TRANSPORT
ISSN 1648-4142 print / ISSN 1648-3480 online
2012 Volume 27(2): 165–170
doi:10.3846/16484142.2012.693894

THE EXTENT OF INFLUENCE OF O–D MATRIX ON THE RESULTS OF PUBLIC TRANSPORT MODELING

Rasa Ušpalytė-Vitkūnienė[1], Vytautas Grigonis[2], Gražvydas Paliulis[3]

Dept of Urban Engineering, Vilnius Gediminas Technical University,
Saulėtekio al. 11, LT-10223 Vilnius, Lithuania
E-mails: [1]rasa.uspalyte@vgtu.lt (corresponding author); [2]vytautas.grigonis@vgtu.lt; [3]msk@vgtu.lt

Submitted 28 October 2011; accepted 27 December 2011

Abstract. A rapidly developing and equality-based society needs a reliable and attractive public transport system. With booming mobility in Lithuania, a great importance has been attached to the sustainable development concept, and public transport has been given priority in the urban transport system. Accessibility and comfort of public transport are essential indicators that guarantee equal travelling opportunities for all people. Transport modelling is the only economical and sufficiently reliable way to carry out a forward assessment of the impact of innovations to be applied to the overall system without involvement of passengers. This paper considers estimation of the origin–destination (O–D) matrix and its size correction. The public transport (PT) system of Vilnius City was taken as a basis for the research. Modelling of Vilnius City public transport was carried out with the help of VISUM software. Modelling of the public transport route network in Vilnius is aimed at improving the quality of life of inhabitants of the city. The O–D matrix is one of the key elements in modelling. Reliability of modelling results is based on reliability and size of the matrix. Although many scientists analyse the problem of estimating an O–D matrix, this paper focuses on the size of the O–D matrix required in order to give reliable results in PT modelling. During the first step, the matrix of 230 transport districts is estimated, which is reduced by 10 percent with every following step. The aim of this article is to find the break point in the size of O–D matrix where the reliability of PT modeling results falls.

Keywords: passenger transportation, origin–destination matrix, mass passenger public transport, modelling, passenger flow.

1. Introduction

A rapidly developing and equality-based society needs a reliable and attractive public transport system. A gradually increasing demand for public mobility attaches a greater importance to the sustainable development concept and public transport is recognized as a priority mode of transport in urban communication systems. Accessibility and comfort of a public transport system is the main indicator ensuring equal travelling opportunities for all people. Transport modelling is the only economical and sufficiently reliable way for an advanced assessment of the impact of innovations to be used in the entire system without involvement of passengers. Every city has a structure that develops on the basis of local inhabitants' needs; it is filled with ideas of local planners and influenced by global and local economic conditions. Consequently, it is crucial to find economically reasonable and sustainable infrastructure solutions to an array of problems that also include traffic flow. A comparison and substantiation of various urban transport infrastructure development strategies in Vilnius and the documentation of these strategies are an essential part of a comprehensive urban planning process.

O–D matrix generation, handling, size and impact on the results of simulation were examined by a number of scientists all over the world in the last few years. But many of these researches were limited to generating private vehicle traffic models. In their studies, Codina and Barceló (2004) examined the ways to avoid the permanent renewal of the O–D matrix, which is expensive. They showed that data update and a likely possibility of small errors can be avoided through believable programming formulations. However, this must be done without reducing the particularity of the matrix.

Such researches were carried out by many foreign scientists (De Grange *et al.* 2010; Nie, Zhang 2008; Marzano *et al.* 2009). O–D matrix calibration characteristics and reliability of the results has demonstrated that grouping of data into larger subdivisions could have an immense influence on the trip distribution parameters, while support to smaller cells (detailed data) allows staying close to the real flow distribution. The accuracy of the O–D matrix may reduce other modeling errors as well.

This paper is focused on analysis of the size of the O–D matrix required to receive reliable results in PT modelling. During the first step, the matrix of 230 transport districts was estimated, which was reduced by 10% with every following step.

The aim of this article is to find the break point in the size of O–D matrix where the reliability of PT modelling results falls.

