as macro-OD and micro-OD structural information. Few studies in the past have applied the macro-OD structural information (captured using correlation coefficient) in OD estimation formulation. However, a method to account for micro-OD structure (captured using ratio of OD flows from the same origin and end into destinations with similar trip attractions) and integrate it into the upper-level objective function formulation was never developed and therefore is the major contribution of this study. The proposed methodology was demonstrated on Brisbane city network using synthetic link flows and partial OD flows from Bluetooth observations. The findings from this study revealed that exploiting the partial OD structural information through macro and micro sub-OD flows improved the quality of OD estimates better than using other candidate formulations.

1.Introduction

Origin-Destination (OD) matrix estimation has always been the topic of research in transport modelling. Several methods have been developed in the last three decades to estimate OD matrices (Van Zuylen and Willumsen, 1980, Cascetta, 1984, Yang et al., 1992, Cascetta and Postorino, 2001, Antoniou et al., 2004a, Michau et al., 2017a).

Traditionally, OD estimation is based on observed traffic counts, a prior OD matrix and a user equilibrium assignment (derived from analytical or simulation model). Since, the number of OD pairs are far greater than the number of mapping relationships between link flows and OD flows, the mathematical problem is under-determined. Thus, there is a possibility of developing poor quality OD estimates if the objective function focusses only on the deviation of observed traffic counts (Antoniou et al., 2016).

One of the ways to improve the quality of OD estimation is to focus on enhancing the structure of OD estimates in addition to matching link flows. The *structure* of OD matrix is defined as the distribution pattern of travel demand between different OD pairs (Djukic, 2014, Behara et al., 2020b). It is often assumed that target matrix contains important *structural* information and can be used to minimise the problem of under-determinacy (Bierlaire and Toint, 1995). However, the solution search space is biased around target OD matrix and it might not improve the quality of OD estimate because target matrix is often constructed from outdated surveys (Yang, 1995, Cascetta and Nguyen, 1988).

Variety of studies exist in the literature that estimated OD matrices indirectly from different other forms of traffic observations; for instance, travel speeds (Jaume and Montero, 2015), travel times (Barcelö et al., 2010), partial OD flows (Dixon and Rilett, 2002), turning proportions (Alibabai and Mahmassani, 2008) and trajectory data (Michau et al., 2017b). While these studies are based on indirect observations of OD flows, few proposed direct inference of OD flows from trajectory data such as taxi trajectories (Mungthanya et al., 2019, Liu et al., 2019, Chu et al., 2019), and cellular probes (Calabrese et al., 2011). Direct and complete OD flow observations demand extensive resources and therefore are very expensive. Thus, limited studies in the past proposed to use direct but partial observations of OD flows but assuming that penetration rates are known (Nasab and Shafahi, 2019, Antoniou et al., 2004b, Antoniou et al., 2006, Michau et al., 2017b) or being estimated (Yang et al., 2017, Iqbal et al., 2014).

To summarise, accounting for the structure of OD in the bi-level estimation process is important, and penetration rates of partial OD flows observed from emerging sources such as Bluetooth that provide up-to-date information are generally unknown. Therefore, there is a great need to develop methods that account for the OD structure without depending on the market penetration rates of partial OD flow observations.

Addressing the above need, the study proposes:

1. Two methods to compare structural information of sub-OD flows without the need to assume or estimate the penetration rates. The methods focus on capturing macro (it is high-level information at the level of OD matrix) and micro (it is a detailed information at the level of individual OD flows) structural knowledge of sub-OD flows.
2. A new upper-level formulation to integrate sub-OD flows into OD estimation problem.

Studies in the past have proposed ways to capture or compare macro structural information of OD matrices. For instance, Hussain et al. (2021), Djukic (2014), and Behara et al. (2020a) used correlation coefficient, and Behara et al. (2020b) proposed Levenshtein distance to compare high-level structure of OD matrices. The earlier works by Behara et al. (2021) and Behara et al. (2020c) designed method to integrate macro-OD structural information into bi-level formulation. Limited research such as the study by Kim et al. (2001) in the past have accounted for the micro-OD structure but as constraints outside objective function. To the best of our knowledge, no study in the past have attempted to include micro-OD structural information within the upper-level formulation. In this context, the method to compare micro structural information of sub-OD flows and further integrating it into the objective function formulation is the major contribution of this study.

This article demonstrates the proposed methodology using traffic simulation on a synthetic Brisbane network. Bluetooth observations provide the sub-OD information in the present study. However, the proposed methodology holds good for sub-OD observed from any other data sources such as GPS, and mobile phone.

The remaining of the paper is organised as follows. Section 2 discusses the methodology proposed in this study; Section 3 presents the experiments and results; and Section 4 concludes the paper.

2. Proposed Methodology

The original objective function of OD estimation is based on the deviations of traffic counts (link flows). Link counts are only point-based measurements and do not significantly contribute towards the quality of OD estimates due to the problem of under-determinacy. Partial observations of OD flows (referred as sub-OD flows) provide some *a-priori* structural knowledge of travel demand. Therefore, a combined formulation that match observed sub-OD flows and link flows with their estimated counterparts might enhance the quality of OD estimates. However, in most cases, sub-OD flows are only a sample, and their market penetration rates are generally unknown. In such situations, there is a great need for alternative ways to exploit this additional information. To this end, we propose:

a) Methods to compare sub-OD flows, and
b) A new upper-level formulation to integrate the above methods into OD estimation problem.

