Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Estimating OD matrices from observed trajectories and link counts

Krishna N.S. Behara[a], Ashish Bhaskar[a], Edward Chung[b]

[a]School of Civil Engineering and Built Environment, Queensland University of Technology, Brisbane, QLD 4000, Australia
[b]Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong

Abstract

Origin-Destination demand estimation has predominantly depended on observed link counts from loop detectors. Although they are convenient to use and cost-efficient, relying on limited known variables (i.e. traffic counts) to estimate a far greater number of OD variables leads to the problem of under-determinacy in OD estimation. Due to this, the quality of OD estimates cannot be much improved unless there is some prior knowledge of structural information. Target OD matrices have been previously used as a source of structural information in the past but they are generally based on outdated surveys. However, with the advent of emerging technologies such as Bluetooth, it is possible to get both loop counts and prior structural information from the same time period. In this light, the study proposes a methodological approach to incorporate prior structural knowledge from observed trajectories as an additional goal function within traditional link-counts based estimation. The study demonstrates a framework to exploit Bluetooth trajectories from synthetic Brisbane network in Aimsun controlled environment. Since the penetration rates of Bluetooth observations is random and unknown, the analysis is performed for different penetration rates and the results highlight some interesting findings on the minimum penetration required for significant improvement in the quality of OD estimates.

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.

Keywords: OD matrix estimation;link counts;trajectories;structural comparison;Bluetooth;Brisbane city

1. Introduction

Origin-Destination (OD) matrix estimation has always been the topic of research in transport modelling. OD matrices for large network can't be measured with limited sensors on the network. A panorama of research techniques have been developed in the last three decades to estimate OD matrices (Van Zuylen and Willumsen 1980; Cascetta 1984; Yang et al. 1992; Antoniou, Ben-Akiva and Koutsopoulos 2004; Michau, Pustelnik, et al. 2017). Traditionally, OD estimation is based on observed traffic counts, a prior OD matrix and user equilibrium assignment (derived from analytical or simulation model). Since, the number of OD pairs to be estimated is far greater than the number of observed traffic counts, it leads to a problem of under-determinacy. 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).

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY

The *structure* of travel demand for any region is defined as the distribution pattern of travel demand along different paths between different OD pairs. OD matrices have similar structure if the travel pattern distribution, represented either by OD matrices or trajectories of trips, is similar. In OD estimation, the quality of OD estimates depends on how efficiently the structural consistency is maintained within the OD matrices that are updated iteratively (Bierlaire and Toint 1995).

In the initial days, the problem of under-determinacy is minimized by introducing target OD matrix within the objective function (Yang 1995; Cascetta and Nguyen 1988). It was assumed that an *a priori* (target) matrix contains important *structural* information in terms of distribution of travel demand across the region. Since actual OD matrix is unobserved, the structural consistency is traditionally compared between target and estimated matrices. However, by doing so, 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.

With the advent of emerging data sources such as Mobile phones, Bluetooth MAC scanners (BMS) etc. it is possible to get partial but seamless traffic data at a larger spatio-temporal scales. Exploiting this rich data from pervasive data sources could relax the only dependence on conventional techniques of OD estimation. Brisbane City Council (BCC) region is equipped with almost 1200 Bluetooth Mac Scanners (BMS) that are installed for area-wide measuring and monitoring traffic state situations. They are primarily meant to estimate travel times and travel speeds in the congested network conditions. However, they also provide rich source of vehicle trajectories that can further be used to estimate good quality OD estimates. Fig. 1 shows the overlay of BMS on BCC region extracted from Google Earth (2018).

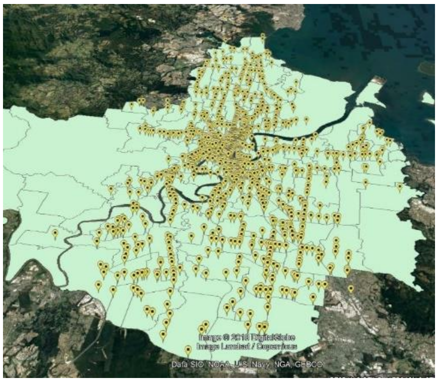

Fig. 1: Bluetooth spatial coverage for Brisbane City Council region

The trajectories of vehicles, observed at larger spatio-temporal scale, provide rich structural information of travel demand (Michau, Nantes, et al. 2017). To preserve structural consistency in OD estimation for improving quality of estimation, either they can be used as a constraint in solution algorithm or as an additional goal within the traditional link-counts based formulation. Since the penetration rate of Bluetooth is quite random (varies with time-of-the day, locations, distance and other factors) and unknown, it is not tangible to scale observed Bluetooth path flows to match with that of estimated ones. Recently, Michau et al. (2015) have shown that the ratio of Bluetooth to loop counts in Brisbane varies between 22 and 30%. Michau et al. (2017) assumed a penetration rate drawn from a Gaussian distribution of mean 30% and standard deviation of 10% for scaling up observed Bluetooth link OD demands. However, it is to be noted that, penetration rate of Bluetooth with respect to traffic counts is different from that with respect to OD demand and high penetration rate of the former might not guarantee high penetration of the latter. Thus, instead of assuming some random penetration rates and applying to observed path flows individually, we consider a macro-level approach by considering all path flows simultaneously. Here, we propose to include *structural comparison* of observed and estimated path flows for different Bluetooth penetration rates. It is assumed that the structure of observed trajectories represents the structure of actual unobserved trajectories and our hypothesis is that "the quality of OD estimates improve with higher penetration rates of Bluetooth observations".

