

Influence of Virtual Road Traffic Sensors of Oporto for Origin-Destination Matrix Estimation

Luciano U. Pando, Ricardo Lüders, Keiko V. O. Fonseca, Marcelo O. Rosa

¹Federal University of Technology – Paraná (UTFPR)
Postal code 80230-901 – Curitiba – PR – Brazil

luciano.pando@ifpr.edu.br, {luders,keiko,mrosa}@utfpr.edu.br

Abstract. *The knowledge of urban mobility patterns is important to maintain good public services as well as to improve city planning. These mobility patterns can be characterized by using expensive fieldwork or analyzing the huge amount of data available from services and environmental monitoring in smart cities. The origin-destination matrix estimation (ODME) aims to estimate the traffic of vehicles between origin and destination areas in the city of Oporto in Portugal from traffic observed from sensors installed at roads. This estimation is stated as an optimization problem and solved by linear programming. The results have shown that the number and location of sensors are important issues to be considered.*

1. Introduction

The origin-destination matrix (ODM) represents flows of vehicles or persons between different origin and destination points. This information is useful for city planning and is usually obtained through expensive and laborious field work which quickly becomes outdated [Chen et al. 2005]. An example of origin destination field work was accomplished in 2016 in Curitiba with an approximate cost of 6 million Brazilian currency [Curitiba 2016]. Alternatives to the ODM estimation (ODME) have been discussed for several decades, and many solutions have been proposed. A survey about the subject can be found in [Bera and Rao 2011].

The ODME is stated as an optimization problem where traffic estimates should match flow measures of vehicles from sensors. This problem is modeled using mixed integer linear problem (MILP) and solved with the GUSEK 0.2.24 solver. Moreover, sensors' position at roads is an important issue. This is referred to as the network sensor location problem (NSLP) in the literature. In this paper, we are interested on testing different strategies for sensor distribution at roads of Oporto and its effects on ODME.

Section 2 states the ODME and NSLP problems as well as performance metrics for ODME. Section 3 presents the MILP model and Section 4 shows the results. Conclusions are show in Section 5.

2. Background

In this section, the problem of origin-destination matrix estimation is presented as well as the underlying network sensor location problem. The performance metrics used for evaluation of ODME are presented as well.

2.1. Origin-destination matrix estimation (ODME)

The origin-destination matrix (ODM) contains information about flow of vehicles between two geographical areas in a city. Let assume that there are n origin-destination (OD) pairs with a flow q_w of vehicles in a particular pair $w \in W$ (the set of all OD pairs). These OD pairs are usually located at the borders of a region [Teknomo and Fernandez 2014]. Moreover, each flow of vehicles for an OD pair distributes over city roads using different routes. Each route is composed by links (or edges) identified as city roads. These links are eventually monitored by traffic sensors which gives the counts of vehicles in each link. As a link can be shared by different routes, counts in each link can be due to more than one OD pair. Hence, a link count is not a direct flow measure for a particular OD pair (except if it is not shared).

The origin-destination matrix estimation (ODME) can be considered as a dual problem of the static traffic assignment (STA) [Bierlaire 2002]. By giving the flow of vehicles in an OD pair, STA gives the flow of vehicles in each link of a route used by this pair. ODME on the contrary, provides the flow of vehicles for OD pairs from the flow observed in links.

The flow of vehicles v_a in a link a is given by (1). It is the sum of n flows q_w multiplied by a given probability p_{wa} of using link a by an OD pair w . Additionally, (2) is a non-negativity constraint of q_w .

$$v_a = \sum_{w \in W} q_w \cdot p_{wa} \quad (1)$$

$$q_w \geq 0 \quad (2)$$

By assuming that a link z is monitored by sensors, an estimate \bar{v}_z of the flow of vehicles in z should approximate the measured flow v_z according to (3). Similarly, \bar{q}_w is approximated to q_w according to (4).

$$\forall z (\bar{v}_z \approx v_z) \quad (3)$$

$$\forall w (\bar{q}_w \approx q_w) \quad (4)$$

According to [Bierlaire 2002], ODME is generally complex because the number of unknown variables (flow of vehicles in OD pairs) is greater than the number of equations (flow of vehicles in links). Other references point out different techniques to solve ODME as [Maher 1983] for Bayesian inference, [Cascetta 1984] and [Bell 1991] for general least squares, [Spiess 1987] for maximum likelihood and [Sherali and Hobeika 1994] for linear programming applications.

