Estimation of dynamic origin-destination demand for urban road networks incorporating a region-level traffic flow model

Ying-Chuan Ni*, Anna M. Schönhärl, Michail A. Makridis, and Anastasios Kouvelas

Traffic Engineering Group, Institute for Transport Planning and Systems, Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Switzerland

Short summary

Representative dynamic origin-destination (OD) demand is crucial for the assessment of network traffic performance. However, challenges for OD estimation include its underdetermination nature, data scarcity, and high computation cost. This paper proposes a mathematical optimization method for dynamic OD estimation (DODE) incorporating a link-level linear mapping method and a region-level traffic flow model. The region production and accumulation are estimated from loop detector data. It is considered an efficient method particularly for a large-scale network. A case study is conducted for a real-world network. After the optimization, we further test the estimated demand with microscopic traffic simulation which has not been calibrated. It is found that approximately 50% of the link count measurements already have GEH statistic values that are below 5. More examinations are required to validate the proposed method. The outcome can also be used as the initial solution in DODE using a simulation-based optimization approach.

Keywords: continuous packet approach; dynamic origin-destination demand; loop detector data; mathematical optimization; region-level traffic flow model; urban road networks

1 Introduction

For traffic flow modeling in a network, traffic demand is typically described by origin-destination (OD) matrices, which specify the amount of travelers or vehicles going from each origin to each destination. The research on OD estimation has started using traffic counts which can be obtained from loop detectors since 1970s, mainly still following the concept of gravity model and regression approaches (Willumsen, 1978). Later on, methods using generalized least squares to match the observed traffic count with a bi-level programming problem formulation were proposed for static OD estimation (Yang et al., 1992). The upper-level of the optimization problem aims to minimize the distance measurements, while the lower-level describes the traffic assignment.

To fully consider the dynamic nature of traffic flow, the problem of dynamic OD demand estimation (DODE), which accounts for time-dependency in the traffic pattern by producing sequential OD matrices, has been discussed since 1990s (Bell, 1991; Cascetta et al., 1993). Based on an assignment matrix-based formulation of the relationship between OD demand and link count in DODE, Toledo and Kolechkina (2013) proposed a method using linear approximation of the assignment matrix to derive the search direction and step size in the solution algorithm. In Frederix et al. (2011, 2013), the non-linearity in the assignment matrix resulted from congestion phenomena was emphasized.

With the advancement in dynamic traffic assignment and traffic simulation tools, DODE studies have started to replace the assignment matrix at the lower-level of the optimization problem with a dynamic traffic model, developing simulation-based optimization (SO) approaches using heuristic solution algorithms. Cipriani et al. (2011) used the DYNAMEQ software coupled with a modified Simultaneous Perturbation Stochastic Approximation (SPSA) method to find the optimal solution to the OD matrices. Later on, various enhancements on SPSA were proposed to reduce the computation burden (Cantelmo et al., 2014; Tympakianaki et al., 2015). Antoniou et al. (2016) further developed a benchmarking platform to compare DODE methods with different input data, objective functions, and solution algorithms. In the SO framework proposed in Osorio (2019), an analytical network model is introduced as the metamodel optimization problem, while a mesoscopic simulator is used to simulate the solution derived from the optimization. Although this type of approaches may be able to describe the traffic dynamics with a great level of granularity, the problem of high computational requirement for a large-scale network and the concern of local optimality still arise.

Region-level traffic flow model which uses network fundamental diagrams (NFDs) to describe the aggregated traffic dynamics has gained a lot of attention nowadays thanks to its supreme computation efficiency (Johari et al., 2021). Dantsuji et al. (2022) first incorporated the concept of region-level traffic dynamics into the DODE problem for urban road networks. However, the study only focused on the estimation of demand on regional paths. The regional demand is distributed to centroids according to pre-determined rules, e.g., equal distribution, gravity model, or population density. Kumarage et al. (2023) further developed a hybrid framework for DODE which incorporates the region-level accumulation-based NFD model and the centroid-level linear mapping proposed by Cascetta et al. (1993) at the upper-level. Although the non-linearity between OD demand and link count is not considered at the centroid-level, it is captured at the region-level through NFD-based traffic modeling. Traffic simulation using the Aimsun software was also incorporated to update information required by the region-level traffic model after every iteration and ensure the applicability of the estimated demand for mesoscopic traffic simulation.

