ON THE EVALUATION OF THE OPORTO PUBLIC TRANSPORTATION NETWORK (STCP)

M. Carmo M. Guedes

Departamento de Matemática, Faculdade de Ciências, Universidade do Porto

Natália Oliveira

Sociedade de Transportes Colectivos do Porto (STCP)

Sérgio Santiago

Sociedade de Transportes Colectivos do Porto (STCP)

Gueorgui Smirnov

Departamento de Matemática e Aplicações, Universidade do Minho

ABSTRACT

A public transportation network serves in adequate way a population if it evolves in time following the existent social reality. Changes made in order to improve service must be analyzed and evaluated. The introduction of modern technology to validate the fare card allowed a quick access to important, although incomplete, data. A data basis with the validation information can be used to construct an origin-destination (OD) matrix which can be used for a service quality analysis. Here it is presented a computer based methodology to evaluate service quality criteria considering what might be interesting for the user. The quality analysis philosophy is the following. First, on the base of automatically gathered data reconstruct the origin-destination (OD) matrix which contains information concerning the number of passengers traveling between zones of a certain region. The OD matrix is used to calculate some criteria characterizing the transportation network quality, such as traveling times, waiting times at a stop or transport occupation. The reconstructed OD matrix always contains errors, which cause errors in the criteria values evaluation. How significant are these errors? A methodology based on statistical analysis for validating the criteria, i.e. for estimating the criteria robustness, is being implemented at the urban bus transport system of Oporto, STCP, allowing the evaluation of the transportation network quality under a number of criteria and guaranteeing rigorous evaluation.

Keywords: origin-destination matrix, transportation network quality evaluation, computational analysis

1. INTRODUCTION

A public transportation network is a complex services system with the social objective of adequately responding to the dislocation needs of the population of a certain region. In particular, in the case of a large urban area which has suffered in recent years great changes in terms of residential zones, many times quite distant from the working and leisure places, it is important that public transportation quickly responds to the changes trying to provide a better service.

Studying public transportation network involves the possibility of measuring its quality in terms of several quantities such as traveling times, number of line changes in a trip, waiting times at stops, vehicle occupation, and so on. It is also important to have an instrument allowing the study of eventual service alterations such as route changes, lines elimination and creation of new ones, augmenting/diminishing of the vehicles number.

A methodology was developed to validate criteria allowing to analyze the performing quality of an urban public transportation network with an emphasis on the user viewpoint. It has been implemented at the urban bus transport system of Oporto, STCP (Sociedade de Transportes Colectivos do Porto, Portugal).

The quality analysis philosophy is basically the following. First, one reconstructs the origin-destination (OD) matrix. This matrix contains information concerning the number of passengers travelling between zones of a certain region (Faria et al.; Macedo et al.). A similar OD matrix appears in studies on the quantity of the cargo moved inside a certain region (Ortúzar and Willumsen). Then the OD matrix is used to calculate some criteria characterizing the transportation network quality, such as traveling times, waiting times at a stop, and number of line changes in a trip, transport occupation. Finally, the most important moment of the evaluation is the criteria validation. Namely, the OD matrix reconstruction always contains errors, which cause errors in the evaluation criteria values. How significant are the errors? The methodology developed is based on statistical analysis, which allows validate the criteria, i.e. to estimate the criteria robustness. With the help of this methodology anyone evaluating a transportation network can accept or reject criteria and guarantee some rigor in the evaluation.

It is difficult and expensive to obtain directly the OD matrix either by measure or by surveys. However, using indirect means, it is possible to get a good approximation of it. In particular, the problem of the OD matrix evaluation for traffic flows has been quite studied (Abrahamsson, Eisenman and List; Gunnar et al.; Kwon and Varaiya; Dixon and Rilett). Also, a more recent study (Zhao et al.) describe how automatic data collection systems can be used to estimate passenger OD matrix in the context of a rail network.