In light of the above-mentioned points, the objectives of this study are:

1. To develop an objective function that can incorporate additional structural information of observed trajectories in OD estimation formulation. Here, the prior structural information is obtained from partial observed Bluetooth path flows, thus ensuring less dependence on target OD matrix.
2. To test the hypothesis that with increase in Bluetooth penetration rate the quality of OD estimates using vehicle trajectories from BMS improve. Since Bluetooth penetration is unknown in reality, the study investigates the optimum penetration rate of trajectories required to improve quality of OD estimation.

2. Literature review

This section mainly focusses on two areas: Firstly, it discusses a few studies that emphasized on the importance of prior structural information of travel demand in OD estimation. Secondly, it reviews some interesting attempts in the past that introduced trajectories information in OD estimation framework.

2.1 Prior structural knowledge of travel demand

In literature, very few studies have tried to incorporate the knowledge of prior structural information for improving the quality of OD estimates. Bierlaire and Toint (1995) pointed out that, the structural information of parking surveys can be used to improve the structure of estimated OD matrices. Since OD matrix is unobserved, structural variation within OD flows are analyzed from the variations observed in link flows. For instance, Willumsen (1984) introduced scale factor (θ) as a method to analyze structural variation of OD flows. Here, θ is computed as total sum of ratios of observed and estimated link counts averaged over all observed links. Yang et al. (1992) used correlation coefficient (ρ) between observed and estimated link flows in addition to scale factor (θ) to explain the structural degradation of OD flows. If X1 and X2 are two OD matrices to be compared, then four different cases are possible from the combinations of θ and ρ as follows:

- Case-1: If $\theta \to 1$ and $\rho \to 1$, the both X1 and X2 are structurally similar to each other.
- Case-2: If $\theta \to 1$ and ρ is small, then X1 and X2 are structurally different with random variations.
- Case-3: If $\theta > 1$ or $\theta < 1$ and $\rho \to 1$, then X1 and X2 have the same structure but total demand in X1 is greater or lower as compared to X2.
- Case-4: If $\theta > 1$ and ρ is small, then there are structurally different at a larger random scale.

The limitation of this approach is that, the statistical indicators (θ and ρ) are comparing link flows to interpret the structural variation in OD matrices that is generally not true because of many to one relationship between them. Kim et al. (2001) defined OD Matrix Structure (ODMS) as the ratio of OD flows to origin flows and used it as a constraint to preserve the structure of OD demand while updating iteratively. Recently, Behara et al. (2017) and Behara et al. (2018) have developed geographical window based SSIM technique for comparing structural similarity of OD matrices, that can further be used in OD estimation techniques for preserving structural consistency of OD matrices.

2.2 Knowledge of observed trajectories in OD demand estimation

Many other studies have focused on improving the quality of OD estimates using observed vehicle trajectories from GPS, Mobile phones, e-tags and Bluetooth etc. For instance, Zipp (1997) proposed to use partial observations of vehicle trajectories in traditional traffic counts-based OD estimation. Here, license plate readers are used for automatic vehicle identification on a motorway corridor and not applicable for networks involving route-choice. Kwon and Varaiya (2005) used sample trajectories from electronic tags and developed statistical methods for better OD estimates but limited to application on freeway networks. Zhang et al.(2010) used cellular probe trajectories to identify the trip-ends and then converted cell-based trips to vehicle trips. Michau et al. (2017) proposed link-dependent OD (LOD) estimation using Bluetooth trajectories and traffic counts. The trajectories are represented in the objective function through LOD formulation. While the new data sources discussed above provide direct travel information in a convenient way, they may not provide detailed contextual and demographic information. Nevertheless they provide seamless data for longer time periods and can be the future of transport modelling if efficient data cleansing and correction mechanisms are implemented.

3. Methodology

The methodology section is further divided into six sub-sections. Firstly, the study site and data description are explained. Next, the proposed approach is explained through conceptual framework, design of experiments, development of objective function, formulation of optimization and finally an approach to check the quality of prior OD and estimated OD matrices from all experiments as compared to true OD is described.