It is worth noting that methods utilizing other data sources, such as probe vehicle trajectories and automatic vehicle identification, and machine learning-based approaches are out of the scope in this study and are therefore not discussed.

This study aims to propose an efficient DODE method by using a region-level traffic flow model in a mathematical optimization problem, which is similar to the work in Kumarage et al. (2023). However, instead of the accumulation-based NFD modeling with a single trip length in each region, this method is able to model the flow propagation on multiple OD paths and eliminate the assumption of uniform network-mean speed. For proof-of-concept, a case study is designed for a medium-sized urban road network.

2 Method

This section first explains the link-level linear mapping method to match the link traffic counts and then introduces the proposed region-level traffic flow model, which would both be integrated into a mathematical optimization problem.

Link-level linear mapping of traffic counts

To load the traffic demand onto road links in the network, the continuous packet approach, which was first introduced in Cascetta et al. (1993), is adopted to consider the phenomenon of delayed arrival of OD demand at links. This subsection provides a detailed and graphical explanation of this approach.

Continuous packet approach assumes that demand is uniformly distributed, i.e., vehicles are equally spaced, within the departure time interval when starting from the origin. A certain proportion of OD demand would experience delayed arrival at the links due to the relatively late departure and should be assigned to the link count at the next time step.

Figure 1 illustrates an example situation, where the travel time from the origin to link 1 and link 2 are 5 minutes and 20 minutes, respectively. The time interval length is set to 15 minutes. For link 1, most of the demand, namely the non-delayed portion, can arrive within the first time interval (0-15 min). However, 1/3 of the demand which departs in the last 5 minutes would arrive at the next time interval (15-30 min) considering the required travel time. This is called the delayed portion, as shown in the second time interval in Figure 1. On the other hand, no demand can arrive at link 2 in the departure time interval due to the origin-to-link travel time that is longer than the time interval length. In this case, the non-delayed portion of the demand would be propagated by one time interval, so is the propagated delayed portion which will hence arrive at link 2 in the third interval (30-45 min).

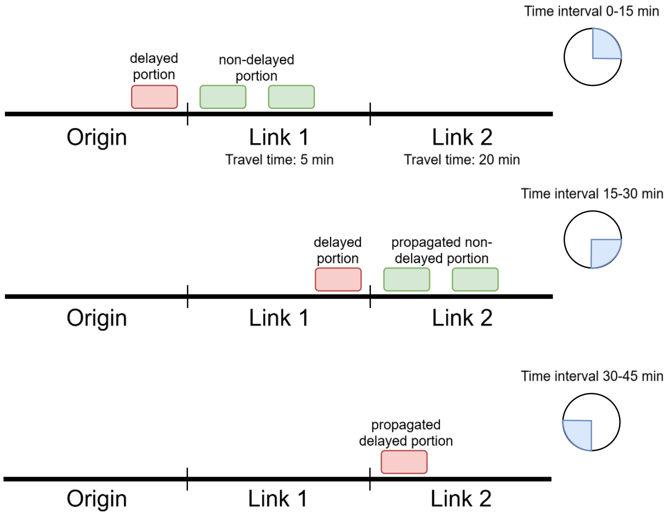

Figure 1: Graphical example of the continuous packet approach

As described with the example, continuous packet approach requires an input parameter, the origin-to-link travel time TT^l_{od}, to calculate the two quantities, the number of time intervals required to arrive at the link β^l_{od} and the delayed portion γ^l_{od}. β^l_{od} is first calculated by Equation 1, where Δk denotes the time interval length, and the function Ψ rounds down the value to the closest integer.

