

Demand modelling for public transport service based on variation of needs in travel¹

ROSSOLOV ALEXANDER

Candidate of technical science, docent, Kharkiv National Automobile and Highway University, str. Petrovskogo 25, 61-002 Kharkiv, Ukraine, tel.: +38 066 804 0001, email: rossolovalex@gmail.com

Streszczenie: W artykule zostało opracowane kompleksowe podejście do modelowania macierzy źródło-cel w oparciu o stochastyczne hipotezy dotyczące generowania popytu na usługi transportu publicznego. Korzystając z analizy kombinatorycznej został opracowany model zmienności popytu i oszacowana możliwa ilość macryc. Opracowano metody modelowania macierzy źródło-cel za pomocą stopniowej symulacji oraz metody Monte Carlo.

Słowa kluczowe: macierz źródło-miejsce docelowe, podróż pasażera, zastosowanie gęstości miejsc pracy, gęstość zaludnienia w mieście

Introduction

The key role of the public municipal passenger transport is to service the demand on travel of the citizens. The biggest network load occurs during morning and evening rush hours when the labor trips are being executed. In those periods of time the city public transport has to provide punctual arrival of the passengers to the destination points to start their labor activity. So it can be said that the main target of the public transport functioning is to deliver the passengers to the necessary city areas with the minimal time expenditure. In such condition one of the most important aspects is the presence of the precise data about demand on transport service in the city. If the designer of the public transport network has the incorrect data about transport demand this will lead to level the effectiveness of the design choices in network constructing.

Related works

There are lots of the different methods, models and approaches to passenger transport demand determination. As usually it is formalized as origin-destination matrix (O-D matrix) which reflects the number of trips between pair of transport zones of the city [1 – 14]. The evolution of the procedure of trips number determination was rather specific. It can be formed two main branches of the researches in this field. The first one is the statistical group which includes growth factor models and regression models. The growth factor models include single coefficient method, the medium coefficient method, Fratar's and Detroit method [3]. The most effective and adapted for calculation procedure is the Detroit method that has following mathematical interpretation:

$$T_{ij} = T_{ij} \cdot k_i \cdot k_j \cdot \frac{M_i + M_j}{2}, \quad (1)$$

where T_{ij}^f – number of trips in forecast period between i and j transport zones, pass.; T_{ij}^c – number of trips between i and j transport zones in current period, pass.; k_i, k_j – accordingly, coefficients of the passenger flow growth in i and j transport zones; M_i, M_j – accordingly, local factors that are determined by iteration calculations.

The calculation procedure is executed by iteration method because of the determination in equal of the general trips sum by rows and column of the O-D matrix:

$$\sum_{j=1}^n T_{ij}^f = O_i, \quad (2)$$

$$\sum_{i=1}^n T_{ij}^f = D_j, \quad (3)$$

where O_i – capacity of the i transport zone on origin, pass.; D_j – capacity of the j transport zone on destination, pass.

In spite of the positive side of the extrapolation models such as calculation algorithm simplicity of the main disadvantage is included in preparation of the basic data and useful in short term forecast. The point is that the volume T_{ij} has to be defined using one of the total survey of passenger trips on the network.

The regression models are constructed on the results of survey on public transport network under condition of carrying out the passive or active experiment. As usually obtained data are handled using the mathematical statistics theory and the researcher as a result may obtain the polynomial model of two types: additive (4) or multiplicative (5,6):

$$T_{ij} = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_i \cdot x_i = a_0 + \sum_{i=1}^k a_i \cdot x_i, \quad (4)$$

$$T_{ij} = b_0 \cdot b_1^{x_1} + b_2^{x_2} + \dots + b_i^{x_i} = b_0 \prod_{i=1}^k b_i^{x_i}, \quad (5)$$

$$T_{ij} = b_0 \cdot x_1^{b_1} + x_2^{b_2} + \dots + x_i^{b_i} = b_0 \prod_{i=1}^k x_i^{b_i}, \quad (6)$$

where a_0, b_0 – absolute term (Y-intercept); a_i, b_i – regression coefficients; x_i – numerical values of factors that influence on resulting function T_{ij} [1].