2. Development of the Vilnius City Public Transport Model

The analysis is focused on Vilnius City, as a case study; i.e., the article presents modelling of transport flows with the help of VISUM 9.3 software and comparison of infrastructure development strategies. This paper studies generation of the origin–destination (O–D) matrix and correction of its size. As a base for research the public transport (PT) system of Vilnius City was taken. Modelling of the Vilnius public transport was carried out with the help of VISUM software. Modelling of a public transport route network in Vilnius is aimed at improving the quality of life of inhabitants of the city. The O–D matrix is the core of modelling.

The Vilnius City public transport model was developed to as precisely as possible represent passenger flows at a peak hour. The model reflects a system of integrated travel with combined travel modes at three levels: a combined-mode choice, a transfer location choice and a route choice. At present, the developed models are used in all alternatives to be modelled.

The Vilnius City public transport network was modelled with the help of VISUM software package, which integrates all of the necessary aspects of private and public transport planning into one comprehensible transport model.

In order to generate passenger flows of Vilnius public transport, an integrated 4-step travel demand modelling – that covers traffic generation, distribution, travel mode choice and route choice – was used. In separate alternatives, the modes and routes of public transport chosen by residents are modelled as a combination of time indicators, such as distance to a stop, time spent in a vehicle, number of transfers and etc, defined by users when choosing a route between transport districts.

The public transport network of Vilnius was formed based on the data of research and analysis of the existing situation of public transport flows in 2011 (Vilniaus miesto susisiekimo… 2002). Vilnius and its suburban areas, serviced by public transport, were divided into 217 transport districts. The Vilnius City public transport model was provided with two-way stops and also with a group of stops in some intersection zones since detail representation of this area has no significance for model accuracy. The public transport network of 2561.32 km was formed including more than 600 stop groups and approx. 750 additional points (a number varies subject to modelled alternatives), which are necessary to as accurately as possible represent the network of routes. The centres of transport districts were joined with stops by 3500 pedestrian paths showing the shortest ways not only to the nearest stops from the centre but also to the stops crossed by other routes or even other transport modes. In order to only enter the possible pedestrian routes, as an auxiliary measure, the basic topographic data should be used around a water body or a green area. Routes of buses, trolley-buses and mini buses with timetables were entered as well. During the further modelling stage, only timetables of the newly entered routes were changed.

During a day, features of the Vilnius public transport network are continuously changing due to demand for public transport and public vehicles get periodically overloaded with passengers. The main problems of public transport emerge with the maximum of passenger flows that is reached during the morning peak hour.

Before modelling theoretical alternatives to public transport systems, the current public transport model was calibrated taking into account the extent it corresponds to the real data recorded during the survey of passenger flows. During verification, cross sections of streets were selected to most accurately represent distribution of passengers within the network.

3. Modelling Possibilities with the Help of VISUM Software Package

Mobility of public transport passengers in the public transport system of the city is more complicated than that in the model since people can freely choose between modes of public transport. A passenger can choose a route and a type of vehicle as well as transfer possibilities by public transport modes. In 1994, Fernandez *et al.* suggested to develop an integrated travel system that combined travel modes in three levels: a combined-mode choice, a transfer location choice and a route choice (Lo *et al.* 2004; Wardman 2004). Currently, travel modes suggested by him are used almost in all popular models aimed at public transport modelling (e.g. Scenes, Litres-2, Expedite, Logit, Lohse, Kirchhoff and etc.).

Among software packages used for public transport modelling, VISION, EMME/2, TRIPS and TRANSYT are the most popular in Europe and USA; while GETRAM and ASCII are more frequently used in Asia. VISION and EMME/2 are the most popular choices around the world. Both of them use an integrated 4-step travel demand modelling that includes traffic flow generation and distribution in public transport network of the city, and travel mode choice (Duff-Riddell, Bester 2005; Murray 2001). Public transport of Vilnius City was modelled with the help of the VISUM software package.

When modelling flows, the integrated method that considers the entire time of a journey was used. Modelling is based on the public transport system provided with a timetable. The specific times of departure and arrival of each vehicle and their intermediate time at stops are considered. Other than diverting passenger flows to the best route, the choice of a route is made on the basis of its Impedance Index (IPD).

