Equilibrium Method for Origination Destination Matrix Estimation Exploited to Urban Traffic Simulation Calibration

Marek Bazan(✉), Tomasz Janiczek, and Łukasz Madej

Faculty of Electronics, Wrocław University of Science and Technology, ul. Janiszewskiego 11/17, 50-372 Wrocław, Poland
{marek.bazan,tomasz.janiczek}@pwr.edu.pl, lukasz-madej@live.com

Abstract. One of methods used to find better light signalization plans for urban regions is a traffic flow numerical simulation. In this paper we present solution of two main issues the user of traffic flow simulation tool encounters while simulating traffic continuously registered in ITS by counting detectors, namely: (a) origination-destination routes setting, (b) calibration of the driver model. We consider two methods for origination-destination matrix estimation. The first is classical Spiess gradient method and the second one is the equilibrium method of Casseta which is suitable to congested scenarios. Resulting origination destination matrices from both methods are input to the traffic simulator to generate traffic between nodes in the city. The traffic generated in the simulator in this manner is used to calibrate the average driver model, to achieve as good recovery of vehicle counts on detectors as possible. As a simulation tool we use an open source ArsNumerica Execution Environment running on top of SUMO simulator. The driver model that is used is stochastic Krauss model. Calibration is done for one of the most congested areas in the city of Wrocław in the morning peak.

Keywords: Origination-destination matrix estimation · Congested traffic simulation · Traffic simulation calibration

1 Introduction

A crucial component of the road traffic simulation process that is based on real data is calculation of the origination destination matrix for the city/region. The input for the methods to estimate origination destination matrix are vehicle counts on locations of the detectors in a road network within the prescribed time.

The output of the estimation process i.e. an estimated origination destination matrix together with the assignment matrix contains the information how to distribute routes of the vehicles between any two nodes of the network to obtain counts of vehicles on the counting detectors that recover real data within the prescribed time.

In this work we consider two methods to calculate origination-destination matrix. The first one is the classical Spiess method [1] and the second one is Cascetta equilibrium method [2]. A review of classical methods of the origination destination matrix estimation can be found in [9] and more modern ones in [8].

J. Mikulski (Ed.): TST 2017, CCIS 715, pp. 16–27, 2017.
DOI: 10.1007/978-3-319-66251-0_2

The third issue for best possible recovery of the counts on detectors for the simulation is calibration of the driver model. In this work we consider the stochastic Krauss model with such parameters as acceleration, deceleration, delay, maximum speed and an average driver reaction time.

In this paper we consider three different methodological areas of the simulation [8] that recovers traffic counts:

– traffic assignment – i.e. supply model,
– origination destination matrix calculation – i.e. demand model,
– simulation with a use of both assignment matrix and the origination destination matrix – i.e. an average diver model. The employment of the numerical calculations in the above three aspects of the road traffic simulation enables us to propose a framework to compare results of the origination destination matrix estimation methods which is nowadays being sought by many researchers (c.f. [9]). A novelty of our approach is usage of the road traffic simulation and the analysis of the recovery of the ground truth traffic. This approach is simpler than the general framework proposed in [9] – giving nevertheless possibility of quantitative comparison of the methods.

We consider two methods for the origination destination matrix estimation – the classical Spiess [1] and more adequate to congested networks equilibrium method by Cascetta et al. [2]. In the notation of [8] both of these methods require link counts and initial OD matrix i.e. they are both updating methods. For the full static load of traffic from the morning peek hour (from 7:00 till 8:00) the Cascetta method as expected works better for saturated networks – it however gives a recovery error about 23% - which is in, as expected, accordance with the literature [7]. The difference between the quality of the obtained matrices by the two methods is shown by checking the hypothesis on the difference between a mean error of the recovery of the ground truth counts.

