# APPLICATION OF DATA DRIVEN APPROACH TO DYNAMIC ORIGIN-DESTINATION MATRIX ESTIMATION (DODME)

X. Ros-Roca[a], J. Barceló[a], L. Montero[a]
[a]Department of Statistics and Operations Research, Universitat Politècnica de Catalunya

## Abstract

Trip patterns in terms of Origin to Destination (OD) traffic flows are a key input to traffic assignment models, namely to Dynamic Traffic Assignment models, where they also must be dynamic, or at least time discretized, to properly approximate the time variability of the demand. OD matrices are not yet fully observable, in the best case the measurements from Information and Communication Technologies (ICT), as GPS vehicle tracking, or mobile phones Call Detail Records (CDR), allow drawing samples that must be suitably expanded to provide estimates of the whole population. Therefore, their estimation must be done resorting to indirect process, usually based on mathematical models.

One of the most appealing mathematical formulation of the OD estimation problem is in terms of bilevel optimization approaches (1), aimed at adjusting an initial target OD, $X^H$, so that it could explain the observed link flow counts $\hat{Y}$ at counting stations in the network, Ros-Roca et al. (2018).

$$\begin{aligned} Min\, Z(X,Y) &= w_1 F_1(X, X^H) + w_2 F_2\left(Y, \hat{Y}\right) \qquad (1) \\ Y &= Assignmt(X) \\ X &\geq 0 \end{aligned}$$

The usual approaches estimate the link flows as the outcome of a traffic assignment at the lower optimization level, which can be formulated in terms of an assignment matrix $A$:

$$Y = A(X)X \qquad (2)$$

which are then used to minimize distance functions between the observed and estimated link flows, $F_2$, and the target and adjusted OD matrix, $F_1$ , at the upper level. Replacing the estimated link flows $Y$ by their relationship (2) with the assignment matrix A, the upper level can then be reformulated as:

$$Min\, Z(X,Y) = w_1 F_1(X, X^H) + w_2 F_2\left(A(X)X, \hat{Y}\right) \qquad (3)$$

The assignment matrix A becomes in thus way the key component of the mathematical model The dynamic formulation of the model decomposes it in time intervals to emulate the flow propagation across the network, assuming that at each iteration of the optimization procedure the relationships between the estimated flows and the adjusted matrix can be set up in terms of a time dependent assignment matrix (2), Toledo and Kolechkina (2013), Frederix et al (2013), Ros-Roca et al. (2019).

$$y_{lt} = \sum_{i \in I} \sum_{j \in J} \sum_{r=1}^{t} a_{ijr}^{lt} x_{ijr} \qquad (4)$$

Where $y_{lt}$ is the estimated flow in link $l$ at time period $t$, $x_{ijr}$ is the flow departing origin $i \in I = \{set\ of\ origins\}$, with destination $j \in J = \{set\ of\ destinations\}$, at time interval $r$, and $a_{ijr}^{lt}$, the assignment matrix, is the fraction of trips from origin $i \in I$ with destination $j \in J$, departing from $i$ at time $r$, that reach link $l$ at time $t$. To account for these time dependencies in the calculation of the assignment matrix the assignment at the lower level is done by a Dynamic Traffic Assignment.

The availability of data collected from samples of GPS tracked vehicles, and other ICT sources, i.e. Bluetooth, if the data collection process is appropriately designed, Ros-Roca et al. (2020), provides the way of generating a discretized time estimate of the target OD matrix $\hat{x}_{ijr}$,in addition a suitable

processing, Janmyr and Wadell (2018), Krishnakumari et al. (2019), Nassir (2014), Ros-Roca et al. (2020), provides a sound empirical estimate of the time dependent assignment matrix $a_{ijr}^{lt}$, which is the key component of the DODME. As in other data driven approaches, Krishnakumari et al. (2019), the expansion of the sampled target matrix to estimate the OD matrix can be done in terms of a process common in transport demand analysis, Ortúzar and Willumsen (2011), from scaling factors per origins, $\alpha_i, i \in I$, and per destinations $\beta_j, j \in J$, such that:

$$x_{ijr} = \alpha_i \beta_j \hat{x}_{ijr}, \forall i \in I, \forall j \in J \tag{5}$$

If $\hat{y}_{lt}, l \in \hat{L} \subset L, t \in T$ are the link flows measured at the counting stations, in a subset $\hat{L} \subset L$ of the network links, the Dynamic Data-Driven OD Matrix Estimation problem can be formulated as the problem of finding the values of the scaling factors $\alpha_i, i \in I$ and $\beta_j, j \in J$, without the need of conducting the traffic assignment at the lower level of (1), exploiting the empirical assignment matrix $a_{ijr}^{lt}$. The optimization problem (1) can then be reformulated as:

$$\min_{\alpha_i,\beta_j} \left\{ w_1 \left[ \sum_{i\in I} \sum_{j\in J} \sum_{r\in T} \left( x_{ijr}^H - \alpha_i \beta_j \hat{x}_{ijr} \right)^2 \right] + w_2 \left[ \sum_{t\in T} \sum_{l\in \hat{L}} \left( \hat{y}_{lt} - \sum_{i\in I} \sum_{j\in J} \sum_{r=1}^{t} a_{ijr}^{lt} \alpha_i \beta_j \hat{x}_{ijr} \right)^2 \right] \right\} \tag{6}$$

Subject to the non-negativity constraints

$$\alpha_i \geq 0, \forall i \in I \tag{7}$$

$$\beta_j \geq 0, \forall j \in J \tag{8}$$

For the proof-of-concept testing of the approach, the nonlinear optimization problem (6)-(8) has been computationally tested in a set of computational experiments with a network modelled in PTV Vissim, using a standard nonlinear optimization algorithm. The microscopic simulation model has been used to run a set of experiments to generate a database of synthetic data that emulates the GPS data collected from tracking individual vehicles.