3.1 Study site and data

The analysis in this study is performed in a controlled environment. A synthetic Brisbane city network is built from open street map imported into Aimsun (Fig. 2) that comprises of 15 centroids, 24 loop detectors (red squares in Fig. 2a) and 51 Bluetooth scanners (blue circles in Fig. 2a). The OD matrix is designed at a zonal level equivalent to Statistical Area 2 (SA2) (ASGS 2017) and is of size 15 x 15. Refer Fig. 2b for the spatial structure of the core of Brisbane city with neighborhood suburbs and primary transport network.

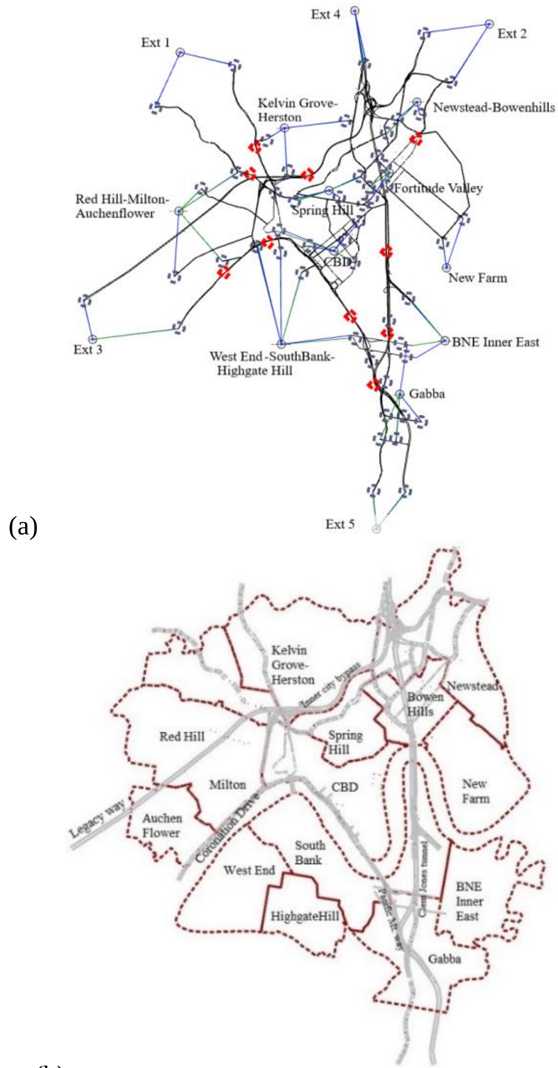

Fig. 2: (a) Study site installed with Bluetooth scanners and loop detectors (b) Spatial structure of Brisbane city core network

3.2 Conceptual framework

The notations used in this study are as follows:

Nomenclature

$\boldsymbol{X}$	OD vector that needs to be estimated *(N*N)*
$\widehat{\boldsymbol{Y}}$	Observed link counts (*L*1*)
$\boldsymbol{Y}$	Estimated link counts (obtained as output of assigning $\boldsymbol{X}$ on the network in Aimsun) (*L*1*)
$\widehat{\boldsymbol{T}}$	Partially observed Bluetooth path flows (*M *1)*
$\widehat{\boldsymbol{P}}$	Total actual path flows (*K*1*)
$\boldsymbol{P}$	Total estimated path flows (*K*1*)
$\widetilde{\boldsymbol{P}}$	Estimated path flows of those paths that have Bluetooth trajectories information (*M*1)*
$\boldsymbol{A}$	Link-proportion matrix that are extracted from Aimsun (*L*N*)
$\boldsymbol{B}$	Path-proportion matrix that are extracted from Aimsun (*M*N)*
Str ($\widehat{\boldsymbol{T}}, \widetilde{\boldsymbol{P}}$)	Structural comparison of observed ($\widehat{\boldsymbol{T}}$) and estimated path flows ($\widetilde{\boldsymbol{P}}$) is a value between -1 and 1

The flowchart in Fig. 3 illustrates the difference between traditional and proposed approaches for OD matrix estimation. In the traditional approach, upper level minimizes the deviation between observed and estimated link flows and user equilibrium traffic assignment runs in the lower level of bi-level formulation. Similar framework is suggested in the proposed approach but with additional structural information in the upper-level formulation. Here, the structural comparison of partially observed Bluetooth path flows ($\widehat{\boldsymbol{T}}$) and estimated path flows ($\widetilde{\boldsymbol{P}}$) is used in addition to link counts deviation.