The remainder of the paper is organized as follows: in the next section we present a notation together with a demonstration of dimensionalities of the vectors and matrices involved, the third section recall the Spiess and Cascetta methods respectively. The fourth describes a procedure of creation of the simulation process to calibrate the driver model which in turn is described in the sixth section. In the fifth section describes a measure to calculate link counts recovery quality which is crucial for comparing results of both methods. In the seventh section we describe a genetic algorithm which is used for a drivel model calibration. The eight section describes numerical results. And finally we conclude the paper.

2 Origination Destination Matrix Problem Formulation

An O-D matrix is a set of values describing traffic from a source point i to a target point j. In the equation below, it becomes vector $\boldsymbol{g}$, with $c = n \cdot (n - 1)$ elements. The main equation binding the vector od measurements with the origin-destination matrix presented below will become

$$\boldsymbol{v} = \boldsymbol{Ug} \tag{1}$$

where:

m – a number of nodes (intersections) in the given topology, actually a number of locations counting detectors,

c – a number of all O–D pairs equal to $m \cdot (m - 1)$.

v – a vector containing values of traffic acquired based on measurements of vehicles traveling between intersections: $\boldsymbol{v} = (v_1, v_2, \ldots, v_m)' \in \boldsymbol{R}^m$,

g – origin destination matrix in the vector form. Each matrix element is equal to the number of vehicles moving from intersection (area) i to intersection (area) j. Index alongside the element means a number of a pair ($i \rightarrow j$): $\boldsymbol{g} = (g_1, g_2, \ldots, g_c)' \in \boldsymbol{R}^c$.

$U \in R^{cxm}$ – assignment matrix, where element $\boldsymbol{u_{ia}}$ describes the fraction of total traffic for O–D pair of i number on a edge where $i = 1, \ldots, c$ and $a = 1, \ldots, m$.

Despite the low complicity of Eq. (1), calculation of $\boldsymbol{g}$ is usually difficult even in spite of the exact knowledge of the measurement vector $\mathbf{v}$ and $\mathbf{U}$ matrix. The main problem while solving Eq. (1) is dimensionality of its components. That is, the $\mathbf{v}$ vector contains m elements, whereas the $\boldsymbol{g}$ vector contains c elements. As the number of all possible combinations of intersection-intersection pairs is usually much bigger than the number of all intersections in the net ($c \gg m$), it is clear, that Eq. (1) is strongly underdetermined. This fact is a reason why there is no unique solution of the equation. The number of possible solutions is infinite.

3 Two Methods of the Origination Destination Matrix

In this section we recall two methods of the origination destination matrix estimation: the Spiess gradient method [1] and Cascetta equilibrium method [2]. For the Spiess method we used assignment method based on Poisson distribution using always a prescribed number of the shortest paths between nodes. As in [6] it was up to 5 shortest paths for each OD pair. For the Cascetta method an assignment process is done by the method itself.

3.1 Spiess Gradient Method

One of the methods used in this paper to determine the matrix OD is gradient method described by Heinz Spiess [1]. This is an iterative method for determining the minimum or maximum of the test function defined by the formula:

$$f : R^n \rightarrow R, \quad x \in R^n$$

The vector x is assigned to several individual values. To start operation of the algorithm is necessary to define the starting point x^0, which allows you to calculate the value of a point in the next iteration, as shown in the following formula:

$$x^{i+1} = x^i - \alpha \nabla f\left(x^i\right),$$

where:

l – number of iteration,

x^l – argument value for l-th iteration,

α – parameter specifies the length of the iteration step, usually values are selected from the interval (0,1).

∇ – gradient operator.

This figure is a general formula describing this method. For the study related to the matrix OD, you must modify the model functions as follows:

$$Z(g) = \frac{1}{2}\sum\nolimits_{a\varepsilon A}\left(v_a - v_a'\right)^2$$

where:

$\boldsymbol{v_a}$ – the measured traffic on the edge of a,

$\boldsymbol{v_a'}$ – the estimated value of the intensity at the edge of a,

a – the edge of the graph,

A – set of the edges,

$\boldsymbol{g}$ – OD matrix in a vector form.