The IPD is described as a combination of time indicators – such as distance to a stop, time spent in a vehicle, number of transfers and etc. – defined by the user when choosing a route between transport districts. When modelling, for each route the percentage of passengers who will choose it $\left(P_i^a\right)$ is calculated in the total travel demand *i* during the selected time interval *a*. The use of every link U_i^a is calculated according to the travel distribution function from IPD_i^a.

The main characteristics of each model obtained are studied during its selection; these are journey time expenditures. When selecting the route of public transport, the following is taken into consideration:

- time spent in a vehicle;
- communication possibility;
- walking time to the starting stop;
- total walking time during the entire journey;
- initial time spent waiting for a vehicle;
- transfer time;
- number of transfers (expressed in time);
- number of transfers (expressed in units and effecting on the cost of a journey).

The IPD of a new model is based on the passenger survey carried out in 2001–2002 in Vilnius City, during which the passengers of public transport pointed out the main reasons for choosing the public transport mode and route, namely:

- time spent waiting for a public vehicle at a starting point and at transfer locations;
- time spent to reach the stop;
- total travel time spent in a vehicle during the entire journey.

Many foreign scientists agree that in order to identify the routes most probably chosen by passengers it is not only necessary to take into consideration the length and time of travel but also look at the minimum cost. Maximum waiting time should only occur at the beginning and at the end of the journey since it can be best controlled by the passenger, while all other weight functions are minimal (Lo *et al.* 2004; Wardman 2004; Eliasson, Mattsson 2000; Horn 2003, 2004). Therefore, the above discussed indicators were given a higher coefficient of significance when generating public transport flows. Besides, travel expenses and passenger behaviour have a significant impact on travel mode and route choice as well. In order to reduce the influence of unpredictable characteristics on simulation results, hybrid simulation, stochastic simulation and genetic algorithms can be used (Li *et al.* 2011; Moore 2011).

Using these methods, the cartograms are generated showing flows on a street network. The cartograms of separate public transport modes are obtained. Since the flows of passengers are obtained in order to generate traffic flow cartograms, it is necessary to take into consideration the average number of passengers using separate public transport modes. When modelling traffic flows, it is possible to predict where the flows will considerably grow and a larger number of buses and trolley-buses on the same line will be required, where to widen streets or divert traffic flows to other streets, and also where the flows will decrease. It is possible to estimate the payback period of newly constructed bus lines, contact networks for trolley-buses or public transport modes.

Calibration of the Vilnius City public transport network was carried out with the help of three selected – a *TSys-based, a Headway-based* and *a Timetable-based* – methods. The *Timetable-based* model uses the Logit, Kirchhoff, BoxCox and Lochse distribution laws interchangeably. In the earlier research performed by the authors of this article, the conformity coefficients were determined by calibrating the Vilnius City public transport model using all the methods stated above. The *TSys-based* and *Headway-based* models for generation of public transport passenger flows were rejected as absolutely unreliable. It was determined that the most reliable method for modelling the Vilnius City public transport was the *Timetable-based* model with the selected Kirchhoff and BoxCox distribution laws, of which the average coefficient of conformity to the base research of 2002 was 0.82 and 0.81, respectively (Ušpalytė-Vitkūnienė 2006).

The *Timetable-based* model with Kirchhoff distribution law to generate travels was selected as the main model for further analysis. Modelling with the use of this law is based on a public transport system and public transport timetables. When modelling by this method, the time of arrival and departure of each individual vehicle is taken into consideration. Rather than diverting passenger flows to the best route, one is chosen depending on its IPD.

Therefore, the *Timetable-based* model with Kirchhoff distribution law was selected for further research.

When modelling for each route the percentage of passengers who will choose this it $\left(P_i^a\right)$ is calculated in the total travel demand *i* during the selected time interval *a*. The use of every link U_i^a is calculated according to the travel distribution function from IPD_i^a:

$$U_i^a = f\left(IPD_i^a\right);\ P_i^a = \frac{U_i^a}{\sum_{j=1}^{n} U_j^a},$$

where: *n* – the total number of links.