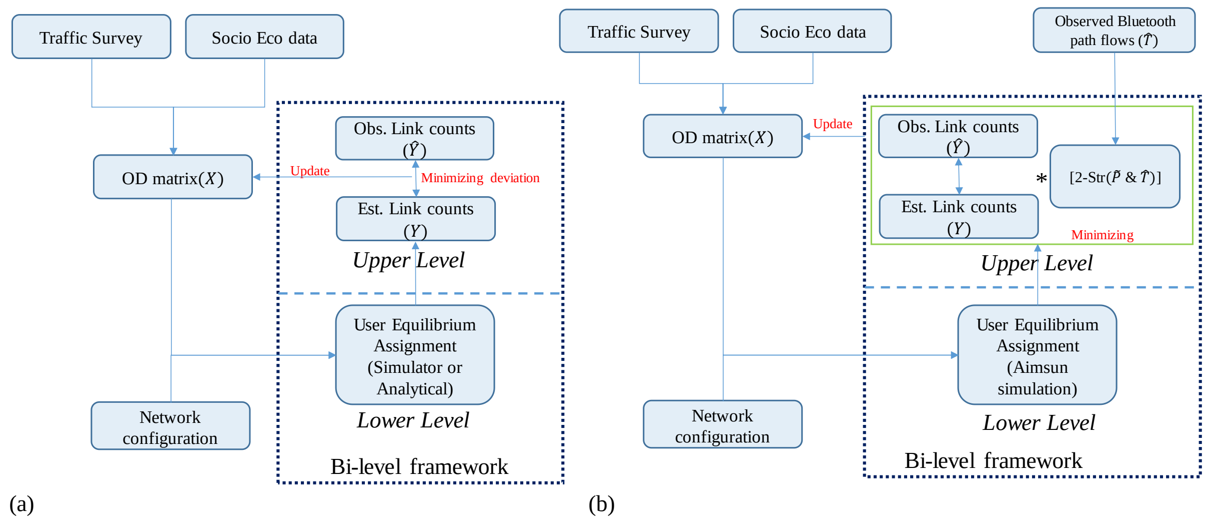

Fig. 3: (a) Traditional vs (b) Proposed approach

3.3 Experiments

Prior OD matrix ($\boldsymbol{X_{prior}}$) is often obtained from outdated surveys. To represent this situation, we have designed it by randomly perturbing true OD as shown in equation (1). Here, *rand ()* function ensures that prior OD varies randomly between 0.7 to 1.0 times true OD demand ($\boldsymbol{X_{true}}$).

$$Xprior = [0.7 + 0.3 * rand()] * Xtrue \tag{1}$$

Proposed approach is tested for different penetration rates of observed Bluetooth trajectories. The penetration rates (η) chosen for this study are 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1 respectively. The most likely paths per OD pair is also a design parameter and is chosen as maximum "1" per OD pair in this study. Randomness in Bluetooth observations is accounted by randomly choosing the corresponding "η" proportion of total trajectories obtained from the assignment of true OD matrix onto Brisbane network. For instance, the total number of actual trajectories for the study network are 5273 and "η" of 0.1 implies that 527 trajectories are chosen randomly from the total actual trajectories.

3.4 Objective function formulation

The optimization formulation comprises of two sub-functions: The deviation of link flows and structural comparison of trajectories. The structural comparison between partially observed Bluetooth ($\hat{\boldsymbol{T}}$) and estimated path flows ($\tilde{\boldsymbol{P}}$) is performed by using Pearson correlation coefficient (referred as *Str* ($\hat{\boldsymbol{T}}$, $\tilde{\boldsymbol{P}}$)) as shown in equations (2a and 2b).

$$Str(\hat{T},\tilde{P}) = \frac{\sum\left((\hat{T}-\mu_{\hat{T}})*(\tilde{P}-\mu_{\tilde{P}})\right)}{\sqrt{\left(\sum(\hat{T}-\mu_{\hat{T}})^2\right)*\left(\sum(\tilde{P}-\mu_{\tilde{P}})^2\right)}} \tag{2a}$$

Where, $\tilde{P} = B * X$ (2b)

In equation (2a), $\mu_{\hat{T}}$ and $\mu_{\tilde{P}}$ are mean of the vectors $\hat{\boldsymbol{T}}$ and $\tilde{\boldsymbol{P}}$ respectively and equation (2b) shows the relationship between path flows ($\tilde{\boldsymbol{P}}$) and OD matrix $\boldsymbol{X}$ using path proportion matrix $\boldsymbol{B}$. The range of *Str*($\hat{\boldsymbol{T}}$, $\tilde{\boldsymbol{P}}$) values lies between -1 and 1. For instance, if structural similarity of trajectories are exactly the same then it has a maximum value of 1 and the extreme opposite case results in -1.