2.2. Network sensor location problem (NSLP)

The network sensor location problem (NSLP) concerns how to define the location and number of sensors. In [Ye and Wen 2017], nine rules have been considered for sensor allocation as follows:

- R1: choose links such that all OD pairs are covered by sensors;
- R2: choose links with high rate q_w/v_a ;
- R3: choose links having all routes covered by sensors (route covering rule);
- R4: choose links with high flow v_a (interception rule);
- R5: choose links with high probability p_{wa} (maximal OD demand fraction rule);
- R6: choose links with big number of routes (maximal net route flow captured);
- R7: choose links with big number of OD pairs (maximal network OD flow);
- R8: avoid redundant links (independence rule);
- R9: minimize the number of sensors (minimal observation rule);

The article [Ye and Wen 2017] considers rules R1, R8 and R9 for all experiments. Tests are made with rule R2 or R4 and R7 simultaneously. The results show that using R4 and R7 simultaneously is better than R2 only. The considered network has 93 links and 210 OD pairs.

In our work, R1 is considered by default. The other rules R2, R4, R7 are tested separately. Differently from [Ye and Wen 2017], R4 and R7 are applied separately, and there is also a combination between R4 and R7, indicated by R4R7 and done due the fact that preliminary results showed better results in these two rules. Another addition of this article is the focus on how the number of sensors interferes in different ways depending on sensor location choice rules.

2.3. Performance metrics

The following metrics are used to evaluate ODME: mean absolute error (MAE), root mean square error (RMSE), total demand deviation (TDD), and determination coefficient (R^2) according to (5), (6), (7), and (8), respectively.

$$MAE = \frac{\sum_{w \in W} |\bar{q}_w - q_w|}{n} \quad (5)$$

$$RMSE = \frac{\sqrt{\frac{1}{n} \sum_{w \in W} (\bar{q}_w - q_w)^2}}{\sum_{w \in W} q_w} \quad (6)$$

$$TDD = \frac{|\sum_{w \in W} \bar{q}_w - \sum_{w \in W} q_w|}{\sum_{w \in W} q_w} \quad (7)$$

$$R^2 = \left(\frac{n(\sum_{w \in W} q_w \bar{q}_w) - (\sum_{w \in W} q_w)(\sum_{w \in W} \bar{q}_w)}{\sqrt{(n \sum_{w \in W} q_w^2 - (\sum_{w \in W} q_w)^2)(n \sum_{w \in W} (\bar{q}_w)^2 - (\sum_{w \in W} \bar{q}_w)^2)}} \right)^2 \quad (8)$$

Both MAE and RMSE are error metrics used to evaluate differences between observed and real flows of OD pairs. On other hand, TDD is an error related to the sum of observed and real OD flows. All these metrics should be minimized for good results.

The determination coefficient R^2 evaluates the proportion of the variance in the dependent variable (estimated OD flows \bar{q}_w) that is predictable from the independent variable (real OD flows q_w) variance, resulting in a output between 0 and 1. The dependence between observed and estimated values is greater as $R^2 \approx 1$. in this case, R^2 should be maximized.

3. MILP for ODME

The ODM is composed by n decision variables representing the traffic flows of n OD pairs. The MILP model for ODME presented here is similar to the one of [Sherali and Hobeika 1994] but using GUSEK 0.2.24 to solve it [Bettoni 2018].