In case of the VISUM software, the Kirchhoff distribution law is described as follows:

$$U_i^a = IPD_i^{a-\beta};\ P_i^a = \frac{IPD_i^{a-\beta}}{\sum_j IPD_j^{a-\beta}}.$$

Coefficient β was introduced to describe the IPD sensitivity and to reflect criteria for the choice of travel mode and route by inhabitants.

When calibrating models, the coefficients β were first of all selected for the distribution laws of Logit, Kirchhoff, BoxCox and Lochse of the *Timetable-based* model, which most optimally conforms to the distribution of passenger flows. Based on obtained results, the further research with β = 4 in Kirchhoff distribution law is described; the coefficient β in Logit, BoxCox and Lochse distribution laws is equal to 0.25, 1 and 1, respectively.

As the morning peak hour reflects the maximum passenger flows and the greatest street capacity problems, it was selected as a time interval of modelling. The model represents the public transport passenger flows, passenger distribution within the route network and their transfer locations. The model provides information about the start and end stops chosen by passengers, loading of public transport routes, passenger journey time between the transport districts, number of transfers in stops and between transport modes and etc.

The correlation coefficient r shows the relationship between one of the factors y (the public transport passenger flows generated by the model) and another variable x (the public transport passenger flows recorded during the survey), thus allowing to evaluate the extent to which the developed public transport system model corresponds to the functioning of Vilnius public transport system (Fig. 1).

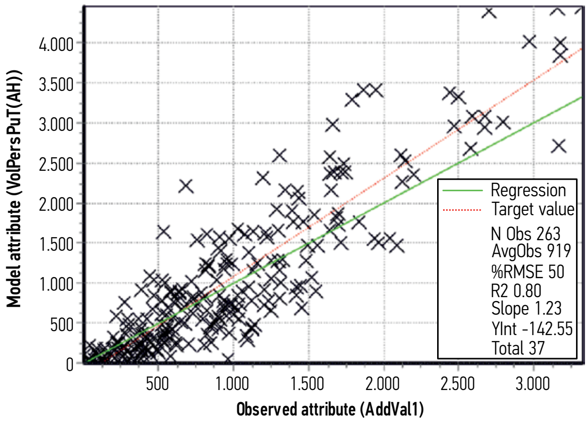

Fig. 1. Correlation between the existing passenger flows and the generated passenger flows in the model

Analysis of the modelling results shows that the correlation coefficient of the existing situation of Vilnius City public transport $r = 0.89$ and this proves that the public transport modelling is reliable enough to be based on when planning the development of public transport system.

4. Relative Dependence Between the Reliability of Modelling Results and the Size of an O–D Matrix

The main objective of public transport modelling is to provide citizens with the top level of public transport services. But the accuracy of modelling results depends on many factors, such as accurate route network, timetables, selection of IPD indices, accuracy and size of an O–D matrix as well as its ability to reflect locations of concentrations of citizens.

Many scientists studied the influence of O–D matrix estimations and accuracy on the reliability of modelling results. All of them agree that an O–D matrix and its size is one of the most important, if not the most important, indices for modelling public transport or any other transport mode. An O–D matrix correction is a common process used in modelling of transport systems and studied by the scientists Cascetta (2001), Marzano *et al.* (2009), Hazelton (2003), Nie and Zhang (2008). They studied the connection between an O–D matrix and obtained results, and determined that the accuracy of an O–D matrix also depends on a passenger survey, which is the basis of the matrix. The blank cells in an O–D matrix can cause non-connections between journeys, therefore, in order to get more accurate results, scientists suggest correcting transport districts of the matrix and increasing modelling intervals up to the entire day as this allows filling in the matrix cells better. The article gives no analysis of the change in estimation results with the increase in a modelling interval; however, it studies the change in modelling results with the decreasing size of an O–D matrix. With each step, the O–D matrix is reduced by 10% of the initial model (where the division of an O–D matrix amounts to an average of 3691 inhabitant per one transport district). The districts were increased depending on the number of residents, the number of journeys and landscape obstacles. The main indices selected for the modelling were: journey time, the number of transfers and the IPD index.