The traditional way of expressing link counts deviation within the objective function is shown in Equation (3).

$$Z = Min\frac{1}{2}\left[Y-\hat{Y}\right]^2 = Min\frac{1}{2}\left[A*X-\hat{Y}\right]^2 \tag{3}$$

The proposed objective function after including the structural comparison of trajectories is shown in Equation (4).

$$Z = Min\frac{1}{2}\left[(Y-\hat{Y})*(2-Str(\hat{T},\tilde{P}))\right]^2 = Min\frac{1}{2}\left[(A*X-\hat{Y})*(2-Str(\hat{T},\tilde{P}))\right]^2 \tag{4}$$

Since *Str*($\hat{\boldsymbol{T}}$, $\tilde{\boldsymbol{P}}$) is a similarity value, it has to be maximized while link counts deviation is minimised. An alternative combination of both functions is possible by minimising | ($\boldsymbol{Y}$ -$\hat{\boldsymbol{Y}}$)*(2-*Str*($\hat{\boldsymbol{T}}$, $\tilde{\boldsymbol{P}}$))|. The stability of this combined formulation is explained as follows: when the structure of two trajectories is same, *Str*($\hat{\boldsymbol{T}}$, $\tilde{\boldsymbol{P}}$) is equal to "1" and objective function reduces to traditional link counts deviation i.e. | ($\boldsymbol{Y}$ -$\hat{\boldsymbol{Y}}$)|. On the other hand, when *Str* reaches its minimum value of "-1", objective function becomes | ($\boldsymbol{Y}$ -$\hat{\boldsymbol{Y}}$)*3|.

3.5 Optimization algorithm

The optimisation method chosen to minimise the objective function in equation (4) is gradient descent algorithm (Spiess 1990). Here, gradient of the equation (4) is computed using equations (5a-5d).

$$\frac{dZ}{dX} = \frac{d\left([(A*X-Y)*(2-Str(\hat{T},\tilde{P}))]^2 \big/ 2 \right)}{dx} \tag{5a}$$

$$\frac{dZ}{dX} = \left[(2 - Str(\hat{T}, \tilde{P})) * A^T - \frac{d(Str(\hat{T}, \tilde{P}))}{dX} * (A * X - Y)^T \right] * \left[(A * X - Y) * (2 - Str(\hat{T}, \tilde{P})) \right] \quad (5b)$$

Where,

$$Str(\hat{T}, \tilde{P}) = \frac{\Sigma\left((\hat{T} - \mu_{\hat{T}}) * (B * X - \mu_{B*X}) \right)}{\sqrt{\left(\Sigma(\hat{T} - \mu_{\hat{T}})^2 \right) * \left(\Sigma(B * X - \mu_{B*X})^2 \right)}} = \frac{\mathbb{R}}{\sqrt{\mathbb{C} * \mathbb{Q}}} \quad (5c)$$

$$\frac{d(Str(\hat{T}, \tilde{P}))}{dX} = \frac{B^T * \left((\hat{T} - \mu_{\hat{T}}) - \frac{\mathbb{R}}{\mathbb{Q}} * (B * X - \mu_{B*X}) \right)}{\sqrt{\mathbb{C} * \mathbb{Q}}} \quad (5d)$$

Convergence criteria is defined as the ratio of difference in Z values of consecutive iterations (i.e. Z_{i-1}- Z_i) and the value in previous iteration (i.e. Z_{i-1}) and its value is chosen as 0.05 in this study. Until convergence, OD matrix (***X***) is updated iteratively using equation (6). Here, X_i and X_{i-1} are the OD matrices in i_{th} and $(i-1)_{th}$ steps of iterative estimation and the rate of convergence depends on the value of step-length/learning rate (λ).

$$X_i = X_{i-1} * \left[1 - \lambda * \frac{dZ}{dX} \right] \qquad \forall \;\; \lambda * \frac{dZ}{dX} < 1 \quad (6)$$

To facilitate smooth convergence, we propose to use bold-driver technique that is commonly used in annealing the learning rate (Vogl et al. 1988; Battiti 1989). According to this approach, if the value of objective function in the $(i-1)^{th}$ step is less than that of value in the current step, i (i.e. $Z_{i-1} < Z_i$) then $\lambda_i = \lambda_{i-1}$*1.05. Otherwise, reset the optimization parameters (i.e. ***X***) to that of $(i-1)^{th}$ iteration and set $\lambda_i = \lambda_{i-1}/2$. The value of λ should always abide by the condition as shown in equation (6b).

The optimisation algorithm is coded in MATLAB and is run for different penetration rates of Bluetooth observations. The sequence of steps for any penetration rate (η) is explained as follows:

- *Step 1:* Load study network with true OD demand ($\boldsymbol{X_{true}}$), run the simulation and retrieve the observed link counts ($\hat{Y}$) and unique vehicle trajectories with path flows ($\hat{P}$)
- Step 2: Choose prior OD demand (***X***)
- *Step 2:* Load the study network with demand ***X*** in Aimsun environment and run the Dynamic User Equilibrium (DUE) assignment
- *Step 3:* Retrieve the link flows (***Y***) and trajectories with their corresponding path flows (***P***)
- *Step 4:* For the penetration rate of η, let's say the observed Bluetooth path flows are $\boldsymbol{\hat{T}}$ and their corresponding estimated path flows from Aimsun are $\boldsymbol{\tilde{P}}$
- *Step 5:* Compute the gradient of proposed objective function using equation (5a-5d)
- *Step 6:* Update the demand (***X***) using equation (6)
- *Step 7:* Check for convergence criteria and if it is not met GOTO *Step 2*. Else terminate the optimisation and value of ***X*** at convergence is $\boldsymbol{X_{est}}$.