Then, in the estimation of matrix flows we use the steepest descent algorithm:

$$g_i^{l+1} = \begin{cases} \hat{g}_i, & l = 0 \\ g_i^l\left[1 - \lambda^l\left[\frac{\partial Z(g)}{\partial g_i}\right]\right], & l = 1, 2, \ldots \end{cases}$$

where:

$\hat{\boldsymbol{g}}_{\boldsymbol{0}}$ – chosen start point,

$\boldsymbol{g_i}$ – i-th component of the vector,

l – number of the iteration,

λ – length of iteration.

You should also designate pattern derivative of the cost function $\boldsymbol{Z(g)}$ in dependence on the length of iteration step:

$$\frac{dZ(\lambda)}{d\lambda} = \sum_{a\varepsilon\hat{A}}\frac{dv_a}{d\lambda}\frac{\partial Z}{\partial v_a} = \sum_{a\varepsilon\hat{A}} v_a'\left(v_a - \hat{v}_a + \lambda v_a'\right)$$

Equating to 0 the previous equation we are looking for the value of λ it meets, which will be determination of the optimal length iteration:

$$\lambda^* = \frac{\sum_{a\varepsilon A} v_a'(\hat{v}_a - v_a)}{\sum_{a\varepsilon A} v_a'^2}.$$

3.2 Starting Point for Spiess Method

In order to circumvent the problem of lack of the starting origination destination matrix $\hat{\boldsymbol{g}}_{\boldsymbol{0}}$ the Tikhonov regularization method is employed. Despite high under determination of the system (1) it is possible to calculate such so that is satisfy it approximately with a certain error. For this purpose the More-Penrose inverse solution to (1) may be derived by multiplying (1) both sides by $\boldsymbol{U}'$

$$\boldsymbol{U}'\boldsymbol{v} = \boldsymbol{U}'\boldsymbol{U}\boldsymbol{g} \tag{2}$$

and then by adding nonzero component to a diagonal of $\boldsymbol{U}'\boldsymbol{U}$ on the right side of the above equation one gets

$$\boldsymbol{U}'\boldsymbol{v} = (\boldsymbol{U}'\boldsymbol{U} + \boldsymbol{\lambda I})\boldsymbol{g}, \tag{3}$$

which allows already to approximate $\boldsymbol{g}$ since the matrix on the right hand side is nonsingular even for very small positive λ's. Therefore one gets

$$\boldsymbol{g_0} = \boldsymbol{U}'(\boldsymbol{U}'\boldsymbol{U} + \boldsymbol{\lambda I})^{-1}\boldsymbol{v} \tag{4}$$

In this form however $\boldsymbol{g_0}$ cannot be used in the process since it may contain negative components. This is not interpretable in the context of the origination destination matrix estimation. To resolve this problem $\hat{\boldsymbol{g}}_{\boldsymbol{0}}$ is calculate from $\boldsymbol{g_0}$ by substituting 0 value where negative component is encountered. This approach gives a starting point which provides a stability of the Spiess iteration process.

3.3 Equilibrium Method

Another method is the one described by Cascetta [2]. We consider situation in which an O–D connections and their location is known and constant. Load of connections and the cost of traffic are dependent of each connection flow and function of travel costs. On the other hand, the flow rate on a particular combination depends on the probability of selection of the specific movement path connecting OD pair in the model. Generally speaking, this model has the task to regulate the interaction between the "supply" of vehicles, and the "demand" of roads, or the number of vehicles that will be able to drive the route most fluent way. The model is presented by following equations (Fig. 1):

$$\begin{aligned}
g^*_{od} &= \Delta^T_{od} c\left(\sum_{od} \Delta_{od} h^*_{od}\right) + g^{NA}_{od} && \forall od\\
V^*_{od} &= -g^*_{od} && \forall od\\
h^*_{od} &= d_{od} p_{od}\left(V^*_{od}\right) && \forall od
\end{aligned}$$

Δ_{od} in the above equations means a row from this matrix corresponding to the pair with an index *od*.