$$\beta^l_{od}(k) = \Psi\left(\frac{TT^l_{od}(k)}{\Delta k}\right) \tag{1}$$

The delayed proportion γ^l_{od} can then be calculated by dividing the origin-to-link travel time by the time interval length and then subtracting the number of time intervals required to arrive at the link, as written in Equation 2.

$$\gamma^l_{od}(k) = \frac{TT^l_{od}(k)}{\Delta k} - \beta^l_{od}(k) \tag{2}$$

The loading of non-delayed and delayed OD demand can be expressed as Equations 3 and 4, respectively. Let x_{od} denote the OD demand and α^l_{od} indicate whether the OD pair passes the link. The amount of demand which arrives at the link at a certain time step y^l_{od} can then be computed accordingly.

$$y^l_{od}(k + \beta^l_{od}(k)) = \alpha^l_{od} \cdot x_{od}(k) \cdot (1 - \gamma^l_{od}(k)) \tag{3}$$

$$y^l_{od}(k + \beta^l_{od}(k) + 1) = \alpha^l_{od} \cdot x_{od}(k) \cdot \gamma^l_{od}(k) \tag{4}$$

Region-level traffic flow modeling

The linear mapping method produces an assignment matrix to describe the correlation between OD demand and link traffic counts. However, as the number of link counts available is smaller than the number of OD paths in a road network, there can be numerous possible solutions that can lead to the same outcome of link counts. This is also known as the underdetermination nature of the DODE problem.

Therefore, to capture the non-linearity which lies within the OD demand and link count caused by congestion and also the precise traffic states in the network, a region-level traffic flow model considering multiple trip lengths within one region, as proposed in Geroliminis (2015) and Mariotte and Leclercq (2019), is applied. Such a model possesses several merits for DODE problems, including its computation efficiency and the lack of necessity for additional calibration. By modeling the traffic dynamics to incorporate additional constraints, the solution space can be reduced to guarantee the uniqueness of the estimated demand.

In the model, subreservoirs are created for each OD in each region (main reservoir), as shown in Figure 2. Compared to the framework proposed in Kumarage et al. (2023), by explicitly modeling the dynamics for every OD path, our method does not require the integration with a simulation tool to update the parameters for region-level traffic flow modeling in every iteration.

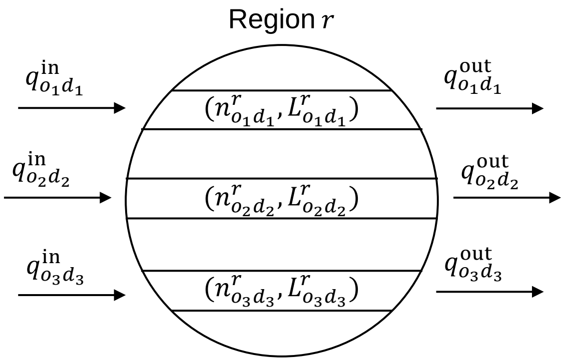

Figure 2: Representation of a single region (main reservoir) and multiple OD paths (subreservoirs)

The propagation of each OD flow in each region can be expressed by the conservation equation, as expressed in Equation 5. t indicates the time step in the model, which may be different from k in the linear mapping. For brevity of notation, the index r which represents the region is omitted here.

$$\Delta n_{od}(t) = q^{\text{in}}_{od}(t) - q^{\text{out}}_{od}(t) \tag{5}$$

For those origin nodes at the network boundary, the inflow is calculated by Equation 6, where Δt denote the time step size in the traffic flow model.

$$q^{\text{in}}_{od}(t) = \frac{\Delta t}{\Delta k} \cdot x_{od}(k) \tag{6}$$

The trip completion rate q^{out}_{od} for each OD in each region can be computed using the assigned production and region-internal route length L_{od}. Different from previous studies regarding NFD-based modeling that assumed a uniform network-mean speed and computed the subreservoir production based on its accumulation n_{od}, this study eliminates such a limitation by simply using a variable η_{od} to represent the proportion of total region production P that is accounted for by each OD pair, as written in Equation 7. The value of η_{od} would be determined by the optimization.