Variables are shown in Table 1. There are $(np + 3 \cdot ns)$ variables. Decision variables are q_w , estimated flows in links are \bar{v}_z , and F_{z+} and F_{z-} are variables used to quantify matching errors.

Table 1. Mixed integer linear programming variables

Variable	Type	Symbol	Amount
Estimated flow per OD pair w	$\mathbb{R} \geq 0$	q_w	np OD pairs
Estimated flow in link z	$\mathbb{R} \geq 0$	\bar{v}_z	ns sensors
Superior deviation in z	$\mathbb{R} \geq 0$	F_{z+}	ns sensors
Inferior deviation in z	$\mathbb{R} \geq 0$	F_{z-}	ns sensors
Total			$(np + 3 \cdot ns)$

The cost function is represented by (9).

$$\min_{q_w} \sum_{z \in Z} (F_{z+} + F_{z-}) \quad (9)$$

Equation (10) requires that a probability p_{wz} of flow q_w should use link z . Each observable link has a restriction of this kind.

$$\bar{v}_z = \sum_{w \in W} p_{wz} \cdot q_w \quad (10)$$

Equation (11) sets a matching between the observed and estimated traffic in link z . F_{z-} is positive when the estimated values are larger than the observed ones and F_{z+} is on the other way around. This restriction acts with (9) and (10) in order to find $\bar{v}_z \approx v_z$ by adjusting q_w decision variables.

$$v_z = (\bar{v}_z + F_{z+} - F_{z-}) \quad (11)$$

Restriction (12) is based on the non-negativity constraint specified in (2).

$$q_w \geq 0 \quad (12)$$

Restrictions (13) and (14) indicate that F_{z+} and F_{z-} are bound to non-negative values.

$$F_{z+} \geq 0 \quad (13)$$

$$F_{z-} \geq 0 \quad (14)$$

For every link sensor, there is a restriction as in (11), (13) and (14), and for every q_w , there is a restriction as in (12).

4. Results

Section 4.1 describes the data used to the ODME problem formulation. Section 4.2 shows results for ODME using MILP and different numbers of sensors.

4.1. Data description

In this work, data comes from the taxi trip database found in [UCI 2015] and the geographic database found in [OpenStreetMap contributors 2018]. After density clustering city areas and identifying routes (map matching) as shown in [Pando and Luders 2017], 13 areas of both origins and destinations have been identified in Oporto yielding to 169 OD pairs. Moreover, the true origin-destination matrix and the probability of route choice have been determined by GPS data of taxi trips.

The [UCI 2015] dataset has eight attributes and 1,710,670 instances. Attribute “polyline” is a sequence of lat-long locations which identifies a taxi trajectory in the network. The attribute “timestamp” is used to group taxi trips in hourly time intervals and to join all trips between 8-9 am allowing to get the true ODM. Attributes “Datatype” and “missing data” are used to select trips for workdays only and trips without missing data, respectively.

The geographic database [OpenStreetMap contributors 2018] defines the city network as a directed graph which is composed by GPS-referenced nodes and edges.

4.2. Influence of the number and position of sensors

The experiments have been made from 200 to 2000 virtual sensors with locations defined by rules R2, R4, R7 or R4R7. The number of sensors was increased by 100 between 200 and 500 sensors, and by 250 for more than 500 sensors.