The matrix reconstruction method depends on the information retrieved and several approaches to obtaining it have been developed (Spiess; Abrahamsson; Gaudry; Sherali et al.; Codina and Barceló; Codina et al.; Ortúzar and Willumsen; Doublas and Benitez). In this work the OD matrix is reconstructed using data from ticket validation by passengers getting in at any stop in the STCP network. This information, pertaining to specific time periods (peak hours, other hours of the day, working days, week-ends, summer, winter), can easily be obtained in an automatic way. A rechargeable fare card, identified with a number, is used to validate a trip when a passenger enters a vehicle making it an entry-only system. The information allows identifying changes on the passenger journey from one point to another in

the region, but does not give precise information about the final destination. Although the ticket number permits to trace patterns of regular behaviour of a certain passenger, the gathered information does not return a full OD matrix which must be approximated using reasonable hypothesis.
In planning a transportation network it is usual to use the idea that the movement is done through shortest time paths. There are many papers on the subject (Bielli et al. and Dial, for example). Under this hypothesis several operational quality criteria can be evaluated.
In this work we consider some specific information retrieved from the STCP ticket data basis to reconstruct the OD matrix and very simple evaluation criteria. However, the same analysis can be fulfilled for a more involved model.
The paper is organized in the following way: section two contains an informal discussion of some problems in the quality evaluation of a transportation network using indirect information and in section three a formal model for such a network is presented. In section four a reconstruction algorithm for the OD matrix is introduced and the validation criteria are considered in section five. Section six is dedicated to the general criteria validation methodology. The computational results are in section seven and the last one contains the conclusions.

2. INFORMAL OUTLINE OF THE PROBLEM

The problem addressed here is the following. Given some information concerning the movement of the transportation network passengers, for example the number of passengers getting in/out at any stop, we have to reconstruct the OD matrix and to evaluate some criteria characterizing the network quality. Moreover it is necessary to study the robustness of the criteria with respect to the OD matrix estimation error in order to preview possible network quality evaluation errors. To illustrate the main aspects of the problem we present a model example. It should be noted that can exist several types of information, but here we are considering only one of them.
Take an oriented line and suppose that at stops 1 and 2 two people were registered in and at stops 3 and 4 two people were registered out. We do not know at which stop the people getting out at stop 3 and 4 got in. Using this information, it is impossible to uniquely determine the OD matrix. All possible trips fitting the given getting in/out information correspond, respectively, to the following OD matrices:

$$\begin{bmatrix} 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

where an element (i, j) gives the number of passengers traveling from stop *i* to stop *j*. One can then conclude that it is not possible to reconstruct without ambiguity the OD matrix, even in a very simple situation.
Let the travel time between two consecutive stops be equal to 1. Then the mean traveling time for the three possible trips is always equal to 2.

Clearly more detailed information is contained in the graph of the function describing the number of persons traveling during a given time. Although the mean times are equal, the graphs are totally different (see Fig. 1).

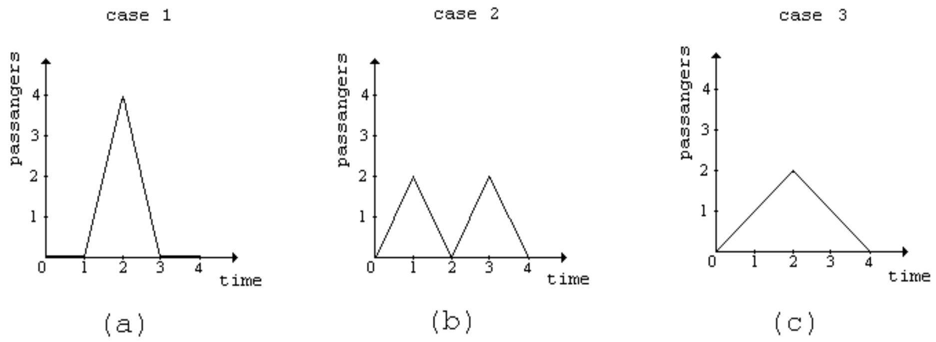

Figure1: Number of passengers as a function of the traveling time in the three cases.

For example, if we get the first matrix as the result of the OD matrix reconstruction in the model problem, then the function describing the number of persons traveling during a given time is given in Fig. 1(a). The 'real' OD matrix may be the second one and the 'real' function describing the number of persons traveling during a given time is that one shown in Fig. 1(b). The distance between the two graphs is significant. Thus the functions describing the number of persons traveling during a given time significantly depend on the OD matrix reconstruction quality, but the mean travel times are more robust. The analysis of all possible situations allows us to determine the transportation network quality evaluation criteria reliability. In a real situation the number of possible cases is enormous and a statistical analysis is needed in order to determine the transportation network quality evaluation criteria reliability.
It is possible to use several evaluation criteria (for example, waiting times at stops, bus occupation) but the main problems of the transportation network quality evaluation and criteria validation we face here, are always the same.