$$q^{\text{out}}_{od}(t) = \eta_{od}(t) \cdot \frac{P_(t)}{L_{od}} \tag{7}$$

By introducing η_{od} as a variable, we add another degree of freedom to the modeling. To avoid enlarging the solution space, the correlation between the subreservoir production and their accumulation needs to be included with the help of region-internal route speed V_{od}, which can be estimated from loop detectors along the route. Equation 8 describes their correlation. By adding this constraint, we can also ensure that the OD paths which actually use the same route within a region can share a similar speed.

$$n_{od}(t) \cdot V_{od}(t) = \eta_{od}(t) \cdot P_(t) \tag{8}$$

Instead of imposing an equality constraint which could make the problem infeasible due to data inaccuracy, we relax the constraint by using a 5 km/h speed range to formulate two inequality constraints, as expressed in Equations 9 and 10. V^{ub}_{od} and V^{lb}_{od} denote the upper- and lower-bounds of the route speed. For region-internal route that does not have any detector data, the free-flow speed $V^{\max}$ can be used as the upper-bound.

$$n_{od}(t) \cdot V^{\text{ub}}_{od}(t) \geq \eta_{od}(t) \cdot P_(t) \tag{9}$$

$$n_{od}(t) \cdot V^{\text{lb}}_{od}(t) \leq \eta_{od}(t) \cdot P_(t) \tag{10}$$

Optimization problem formulation

As the objective function written in Equation 11, the mathematical optimization aims to derive the OD demand x_{od} by minimizing the normalized square error of the link counts and the region accumulations, which can be obtained from the link-level mapping and region-level traffic model, respectively.

$$\underset{x_{od}}{\text{minimize}} \sum_{l=1}^{L} \sum_{k=1}^{K} \left(\frac{\hat{c}_l(k) - c_l(k)}{\hat{c}_l(k)}\right)^2 + \sum_{r=1}^{R} \omega_r \cdot \sum_{t=1}^{T} \left(\frac{\hat{n}_r(t) - n_r(t)}{\hat{n}_r(t)}\right)^2 \tag{11}$$

$\hat{c}_l$ and c_l are the measured and computed link count, respectively. Equation 12 derives the computed link count by summing over the amount of demand contributed by each OD that arrives at the link y^l_{od}, as calculated by the linear mapping method.

$$c_l(k) = \sum_{od} y^l_{od}(k) \tag{12}$$

The measured region accumulation is denoted by $\hat{n}_r$, while n_r is computed by summing the accumulations of OD paths n^r_{od} within that region. A weight factor ω_r is applied to each region so that the magnitude of influence from the region accumulation errors can be balanced with the influence of link count errors.

Besides the constraints already mentioned in the previous two subsections, a capacity constraint needs to be added for origin nodes. Therefore, the summation of the inflow q^{in}_{od} of all ODs that use the origin node i as the entry into the network (in path set Z_i) can be restricted by its capacity $Q^{\max}_i$, as written in Equation 13.

$$\sum_{od \in Z_i} q^{\text{in}}_{od}(t) \leq Q^{\max}_i \tag{13}$$

Although the formulated optimization problem contains quadratic terms in the objective function, all the constraints are linear. Such a quadratic programming problem can be easily solved with either simplex method or interior point method.

3 Case study

A four-region partition of the city center of Zurich is selected as the case study network, as shown on the left of Figure 3. The OD nodes and detector locations can be seen on the right of Figure 3. In the network, 31 origin nodes and 29 destination nodes are identified, while 10 of them being internal OD nodes so that the demand of parking trips can be considered. Among all the road links in the network, 191 links have traffic counts available.