$$
\Delta = \begin{array}{c|cccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
1,2 & 1 & 1 & 0 & 0 & 0 & 0 \\
1,3 & 0 & 0 & 1 & 0 & 0 & 0 \\
2,3 & 1 & 0 & 0 & 1 & 0 & 0 \\
2,4 & 0 & 1 & 0 & 0 & 1 & 0 \\
3,4 & 1 & 0 & 1 & 1 & 0 & 1
\end{array}
$$

Fig. 1. Example of connections matrix, columns – number of junction, rows – OD pair [2]

After substituting:

$$
\begin{aligned}
g^{*}_{od} &= \Delta^{T}_{od} c \left(\sum_{od} \Delta_{od} h^{*}_{od} \right) + g^{NA*}_{od} \quad && \forall od \\
h^{*}_{od} &= d_{od} p_{od} \left(V^{*}_{od} \right) && \forall od
\end{aligned}
$$

We have the following variables:

g^{*}_{od} – vector movement cost for a given path OD, interpreted as travel time, index NA means non additive costs,

h^{*}_{od} – flow on the path for an OD pair, the number of vehicles on the track,

d_{od} – the expected flow on the path for a given pair of OD.

p_{od} – the probability of choosing a path for a given pair of OD

c – the cost of connections,

Δ^{T}_{od} – connections matrix.

The values of variables marked with an asterisk are determined at the time of calculation, while the others are known.

3.4 Starting Point for the Equilibrium Method

As a starting point of the equilibrium method we used the solution of the Spiess iteration process. As it may be seen from the analysis of the results of the worked out examples the equilibrium method can substantially improve this solution.

4 Procedure of Simulation Creation

At the beginning it is necessary to prepare the map of the area. As nodes was the intersection where the cameras are located in Wroclaw Intelligent Transport System, and thus it is possible to read data on the movement of vehicles in these places. The number of vehicles was read during the morning rush, on the work day beyond the summer break. Intersections main, were considered transit centers, that means those where new vehicles cannot appear – they can only enter and leave. Powering the model performed with the nodes of up to major intersections, is declared to them the flow of vehicles. Amount of vehicles on specified connections can be calculated by one of described models. After that it is possible to perform simulation. In our work we used traffic simulator SUMO [3] and the ArsNumerica wrapper [4, 5] for it enabling to perform calibration. We prepared simulation map based on one used to calculations. There was used constant traffic lights settings.

5 Measuring Link Counts Recovery Quality

Basic way to establish quality of model was calculating difference between number of vehicles which should arrive to specific output and number of vehicles which actually arrived to this point. However, this method isn't very accurate. That's why we used also Krauss model to more specific calculation of quality function for both methods:

$$F = \sum\nolimits_j \left(\frac{DDD_j - DSS_j}{DDD_j} \right) \Big/ N \tag{5}$$

where:

DDD_j – read data in simulation for detector,
DSS_j – known data for detector,
K – tested detector,
N – number of detectors.

6 Krauss Calibration Model

In order to compare the quality of action set out methods of determining the matrix flows were carried out a survey to calibrate Krauss model [3]. This will determine the optimal parameters of the model, which are interpreted as the most appropriate way driver behavior in traffic. However, the main benefit of this study is to answer the question which come to mind: which method provides more capacity in traffic?

To carry out this study has been prepared runtime environment, consisting of several sub-programs:

– M++atlab script - is responsible for calculating successive input values for the calibrated model, using a genetic algorithm global optimization
– Traffic Program - a program implemented in C++, whose task is to type in the startup files simulator SUMO model parameters, as well as the calculation of the value of the objective function for each iteration,
– SUMOConnector program - a program implemented in C++, corresponding to the communication with the simulator. Its function is to read in a dynamic way during the simulation of the traffic measurement detectors and write to a file,
– SUMO - simulator processing received data on the traffic, the model parameters Krauss and road network. It allows you to obtain the necessary output.

The present model is implemented in the simulator used, which allows for testing using the developed runtime without additional implementation. The survey will proceed by iteratively calculating successive values of model parameters Krauss on the basis of which will be carried out simulation.