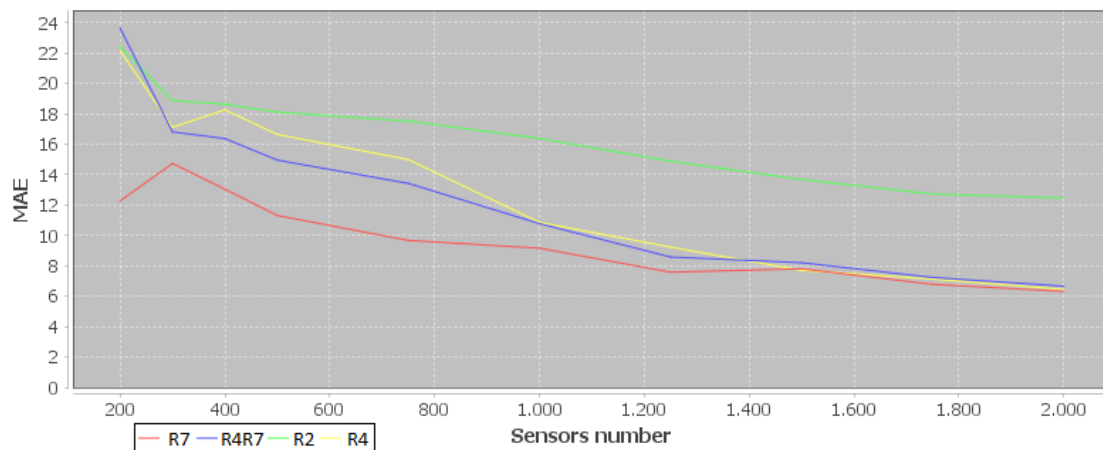


Figure 1. MAE for different rules and number of sensors.

According to Fig. 1 (MAE), rule R7 yields better results for any number of sensors except 1500 sensors. Particularly, for less than 1500 sensors. The gap between R7 and R4R7 or R4 is reduced for a large number of sensors. R4R7 has the worst result for 200 sensors, but it is better than R4 until 1500 sensors, where R4 is better. R2 has the worst result for every number of sensors except for 200 sensors. The results for R2 are close to

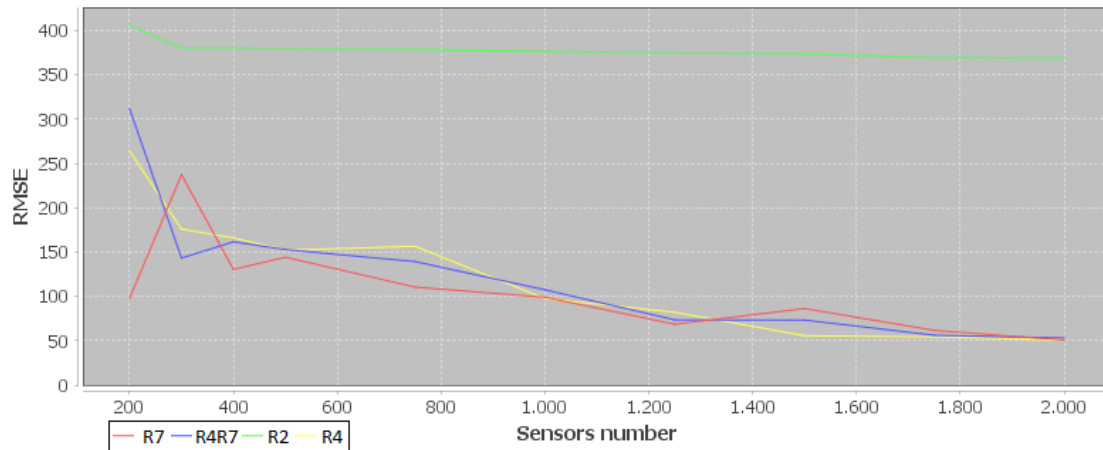


Figure 2. RMSE using different sets of rules and numbers of sensors.

R4 and R4R7 for small number of sensors but R4 and R4R7 are better for a larger number of sensors than R2.

In Fig. 2 (RMSE), R7 has better results for 200 sensors and between 400 and 1250 sensors. Results for R7 are worse for 300 sensors than for 200 sensors which is unexpected since the result was worse with more information added. R4R7 has better results for 300 sensors, and R4 has better results for 1500 sensors and above. R2 has the worst result for any number of sensors, showing a slightly decline in RMSE from above 400 to close to 330 for 200 and 2000 sensors, respectively.

