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

Estimation of origin-destination flows in large scale traffic networks

Sakitha Kumarage, Mehmet Yildirimoglu, Zuduo Zheng

School of Civil Engineering, University of Queensland, Australia
Email for correspondence: m.yildirimoglu@uq.edu.au

1 Background and Aims

Reliable estimation of dynamic Origin- Destination (OD) flows in a traffic network is a crucial first step in developing an accurate traffic model, which can be later be used for the evaluation of various policies and management strategies. There has been a myriad of studies on the OD estimation problem. Most of the existing studies start from the premise that *a priori* OD information can be gathered through surveys and/or mathematical models, (*se*e de Dios Ortúzar and Willumsen (2011) for details). However, such surveys may be outdated and limited, and mathematical models may not be advanced enough to tackle the inherent complexity of the underlying demand patterns. Therefore, *a priori* information on OD flows is often considered erroneous and does not adequately represent the existing demand patterns. OD flows can be improved and updated to develop posterior estimations based on link counts that can easily be collected by induction-based loop detectors, see e.g., (Bell, 1983; Fisk and Boyce, 1983; Zuylen and Willumsen, 1980; Cascetta et al., 1993). The core of the estimation problem is to find the share of trips passing through a link entering from any origin and headed towards any destination such that the sum of trips would match the observed link counts.

The basic theory of static OD flow estimation problem is formulated as an error minimization between observed and estimated values of link counts and *a priori* OD flows. A traffic assignment model which considers microscopic elements of the network such as link cost, path connectivity and path choice are used to estimate the resultant link counts from estimated OD flows. Therefore, the objective function of the minimization problem consists of two components; 1) the gap between estimated OD matrix resulted by OD estimation and *a priori* OD matrix, 2) the error between flows resulting from the estimated OD flow and observed link counts. The static OD estimation problem is extended to a dynamic (time-varying) scenario by discretizing the time horizon into finite number of time steps (j). Then, time-varying link counts are measured and OD flows are estimated for each time step. The complexity arises when dealing with these time-varying link counts as the demand of time step t_0 can cause link count at any time step $t \geq t_0$. Hence, the traffic assignment becomes a dynamic problem where dynamic route choice and dynamic link costs have to be incorporated, which makes the estimation problem a more complex and computationally expensive iterative process. The general formulation of OD flow estimation in dynamic case could be presented as;

$$\left(d^*_{[1]},.d^*_{[j]}\right) = argmin_{x_{[1]},.,x_{[j]}} \left\{g_1\left(x_{[1]},.,x_{[j]},\widehat{d}_{[1]},.,\widehat{d}_{[j]}\right) + g_2(x_{[1]},.,x_{[j]},\widehat{F}_{[1]},.,\widehat{F}_{[j]})\right\} \tag{1}$$

Note, $(g_1(.))$ is the measure of the distance between estimated OD flows $(x_{[1]}..x_{[j]})$ and the *a priori*OD flows $(\widehat{d}_{[1]},.,\widehat{d}_{[j]})$, and $(g_2(.))$ is the measure of the distance between the link counts resulting from the estimated OD flows $(x_{[1]}..x_{[j]})$ and observed link counts $(\widehat{F_{[1]}}...\widehat{F_{[j]}})$.

Many estimation methods were developed based on the above fundamental concept. The available techniques could be broadly divided into offline methods and online methods. The online methods target for real-time continuous time horizon OD flow estimation, while offline methods are focusing on the finite time horizon estimations. We are interested in offline studies here. Most offline dynamic OD flow estimation methods develop a bi-level framework where the first level accounts for OD flow estimation and second level accounts for the dynamic traffic assignment (DTA) model. Marzano et al. (2008) observed that many existing methods face scalability issues when the number of OD pairs significantly larger than the number of links with link counts in a network, which is a common occurrence in medium to large scale networks. Further, the granularity of required information for microscopic and mesoscopic DTA models imposes a massive computational burden on the OD estimation problem, which challenges its practicality in the medium to large scale networks. In this study, we propose a hybrid OD flow estimation model that integrates a region-level OD estimation model with a traditional Traffic Analysis Zone (TAZ)-level (or centroid-level) OD estimation problem. This is a promising direction that has the potential to overcome the scalability issues and computational complexities faced by existing methods.

2 Problem Definition and Formulation

Here, we propose a hybrid OD flow estimation method by combining region-level and centroid-level OD estimation problems. We run the region-level OD estimation by partitioning a large-scale network into neighbourhoods with homogeneous traffic conditions. We expect that incorporating an aggregate level model to describe traffic dynamics will guide the centroid-level OD estimation and eliminate scalability issues and cut down computational complexities.