3.6 Quality check of estimated OD matrices

The quality of estimated OD matrices from all experiments are compared using the two statistical measures namely Root Mean Square Error (RMSE) and Str_{OD}. The computation of RMSE between $\boldsymbol{X_{true}}$ and $\boldsymbol{X_{est}}$ is shown in equation (7). The formulation for Str_{OD} is also based on Pearson correlation coefficient and it compares the structure of any two OD matrices. For structural comparison of true OD matrix ($\boldsymbol{X_{true}}$) and estimated OD matrix ($\boldsymbol{X_{est}}$), Str_{OD} is expressed as shown in equation (8). Here, $\mu_{X_{true}}$ and $\mu_{X_{est}}$ are the mean demand values of true ($\boldsymbol{X_{true}}$) and estimated ($\boldsymbol{X_{est}}$) OD demand vectors.

$$RMSE(x_{true}, x_{est}) = \sqrt{\frac{\sum (x_{true} - x_{est})^2}{N}} \quad (7)$$

$$Str_{OD}(X_{true}, X_{est}) = \frac{\Sigma\left((X_{true} - \mu_{X_{true}}) * (X_{est} - \mu_{X_{est}})\right)}{\sqrt{\left(\Sigma(X_{true} - \mu_{X_{true}})^2\right) * \left(\Sigma(X_{est} - \mu_{X_{est}})^2\right)}} \quad (8)$$

4. Results and discussion

The results of experiments conducted for different penetration rates are illustrated through Fig. 4-Fig. 7. The discussion of the results is based on the scores of RMSE and Str_{OD} averaged over three replications per each experiment and are explained as follows:

1. *Traditional approach can lead to structural degradation of OD estimates*: It is important to note that, although RMSE values (Fig. 4) for traditional approach shows improvement as compared to Prior OD (from 14.02 to 12.42) but the structure of OD matrix i.e. Str_{OD} (refer Fig. 5) has degraded from 0.8142 to 0.8091. From Fig. 6, it is illustrated that the structural degradation is -0.63% as compared to chosen Prior OD. This highlights the problem of under-determinacy for OD estimated in the traditional methods that is based on deviation of link counts only.
2. *Prior structural knowledge improves quality of OD estimates*: By using prior structural knowledge from Bluetooth trajectories, it is observed that the quality of OD estimates improved significantly as the Bluetooth penetration rate increase. For instance, RMSE values (Fig. 4) and Str_{OD} values (Fig. 5) for η=0.1 to 1.0 are improved from 11.84 to 10.84 and from 0.8339 to 0.8591 respectively.
3. *Marginal improvements after a certain Bluetooth penetration rate:* As compared to traditional approach, the percentage improvement in RMSE is 4.67% and Str_{OD} is 3.07% for η = 0.1 and they improve to 12.75% and 6.18% respectively for η = 1 (refer Fig. 7). However, marginal improvements in RMSE are observed for penetration rates (η) from 0.4 to 0.6 and from 0.7 to 1.0 (refer RMSE plot in Fig. 7). And the percentage improvement for Str_{OD} is almost stable for η greater than 0.4 (refer Str_{OD} plot in Fig. 7). This implies, minimum penetration rate of η =0.4 is good enough to exploit structural information of trajectories to improve the quality of OD estimates. This also highlights some potential financial benefits such as: it helps in minimising the infrastructure cost by identifying the locations and optimum number of BMS that are required to be installed in order to capture 40% of actual trajectories.
4. *Proposed approach works efficiently well with limited trajectories data*: It is important to mention here that Str_{OD} formulation is a macroscopic level approach and applying Str_{OD} on a small random sample of trajectories yields almost similar results as that of complete information of trajectories. This is reflected from the marginal improvement in quality of OD estimates for η > 0.4.