The choice between two types of regression function is made on numerical value of the coefficient of determination. But it has to be noticed that regression analysis is based on least square method and oscillatory amplitude of the resulting function influences on value of the coefficient of determination. In this case the higher value of the coefficient of determination does not guarantee the minimal

¹ ©Urban and Regional Transport, 2016.

value of the approximation coefficient that in fact reflects the quality of the model. Also it has to be told that the list of factors x_i is formed by the researcher and it can be passed by some important components. Besides that some factors have the qualitative character that complicates the regression model constructing. And finally it is hard to construct the regression model that will precisely describe the trips around the town because of plenty factors that influence on it.

The second group of the O-D matrix modelling includes different methods that have in their bases physical hypothesis about trips forming. Because of that the second group is usually called as “synthetical” which means the O-D matrix modelling without grand-scale surveys but with the help of small quantity of transport factors [1, 2, 6]. By way of hypothesis may be used such as: gravitation law, the law of entropy variation, limits of the resettlement and others. Most of them can be characterized as the deterministic models. This shows their mathematical formalization. For example, the gravitation group models have the following statement of problem:

$$T_{ij} = O_i \cdot D_j \cdot k \cdot f(c_{ij}), \quad (7)$$

where k – calibration factor; $f(c_{ij})$ – gravitation function.

The statement (7) has the relevant working approximation:

$$T_{ij} = O_i \frac{D_j \cdot c_{ij} \cdot k_j}{\sum_m^n D_m \cdot c_{im} \cdot k_m}, \quad (8)$$

where n – number of the transport zones.

As procedure of modelling is invariant the searches for accuracy of calculation have been focused on the type of the attraction function. The full range of them can be systematized in four groups: classical (9), exponential (10), combined (11) and based on law of random functions (12).

$$f(c_{ij}) = c_{ij}^{-n}, \quad (9)$$

$$f(c_{ij}) = \exp(-\beta \cdot c_{ij}), \quad (10)$$

$$f(c_{ij}) = c_{ij}^n \cdot \exp(-\beta \cdot c_{ij}), \quad (11)$$

$$f(c_{ij}) = \frac{4 \cdot t^2}{\alpha^3 \cdot \sqrt{\pi}} \cdot e^{-\frac{t^2}{\alpha^2}}, \quad (12)$$

where β – empirical coefficient; t – the trip time between i and j transport zones; α – Maxwell distribution parameter.

As usually in case of c_{ij} it is assumed the distance between i and j transport zones or trip time.

Another widely used model is based on hypothesis about trips forming analogy with the thermodynamic systems processes. It is supposed the system becomes stable when its entropy reaches the maximum value. The mathematical statement is the following:

$$S = \left(- \sum_{i=1}^n \sum_{j=1}^n T_{ij} \cdot \ln(T_{ij}) \right) \rightarrow \max, \quad (13)$$

where S – the system entropy.

According to [1] the (13) has similar to gravitation model practical application:

$$T_{ij} = A_i \cdot B_j \cdot O_i \cdot D_j \cdot \exp(-\beta \cdot c_{ij}) \quad (14)$$

where

$$A_i = \left(\sum_j B_j \cdot D_j \cdot \exp(-\beta \cdot c_{ij}) \right)^{-1}, \quad (15)$$

$$B_j = \left(\sum_i A_i \cdot O_i \cdot \exp(-\beta \cdot c_{ij}) \right)^{-1}. \quad (16)$$

It is clearly seen that entropy model may be considered as the special case of the gravitation approach (10) with its advantages and the list of disadvantages.

Another approach to O-D matrix modelling is based on hypothesis for labor resettlement of city inhabitants [14]:

$$C_{i(t_1, t_2)} = N_i \int_{t_1}^{t_2} \frac{1}{T} \cdot \ln \frac{T}{t} dt, \quad (17)$$

where $C_{i(t_1, t_2)}$ – the passenger flow among isochrones t_1 and t_2 ; N_i – object capacity of attraction; t – time expenditure on trip realization; T – extreme value of the resettlement.

The main difference Shelejhovskij’s model from Wilson’s and models of gravitation group is not calibrated calculation algorithm but the model parameters are the same. The main determining factors are the transport parameters and according to regularities that obtained on $f(c_{ij})$ changing it can be seen that probability of trip realization is being reduced when the trip distance is being increased. The hard determination of trips forming process does not correlate to real distribution of trips on the city network. That is why it is necessary to form the stochastic approach to O-D matrix modelling.