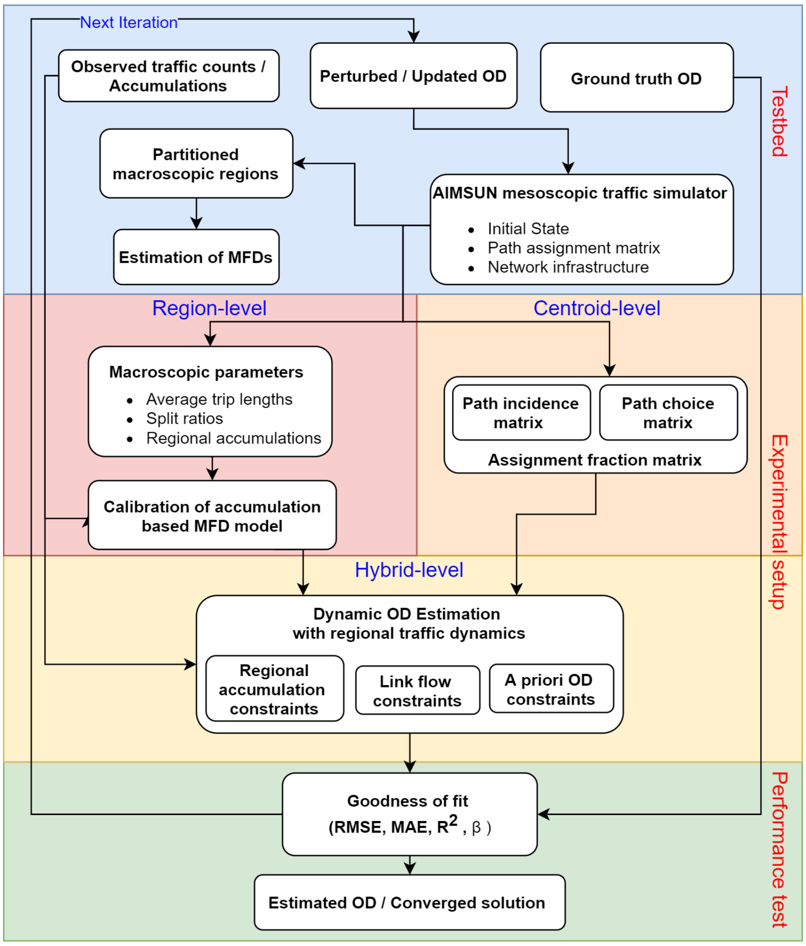

Figure 1: Modelling framework

Our method follows a three-stage modelling framework introduced by the benchmark study of Antoniou et al. (2016) on OD flow estimation. This modelling framework integrates all modelling components within three hierarchical layers such as test-bed setup, experiment setup and performance setup as shown in Figure

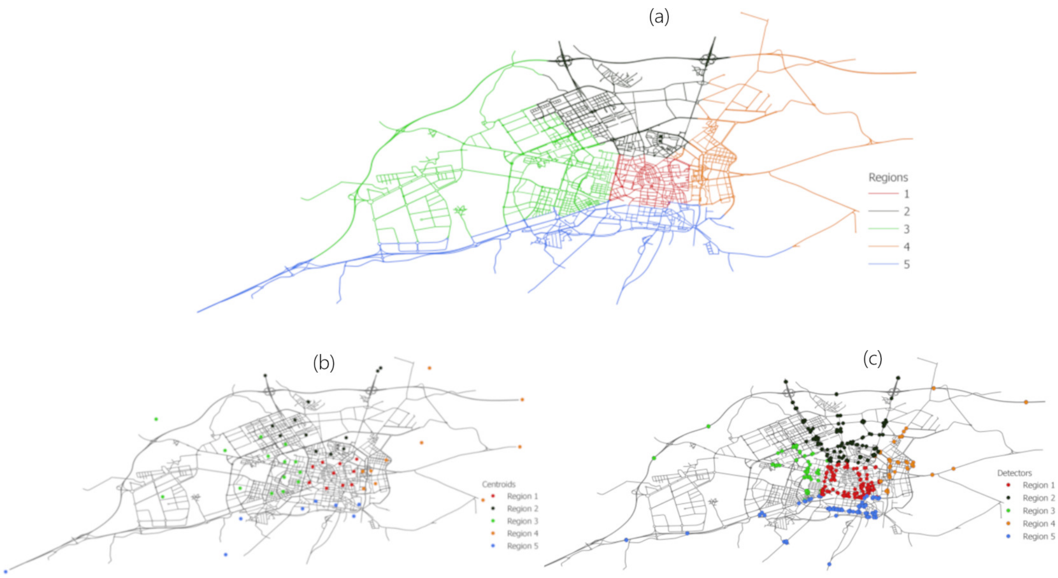

Figure 2: (a) - Proposed regions; (b) - centroid level ODs; (c) - traffic detectors