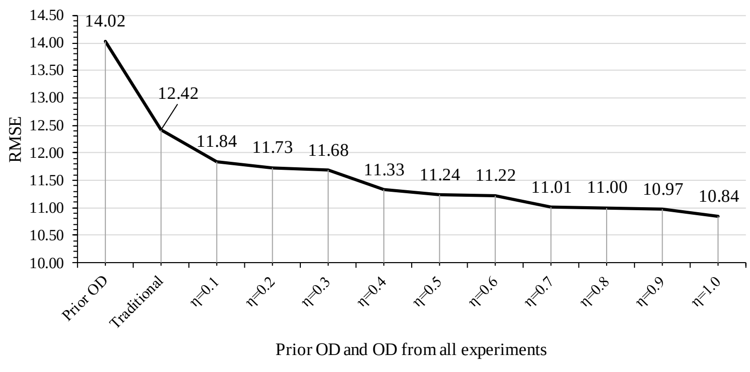

Fig. 4: RMSE scores of experiments as compared to that of Prior OD

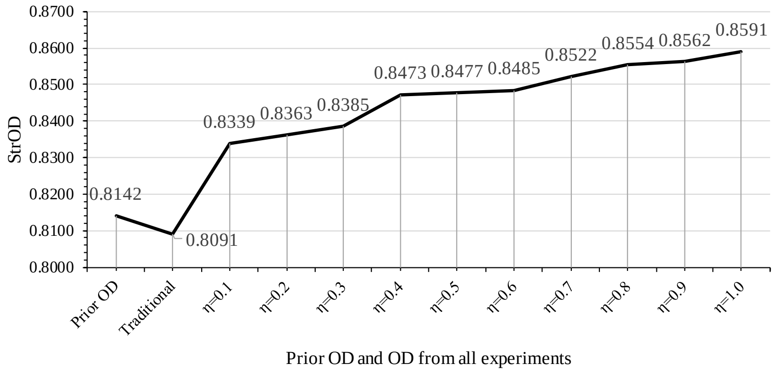

Fig. 5: Str$_{OD}$ scores of experiments as compared to that of Prior OD

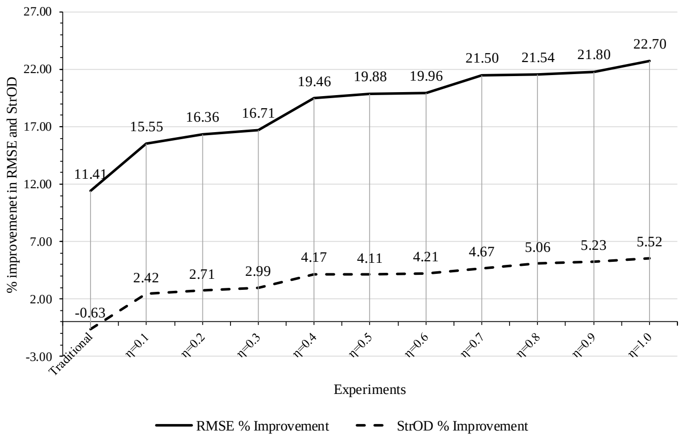

Fig. 6: Percentage improvement in RMSE and Str$_{OD}$ as compared to Prior OD

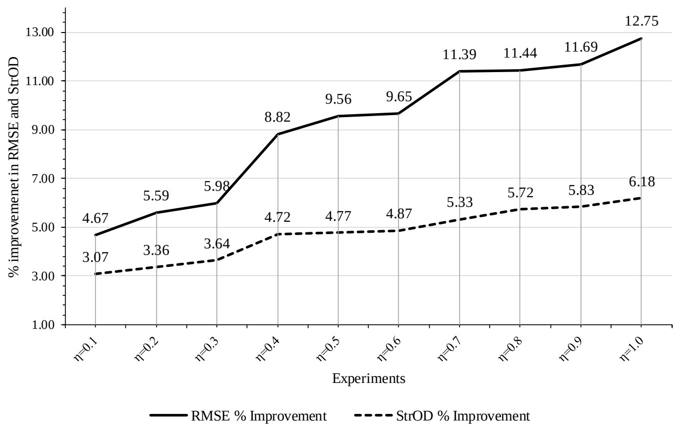

Fig. 7: Percentage improvement in RMSE and StrOD as compared to OD from traditional method

5. Conclusion

The study proposes an approach to improve the quality of OD estimates by using prior structural knowledge of trajectories information in addition to observed link counts. It presents a case study on synthetic Brisbane network with link counts data obtained from loop detectors and trajectories information retrieved from Bluetooth Mac Scanners. The two major contributions of this study are:

Firstly, it proposes a macro-level *structural* comparison technique an alternative approach to exploit observed trajectories of unknown penetration rates for large scale urban networks. It proves that use of prior structural information from trajectories helps improve the quality of OD estimates as compared to traditional approach.

Secondly, it tests the approach for different penetration rates and suggests the minimum penetration rate of trajectories information required to achieve better quality of OD estimates. This implies that path flows from a few but critical paths such as motorways and major arterials is sufficient enough to improve the quality of OD estimates.

The proposed methodology demonstrates exploitation of trajectories information from Bluetooth data sets because Brisbane City has wide spatial coverage of BMS. However, the approach is generic in nature and can incorporate trajectories knowledge from any other data sources such as GPS, Mobile phone etc. The study is based on one typical case of Prior OD and it is a proof of concept. As a part of future work, thorough testing and sensitivity analysis is needed.

Acknowledgements

The authors are thankful to Brisbane City Council and Queensland University of Technology, Brisbane for providing the Bluetooth information for Brisbane city region. The conclusions of this paper reflect understandings of the authors, who are responsible for the accuracy of the model.