Mathematical statement of O-D matrix modelling

According to transport demand for public network service investigation the O-D matrix is considered as a function of such variables: transport zones capacities and a list of random factors which transport engineer may don’t know. Mathematically this can be formed as:

$$H = f(O_i, D_j, \mu), \quad (18)$$

where H – origin-destination matrix; μ – list of random factors.

The set of constrains is the following:

$$\left\{ \begin{array}{l} \sum_{j=1}^r T_{ij} = O_i, \\ \sum_{i=1}^r T_{ij} = D_j, \\ \sum_{i=1}^r O_i = \sum_{j=1}^r D_j = Q, \\ T_{ij} \in Z; T_{ij}, O_i, D_j \geq 0, \end{array} \right. \quad (19)$$

where Q – the volume of the passengers that have been serviced in specified period.

Under mentioned above condition the number of O-D matrix variants may be more than one. If make the transformation of O-D matrix by cycle changing random value of the trips, as a result it will be obtained the entirely new state of the O-D matrix. It has to be noticed that the volume of the general trips will be the same. The sample of the such transformation is showed on pic.1.

Transport area	1	2	...	j	...	r	O
1	$T_{11} - \min(T_{11}; T_{22}; T_{ij})$	T_{12}	...	$T_{1j} + \min(T_{11}; T_{22}; T_{ij})$...	T_{1r}	O_1
2	$T_{21} + \min(T_{11}; T_{22}; T_{ij})$	$T_{22} - \min(T_{11}; T_{22}; T_{ij})$...	T_{2j}	...	T_{2r}	O_2
...
i	T_{i1}	$T_{i2} + \min(T_{11}; T_{22}; T_{ij})$...	$T_{ij} - \min(T_{11}; T_{22}; T_{ij})$...	T_{ir}	O_i
...
r	T_{r1}	T_{r2}	...	T_{rj}	...	T_{rr}	O_r
D	D_1	D_2	...	D_j	...	D_r	--

Pic.1. The sample of O-D matrix transformation

To estimate the state of the O-D matrix it is necessary to use some criterion. Taking into account the matrix of the distances between transport zones that is fixed it was chosen as the function the volume of the passenger turnover. Making it optimization we can obtain two extreme variants of the O-D matrix (minimal and maximum volume of the passenger turnover). And it will be all possible variants of the O-D matrix in the obtained interval:

$$\left. \begin{array}{l} b_i \in [b_{\min}; b_{\max}] \\ b_{\min} \equiv H_{\min} \\ b_{\max} \equiv H_{\max} \end{array} \right\} \Rightarrow H_i \in [H_{\min}; H_{\max}], \quad (20)$$

where b – the volume of the passenger turnover.

So, it is necessary to determine the most probable states of the O-D matrix. This can be done by stochastic modeling of O-D matrix transformation. But on the first step it is needed to estimate the rate of the transport factors influence on character of the trips forming.

The detection of social and transport factors influence on trips forming

To define the level of transport and social factors influence on trips character it has been worked up the method of social survey that based on Likert scale. Every respondent have to estimate the weight of transport and social factors from the list according to the scale from 1 to 5. The maximum grade corresponds to high influence level and, accordingly, the minimal grade reflects the low influence level of the factor. To define the rate of the factor weight it has been introduced the following criterion:

$$r = \frac{\sum_{i=1}^n (\Delta_{ik} - \bar{\Delta}_k) \cdot \left(\sum_{j=1}^m B_{ji} - \overline{\sum_{j=1}^m B_j} \right)}{\sqrt{\sum_{i=1}^n (\Delta_{ik} - \bar{\Delta}_k)^2 \cdot \sum_{i=1}^n \left(\sum_{j=1}^m B_{ji} - \overline{\sum_{j=1}^m B_j} \right)^2}}, \quad (21)$$

where Δ_{ik} – the value of difference between point rating for k factor in i questionnaire; $\bar{\Delta}_k$ – the average estimation of difference between point rating for k factor; $\sum_{j=1}^m B_{ji}$ – the sum of the rates in i questionnaire for every factors group (transport and social); $\overline{\sum_{j=1}^m B_j}$ – the average estimation of the sum of the rates.