When increasing the size of one zone, it could be expected that the time of a journey would increase as well, since the time spent to reach the stop and the destination point from the stop increases and the number of transfers decreases, since certain journeys that include a transfer are integrated into the increased zones. The general IPD index, describing the entire trip, would correspondingly increase. Modelling results show (Table) that in the beginning, the journey time increases insignificantly; however, at a threshold of 6000 residents, the journey time starts increasing more rapidly. A more significant increase in the number of transfers was recorded at the threshold of 7500 residents per one transport district.

Table. Changes in relative coefficients

No.	Number of districts	Average number of residents in one district	Journey time, in arbitrary units	Number of transfers	IPD
1	230	3691	1.00	1.000	1.00
2	207	4101	1.01	0.993	1.01
3	184	4614	1.01	0.985	1.02
4	161	5273	1.01	0.982	1.02
5	138	6152	1.02	0.976	1.03
6	115	7382	1.03	0.969	1.05
7	92	9228	1.05	0.955	1.08
8	69	12304	1.10	0.924	1.26

Modelling results are mostly influenced by the IPD index, described as a combination of time indicators defined by the user when choosing the route between transportation districts. This turning-point is clearly seen in the obtained results. Behind the turning-point, the indices of journey time and the number of transfers deviate from their base indices by approx. 2÷5% per step. The turning-point of the IPD coefficients, having reached 8700 residents per one district, is clearly seen in Fig. 2.

When constructing an O–D matrix, its size is affected by the length of time and the amount of financial investments. The smaller the O–D matrix by the size of districts, the less accuracy of passenger surveys is required; and the lower the scope of surveys, the better filling of O–D matrix cells. Therefore, it is very important to know the limits of the size of a transport district, for which errors of results could still be predicted.

Based on the obtained results, it is suggested for the O–D matrix to avoid transport districts with more than 8700 residents as much as possible. Having exceeded this number, errors of results become hardly predictable.

Aiming at comprehensive research, it can hardly be assumed that the Kirchhoff distribution law is the most accurate and reliable when the number of residents of one transport district is increased in the course of modelling. It is necessary to analyse the other modelling alternatives under the same conditions based on the Logit, BoxCox and Lochse distribution laws.

If to compare results obtained by modelling of the Vilnius City public transport based on four different distribution laws (Fig. 3), it is evident that in case of the BoxCox alternative, the IPD index increases by a 5% reliability limit at the same time as the number of residents per one transport district reaches 3700. The indices of transfer and journey time in all modelling stages remain within a 5% reliability limit, though, it could be concluded that the modelling, based on BoxCox distribution law, is unpredictable due to unstable reliability of relative indices, therefore, this distribution law is not recommended for modelling of public transport passenger flows.

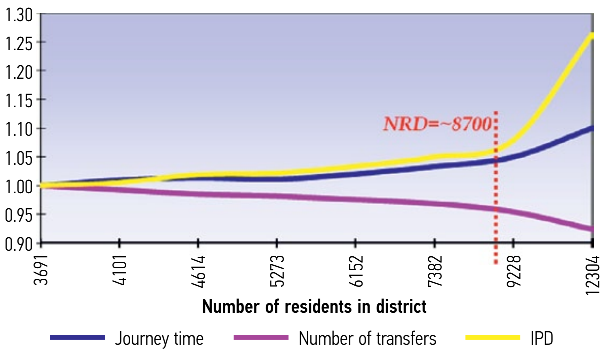

Fig. 2. Dependence of relative indices on the size of O–D matrix zones

The analysis of research results shows that in public transport models, based on Lochse distribution law, the index of transfer remains highly reliable and plunges only at the level of 9000 residents per one transport district.

Having studied and assessed all the relative indices, it could be concluded that reliability of Lochse distribution law, when increasing modelled transport districts, remains the highest and the most stable for a long time.