Figure 3: Case study network with a four-region partition (left) and OD nodes and loop detector locations (right)

The loop detector data collected during the morning period (06:00 - 11:00) on the 7th of November (Monday), 2022 are used. The time interval length used in the linear mapping and time step size in the traffic flow model are 15 and 3 minutes, respectively. Hence, the aggregation interval for the estimation of region productions and accumulations would also be 3 minutes. Figure 4 plots the region production and accumulation estimated based on the data for each region. It is worth noting that the fitting of the region-level traffic dynamics to create functional-form NFDs is not required as this method directly uses the estimated values.

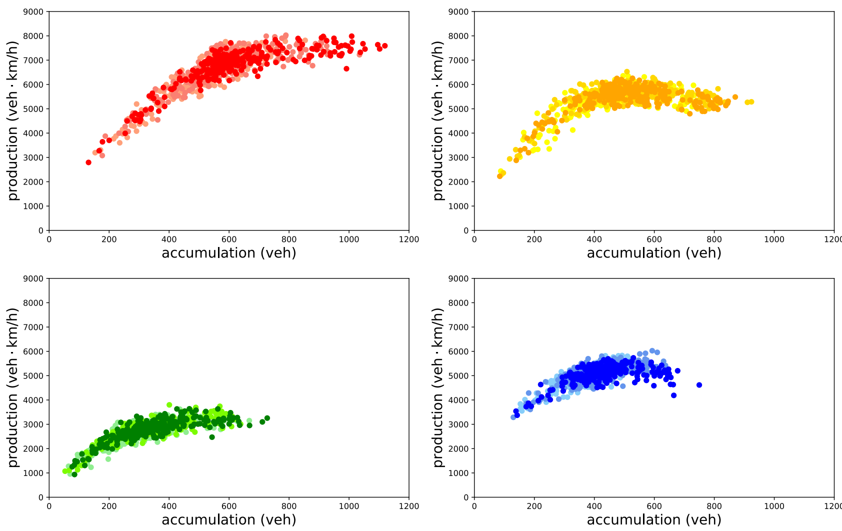

Figure 4: Region productions and accumulations

It is assumed that there is no route choice behavior within each region since the partition is sufficiently small. Each OD path only uses one region-internal route. Also, there is no regional path choice behavior in this small-scale case study network. After preliminarily eliminating the unreasonable ODs by clustering adjacent nodes into several traffic analysis zones (TAZs) and only maintaining all the shortest OD paths from and to each TAZ, 475 OD paths are identified. The demand of these OD paths at every 15-minute time interval during the 5-hour period is the final output of this study.

There are a few parameters that need to be pre-defined for the optimization problem. The origin-to-link travel time TT^l_{od} at every time interval and the route-mean speed V^r_{od} at every time step are estimated by aggregating the flow and occupancy information collected by detectors along the regional OD paths and region-internal routes. As the case study is conducted for an urban network, the capacity $Q^{\max}$ of interrupted traffic flow is set to 600 veh/h/lane to account for the influence of traffic signals. For internal origins, which are often parking space or minor streets, we assume their capacity to be one-third of this value.

4 Results and discussion

Figure 5 shows the evolutions of region accumulations estimated from the data and computed by the model in the optimization problem. The compliance between the two curves indicates that the estimated demand results in region-level traffic states that match the ground-truth.

Using the estimated OD demand as input, we run a microscopic traffic simulation scenario in Simulation of Urban MObility (SUMO) (Lopez et al., 2018). It is worth mentioning that the simulation is not calibrated and does not align with the setup in the region-level traffic flow model or ground-truth. Given this inconsistency, the results still show that around 48% of the GEH values analyzed for the links with count measurements are below 5 (satisfactory threshold), while another 20% of the GEH values between 5 and 10 (acceptable threshold).

The optimization shows outstanding computation efficiency as it only takes approximately 100 seconds on a device with a 4-core Intel Core i7 CPU running at 2.80 GHz with 16.0 GB RAM. Therefore, we consider such an outcome promising and worth further exploration.