The proposed methodology is illustrated in Figure 1. The new upper-level formulation (Z($\mathbf{x}$)) includes deviations of traffic counts ($\tilde{\mathbf{y}}$ and $\mathbf{y}$) and partial or sub-OD flows ($\tilde{\mathbf{b}}$ and $\mathbf{b}$).

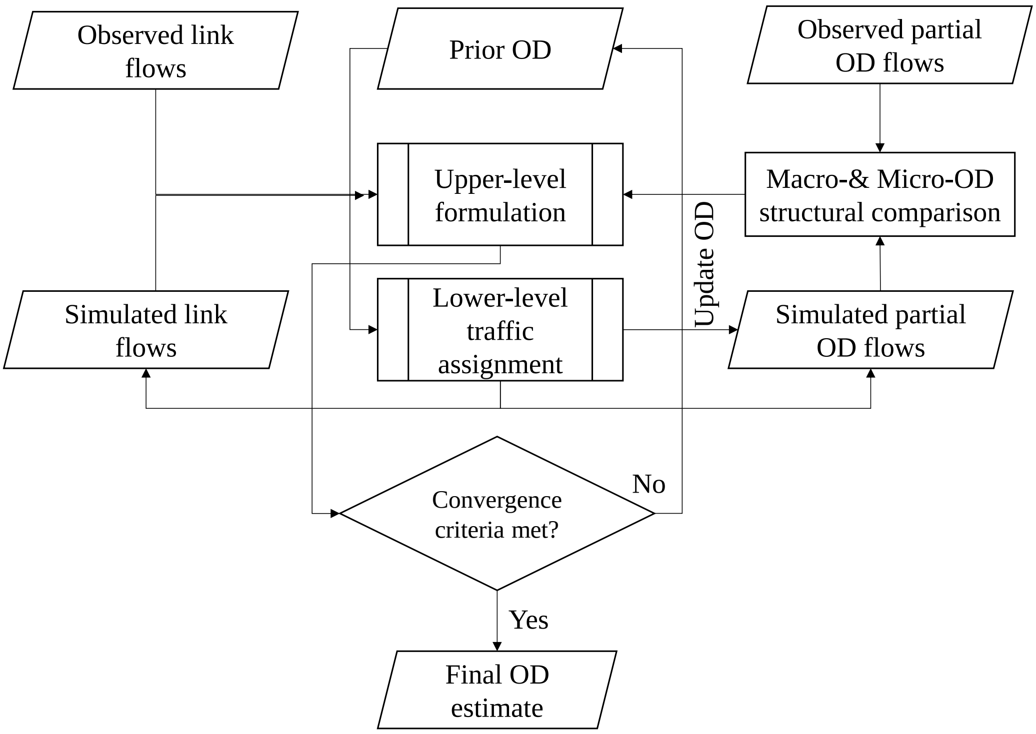


Figure 1: OD estimation algorithm

2.1. Structural comparison of sub-OD flows

The concept of "OD structure" is used in the comparison of sub-OD flows. In literature, the structure of OD matrix is defined as the distribution pattern of OD flows, and is generally captured through the coefficient of variation between the normalized OD flows (Djukic, 2014) or using Levenshtein distance (Behara et al., 2020b). In this paper, we exploit the structural information from sub-OD flows from two different levels as discussed below.

2.1.1. Macro sub-OD structural information

Here, we propose to capture sub-OD structural information from the partial "skeleton" of OD matrix. For instance, grey shaded cells of OD matrix shown in figure illustrates partial "skeleton "of OD matrix for which partial OD flows are known *a-priori*.

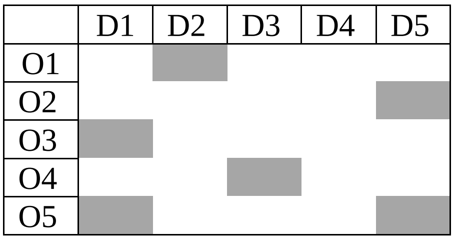


Figure 2: Demonstration of macro-OD structure

Let $\mathbf{x}^*$ refers to true OD vector, $\mathbf{x}$ is the estimated OD vector, $\tilde{\mathbf{b}}$ is OD vector with partial observations of OD flows (*e.g.* grey shaded OD pairs in Figure 2), $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{x}}^*$ are the portion of $\mathbf{x}$ and $\mathbf{x}^*$ that correspond to OD pairs used in $\tilde{\mathbf{b}}$. We assume that the structure $\tilde{\mathbf{b}}$ is a proxy for the structure of $\tilde{\mathbf{x}}^*$. Structural comparison of $\tilde{\mathbf{b}}$ and $\tilde{\mathbf{x}}$ should be equivalent to structural comparison of $\tilde{\mathbf{x}}^*$ and $\tilde{\mathbf{x}}$; and therefore, in the absence of $\tilde{\mathbf{x}}^*$ we can use $\tilde{\mathbf{b}}$ from emerging data sources such as Bluetooth. We have used Pearson correlation coefficient (ρ) for the structural comparison of OD flows in our study. A higher correlation implies that both vectors (i. e. $\tilde{\mathbf{b}}$ and $\tilde{\mathbf{x}}$) are structurally closer to each other. The macro structural comparison of sub-OD flows is expressed in the Equation (1).