References

Antoniou, Constantinos, Jaume Barceló, Martijn Breen, Manuel Bullejos, Jordi Casas, Ernesto Cipriani, Biagio Ciuffo, Tamara Djukic, Serge Hoogendoorn and Vittorio Marzano. 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, Constantinos, Moshe Ben-Akiva and Haris Koutsopoulos. 2004. "Incorporating automated vehicle identification data into origin-destination estimation." *Transportation Research Record: Journal of the Transportation Research Board* (1882): 37-44.

ASGS. 2017. "Australian Statistical Geography Standard (ASGS)." http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+(ASGS).

Battiti, Roberto. 1989. "Accelerated backpropagation learning: Two optimization methods." *Complex systems* 3 (4): 331-342.

Behara, Krishna NS, Ashish Bhaskar and Edward Chung. 2017. "Insights into geographical window based SSIM for comparison of OD matrices." In *Australasian Transport Research Forum 2017 Proceedings, Auckland, New Zealand, 27 – 29 November 2017*, edited.

Behara, Krishna NS, Ashish Bhaskar and Edward Chung. 2018. "Classification of typical Bluetooth OD matrices based on structural similarity of travel patterns-Case study on Brisbane city." In *Transportation Research Board 97th Annual Meeting, Washington DC, United States, 7- 11 January 2018*, edited.

Bierlaire, Michel and Ph L Toint. 1995. "Meuse: An origin-destination matrix estimator that exploits structure." *Transportation Research Part B: Methodological* 29 (1): 47-60.

Cascetta, Ennio. 1984. "Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator." *Transportation Research Part B: Methodological* 18 (4-5): 289-299.

Cascetta, Ennio and Sang Nguyen. 1988. "A unified framework for estimating or updating origin/destination matrices from traffic counts." *Transportation Research Part B: Methodological* 22 (6): 437-455.

Google. 2018. "Brisbane City Council region".

Kim, Hyunmyung, Seungkirl Baek and Yongtaek Lim. 2001. "Origin-destination matrices estimated with a genetic algorithm from link traffic counts." *Transportation Research Record: Journal of the Transportation Research Board* (1771): 156-163.

Kwon, Jaimyoung and Pravin Varaiya. 2005. "Real-time estimation of origin-destination matrices with partial trajectories from electronic toll collection tag data." *Transportation Research Record: Journal of the Transportation Research Board* (1923): 119-126.

Michau, Gabriel, Pierre Borgnat, Nelly Pustelnik, Patrice Abry, Alfredo Nantes and Edward Chung. 2015. "Estimating link-dependent origin-destination matrices from sample trajectories and traffic counts." In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, edited, 5480-5484: IEEE.

Michau, Gabriel, Alfredo Nantes, Ashish Bhaskar, Edward Chung, Patrice Abry and Pierre Borgnat. 2017. "Bluetooth data in an urban context: Retrieving vehicle trajectories." *IEEE Transactions on Intelligent Transportation Systems* 18 (9): 2377-2386.

Michau, Gabriel, Nelly Pustelnik, Pierre Borgnat, Patrice Abry, Alfredo Nantes, Ashish Bhaskar and Edward Chung. 2017. "A primal-dual algorithm for link dependent origin destination matrix estimation." *IEEE Transactions on Signal and Information Processing over Networks* 3 (1): 104-113.

Spiess, Heinz. 1990. "A gradient approach for the OD matrix adjustment problem." *CENTRE DE RECHERCHE SUR LES TRANSPORTS PUBLICATION* 1 (693): 2.

Van Der Zijpp, Nanne. 1997. "Dynamic origin-destination matrix estimation from traffic counts and automated vehicle identification data." *Transportation Research Record: Journal of the Transportation Research Board* (1607): 87-94.

Van Zuylen, Henk J and Luis G Willumsen. 1980. "The most likely trip matrix estimated from traffic counts." *Transportation Research Part B: Methodological* 14 (3): 281-293.

Vogl, Thomas P, JK Mangis, AK Rigler, WT Zink and DL Alkon. 1988. "Accelerating the convergence of the back-propagation method." *Biological cybernetics* 59 (4-5): 257-263.

Willumsen, LG. 1984. "Estimating time-dependent trip matrices from traffic counts." In *Ninth International Symposium on Transportation and Traffic Theory*, edited, 397-411: VNU Science Press Utrecht.

Yang, Hai. 1995. "Heuristic algorithms for the bilevel origin-destination matrix estimation problem." *Transportation Research Part B: Methodological* 29 (4): 231-242.

Yang, Hai, Tsuna Sasaki, Yasunori Iida and Yasuo Asakura. 1992. "Estimation of origin-destination matrices from link traffic counts on congested networks." *Transportation Research Part B: Methodological* 26 (6): 417-434.

Zhang, Yi, Xiao Qin, Shen Dong and Bin Ran. 2010. "Daily OD matrix estimation using cellular probe data." In *89th Annual Meeting Transportation Research Board*, edited.