Simply judging from the results of the case study conducted in this paper, we can only conclude

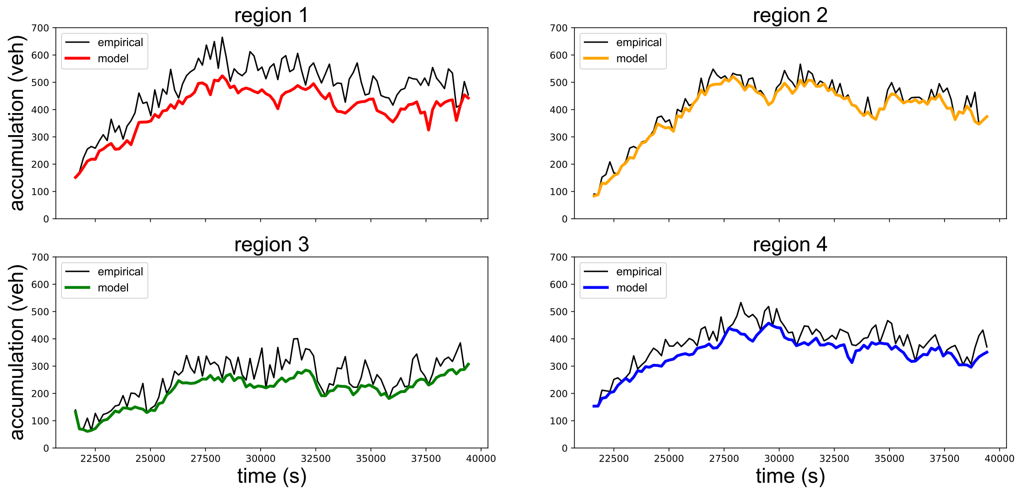

Figure 5: Evolutions of region productions and accumulations

that the proposed DODE method is valid for region-level traffic flow modeling using the similar approach introduced in subsection 2. To comprehensively examine its validity, several self-designed demand scenarios need to be used, as conducted in Kumarage et al. (2023).

5 Conclusions

The DODE method proposed in this paper describes the correlation between OD demand and link count through a link-level linear mapping method and utilizes a novel region-level traffic flow model to reduce the solution space and capture the non-linearity resulted from traffic congestion dynamics in the road network. By combining these two components, a mathematical optimization problem is formulated to estimate the demand.

From the case study results, it is found that the region-level traffic states produced by the traffic flow model according to the estimated demand comply with empirical data, and the simulated link count measurements show promising outcome. Most importantly, the computation efficiency of the proposed method is demonstrated, which means that it can be applied to a large-scale network to resolve the scalability concern of DODE problems.

More self-designed case studies with various demand scenarios are required to validate the proposed method. Future work will focus on integrating reliable travel time and route-mean speed estimation methods with loop detector data and also testing the outcome in a well-calibrated microscopic traffic simulation environment which has better consistency with the region-level traffic flow model. In addition, when applying the method to a large-scale network, region-level path choice behavior will need to be considered.

One can also consider using the estimated OD demand as the initial solution for SO approaches which integrate a simulation tool and a heuristic solution algorithm, such as SPSA. We believe that the number of iterations required to reach convergence can be significantly reduced.

Acknowledgements

The authors would like to thank Dr. Lukas Ambühl and Transcality AG for having an inspiring discussion with us and sharing the SUMO network and loop detector data of the city of Zürich.

References

Antoniou, C., Barceló, J., Breen, M., Bullejos, M., Casas, J., Cipriani, E., Ciuffo, B., Djukic, T., Hoogendoorn, S., Marzano, V., Montero, L., Nigro, M., Perarnau, J., Punzo, V., Toledo, T., & van Lint, H. (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. https://doi.org/10.1016/j.trc.2015.08.009

Bell, M. G. (1991). The real time estimation of origin-destination flows in the presence of platoon dispersion. *Transportation Research Part B: Methodological*, *25*(2-3), 115–125. https://doi.org/10.1016/0191-2615(91)90018-E

Cantelmo, G., Cipriani, E., Gemma, A., & Nigro, M. (2014). An Adaptive Bi-Level Gradient Procedure for the Estimation of Dynamic Traffic Demand. *IEEE Transactions on Intelligent Transportation Systems*, *15*(3), 1348–1361. https://doi.org/10.1109/TITS.2014.2299734