According to criterion (21) it is made the estimation of the correlation between total score in the questionnaire and the score difference by the being analyzed factor. The value Δ_{ik} characterizes the level of the k factor contribution to the general sum of rates in i questionnaire and allows to trace the correlation ratio. The conclusion about significant influence of the factor on trip character forming is made if the value of the r criterion goes to 1. The survey has been made on the 17 different enterprises of business, education, financial, industrial and medicine branches. The total number of the questionnaire is 387. The results are showed in the table 1.

The analyses of the obtained data allowed to make the conclusion that the transport factors make the bigger influence on labor trips than social. But these factors are not dominant. According to that it has been made the derivation that trips are formed under stochastic influence of the big list of factors.

Table 1

The results of the social survey		
Factor type	The factors	The value of the criterion r
Social	Salary size	0,293
	Social security	0,332
	Career prospects	0,057
	Work time schedule	0,4
	Social status of the work	0,405
	Additional income, bonuses	0,182
Transport	Trip time	0,698
	Different variants of trip realization	0,563
	Trip total costs	0,706
	Number of transfers	0,665
	Travel privilege	0,381

Theoretical and experimental researches of O-D matrix states

The modelling of the most probable states of the O-D matrix executes step by step changing of the it “start” condition which is evaluation by formulas:

$$p_i = O_i / Q, \tag{22}$$

$$p_j = D_j / Q, \tag{23}$$

$$T_{ij} = Q \cdot p_i \cdot p_j, \tag{24}$$

The calculation of trips value using formulas (22) – (24) levels the necessity of obtained data calibration. It can be mathematically showed making the following transformations:

$$\begin{aligned} \sum_{i=1}^r T_{ij} &= \sum_{i=1}^r Q \cdot p_i \cdot p_j = Q \cdot p_j \cdot \sum_{i=1}^r p_i = Q \cdot p_j \cdot \sum_{i=1}^r \frac{O_i}{Q} = \\ &= Q \cdot p_j \cdot \frac{1}{Q} \cdot \sum_{i=1}^r O_i = p_j \cdot \sum_{i=1}^r O_i = p_j \cdot Q = D_j. \end{aligned} \tag{25}$$

So, by mathematical expressions (22) – (24) it is obtained the “start” condition of the O-D matrix. It transformation is proposed to make by such way. On the first step it is defined the value of the trips that will be transformed on a cycle:

$$\Delta T_{z_{max}} = \min(T_1, T_3, \dots, T_{d-1}). \tag{26}$$

On the second step it is being made the estimation of the turnover:

$$b_t = b_0 + \sum_{z=1}^p \Delta T_z \cdot \Delta L_z, \tag{27}$$

under condition

$$\Delta T_z \in [0; \Delta T_{z_{max}}], \tag{28}$$

where b_0 – the turnover that reflects the “start” state of the O-D matrix; Δb_2 – the number of trips that transforming on a cycle (the stochastic value); p – the number of the cycle transformations.

Because of O-D matrix transformation it is changed the volume of the turnover and in turn is made the change of the average trip distance. It can be estimated by following formula:

$$l_t = l_0 + \frac{\sum_{z=1}^p \Delta L_z \cdot \Delta h_z}{Q}, \tag{29}$$

under condition

$$\Delta L_z = \sum_{f=1}^d (-1)^f \cdot l_f, \tag{30}$$

where l_0 – the average trip distance in the “start” state of the O-D matrix; Δb_2 – the change of the trip distance after one cycle transformation is realized; d – the number of the

cells that are involved in the cycle transformation; l_f – the trip distance for the f trip that is involved in the cycle transformation.