$$\rho(\tilde{\mathbf{b}},\tilde{\mathbf{x}}) = \frac{(\tilde{\mathbf{b}}-\mu_{\tilde{\mathbf{b}}})^{\mathrm{T}}(\tilde{\mathbf{x}}-\mu_{\tilde{\mathbf{x}}})}{\sqrt{(\tilde{\mathbf{b}}-\mu_{\tilde{\mathbf{b}}})^{\mathrm{T}}(\tilde{\mathbf{b}}-\mu_{\tilde{\mathbf{b}}})}\sqrt{(\tilde{\mathbf{x}}-\mu_{\tilde{\mathbf{x}}})^{\mathrm{T}}(\tilde{\mathbf{x}}-\mu_{\tilde{\mathbf{x}}})}} \tag{1}$$

2.1.2 Micro sub-OD structural information

In this study, we propose to capture OD structural information at a micro scale; that is, at individual OD level. The ratio of OD flows from the same origin to destinations of similar trip attraction characteristics is defined as the micro-OD structure. Figure 3 shows an example demonstrating partial and complete OD flows from the same origin O1 and similar work-based destinations (D1 and D2). This study assumes that the ratio of complete to partial OD flows originating from the same zone to similar destinations remain same (in this example, 500/50 = 1000/100 = 10).

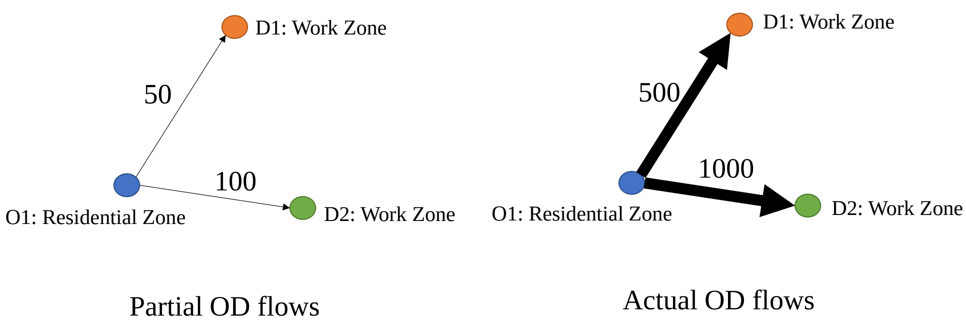


Figure 3: Demonstration of micro-OD structure

To demonstrate the ratio of OD flows, let $\tilde{\mathbf{b}}$ and $\mathbf{b}$ be the observed and simulated/estimated sub-OD vectors; and $\tilde{b}_{ij}$ and b_{ij} correspond to flows between OD pair from origin, i to destination, j. Consider $\tilde{b}_{i_1j_1}$ and $b_{i_1j_1}$ between i_1 and j_1; and $\tilde{b}_{i_1j_2}$ and $b_{i_1j_2}$ between i_1 and j_2. Since, $\tilde{b}_{ij}$ is only a fraction of b_{ij}, let the relation between both is represented using a scaling factor, α_{ij}, as shown in Equation (2).

$$\tilde{b}_{ij} = \alpha_{ij} b_{ij} \tag{2}$$

Using, Equation (2), $\tilde{b}_{i_1j_1}$ and $b_{i_1j_1}$ are related using $\alpha_{i_1j_1}$, and $\tilde{b}_{i_1j_2}$ and $b_{i_1j_2}$ using $\alpha_{i_1j_2}$ as shown in Equations (2a) and (2b), respectively.

$$\tilde{b}_{i_1j_1} = \alpha_{i_1j_1} b_{i_1j_1} \tag{2a}$$

$$\tilde{b}_{i_1j_2} = \alpha_{i_1j_2} b_{i_1j_2} \tag{2b}$$

Assuming $\alpha_{i_1j_1} \cong \alpha_{i_1j_2}$, the ratio of $\tilde{b}_{i_1j_1}$ and $\tilde{b}_{i_1j_2}$ can be represented as shown in Equation 3, and the relationship between them as presented in Equation 3a.

$$\frac{\tilde{b}_{i_1j_1}}{\tilde{b}_{i_1j_2}} = \frac{\alpha_{i_1j_1} b_{i_1j_1}}{\alpha_{i_1j_2} b_{i_1j_2}} = \ddot{\Upsilon}_{i_1j_1j_2} \tag{3}$$

$$\text{if } \alpha_{i_1j_1} \cong \alpha_{i_1j_2}$$

$$b_{i_1j_1} = \ddot{\Upsilon}_{i_1j_1j_2} b_{i_1j_2} \tag{3a}$$

The $\ddot{\Upsilon}_{i_1j_1j_2}$ in Equation 3 is the OD demand ratio and refers to the micro-OD structural information for the OD pairs i_1 and j_1, and i_1 and j_2. The vector $\ddot{\Upsilon}$ constitutes the ratio of OD flows between different sets of OD pairs and captures the micro-OD structural information.

2.2 Integration of sub-OD structural information into upper-level formulation

Traditional OD estimation methods attempt to minimise the deviations between traffic counts in the upper-level formulation. The structural information (both macro and micro) of sub-OD flows discussed in the previous section are two additional objectives that need to be integrated with the deviations of traffic counts.