7 Genetic Algorithm to Solve Calibration Task

Like it was mentioned in previous section during calculations with usage of the Krauss model the genetic algorithm was used. It was implemented in the same way as in [4, 5]. The Krauss model simulation was performed every iteration for new values of this model parameters i.e. acceleration, delay, maximum speed, reaction time.

8 Tested Area

Performed tests was executed on part of city of Wroclaw. It is known as *area3* and covers Grunwaldzki Square and surrounding area. Schematic picture of this part of city is presented below (Figs. 2, 3).

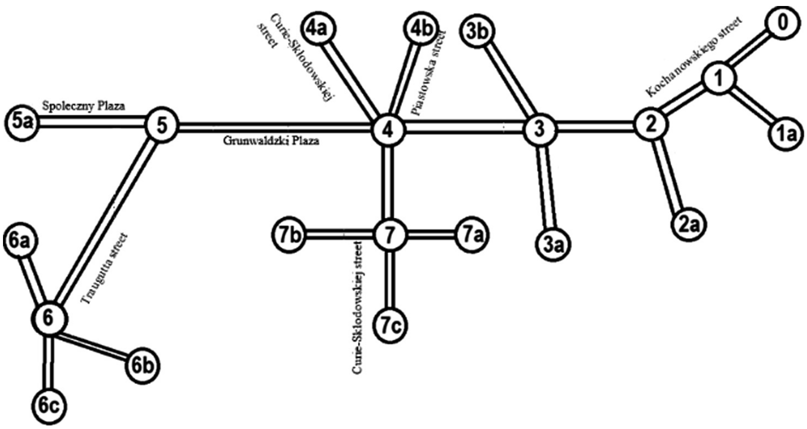

Fig. 2. Schematic view of the simulated network – it corresponds to the one of the most congested region in the center of Wrocław city (Poland) – surroundings of the Reagan roundabout [own study]

8.1 Tests of OD Matrix Calculating Methods

At first we performed calculations of OD matrices by both methods. Next, we prepared simulations with obtained results like it was described in previous sections. In this test we measured number of vehicles, which achieved their targets in one simulation. Each simulation was depicting 1 h of traffic. Results are presented in Fig. 4.

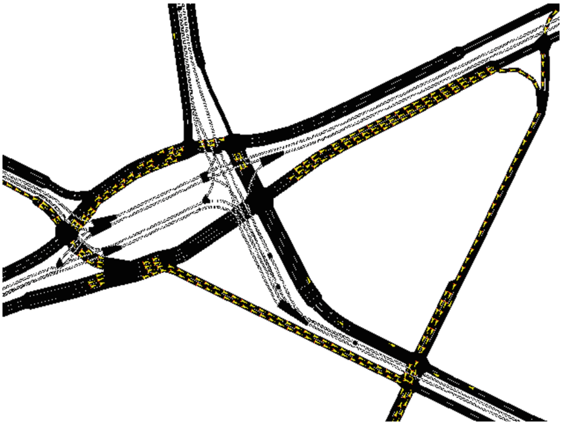

Fig. 3. Example view of the simulation - a zoom of the Reagan roundabout and its surroundings from SUMO [3]

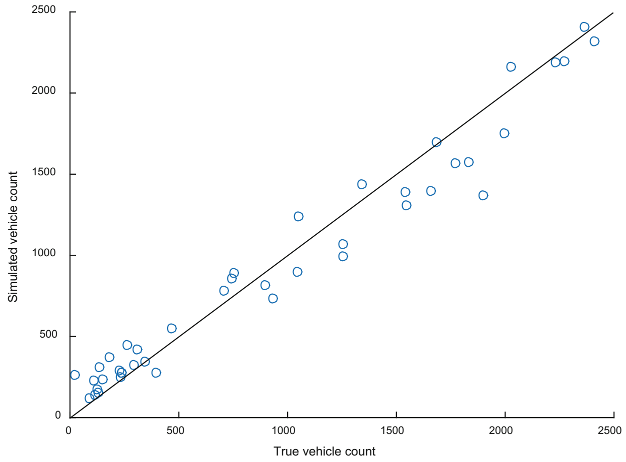

Fig. 4. Relation between true and simulated vehicle count on detectors for equilibrium model. [own study]

Second test was performed with usage of Krauss calibration. In this test we performed 30000 continuous simulations for each method. Krauss model parameters was set by genetic algorithm for every iteration. Results are presented by plot in Fig. 6.