From a practical point of view an important question is what happens when the available commercial GPS data from data providers, cannot generate a reliable OD matrix for some practical reasons (i.e. for privacy requirements changes in the identities assigned to the GPS devices change arbitrarily splitting long trips into smaller short trips that do not represent the OD patterns; or when collected data mix fleets with very different mobility patterns as for example, passenger cars and commercial fleet vehicles). But the appropriate treatment of the waypoint supplied by GPC allows, Ros-Roca et al. (2020), the calculation of a reliable empirical assignment matrix $a_{ijr}^{lt}$ Then if a reliable historical OD can be obtained from other sources, Montero et al. (2019), the objective function of the optimization problem (6), (7), (8) can be reformulated as:

$$\min_{\alpha_i,\beta_j} \sum_{t\in T} \sum_{l\in \hat{L}} \left( \hat{y}_{lt} - \sum_{i\in I} \sum_{j\in J} \sum_{r=1}^{t} a_{ijr}^{lt} \alpha_i \beta_j \hat{x}_{ijr} \right)^2 \tag{9}$$

This paper investigates the computational performance of the proposed approach in a relevant subnetwork of the city of Barcelona, the CBD area, composed of (description: 9250 links, 202 zones), for which are available an Historical OD obtained from the processing of CDR from Orange Telecom Operator provided by KINEO, Montero et al. (2019), one month of GPS waypoints from INRIX (38M waypoints and 75k trips for 8-10AM period on weekdays), and a set of link flow counts from 70 counting stations.

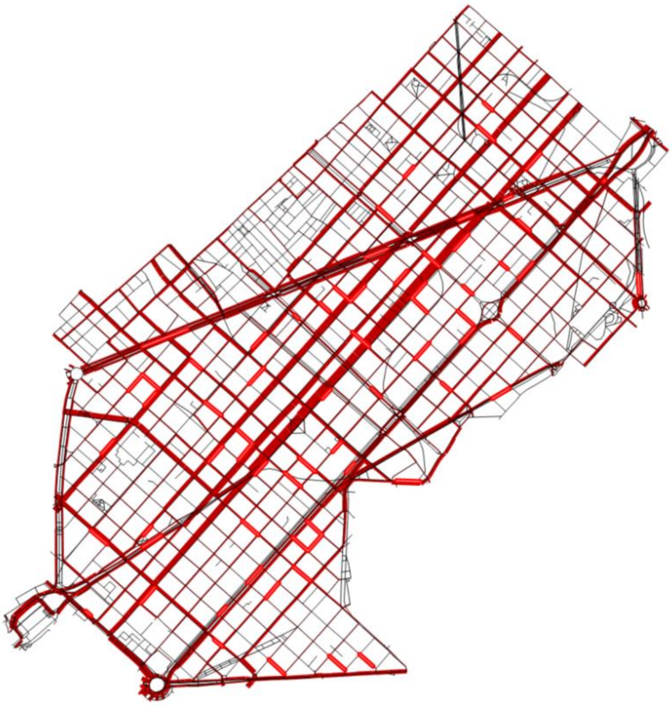

Figure 1: Test site: Barcelona's network

**References**


Frederix, R., Viti F. and Tampère C. (2013). "Dynamic Origin-Destination Estimation in Congested Networks: Theoretical Findings and Implications in Practice." *Transportmetrica A: Transport Science* 9(6):494–513.

Janmyr, J. and Wadell D. (2018), Analysis of vehicle route choice during incidents, Ms Thesis, University of Linköping, Department of Science and Technology, LiU-ITN-TEK-A--18/020—SE

Krishnakumari P., van Lint H., Djukic T. and Cats, O. (2019), A data driven method for OD matrix estimation, Transportation Research C, https://doi.org/10.1016/j.trc.2019.05.014

Montero L., Ros-Roca X., Herranz R., Barceló J.. (2019), Fusing mobile phone data with other data sources to generate input OD matrices for transport models, Transportation Research Procedia (www.sciencedirect.com), Volume 37,pp.417-424.

Nassir N., J. Ziebarth, E. Sall, and L. Zorn, "Choice Set Generation Algorithm Suitable for Measuring Route Choice Accessibility," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2430, no. 1, pp. 170–181, 2014.

Ortúzar J. D. and Willumsen L.G. (2011), *Modelling Transport*, John Wiley

Ros-Roca X., Montero L. , Barceló J., Nökel K., Gentile G. (2020), Transport Analytics approaches to the Dynamic Origin-Destination Estimation Problem, paper presented at MFTS 2020 (Luxemburg), submitted for publication in Transportation Research Series C.

Ros-Roca X., Montero L. and Barceló J. (2019), Investigating the Quality of Spiess-Like and SPSA approaches for Dynamic OD Matrix Estimation, accepted for publication in Transportmetrica.

X. Ros-Roca, L. Montero, A. Schneck, J. Barceló (2018), Investigating the performance of SPSA in Simulation-Optimization approaches to transportation problems, International Symposium of Transport Simulation (ISTS'18), Transportation Research Procedia, Volume 34, pp. 83-90. www.sciencedirect.com

Toledo, T. and Kolechkina T., 2013. Estimation of dynamic origin-destination matrices using linear assignment matrix approximations. IEEE Transactions on ITS 14(2), pp 618–626.