The upper-level formulation with deviations of traffic counts and macro structural OD comparison is shown in Equation (4). The macro structural comparison of $\tilde{\mathbf{b}}$ and $\mathbf{b}$ is a scalar value between -1 and 1 and is considered as a penalty/scaling factor to the deviations of observed ($\tilde{\mathbf{y}}$) and estimated ($\mathbf{y}$) link flows. The Equation (4) has two objectives to be optimised and is same as the formulation presented in the study by Behara et al. (2021).

$$\min_{\mathbf{x}} Z(\mathbf{x}) = \min_{\mathbf{x}} \frac{1}{2}\Big(\big(c_1 + (\mathbf{y}-\tilde{\mathbf{y}})^T(\mathbf{y}-\tilde{\mathbf{y}})\big)\Big)\Big(\big(c_2 + f(\mathbf{b},\tilde{\mathbf{b}})\big)^T \big(c_2 + f(\mathbf{b},\tilde{\mathbf{b}})\big)\Big) \tag{4}$$

$$f(\mathbf{b},\tilde{\mathbf{b}}) = \frac{1-\rho(\mathbf{b},\tilde{\mathbf{b}})}{2} \tag{4a}$$

$$\text{such that } \mathbf{y} = \mathbf{P}\mathbf{x}\ ;\ \ \mathbf{b} = \mathbf{I}\mathbf{x} \tag{4b}$$

$$\rho(\mathbf{b},\tilde{\mathbf{b}}) = \frac{(\tilde{\mathbf{b}}-\boldsymbol{\mu}_{\tilde{\mathbf{b}}})^{\mathrm{T}}(\mathbf{b}-\boldsymbol{\mu}_{\mathbf{b}})}{\sqrt{(\tilde{\mathbf{b}}-\boldsymbol{\mu}_{\tilde{\mathbf{b}}})^{\mathrm{T}}(\tilde{\mathbf{b}}-\boldsymbol{\mu}_{\tilde{\mathbf{b}}})}\sqrt{(\mathbf{b}-\boldsymbol{\mu}_{\mathbf{b}})^{\mathrm{T}}(\mathbf{b}-\boldsymbol{\mu}_{\mathbf{b}})}} \tag{4c}$$

In Equation (4), **P** is the link proportion matrix with dimensions L x N where L and N indicate number of selected links and total number of OD pairs, respectively. The incidence matrix **I** maps the OD pairs for which partial OD flows are known *a-priori*.

To include an additional micro structural information of sub-OD flows, we propose Equation (5) where the sub-OD flows estimated using a known ratio of OD flows are matched with their observed counterparts. The Equation (5) has three objectives to be optimised compared to two in Equation (4).

$$\min_{\mathbf{x}} Z(\mathbf{x}) = \min_{\mathbf{x}} \alpha\left(\left(c_1 + (\mathbf{y}-\tilde{\mathbf{y}})^{\mathrm{T}}(\mathbf{y}-\tilde{\mathbf{y}})\right)\left(c_2 + f\left(\mathbf{b},\tilde{\mathbf{b}}\right)\right)^{\mathrm{T}}\left(c_2 + f\left(\mathbf{b},\tilde{\mathbf{b}}\right)\right)\right) + (1-\alpha)\left((\mathbf{Fb})^{\mathrm{T}}(\mathbf{Fb})\right) \tag{5}$$

In Equation (5), $f\left(\mathbf{b},\tilde{\mathbf{b}}\right)$, $\mathbf{y}$ and $\mathbf{b}$, and $\rho\left(\mathbf{b},\tilde{\mathbf{b}}\right)$ are same the previous Equations 4(a), 4(b), and 4(c), respectively. The dimension of **F** is $|\mathbf{b}|$ x $|\mathbf{b}|$ where $|\mathbf{b}|$ is the number of OD pairs for which ratio of OD flows is known *a-priori*. The matrix **F** multiplied with estimated OD vector **b** calculates the deviation between two estimates of partial OD demands. The first estimate is calculated using known *a-priori* ratio of sample OD flows, and the second estimate is an output from an iterative traffic assignment. For example, if $\tilde{b}_{i_1j_1}$ and $\tilde{b}_{i_1j_2}$ are the observed partial OD flows between i_1 and j_1, and i_1 and j_2; and $\ddot{Y}_{i_1j_1j_2}$ is $\frac{\tilde{b}_{i_1j_1}}{\tilde{b}_{i_1j_2}}$; then $\ddot{Y}_{i_1j_1j_2}b_{i_1j_2}$ is an another estimate of $b_{i_1j_1}$ The difference between $b_{i_1j_1}$ (from simulation) and $\ddot{Y}_{i_1j_1j_2}b_{i_1j_2}$ is an element in the matrix, $\mathbf{Fb}$. In other words, $b_{i_1j_1} \in \mathbf{b}$ and $1-\frac{\ddot{Y}_{i_1j_1j_2}b_{i_1j_2}}{b_{i_1j_1}} \in \mathbf{F}$. The condition for using $\ddot{Y}_{i_1j_1j_2}$ is that $b_{i_1j_1}$ and $b_{i_1j_2}$ should originate from the same origin and end into similar destinations.

The terms c_1 (<<1) and c_2 (<<1) in the Equations (4) and (5) are the constants meant to stabilise the objective function when $(\mathbf{y}-\tilde{\mathbf{y}})^{\mathrm{T}}(\mathbf{y}-\tilde{\mathbf{y}})$ or $f\left(\mathbf{b},\tilde{\mathbf{b}}\right)$ becomes zero.