1. The test-bed layer establishes the preliminary components required for OD estimation such as observed link counts, ground truth OD, *a priori* OD, traffic simulation experiment, and assemble macroscopic components such as partitioning of the network and MFD estimations. We will be using the traffic network of Vitoria (Spain) to develop the methodology as it is suggested as a benchmark network for OD flow estimation, and has been used in the works of Antoniou et al. (2016); Djukic et al. (2011); Masip et al. (2018). The network has 57 centroids, 3249 OD pairs, 2884 nodes, 5799 links and expands to a modelled network of 600km. The *a priori* OD matrix will be obtained by perturbing the ground truth OD matrix with a uniformly distributed random noise (Antoniou et al., 2016).

The second layer of the modelling framework focus on developing the experimental setup. Here we have three allotments as region-level, centroid-level and hybrid OD estimation level. The centroid-level relates to the components in the (*or* conventional) OD estimation problem and the region-level relates to the collection of components (macroscopic parameters) that contribute to regional OD estimation based on regional traffic dynamics. Later, we present the hybrid OD estimation level where we discuss the formulation of OD estimation problem combining both region- and centroid-level components.

The regional components of the problem relies on the regional representation of traffic. On this regard, macroscopic fundamental diagram (MFD) presented by Geroliminis and Daganzo (2008) is adopted in this study. MFD builds an uni-modal, low-scatter and demand insensitive relationship between network flows and network densities for homogeneous traffic regions.

For the implementation of MFD modelling, the urban network is divided into five regions (neighbourhoods) with homogeneous traffic conditions, as shown in Figure 2-(a). MFDs for each region are estimated from the flow and density data collected from the loop detectors in the test-bed. The Figure 2-(b) shows the map of centroid level ODs and Figure 2-(c) shows the spread of traffic detectors in the network. The partitioning of the network into regions instigates a regional route choice phenomenon which occurs when there are two or more regional paths between two regions.Therefore, the traffic dynamics are handled by the multi-region accumulation-based model, which consider regional route choice. For more details on multi region MFD dynamics refer to Yildirimoglu et al. (2015).

Parallel to regional parameters, the centroid-level variables required for OD estimation are extracted. The path assignment required for conventional OD estimation problem will be extracted from the simulation environment (*see* Cascetta et al. (1993) for more details) . The benchmark platform followed in this study assumes that the OD routes and choice preferences generated by a dynamic user equilibrium will not deviate significantly over OD estimation.

The hybrid OD estimation problem is solved upon assembling region- and centroid-level components of the modeling framework. The optimization problem to estimate regional OD flows from link counts is formulated as a minimization problem as follows;

$$\underset{q_{ij}}{\text{minimize}} \quad \sum_{t=1:T} \left\{ \sum_{l=1:L} \left(\frac{f_l^{\mathrm{o}}(t) - f_l^{\mathrm{c}}(t)}{f_l^{\mathrm{o}}(t)} \right)^2 \right\} \tag{2a}$$

$$\text{subject to} \quad f_l^{\mathrm{c}}(t) = \sum_{\lambda=t-\eta:t} \sum_{i,j\in od} m_{l,ij}(\lambda) * q_{ij}(\lambda) \quad \forall\, l, \tag{2b}$$

$$Q_{IJ}(t) = \sum_{i\in I} \sum_{j\in J} q_{ij}(t) \quad \forall\, (I,J), \tag{2c}$$

$$N_I^{\mathrm{M}}(t+1) = g(Q_{IJ}(t), N_I^{\mathrm{M}}(t), L_{IH}(t), \Theta_{H,J}^{I}(t))) \quad \forall\, (I,t), \tag{2d}$$

$$N_I^{\mathrm{O}}(t) * \alpha_{lb} \le N_I^{\mathrm{M}}(t) \le N_I^{\mathrm{O}}(t) * \alpha_{ub} \quad \forall\, (I,t), \tag{2e}$$

$$q_{i,j}^{\mathrm{p}}(t) * \beta_{lb} \le q_{i,j}(t) \le q_{i,j}^{\mathrm{p}}(t) * \beta_{ub} \quad \forall\, (i,j) \tag{2f}$$