3. FORMAL MODEL OF A TRANSPORTATION NETWORK

Now a description of a transportation network formal model allowing a mathematical treatment of the problem is presented.

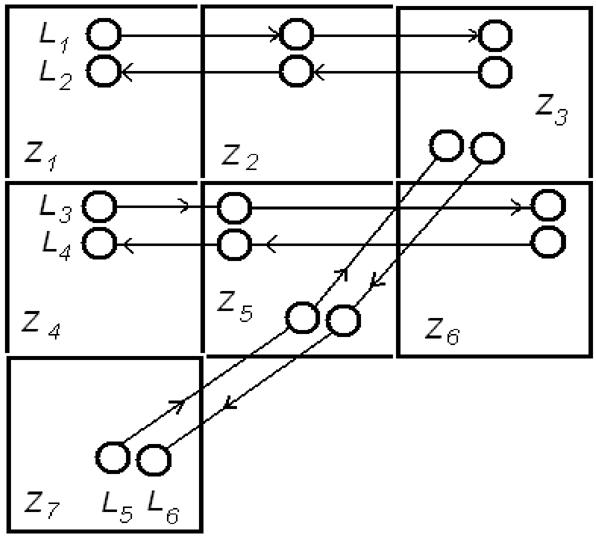

Figure 2: A transportation network.

Each vehicle is equipped with a device such that any passenger, boarding it, registers an identifiable ticket. If the time between two consecutive boardings is less than a certain fixed time, it is supposed that a line change has occurred. This gives the possibility of knowing the sequence of lines that someone takes in a certain route.

The formal model of an urban transportation network proposed here is the following. Consider a city divided in N zones, Z_k, k = 1,…, N. The transport lines are paths, L_m, m = 1,…, M, of an oriented network graph in which the nodes represent stops. In each zone there exists at maximum one stop of each line, as it is exemplified in Fig. 2. A passenger can only change from line L_m to the line L_n when there is a stop of both lines in the same zone. A route is defined as a sequence of lines which a passenger takes going from one zone to another. For all routes the number of passengers getting in at a certain stop is known. This number can easily be obtained by the ticket information. Formally this information is represented as a set of Q sequences of P_q, q = 1,…,Q, ternaries of the form ($Z_{k_p(q)}$, $L_{m_p(q)}$, w_q), p = 1,…,P_q, where w_q is the number of passengers that board line $L_{m_1(q)}$ in zone $Z_{k_1(q)}$, then board line $L_{m_2(q)}$ in zone $Z_{k_2(q)}$ and so on until the line $L_{m_{P_q}(q)}$.

The OD matrix is denoted by X where the elements X_{ij}, i, j = 1,…,N, represent the number of passengers journeying from zone Z_i to zone Z_j. To simplify it is supposed that X_{ii} is equal to zero.

4. OD MATRIX RECONSTRUCTION

In order to reconstruct (of course approximately) the OD matrix, the following algorithm can be used. Set X=0. Then identify all the routes whose last lines have 1, 2, 3,… stops of possible alighting. Order the routes by increasing order of the number of possible stops. To the ordered routes apply the following procedure: uniformly distribute the w_q passengers through the possible alighting stops of the last line $L_{m_{P_q}(q)}$, modifying $X_{k_1(q),i}$ in the following way

$$X_{k_1(q),i} = X_{k_1(q),i} + 1$$

Here i represents a stop of line $L_{m_{P_q}(q)}$ in zone $Z_{k_{P_q}(q)}$.

The algorithm ends after passing through all the routes.

The described process does not yet use all the information that can be collected without great expenses. When there is additional information about possible passengers alighting, the algorithm can be improved. For example, distributing the w_q passengers among the I stops of possible alighting, in line $L_{m_{P_q}(q)}$, as the solution w^i, i = 1,…,I, to the following minimization problem:

$$\sum_{i=1}^{I} | w^i - w_q w_0^i | \to \min,$$

$$\sum_{i=1}^{I} w^i = w_q \ ,$$

$$0 \leq w^i \ , \ i = 1,\ldots,I,$$

where $w_0^i \geq 0$, i = 1,…,I, verify

$$\sum_{i=1}^{I} w_0^i = 1.$$

The coefficients w_0^i are the anticipated percentages of the w_q passengers getting out at the i stops in that line. The OD matrix is modified in the following way:

$$X_{k_1(q),i} = X_{k_1(q),i} + w^i.$$

The described process gives an approximated OD matrix.