2.3 Procedure to implement the proposed methodology

To execute the framework illustrated in Fig. 2 under controlled environment, we need to run upper-level and lower optimisations iteratively. For the current analysis, the codes for the optimisation are written in Matlab, and lower level traffic assignment is optimised using Aimsun next (2019). A Python script is written to integrate the optimisation model (in Matlab) with the traffic assignment (in Aimsun). However, Matlab is the primary platform that writes OD data into Aimsun OD format, runs the simulation, executes the Python script, and reads the simulation outputs for further optimisation process. This integration of Aimsun with Matlab is similar to the one presented in Antoniou et al., (2016).

The step-by-step procedure for OD estimation is outlined in the following:

Step-1: Prior inputs

Obtain the observed sub-OD flows ($\tilde{\mathbf{b}}$), ratio of OD flows ($\ddot{\Upsilon}$) , and observed link flows ($\tilde{\mathbf{y}}$). Set $k=1$; $\mathbf{x_k} = \tilde{\mathbf{x}}$.

Step-2: Traffic assignment (lower-level optimisation)

Load the study network in Aimsun next (2019) with demand, $\mathbf{x_k}$, and run traffic assignment (either stochastic route choice or dynamic user equilibrium). The outputs of the traffic assignment process include link flows ($\mathbf{y_k}$), sub-OD flows ($\mathbf{b_k}$), and link-proportion matrix ($\mathbf{P_k}$).

Step-3: Minimising the upper-level formulation

The Equation (6) presents the gradient of $Z(\mathbf{x})$ and is used to minimise the proposed upper-level formulation (Equation (5)). The term α in Equation (5) is assumed to be $\frac{1}{2}$ in the Equation (6).

$$\frac{\partial Z(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial\left(\frac{1}{2}\left(\left(c_1 + (\mathbf{y}-\tilde{\mathbf{y}})^{\mathrm{T}}(\mathbf{y}-\tilde{\mathbf{y}})\right)\left(c_2 + f(\mathbf{b},\tilde{\mathbf{b}})\right)^{\mathrm{T}}\left(c_2 + f(\mathbf{b},\tilde{\mathbf{b}})\right)\right) + \frac{1}{2}\left((\mathbf{Fb})^{\mathrm{T}}(\mathbf{Fb})\right)\right)}{\partial \mathbf{x}} \tag{6}$$

Step-4: Check for convergence

In this step, we check for termination criterion. If the criterion is not met, we set $k = k+1$; update the demand ($\mathbf{x_k}$) for the next iteration using ***Step-5*** (refer Equation 7). Else, go to ***Step-6***.

Step-5: OD matrix updating step

The Equation (7) presents the method to update OD vector from $\mathbf{x}_\mathrm{k}$ to $\mathbf{x}_{\mathrm{k+1}}$. This updating step involves search direction and step-size (λ). The search direction is determined by the gradient of $Z(\mathbf{x})$. On the other hand, the step-size (λ) parameter determines the number of iterations required for the convergence. Lower values of λ ensure that the path of the gradient is smooth but computationally expensive. Higher values of λ can lead to higher values of the objective function, and the convergence could be affected.

$$\mathbf{x}_{\mathrm{k+1}} = \mathbf{x}_\mathrm{k} \circ \left(e - \lambda_k \circ \frac{\partial Z(\mathbf{x})}{\partial \mathbf{x}}\right) \tag{7}$$

$$\lambda_k \circ \frac{\partial Z(\mathbf{x})}{\partial \mathbf{x}} < 1 \tag{7a}$$

In the Equation (7), e is vector of 1s and has dimensions same as $\mathbf{x}$. The Hadamard product "∘" is used for element wise multiplication between λ_k and the gradient, and $\mathbf{x}_\mathrm{k}$ and $\left(e - \lambda_k \circ \frac{\partial Z(\mathbf{x})}{\partial \mathbf{x}}\right)$. The optimum λ_k is calculated as the solution to Equation (6) in every iteration.

Step-6: Termination of the OD estimation process

In this step we terminate the OD estimation process based on maximum relative change in the elements of estimated OD flows at successive iterations (Maher et al., 2001), and value of $\mathbf{x_k}$ is the final estimated OD vector.

3. Experiments and results

The study site is for this research is the core of Brisbane city network and chosen from Behara et al. (2020c). The number of OD pairs were 210, and number of selected links with detectors were 24. The experiments in this study were conducted for four different formulations and four different cases as listed below.

The four different formulations include:

- Formulation-1: Only link flows were used in the upper-level formulation. This is the traditional method of OD estimation
- Formulation-2: A combination of link flows and micro-OD structural information in the upper-level formulation
- Formulation-3: A combination of link flows and macro-OD structural information in the upper-level formulation (refer to Equation (4))
- Formulation-4: A combination of link flows, macro-OD, and micro-OD structural information in the upper-level formulation (refer to Equation (5))

The four different cases represent different percentages of OD pairs for which partial OD flows from Bluetooth are available *a-priori* and they include:

- Case-1: 25% of total OD pairs
- Case-2: 50% of total OD pairs
- Case-3: 75% of total OD pairs
- Case-4: 100% of total OD pairs

The final OD estimates were assessed by comparing their structural similarity with that of true OD using Geographical Window based Structural Similarity Index (GSSI) as shown in Equation (8).