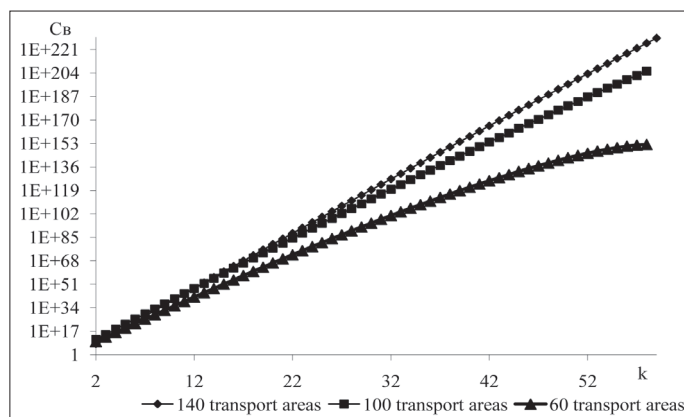
The target of the cycle matrix transformation is to investigate all possible states of the transport demand and obtain the numerical values of the passenger turnover and the avarege trip distance. In such way it can be defined the statistical low of these two stochastic values distribution. But on the first step of modelling it is necessary to make the estimation of possible number of the O-D matrix cycle transformations. This will allow to determine the approximate modelling time and the general labor intensity of the investigation. Using the methodology of the combinatorial theory by induction method was worked up the mathematical model (31) that reflects the number of possible states of the O-D matrix. It has to be noticed that in case of one O-D matrix state is considered the matrix that obtained after one cycle transformation.

$$C_B = \frac{\prod_{i=0}^{k-1} (m-i) \cdot (n-i)}{k} \cdot \left(1 - \left(1 - (1-p_0)^k\right)\right), \tag{31}$$

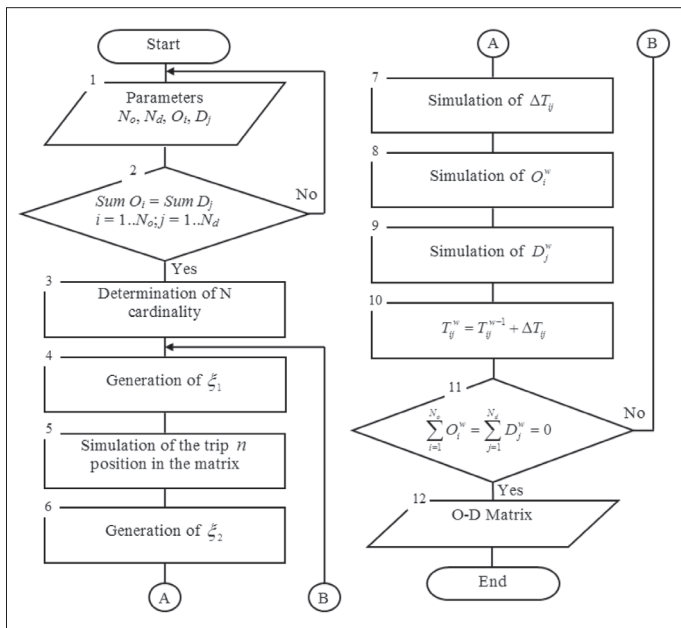
where C_B – number of possible states of the O-D matrix; m, n – number of columns and rows in the matrix, accordingly; k – quantity of the cells with odd numbers that are involved in the cycle transformation; p_0 the probability of the zero value trip appearance in the matrix.

It has been taken into account the probability because if the trip with the zero value is engaged in the transformation cycle the state of the matrix will stay the same. The value C_B has been modeled under such conditions: $m, n \in [60; 120]$, $k \in [2; 60]$. The results of the simulation procedure are showed on pic. 2.

As we can see with increasing the quantity of the transport zones the number of the possible states of the O-D matrix is significant growing too. The obtained volumes of the cycle transformation are huge that is why the step-by-step matrix transformation in practice is particularly impossible. That is why it was proposed to simulate the possible states of the O-D matrix with the help of the Monte Carlo method. In this case every matrix state will not be depended from the previous one. So the simulation algorithm is the following (pic. 3):



Pic. 2. The results of the O-D matrix states simulation



Pic. 3. The algorithm of the O-D matrix state simulation

The Monte Carlo method is applied in the stages 4 and 6 where it is being generated the stochastic values of the uniform law in the interval (0;1). This procedure provides the equally likely possibility of the trips inclusion to the sampling frame. The simulation of the n trip position in the matrix is being executed by the following formula:

$$n = \xi_1 \cdot N_o \cdot N_d, n \in [1; |N|]. \quad (32)$$

On the seventh step of the algorithm it is being simulated the value of the trips:

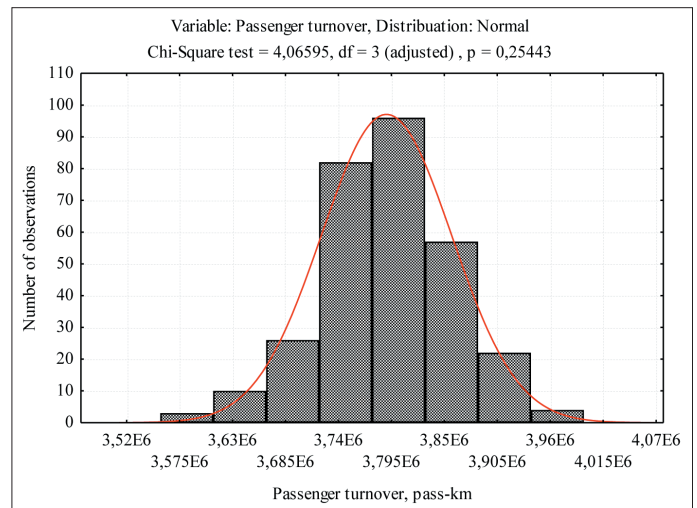
$$\Delta T_{ij} = ROUND(\xi_2 \cdot \min(O_i^{w-1}; D_j^{w-1})). \quad (33)$$

After every iteration it is obtained the value ΔT_{ij} for the position n in the matrix. And if the positions number n in some iterations coincided the final value of the trips will be the following:

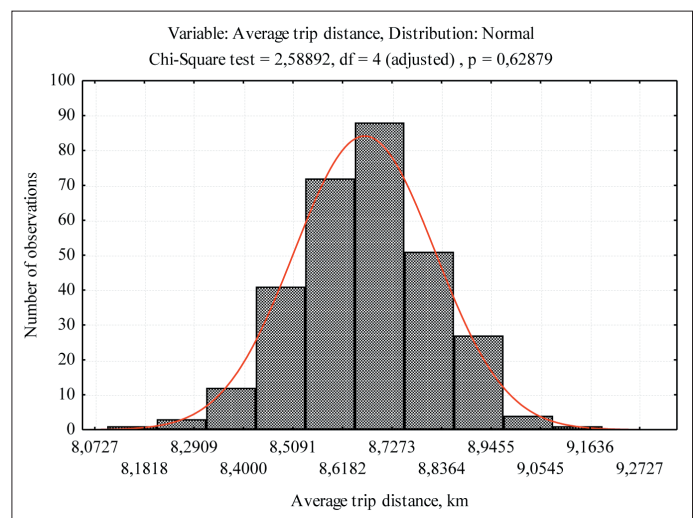
$$T_{ij}^w = T_{ij}^{w-1} + \Delta T_{ij}. \quad (34)$$

The procedure according to showed above algorithm has been software-implemented in the *VBA MS Excel* and it was simulated the 384 states of the O-D matrix which have been taken as the base for determination of passenger turnover and average values of the trip distances. As the experimental object it has been taken the one of the Ukrainian biggest cities – Kharkov (population – 1.4 million people).

The obtained values of that two parameters in fact are the stochastic so it can be defined their statistical lows. Using the application program package *Statistica* it has been checked the hypothesis about the normal distribution of the passenger turnover and the average values of the trips distances. The results of this procedure are presented on the pic. 4 and 5.



Pic. 4. Passenger turnover distribution



Pic. 5. Average trips distances distribution

As we can see the hypothesis about normal distribution of the turnover and the average trips distances have not been rejected. So, according to that it can be defined the confidence interval for the average trip distance which will reflect the most probable states of the O-D matrix.

$$\bar{l}_i(p) = \mu_l \pm t_p \cdot \sigma_l, \quad (35)$$

where μ_l – the mean of distribution of the average trip distance; t_p – fractile of normal distribution; σ_l – standard deviation of the average trip distance.

Summary

As a results of the worked up investigations it can be said the following. First of all widely used approaches of the O-D matrix modelling are constructed on the deterministic hypothesis about trips forming that do not correlate with the real forces and characters of the trips in the biggest cities. The second one it has to be told that the trips have the variable-based nature and to describe their distribution on the city network it has to be used the simulation modelling.