5. NETWORK PERFORMANCE EVALUATION

Once known the OD matrix it can be used to evaluate the network. There are many criteria to evaluate a transportation network valuing the user point of view. Here two of such criteria are considered, the travel time distribution and its mean value.

When the travel times between consecutive stops of each line are known it is possible to determine the shortest time routes between zones Z_i and Z_j. It is supposed that the passengers choose this type of route. From this information it is possible to determine function, $v(t)$, the number of people with travel time t and the mean travelling time value.

It should be noted that whatever the chosen criteria, the two used here or other more complex ones, the important is that the general methodology has to be the same, and eventually the results will be similar to those presented below.

6. CRITERIA VALIDATION METHODOLOGY

Let X be the OD matrix and let X^R be its reconstruction. We calculate a criterion of evaluation $c = c(X)$ and also the criterion value using the reconstructed OD matrix X^R, to obtain $c^R = c(X^R)$, which can be different from $c = c(X)$. If the difference $c - c^R$ is significant, the criterion is not useful. How to evaluate the difference? This can be done with a statistical study. Namely, we generate a sequence of random matrices $X_k, k = 1,2,..., K,$ satisfying some natural restrictions, like non-negativeness, restrictions on the maximal values of elements, etc. The set of these matrices can be much larger than the set of matrices which are in agreement with the gathered ticket information. Notice that to generate uniformly distributed random matrices satisfying the ticket information is a very hard problem because the function that makes the correspondence of a matrix to the gathered information is strongly non-linear. So we need an approach that overcomes that difficulty at cost of generating a larger set of matrices.

Thus, we simulate the input information for the OD matrix reconstruction and use a reconstruction algorithm to calculate a sequence of matrices $X_k^R, k = 1,2,..., K$. (The value K should be determined experimentally.) Next, we use the reconstructed matrices to evaluate the random variable ψ :

$$\psi = \frac{| c(X) - c(X^R) |_C}{| c(X) |_C},$$

where $| c |_C$ is the norm in the space of the function $c = c(X)$ values. This random variable characterizes a relative error for the criterion $c = c(X)$. A similar random variable, $\tilde{\psi}$, can be defined when X is a matrix satisfying the ticket information, that is with more restrictions than those considered to define ψ. Let m ($\tilde{m}$) and s ($\tilde{s}$) be the mean value and standard deviation, respectively, of ψ ($\tilde{\psi}$). It should be noted that m is not a good parameter to reach conclusions about a criterion because $\tilde{m}$ can be greater than m, as is shown in Fig. 3. But defining

$$\psi_{\max} = \sup \{\psi \mid f(\psi) > 0\},$$

that is the greatest ψ value such that $f(\psi)$ is nonzero, and similarly

$$\tilde{\psi}_{\max} = \sup \{\tilde{\psi} \mid f(\tilde{\psi}) > 0\},$$

we always have $\tilde{\psi}_{\max} < \psi_{\max}$.

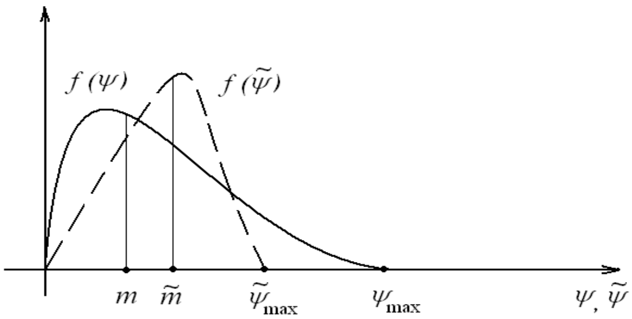

Figure 3.The probability density functions of ψ and $\tilde{\psi}$.

The value ψ_{max} can be obtained from the graph of *f* (ψ) but it is possible to give a quantitative expression to obtain it approximately. Usually one would expect $\psi_{max} \approx m + 3s$ for well behaved distributions. In computational experiments presented in the next section we have

$$m + 3s < \psi_{max} < m + 4s$$

in the case under consideration. And thus, for example, if $\psi_{max} \approx m + 3s$ is rather small, say 2% or 3%, one can conclude that the relative error in the criterion evaluation is acceptable and therefore the criterion can be utilized to evaluate the transportation network quality. Otherwise the criterion should not be used.