$$\mathrm{GSSI}(\mathbf{x}, \mathbf{x}^*) = \frac{1}{S}\sum_{s=1}^{s=S} \frac{(2\mu_{\mathbf{x_s}}\mu_{\mathbf{x^*}_\mathbf{s}} + \epsilon_1)(2\sigma_{\mathbf{x_s}\mathbf{x^*}_\mathbf{s}} + \epsilon_2)}{(\mu^2_{\mathbf{x_s}} + \mu^2_{\mathbf{x^*}_\mathbf{s}} + \epsilon_1)(\sigma^2_{\mathbf{x_s}} + \sigma^2_{\mathbf{x^*}_\mathbf{s}} + \epsilon_2)} \tag{8}$$

Where, S is the total number geographical windows[1] for the OD matrices; and $\mathbf{x_s}$ and $\mathbf{x^*}_\mathbf{s}$ are group of OD flows in s^{th} geographical window in $\mathbf{x}$ and $\mathbf{x}^*$, respectively. Mean and standard deviation of $\mathbf{x_s}$ and $\mathbf{x^*}_\mathbf{s}$ are $\mu_{\mathbf{x_s}}$, $\mu_{\mathbf{x^*}_\mathbf{s}}$, and $\sigma_{\mathbf{x_s}}$, $\sigma_{\mathbf{x^*}_\mathbf{s}}$; and covariance of $\mathbf{x_s}$ and $\mathbf{x^*}_\mathbf{s}$ is $\sigma_{\mathbf{x_s}\mathbf{x^*}_\mathbf{s}}$, respectively. For the present study the number of geographical windows (S) is assumed to be 1.

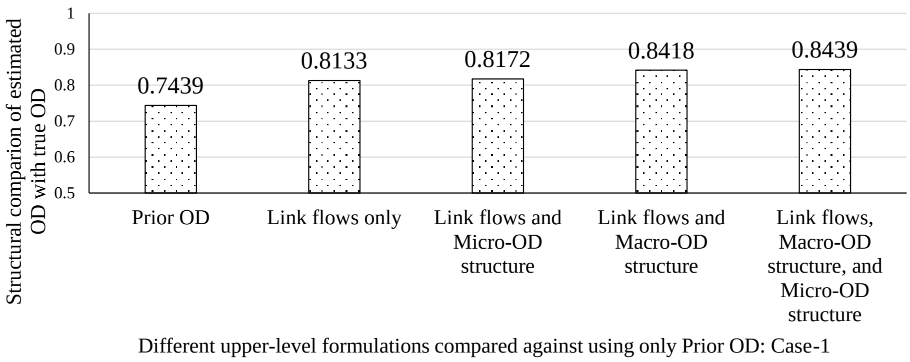
The Figure 4(a) presents $\mathrm{GSSI}(\mathbf{x}, \mathbf{x}^*)$ for different upper-level formulations provided a-priori knowledge about partial OD flows was available for 25% of the OD pairs. It is evident that the objective function that included link flows, macro-, and micro-OD structural information

[1] GSSI computes statistics on group of OD pairs belonging to same geographical windows. These windows are defined based on higher zonal level boundaries. Refer BEHARA, K. N., BHASKAR, A. & CHUNG, E. 2020b. A novel approach for the structural comparison of origin-destination matrices: Levenshtein distance. *Transportation Research Part C: Emerging Technologies,* 111**,** 513-530. for more details.

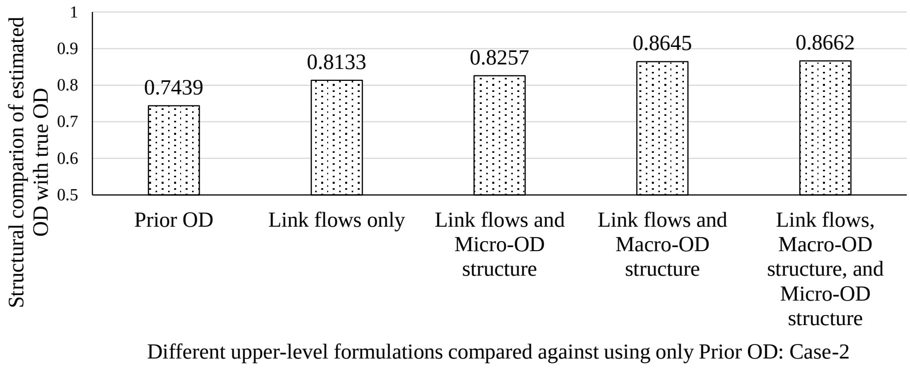
resulted in a better quality of OD estimates compared to other candidate formulations as well as the Prior OD.

The Figure 4(b) shows that as the percentage of OD pairs with *a-priori* partial OD flows increased from 25% to 50%, the improvement was even better. For instance, the Formulation-4 resulted in a $\text{GSSI}(\mathbf{x}, \mathbf{x}^*)$ of 0.8662 which was 2.64% improvement compared to 0.8439 using same formulation in Case-1.