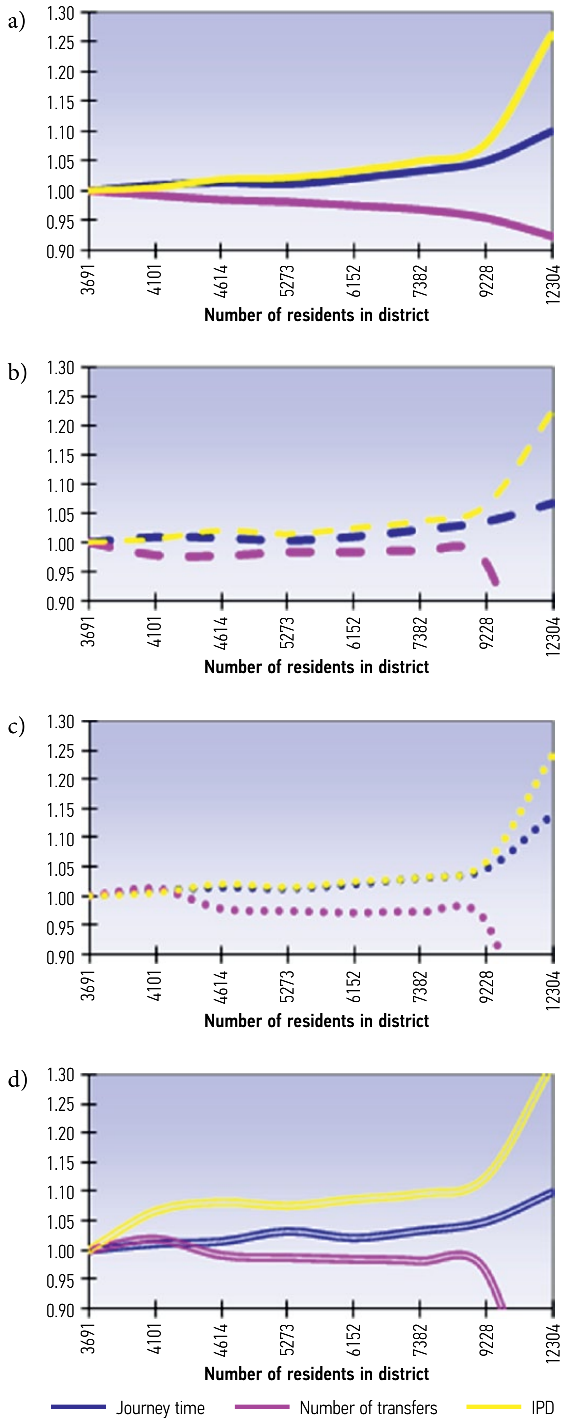

Fig. 3. The ratio of relative indices and the size of O–D matrix zones when modelling with the distribution laws: a – Kirchhoff; b – Lochse; c – Logit; d – BoxCox

Therefore, as there is no possibility of dividing the territory to be modelled into smaller transport districts, the most reliable indices are obtained when modelling with the Lochse distribution law. However, it should be emphasised that even in case of this distribution law, the limit of reliability still exists at 9000 residents per transport district.

5. Conclusions

1. Mobility of public transport passengers in the urban public transport system is more complicated than that in the model since people can freely choose between modes of public transport. A passenger can choose a route and a type of vehicle, as well as transfer possibilities by public transport modes.
2. The coefficient of journey time varies insignificantly up to the threshold of 6000 residents per transport district, and then it starts increasing more rapidly. A more significant increase in the number of transfers is observed at a threshold of 7500 residents per transport district.
3. The turning-point of IPD is the most obvious. If indices of journey time and the number of transfers after their turning-point deviate from their base indices by about 2÷5% per step, the coefficients of IPD, having reached 8700 residents per district, can no longer be predicted or evaluated.
4. Based on the results obtained, it is suggested for the O–D matrix to avoid transport districts with more than 8700 residents as much as possible. Having exceeded this number, errors of results become hardly predictable.
5. Modelling, based on BoxCox distribution law, is unpredictable due to unstable reliability of relative indices, therefore, this distribution law is not recommended for modelling of public transport passenger flows.
6. Reliability of Lochse distribution law, when increasing modelled transport districts, remains the highest and the most stable for a long time. Therefore, as there is no possibility to divide the modelled territory into smaller transport districts, the most reliable indices are obtained when modelling with the Lochse distribution law. However, it should be emphasized that even in case of this distribution law, the limit of reliability still exists at 9000 residents per transport district.