7. COMPUTATIONAL EXPERIMENTS

This methodology was implemented at the urban STCP, considering 253 zones as well as a system of lines connecting the stops, following the principle of only one stop in each zone. In reality this stop is a mega-stop, combining a certain number of consecutive stops. The important is that in each zone a person is at a walking distance of whatever place he wants to go to. Travel times between two consecutive stops and the time intervals between two passages of transport, for each line, were given. In any experience a matrix, 253×253, was randomly generated and considered as the true OD matrix. Using the shortest time route hypothesis, the function $v(t)$, the number of people with travel time t, as well as the mean time value, was calculated, and an input information for the OD matrix reconstruction algorithm was obtained. Based on these data, and with the help of the previously described algorithm, a reconstruction of the OD matrix was done. Next, for the reconstructed matrix, the evaluation was done calculating, as previously, the functions $v^R(t)$, the number of people travelling in time t as well as the respective mean time value. The values of the following variables (the relative errors) were calculated:

$$\xi_v = \frac{\int_0^\infty | v(t) - v^R(t) | \, dt}{\int_0^\infty v(t) dt}$$

$$\mu_v = \frac{\left| \int_0^\infty tv(t)dt - \int_0^\infty tv^R(t)dt \right|}{\int_0^\infty tv(t)dt}$$

After a series of such experiments the empirical probability density functions of these relative errors were obtained as well as their respective mean values, m and standard deviations, s. This experiment was repeated enough times to guarantee the stabilization of m and s. The results are presented in Fig. 4 and 5. The analysis of these graphs allows estimate the reliability of the chosen transportation network quality evaluation criteria and to accept or reject them. In all experiments we have

$$m + 3s < \psi_{\max} < m + 4s$$

Therefore the value $\psi_{\max} \approx m + 3s$ can be used as an indicator of a criterion validation.

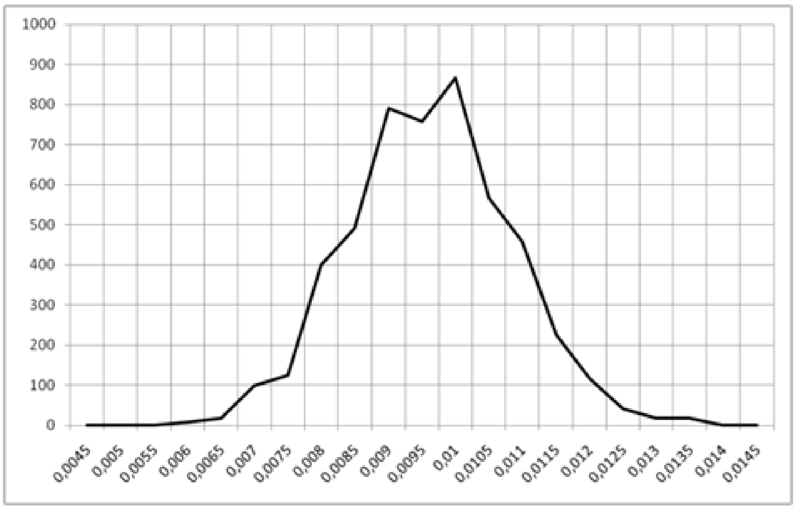

Figure 4: Empirical probability density function of variable ξ_v ;

$m\ (\xi_v) = 0.0094$, $s\ (\xi_v) = 0.0012$.

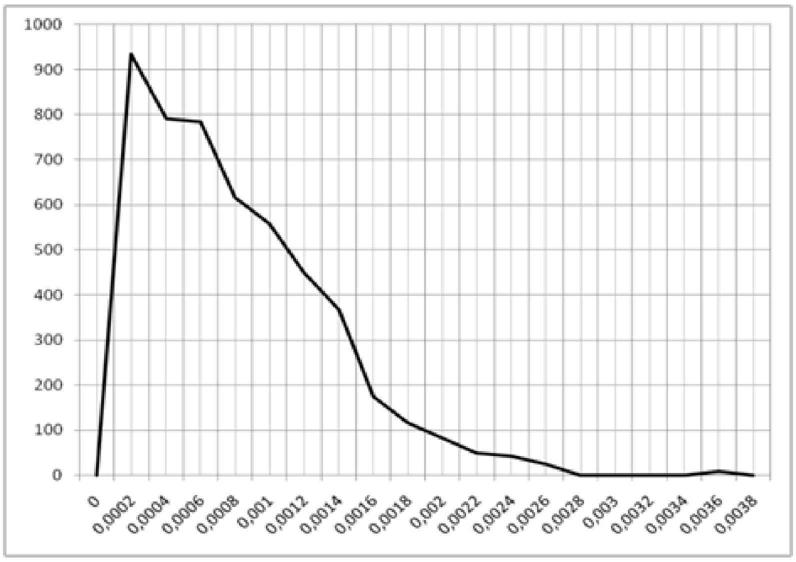

Figure 5: Empirical probability density function of variable μ_v ;

m (μ_v) = 0.00071, s (μ_v) = 0.00053.