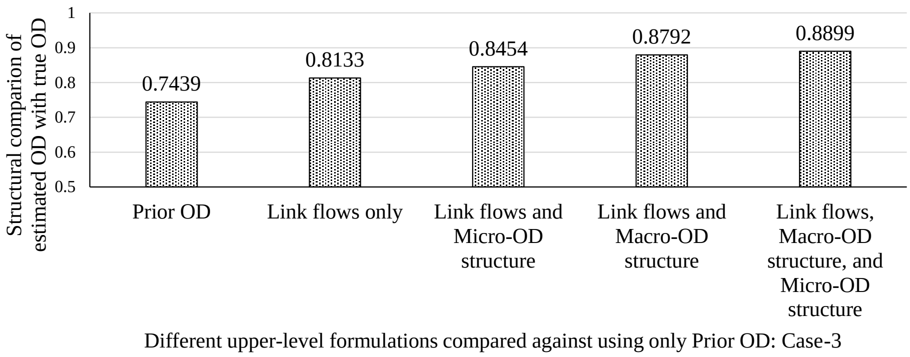
The Case-3 and Case-4 showed further enhancement in the quality of OD estimates as shown in Figure 4(c), and 4(d), respectively. Compared to Case-1, the percentage improvement in the quality of final OD matrices estimated using Formulation-4 in Case-3 and Case-4 were 5.33% and 9.54%, respectively. Therefore, higher the availability of *a-priori* partial OD flows, higher is the improvement in OD estimation.



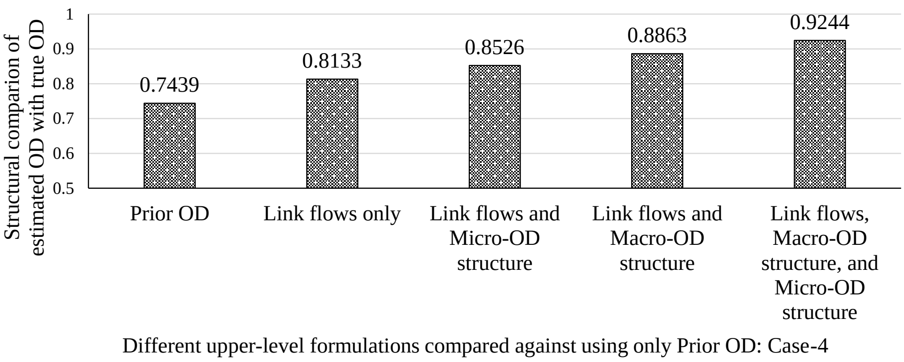
(a)



(b)



(c)



(d)

Figure 4: Results from different combination of upper-level formulations and percentages of a-priori partial OD structural information

4. Conclusion

One of the ways to enhance the quality of OD estimates is to simultaneously improve the structure of OD as well as the individual OD flows. The traditional method of matching link flows in the upper-level formulation fails to improve the quality of OD structure because mathematically the formulation is under-determinate. Other types of traffic information such as target OD contains important *structural* information; however, it is often constructed from outdated surveys. Various methods exist in literature to include partial OD flows in the upper-level formulation but with assumption that their market penetration rates are known. The sample rates of partial OD flows are not known in practice. Therefore, to improve the quality of OD estimates it is utmost important to exploit the *a-priori* knowledge of OD structural information to the maximum extent possible. To this end, this study introduces the concept of macro- and micro-OD structure and integrate them using a new upper-level formulation. The macro-OD structure is a high-level information captured using correlation coefficient. The micro-OD structure is at a more detailed level and captured through the ratio of OD flows from the same origin and ending into destinations with similar trip attraction characteristics. Integration of the micro-OD structure into the bi-level OD estimation formulation is the major contribution of this study. The proposed methodology was tested on a synthetic Brisbane network using partial OD flows (referred as sub-OD flows in this study) from Bluetooth observations. The findings revealed that using link flows, macro-OD structure, and micro-OD structure together in the upper-level formulation improved the quality of OD estimates compared to other candidate formulations.

References

AIMSUN 2019. Aimsun Next 8.4 User's Manual. *Aimsun Next 8.3.* Aimsun, Barcelona, Spain.

ALIBABAI, H. & MAHMASSANI, H. Dynamic origin-destination demand estimation using turning movement counts. Transportation Research Record: Journal of the Transportation Research Board, 2008. 39-48.

ANTONIOU, C., BARCELÓ, J., BREEN, M., BULLEJOS, M., CASAS, J., CIPRIANI, E., CIUFFO, B., DJUKIC, T., HOOGENDOORN, S. & MARZANO, V. 2016. Towards a generic benchmarking platform for origin–destination flows estimation/updating algorithms: Design, demonstration and validation. *Transportation Research Part C: Emerging Technologies,* 66**,** 79-98.

ANTONIOU, C., BEN-AKIVA, M. & KOUTSOPOULOS, H. 2004a. Incorporating automated vehicle identification data into origin-destination estimation. *Transportation Research Record: Journal of the Transportation Research Board***,** 37-44.

ANTONIOU, C., BEN-AKIVA, M. & KOUTSOPOULOS, H. N. Incorporating automated vehicle identification data into origin-destination estimation. Transportation Research Record, 2004b. 37-44.

ANTONIOU, C., BEN-AKIVA, M. & KOUTSOPOULOS, H. N. Dynamic traffic demand prediction using conventional and emerging data sources. IEE Proceedings-Intelligent Transport Systems, 2006. IET, 97-104.

BARCELÖ, J., MONTERO, L., MARQUÉS, L. & CARMONA, C. 2010. Travel time forecasting and dynamic origin-destination estimation for freeways based on bluetooth traffic monitoring. *Transportation research record,* 2175**,** 19-27.

BEHARA, K. N., BHASKAR, A. & CHUNG, E. 2020a. Geographical window based structural similarity index for origin-destination matrices comparison. *Journal of Intelligent Transportation Systems***,** 1-22.