8.2 Test of Hypothesis

In last test we checked hypothesis that calibration error is significantly smaller for equilibrium method than Spiess method. We take 30 consecutive results of objective function for each method to check if result achieved for equilibrium calibration is significantly smaller. For trust level 0.95 and 28 degrees of freedom we get critical value of 1.701 and critical area $\langle 1,701, +\infty \rangle$. Using Eqs. (3) and (4) as result we obtained value 122.50, which is higher than critical value and belong to critical area. On this basis we can tell that calibration error is significantly smaller for equilibrium method:

$$S_{X1-X2} = \sqrt{\frac{(N_1 - 1) * s_1^2 + (N_2 - 1) * s_2^2}{N_1 + N_2 - 2}} * \left(\frac{1}{N_1} + \frac{1}{N_2}\right) \tag{6}$$

$$T = \frac{X_1 - X_2}{S_{X1-X2}} \tag{7}$$

where:

$\boldsymbol{X}_1$ – mean value for Spiess method,
$\boldsymbol{X}_2$ – mean value for equilibrium method,
$\boldsymbol{S}_1$ – variance for Spiess method,
$\boldsymbol{S}_2$ – variance for equilibrium method,
$\boldsymbol{N}_1$ – size of first group of values,
$\boldsymbol{N}_2$ – size of second group of values,
$\boldsymbol{T}_1, \boldsymbol{S}_{X_1-X_2}$ – estimators.

8.3 Plots

See Figs. 5, and 7.

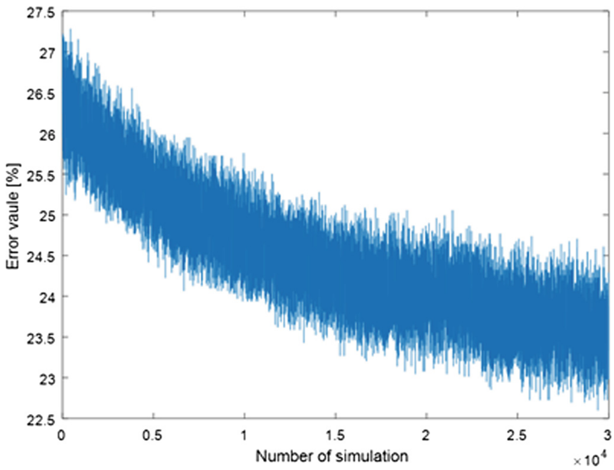

Fig. 5. Error value in Krauss calibration for equilibrium model. [own study]

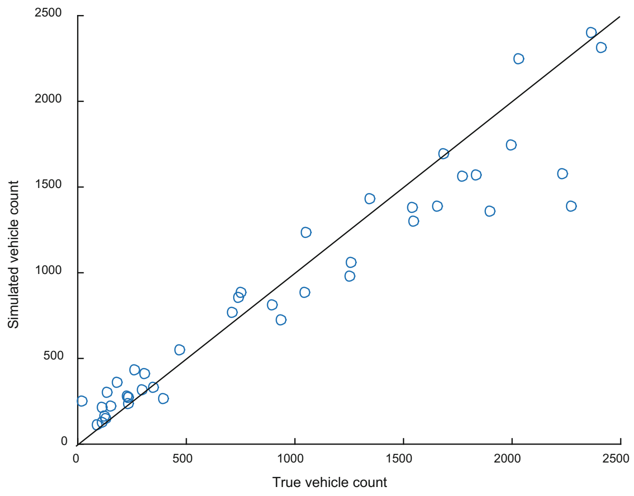

Fig. 6. Relation between true and simulated vehicle count on detectors for Spiess *model* [own study]