The objective function given in Equation (2a) targets to minimize the normalized error between the observed link counts (f_l^{o}) and calculated link counts (f_l^{c}) for all the traffic detectors $(1:\mathrm{L})$ and for all the time steps $(1:\mathrm{T})$. The decision variable of the objective function is q_{ij} where i, j represents the origin and destination respectively. In summary, Equation (2b) provides link counts ($f_l^{\mathrm{c}}(t)$) resulting from traffic assignment matrix ($m_{l,ij}$) and estimated OD $q_{i,j}$, which is the analytical approximation to traffic assignment at link level. Equation (2c) aggregates centroid-level ODs into region-level ODs Q_{IJ}. Here I, J represents origin and destination regions. Equation (2d) defines the MFD traffic dynamics where accumulation in the next time step $N_I^{\mathrm{M}}(t+1)$ is given by the demand $(Q_{I,J}(t))$, accumulation ($N_I^{\mathrm{M}}(t)$), average trip length ($L_{IH}(t)$), and split ratios ($\Theta_{H,J}^{I}(t)$) in current time step t. Equation (2e) defines the bounds for the regional accumulations given by MFD dynamics (N_I^{M}) with respect to the observed accumulations (N_I^{O}). This constraint helps to bound the solution space of $q_{i,j}$ such that centroid-level ODs are guided/shaped by the regional observations. Finally Equation (2f) generates constraints to the decision variable $q_{i,j}$ where the OD estimates are limited to a boundary defined by the *a priori* OD estimate $q_{i,j}^{\mathrm{p}}$. Note that $\alpha_{lb}, \alpha_{ub}, \beta_{lb}, \beta_{ub}$ are non-negative constants. This complex optimization problem could be categorized as non-linear, non-convex optimization problem and we use iterative non-linear solvers (interior point solvers) to derive the optimal solution.

The third layer of our framework focus on goodness of fit test in estimates.While literature suggests a vast number of tests in dealing with the goodness of fit of OD flows and link counts, we will be using (1) Coefficient of determination (R^2) obtained by the regression of estimated ODs and ground truth ODs, (2) regression coefficient (β) of estimated ODs and ground truth ODs, (3) root mean squared error (RMSE) between observed link counts and estimated link counts, (4) mean absolute error (MAE) between observed link counts and estimated link counts. The preliminary results obtained by implementing the proposed OD estimation framework are successful. Figure 3 presents preliminary results of OD estimation for an uncongested scenario D7 which is a bechmarked scenario in Antoniou et al. (2016). The perturbed OD is underestimated by 15% on average with a random noise as shown in Figure 3-(a). We were able to obtain a significantly improved OD estimate after applying the hybrid OD estimation framework. Figure 3-(b) shows the agreement of estimated ODs with ground truth ODs where we observe $\beta = 0.99$ with a significantly high $R^2 = 0.95$. We see a very high match with high R^2 between simulated link counts with observed link counts as shown in Figure 3-(c). Further, Figure 3-(d) shows the reduction MAE in link counts over iterations, which demonstrate the capability of the hybrid OD estimation to derive the optimal solution within few iterations. In overall, the proposed experimental setup framework builds on two components; (i) guidance of regional OD estimation problem which relies on analytically tractable MFD modelling, and which therefore offers significant computational advantages, and (ii) traditional