BEHARA, K. N., BHASKAR, A. & CHUNG, E. 2020b. A novel approach for the structural comparison of origin-destination matrices: Levenshtein distance. *Transportation Research Part C: Emerging Technologies,* 111**,** 513-530.

BEHARA, K. N., BHASKAR, A. & CHUNG, E. 2020c. A novel methodology to assimilate sub-path flows in bi-level OD matrix estimation process. *IEEE Transactions on Intelligent Transportation Systems***,** 1-11.

BEHARA, K. N., BHASKAR, A. & CHUNG, E. 2021. Single-level approach to estimate origin-destination matrix: exploiting turning proportions and partial OD flows. *Transportation Letters***,** 1-12.

BIERLAIRE, M. & TOINT, P. L. 1995. Meuse: An origin-destination matrix estimator that exploits structure. *Transportation Research Part B: Methodological,* 29**,** 47-60.

CALABRESE, F., DI LORENZO, G., LIU, L. & RATTI, C. Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area. IEEE Pervasive Computing, 2011.

CASCETTA, E. 1984. Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. *Transportation Research Part B: Methodological,* 18**,** 289-299.

CASCETTA, E. & NGUYEN, S. 1988. A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research Part B: Methodological,* 22**,** 437-455.

CASCETTA, E. & POSTORINO, M. N. 2001. Fixed point approaches to the estimation of O/D matrices using traffic counts on congested networks. *Transportation science,* 35**,** 134-147.

CHU, K.-F., LAM, A. Y. & LI, V. O. 2019. Deep multi-scale convolutional LSTM network for travel demand and origin-destination predictions. *IEEE Transactions on Intelligent Transportation Systems*.

DIXON, M. P. & RILETT, L. Real-Time OD Estimation Using Automatic Vehicle Identification and Traffic Count Data. Computer-Aided Civil and Infrastructure Engineering, 2002. 7-21.

DJUKIC, T. 2014. *Dynamic OD demand estimation and prediction for dynamic traffic management.* PhD Thesis, TU Delft, Delft University of Technology.

HUSSAIN, E., BEHARA, K. N., BHASKAR, A. & CHUNG, E. 2021. A Framework for the Comparative Analysis of Multi-Modal Travel Demand: Case Study on Brisbane Network. *IEEE Transactions on Intelligent Transportation Systems*.

IQBAL, M. S., CHOUDHURY, C. F., WANG, P. & GONZÁLEZ, M. C. 2014. Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies,* 40**,** 63-74.

JAUME, B. & MONTERO, L. An integrated Computational framework for the estimation of dynamic OD trip matrices. 2015 IEEE 18th International Conference on Intelligent Transportation Systems, 2015. IEEE, 612-619.

KIM, H., BAEK, S. & LIM, Y. 2001. Origin-destination matrices estimated with a genetic algorithm from link traffic counts. *Transportation Research Record: Journal of the Transportation Research Board***,** 156-163.

LIU, L., QIU, Z., LI, G., WANG, Q., OUYANG, W. & LIN, L. 2019. Contextualized Spatial–Temporal Network for Taxi Origin-Destination Demand Prediction. *IEEE Transactions on Intelligent Transportation Systems,* 20**,** 3875-3887.

MAHER, M. J., ZHANG, X. & VAN VLIET, D. 2001. A bi-level programming approach for trip matrix estimation and traffic control problems with stochastic user equilibrium link flows. *Transportation Research Part B: Methodological,* 35**,** 23-40.

MICHAU, G., PUSTELNIK, N., BORGNAT, P., ABRY, P., NANTES, A., BHASKAR, A. & CHUNG, E. 2017a. A primal-dual algorithm for link dependent origin destination matrix estimation. *IEEE Transactions on Signal and Information Processing over Networks,* 3**,** 104-113.

MICHAU, G., PUSTELNIK, N., BORGNAT, P., ABRY, P., NANTES, A., BHASKAR, A. & CHUNG, E. A primal-dual algorithm for link dependent origin destination matrix estimation. IEEE Transactions on Signal and Information Processing over Networks, 2017b. 104-113.

MUNGTHANYA, W., PHITHAKKITNUKOON, S., DEMISSIE, M. G., KATTAN, L., VELOSO, M., BENTO, C. & RATTI, C. 2019. Constructing time-dependent origin-destination matrices with adaptive zoning scheme and measuring their similarities with taxi trajectory data. *IEEE Access,* 7**,** 77723-77737.

NASAB, M. R. & SHAFAHI, Y. 2019. Estimation of origin–destination matrices using link counts and partial path data. *Transportation***,** 1-28.

VAN ZUYLEN, H. J. & WILLUMSEN, L. G. 1980. The most likely trip matrix estimated from traffic counts. *Transportation Research Part B: Methodological,* 14**,** 281-293.

YANG, H. 1995. Heuristic algorithms for the bilevel origin-destination matrix estimation problem. *Transportation Research Part B: Methodological,* 29**,** 231-242.

YANG, H., SASAKI, T., IIDA, Y. & ASAKURA, Y. 1992. Estimation of origin-destination matrices from link traffic counts on congested networks. *Transportation Research Part B: Methodological,* 26**,** 417-434.

YANG, X., LU, Y. & HAO, W. 2017. Origin-destination estimation using probe vehicle trajectory and link counts. *Journal of Advanced Transportation,* 2017.