Cascetta, E., Inaudi, D., & Marquis, G. (1993). Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts. *Transportation Science*, *27*(4), 363–373. https://doi.org/10.1287/trsc.27.4.363

Cipriani, E., Florian, M., Mahut, M., & Nigro, M. (2011). A gradient approximation approach for adjusting temporal origin–destination matrices. *Transportation Research Part C: Emerging Technologies*, *19*(2), 270–282. https://doi.org/10.1016/j.trc.2010.05.013

Dantsuji, T., Hoang, N. H., Zheng, N., & Vu, H. L. (2022). A novel metamodel-based framework for large-scale dynamic origin–destination demand calibration. *Transportation Research Part C: Emerging Technologies*, *136*, 103545. https://doi.org/10.1016/j.trc.2021.103545

Frederix, R., Viti, F., Corthout, R., & Tampère, C. M. J. (2011). New Gradient Approximation Method for Dynamic Origin–Destination Matrix Estimation on Congested Networks. *Transportation Research Record: Journal of the Transportation Research Board*, *2263*(1), 19–25. https://doi.org/10.3141/2263-03

Frederix, R., Viti, F., & Tampère, C. M. (2013). Dynamic origin–destination estimation in congested networks: theoretical findings and implications in practice. *Transportmetrica A: Transport Science*, *9*(6), 494–513. https://doi.org/10.1080/18128602.2011.619587

Geroliminis, N. (2015). Cruising-for-parking in congested cities with an MFD representation. *Economics of Transportation*, *4*(3), 156–165. https://doi.org/10.1016/j.ecotra.2015.04.001

Johari, M., Keyvan-Ekbatani, M., Leclercq, L., Ngoduy, D., & Mahmassani, H. S. (2021). Macroscopic network-level traffic models: Bridging fifty years of development toward the next era. *Transportation Research Part C: Emerging Technologies*, *131*, 103334. https://doi.org/10.1016/j.trc.2021.103334

Kumarage, S., Yildirimoglu, M., & Zheng, Z. (2023). A hybrid modelling framework for the estimation of dynamic origin–destination flows. *Transportation Research Part B: Methodological*, *176*, 102804. https://doi.org/10.1016/j.trb.2023.102804

Lopez, P. A., Wiessner, E., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flotterod, Y.-P., Hilbrich, R., Lucken, L., Rummel, J., & Wagner, P. (2018). Microscopic Traffic Simulation using SUMO. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2575–2582. https://doi.org/10.1109/ITSC.2018.8569938

Mariotte, G., & Leclercq, L. (2019). Flow exchanges in multi-reservoir systems with spillbacks. *Transportation Research Part B: Methodological*, *122*, 327–349. https://doi.org/10.1016/j.trb.2019.02.014

Osorio, C. (2019). Dynamic origin-destination matrix calibration for large-scale network simulators. *Transportation Research Part C: Emerging Technologies*, *98*, 186–206. https://doi.org/10.1016/j.trc.2018.09.023

Toledo, T., & Kolechkina, T. (2013). Estimation of Dynamic Origin–Destination Matrices Using Linear Assignment Matrix Approximations. *IEEE Transactions on Intelligent Transportation Systems*, *14*(2), 618–626. https://doi.org/10.1109/TITS.2012.2226211

Tympakianaki, A., Koutsopoulos, H. N., & Jenelius, E. (2015). c-SPSA: Cluster-wise simultaneous perturbation stochastic approximation algorithm and its application to dynamic origin–destination matrix estimation. *Transportation Research Part C: Emerging Technologies*, *55*, 231–245. https://doi.org/10.1016/j.trc.2015.01.016

Willumsen, L. (1978). *Estimation of an O-D matrix from traffic counts: a review* (tech. rep.). Institute for Transport Studies, University of Leeds.

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*(6), 417–434. https://doi.org/10.1016/0191-2615(92)90008-K