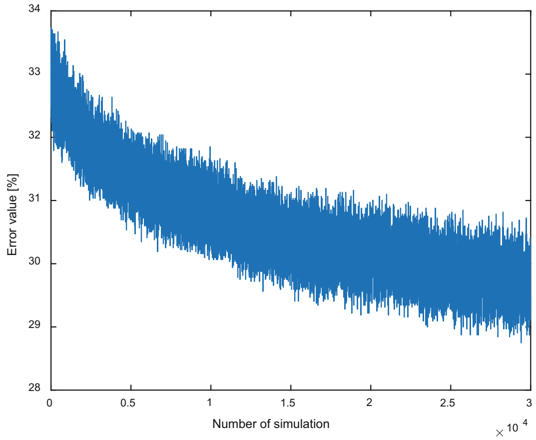

Fig. 7. Error value in Krauss calibration for equilibrium model [own study]

9 Conclusion

Obtained results are intuitively in accordance with the literature, i.e. better results were obtained for Cascetta method for congested networks. Cascetta method approaches 20% global error in traffic count recovery.

The outlook for future research to beat 20% barrier is to implement two things

– fully dynamic operation of intersections signalization to be in accordance with ITS microprograms that work on most of the intersections,

– dynamic assignment model more tightly coupled with origination-destination matrix estimation method, this is may be achieved with bi-level programming methods (c.f. [10]).

In this paper we introduced a methodology to compare methods for estimation of the origination destination matrix. The approach consists of calculation of the origination destination matrix by the methods being compared and then to start the calibration process of the driver model using a genetic algorithm. The comparison is performed by calculation of the relative mean error of the link counts recover on the calibrated models. This approach seems to be simpler than proposed in [9] and gives as a spin-off the behavior of the origination destination matrices calculated by the methods in the calibrated simulation.

Acknowledgement. The work presented in this paper was partially financed from Grant 0401/0230/16.

References

1. Spiess, H.: A Gradient Approach for the O-D Matrix Adjustment Problem, EMME/2 Support Center, Switzerland (1990)
2. Cascetta, E.: Transportation Systems Analysis. Model and Applications. Springer Optimisations and ITS Applications, 2nd edn. Springer, Heidelberg (2009)
3. Krauss, S.: Microscopic modeling of traffic flow: investigation of collision free vehicle dynamics. DLR, Ph.D. thesis (1998)
4. Bazan, M., et al.: Intelligent transportation system auditing using micro-simulation. Arch. Transp. Syst. Telemat. **8**(4), 3–8 (2015)
5. Bazan, M., et al.: Green wave optimisation. Arch. Transp. Syst. Telemat. **9**(3), 3–8 (2016)
6. Bazan, M., et al.: Two methods of calculation of the origination destination matrix of an urban area, Raport W04/P-007/15, Wrocław University of Technology (2015)
7. Treiber, M., Kersting, A.: Trafic Flow Dynamics, Data, Models and Simulation. Springer, Heidelberg (2013)
8. Toledo, T., et al.: Network model calibration studiem. In: Traffic Simulation and Data: Validation methods and Applications, pp. 141–162. CRC Press, Boca Raton (2015)
9. Antoniou, C., et al.: Towards a generic benchmarking platform for origin–destination flows estimation/updating algorithms: design, demonstration and validation. Transp. Res. Part C **66**, 79–98 (2016)
10. Tavana, H.: Internally consistent estimation of dynamic network origin–destination flows from intelligent transportation systems data using bi-level optimization. Ph.D. thesis, University of Texas, Austin (2001)

Smart Solutions in Today's Transport
17th International Conference on Transport Systems
Telematics, TST 2017, Katowice – Ustroń, Poland, April
5-8, 2017, Selected Papers
Mikulski, J. (Ed.)
2017, XII, 508 p. 263 illus., Softcover
ISBN: 978-3-319-66250-3