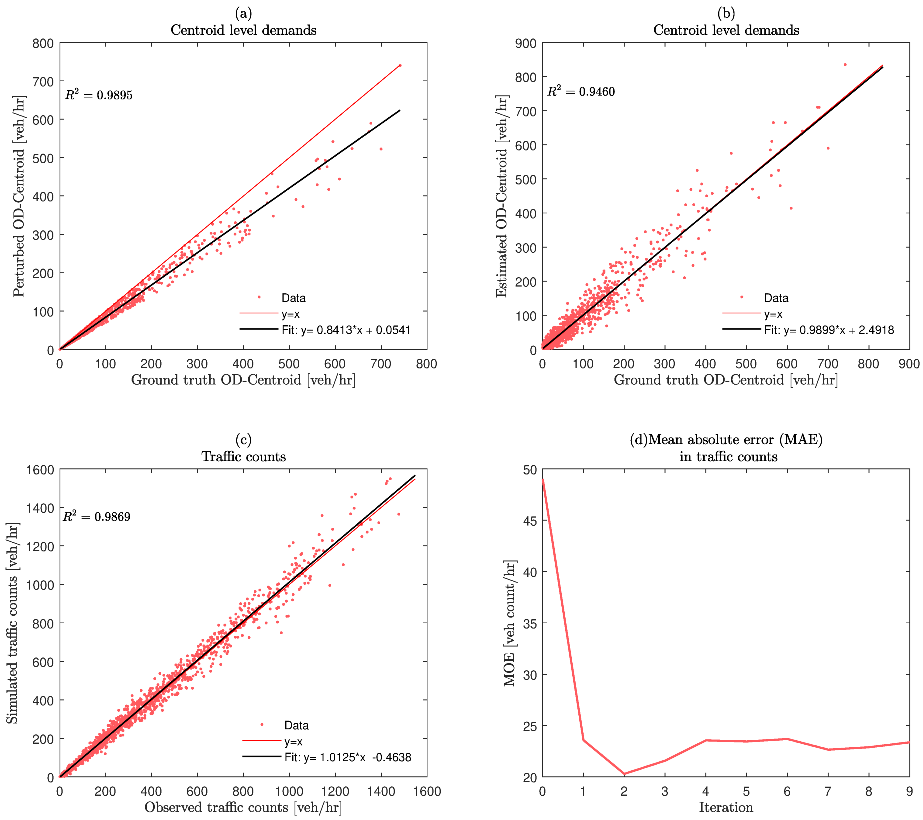

Figure 3: Preliminary results for uncongested demand scenario D7 (-15%)

centroid-level OD estimation problem involving an additional constraints that links the two components. The motivation behind the proposed integrated structure is to produce high-level guidance towards the descent direction with the regional problem, which is relatively easy to solve, and thereby to call for fewer iterations in the overall OD estimation problem. We will further conduct experiments on several scenarios involving congested traffic conditions, and report the results in a future publication with further elaborations on theory.

References

Constantinos Antoniou, Jaume Barceló, Martijn Breen, Manuel Bullejos, Jordi Casas, Ernesto Cipriani, Biagio Ciuffo, Tamara Djukic, Serge Hoogendoorn, Vittorio Marzano, Lídia Montero, Marialisa Nigro, Josep Perarnau, Vincenzo Punzo, Tomer Toledo, and Hans van Lint. Towards a generic benchmarking platform for origin-destination flows estimation/updating algorithms: Design, demonstration and validation. *Transportation Research Part C: Emerging Technologies*, 2016. doi: 10.1016/j.trc.2015.08.009.

M. G. H. Bell. The Estimation of an Origin-Destination Matrix from Traffic Counts. *Transportation Science*, 17:198–217, 1983. doi: 10.1287/trsc.17.2.198.

Ennio Cascetta, Domenico Inaudi, and Gérald Gerald Marquis. Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts. *Transportation Science*, 27:363–373, 1993. doi: 10.1287/trsc.27.4.363.

Juan de Dios Ortúzar and Luis G. Willumsen. Modelling transport. mar 2011. doi: 10.1002/9781119993308.

Tamara Djukic, J W C van Lint, and S P Hoogendoorn. Efficient Methodology for Benchmarking Dynamic Origin-Destination Demand Estimation Methods. *Transportation Research Record*, pages 35–44, 2011. doi: 10.3141/2263-05.

C. S. Fisk and D. E. Boyce. A note on trip matrix estimation from link traffic count data. *Transportation Research Part B: Methodological*, 17:245–250, 1983. doi: 10.1016/0191-2615(83)90018-8.

Nikolas Geroliminis and Carlos F. Daganzo. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42:759–770, 2008. doi: 10.1016/j.trb.2008.02.002.

Vittorio Marzano, Andrea Papola, and Fulvio Simonelli. Effectiveness of Origin-Destination Matrix Correction Procedure Using Traffic Counts. *Transportation Research Record*, 2085:57–66, 2008. doi: 10.3141/2085-07.

David Masip, Tamara Djukic, Martijn Breen, and Jordi Casas. Efficient OD matrix estimation based on metamodel for nonlinear assignment function. *Australasian Transport Research Forum Proceedings*, 2018.

Mehmet Yildirimoglu, Mohsen Ramezani, and Nikolas Geroliminis. Equilibrium analysis and route guidance in large-scale networks with MFD dynamics. *Transportation Research Part C: Emerging Technologies*, 59:404–420, 2015. doi: 10.1016/j.trc.2015.05.009.

Henk J. Van Zuylen and Luis G. Willumsen. The most likely trip matrix estimated from traffic counts. *Transportation Research Part B: Methodological*, 14:281–293, 1980. doi: 10.1016/0191-2615(80)90008-9.