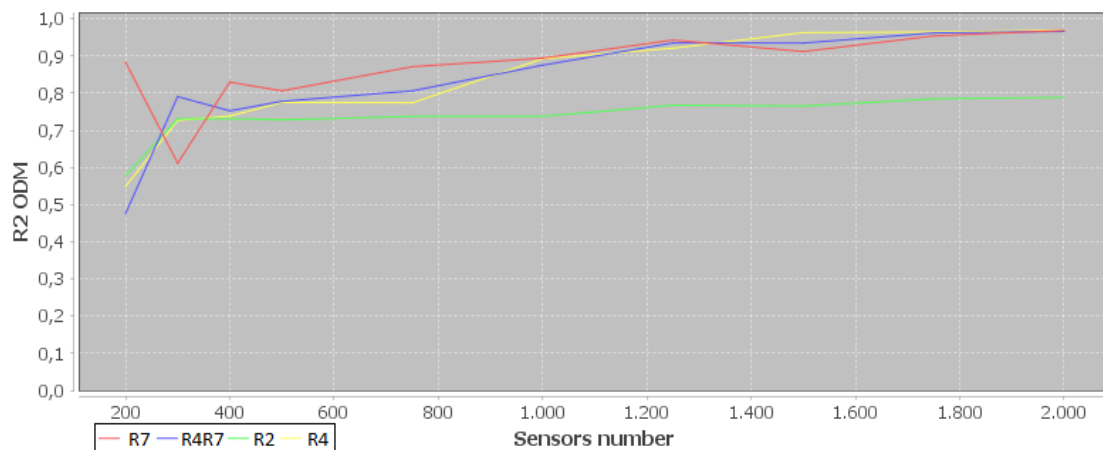


Figure 3. R^2 using different sets of rules and numbers of sensors.

Fig. 3 (R^2) has a similar information of Fig. 2. R7 also has better results for 200 sensors and between 400 and 1250 sensors in this metric. Results for R7 are also worse for 300 sensors than for 200 sensors. R4R7 has better results for 300 sensors, and R4 has better results for 1500 sensors and above. R2 has the worst result except for 200 or 300 sensors.

Fig. 4 (TDD) shows different behavior when compared to the three previous analysis. R4R7 has better results most of the time. R4, R7, and R4R7 are close to zero

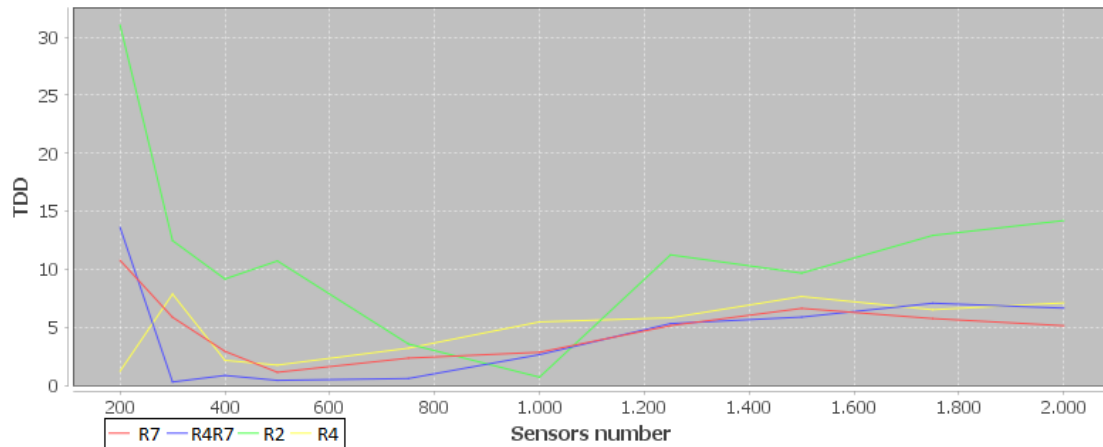


Figure 4. TDD using different sets of rules and numbers of sensors.

between 400 and 800 sensors, while the minimum TDD value for R2 occurs with 1000 sensors, which is exactly the number of sensors used in Section 4.2.