8. CONCLUSIONS

In this work a methodology for a transportation network quality evaluation is presented. The approach is based on the OD matrix reconstruction algorithm. Evaluation criteria, the travel time distribution and the mean travel time, are determined under the hypothesis of passengers choosing the shortest time paths. The robustness of these criteria, i.e. the influence of the OD matrix reconstruction error on these evaluation criteria, was statistically studied and the indicator of a criterion validation was introduced. It should be noted that the evaluation criteria and the algorithm for the reconstruction of the OD matrix may be different. This general method of public transportation network quality evaluation is implemented at STCP and is ready to be used.

REFERENCES

Abrahamsson, T. (1998). Estimation of origin-destination matrices using traffic counts - A literature survey. Interim Report IR-98-021, International Institute for Applied System Analysis.

Bielli, M., A. Boulmakoul, and H. Mouncif (2006). Object modeling and path computation for multimodal travel systems. European Journal of Operational Research. 175, 1705-1730.

Codina, E. and J. Barceló (2004). Adjustment of O-D trip matrices from observed volumes: an algorithmic approach based on conjugate directions. European Journal of Operational Research, 155, 535-557.

Codina, E., R. Garcia, and A. Marín (2006). New algorithmic alternatives for the O-D matrix adjustment problem on traffic networks. European Journal of Operational Research, 175, 1484-1500.

Dial, R. (2006). An efficient algorithm for building min-path trees for all origins in a multi-class network. Transportation Research Part B, 40, 851-856.

Dixon, M. P. and L. R. Rilett (2005). Population origin-destination estimation using automatic vehicle identification and volume data, Journal of Transportation Engineering, 131, 75-82.

Doublas, J. and F. Benitez (2005). An approach to estimating and updating origin-destination matrices based upon traffic counts preserving the prior structure of a survey matrix. Transportation Research Part B, 39, 565-591.

Eisenman, S. and G. List (2004). Using probe data to estimate OD matrices. IEEE Intelligent Transportation Systems Conference, Washington D.C., October 2004.

Faria, C., J. Vieira, J. Sorratini, and M. Macedo (2004). Matriz origem/destino - um instrumento para a optimização do transporte público urbano por ônibus. Métodos Computacionais em Engenharia, APMTAC, Lisboa, Portugal.

Gaudry, M. (2000). The four approaches to origin-destination matrix estimation for consideration by the MYSTIC Research Consortium. Bureau d'économie théorique et appliquée, Working Paper No 2000-10.

Gunnar, A., M. Johansson, and T. Telkamp (2004). Traffic matrix estimation on a large IP backbone - A comparison on real data. IMC'04, Taormina, Sicília, Itália.

Kwon, J. and P. Varaiya (2005). Real-time estimation of origin-destination matrices with partial trajectories from electronic toll collection tag data. Annual Meeting Transportation Research Board, Washington D.C., January 2005.

Macedo, M., J. Sorratini, C. Faria, and W. Carvalho (2004). Avaliação de um Sistema Integrado de Transporte Coletivo Urbano, Métodos Computacionais em Engenharia, APMTAC, Lisboa, Portugal.

Ortúzar, J. and L. Willumsen (2004). Modelling Transport. John Wiley & Sons, New York.

Sherali, H.D., A. Narayanan, and R. Sivanandan (2003). Estimation of origin-destination trip-tables based on a partial set of traffic link volumes, Transportation Research. Part B: Methodological, 37, 815-836.

Spiess, H. (1990). A gradient approach for the O-D matrix adjustment problem. Publication 693, Centre for Research on Transportation, University of Montreal.

Zhao, J., A. Rahbee, and N. H. M. Wilson (2007). Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. Computer-Aided Civil and Infrastructure Engineering, 22, 376-387.