References

Cascetta, E. 2001. *Transportation Systems Engineering: Theory and Methods*. 1st edition. Springer. 736 p.

Codina, E.; Barceló, J. 2004. Adjustment of O–D trip matrices from observed volumes: An algorithmic approach based on conjugate directions, *European Journal of Operational Research* 155(3): 535–557. http://dx.doi.org/10.1016/j.ejor.2003.08.004

De Grange, L.; Fernández, E.; De Cea, J. 2010. A consolidated model of trip distribution, *Transportation Research Part E: Logistics and Transportation Review* 46(1): 61–75. http://dx.doi.org/10.1016/j.tre.2009.06.001

Duff-Riddell, W. R.; Bester, C. J. 2005. Network modeling approach to transit network design, *Journal of Urban Planning and Development* 131(2): 87–97. http://dx.doi.org/10.1061/(ASCE)0733-9488(2005)131:2(87)

Eliasson, J.; Mattsson, L.-G. 2000. A model for integrated analysis of household location and travel choices, *Transportation Research Part A: Policy and Practice* 34(5): 375–394. http://dx.doi.org/10.1016/S0965-8564(99)00038-5

Hazelton, M. L. 2003. Some comments on origin–destination matrix estimation, *Transportation Research Part A: Policy and Practice* 37(10): 811–822. http://dx.doi.org/10.1016/S0965-8564(03)00044-2

Horn, M. E. T. 2003. An extended model and procedural framework for planning multi-modal passenger journeys, *Transportation Research Part B: Methodological* 37(7): 641–660. http://dx.doi.org/10.1016/S0191-2615(02)00043-7

Horn, M. E. T. 2004. Procedures for planning multi-leg journeys with fixed-route and demand-responsive passenger transport services, *Transportation Research Part C: Emerging Technologies* 12(1): 33–55. http://dx.doi.org/10.1016/j.trc.2002.08.001

Lo, H. K.; Yip, C.-W.; Wan, Q. K. 2004. Modeling competitive multi-modal transit services: a nested logit approach, *Transportation Research Part C: Emerging Technologies* 12(3–4): 251–272. http://dx.doi.org/10.1016/j.trc.2004.07.011

Li, X.; Qin, Z.; Yang, L.; Li, K. 2011. Entropy maximization model for the trip distribution problem with fuzzy and random parameters, *Journal of Computational and Applied Mathematics* 235(8): 1906–1913. http://dx.doi.org/10.1016/j.cam.2010.09.004

Marzano, V.; Papola, A.; Simonelli, F. 2009. Limits and perspectives of effective O–D matrix correction using traffic counts, *Transportation Research Part C: Emerging Technologies* 17(2): 120–132. http://dx.doi.org/10.1016/j.trc.2008.09.001

Moore, S. 2011. Understanding and managing anti-social behaviour on public transport through value change: The considerate travel campaign, *Transport Policy* 18(1): 53–59. http://dx.doi.org/10.1016/j.tranpol.2010.05.008

Murray, A. T. 2001. Strategic analysis of public transport coverage, *Socio-Economic Planning Sciences* 35(3): 175–188. http://dx.doi.org/10.1016/S0038-0121(01)00004-0

Nie, Y.; Zhang, H. M. 2008. A variational inequality formulation for inferring dynamic origin–destination travel demands, *Transportation Research Part B: Methodological* 42(7–8): 635–662. http://dx.doi.org/10.1016/j.trb.2008.01.001

Ušpalytė-Vitkūnienė, R. 2006. Modelling of Vilnius public transport route network, *Technological and Economic Development of Economy* 12(4): 334–340.

Vilniaus miesto susisiekimo infrastruktūros (tramvajaus) plėtojimo specialusis planas. Keleivių ryšių anketinė apklausa. Vilnius: SĮ „Susisiekimo paslaugos“, 2002. 104 p. (in Lithuanian).

Wardman, M. 2004. Public transport values of time, *Transport Policy* 11(4): 363–377. http://dx.doi.org/10.1016/j.tranpol.2004.05.001