R7 has the better results in RMSE and R^2 metrics between 400 and 1250 sensors. It has the better results for any number of sensors except with 1500 sensors for MAE as well as the better results in TDD for 1500 sensors and above. R4 is better above 1250 sensors for RMSE and R^2 . R4R7 is only better for TDD between 300 and 750 sensors, and with 300 sensors for RMSE.

5. Conclusion

This work presented the effects of using different rules of distribution of sensors at roads of Oporto in Portugal in the quality of ODME. The quality of ODME was evaluated by four different metrics: MAE, RMSE, TDD and R^2 . The better set of rules may vary depending on the number of sensors and metrics used. For MAE, RMSE and R^2 metrics, experiments with more sensors produced better results using any set of rules. TDD is better mainly between 400 and 800 sensors, but it presents poor results after 1000 sensors. Rule R7 is better for two metrics and any number of sensors except for 300 sensors. For 1500 sensors and above, R4 has better results for metrics RMSE and R^2 but R7 has better results for the other two metrics, i.e., a tie occurs between R4 and R7. For 300 sensors, the best performing set of rules was R4R7 in three of four metrics.

For future work, the introduction of probe vehicles could add information to improve ODM and assignment matrix estimations. The influence of number and location of sensors can also be analyzed for other ODME techniques as for instance general least squares.

References

- Bell, M. G. H. (1991). The estimation of origin-destination matrices by constrained generalised least squares. *Transportation Research Part B: Methodological*, 25B(1):13–22.
- Bera, S. and Rao, K. V. K. (2011). Estimation of origin-destination matrix from traffic counts : the state of the art. *European Transport*, 49:3–23.

- Bettoni, L. (2018). Gusek description. <http://gusek.sourceforge.net/gusek.html>. Access in 2018-06-08.
- Bierlaire, M. (2002). The total demand scale : a new measure of quality for static and dynamic origin – destination trip tables. *Transportation Research Part B*, 36:837–850.
- Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data. A generalized least squares estimator. *Transportation Research Part B*, I(415):289–299.
- Chen, A., Chootinan, P., and Recker, W. W. (2005). Examining the quality of synthetic Origin Destination Trip table estimated by path flow estimator. *Journal of Transportation Engineering*, 131(July):506–513.
- Curitiba (2016). Pesquisa origem destino curitiba. Access in 2018, May.
- Maher, M. J. (1983). Inferences on trip matrices from observations on link volumes: A Bayesian statistical approach. *Transportation Research Part B*, 17(6).
- OpenStreetMap contributors (2018). Planet dump retrieved from <https://planet.osm.org>. Access in 2018, January.
- Pando, L. U. and Luders, R. (2017). Estimation of origin-destination matrix from traffic counts in the city of porto with pso and taxi trip data. volume 1, pages 901–911, Uberlandia, Brasil. Encontro Nacional de Inteligência Artificial.
- Sherali, H. D. and Hobeika, A. G. (1994). A linear programming approach for synthesizing origin–destination trip tables from link traffic volumes. *Transportation Research Part B*, 28(3):213–233.
- Spiess, H. (1987). A maximum likelihood model for estimating origin-destination matrices. *Transportation Research Part B*, 21(5):395–412.
- Teknomo, K. and Fernandez, P. (2014). A theoretical foundation for the relationship between generalized origin-destination matrix and flow matrix based on ordinal graph trajectories. *Journal of Advanced Transportation*, 48(6):608–626.
- UCI (2015). Taxi trips dataset of porto, portugal. Access in 2016, November.
- Ye, P. and Wen, D. (2017). Optimal traffic sensor location for origin – destination estimation using a compressed sensing framework. *Journal of Advanced Transportation*, 18(7):1857–1866.