Dokończenie tekstu na stronie 12

we created factors from the distances, and corrugated the original time series distances with them. The stop points which are close to each other and shows similarity grouped together, while the remote and dissimilar ones have not.

We were thinking about what would be the results, if we took into account only the physical distances and made the clusters that way. The results show, that however there are essential differences between the two methods, the effects of physical distances are too strong. These effects distort the attribute parameters and the results, so we rejected this method in this form.

As it can be seen on the figure the coverage of different colours and planning zones (drawn on the basis of local knowledge) are quite good, although the colouring was made just with mathematical methods without any local knowledge as shown above.

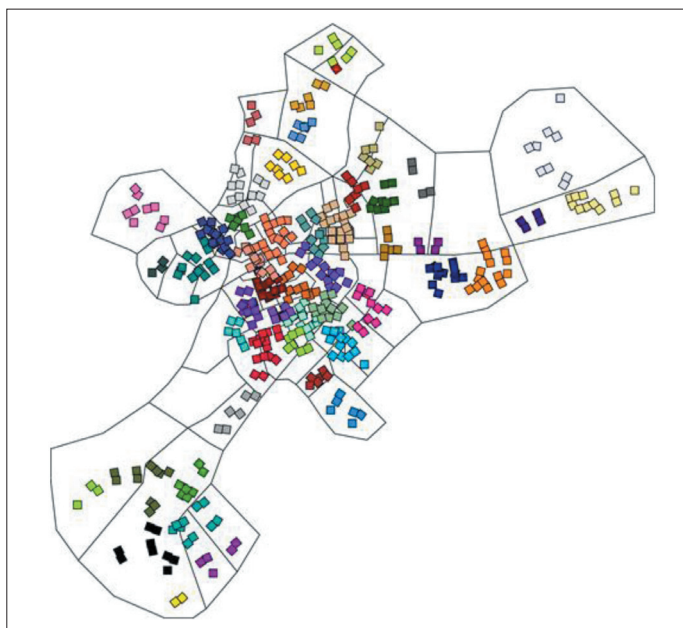


Fig. 6. The clusters of stop points, corrugated with physical distances

Conclusion

Continually developing world of today, the amounts of different kind of data are growing in every aspect of life. In this paper we showed an opportunity to examine and use this data, which could be generated in public transport every day. We clustered the stop points into groups, based on one day's data of passengers boarding. In a city, where the smart cards are in daily usage, our investigation would show a more accurate picture, thanks to the bigger data set. In the examination we used the Euclidean distance method for time series similarity measuring and the Ward method for clustering the stop points. The result did not show the perfect shape of the zones, but definitely gives us a reason for further investigation. For the accurate picture it is necessary to make the correct calibration so in the further work we want to deal with these issues. We also want to take into account the alighting number of passengers and parallel to this to try other calculating methods.

References

1. Csiszár Cs., Földes D., *Analysis and Modelling Methods of Urban Integrated Information System of Transportation*. Smart Cities Symposium, 24-25 June 2015. Prague, Czech Republic, pp. 1-10 DOI:10.1109/SCSP.2015.7181574, ISBN: 978-1-4673-6727-1 <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7181574>
2. http://psycho.unideb.hu/statisztikav1.0/pages/p_5_10.xml (2016.01.06)
3. Pang-Ning Tan, Michael Steinbach, Vipin Kumar, *Bevezetés az adatbányászathoz*, Panem Kft., Magyarország, 2011.
4. Buza K., *Fusion Methods for Time-Series Classification*, University of Hildesheim, 2011.
5. Bodon F., Buza K., *Adatbányászat*, 2014.
6. Abonyi J. (szerk), *Adatbányászat a hatékonyság eszköze*, ComputerBooks, Budapest, 2006.
7. Obádovics C., *Klaszteranalízis*, Eszterházy Károly Főiskola, Eger, 2009.
8. *A language and environment for statistical computing*, Foundation for Statistical Computing, Bécs, 2008.

Dokończenie tekstu ze strony 8

And the third one – the value of the array data forms the big number of degrees of freedom that is why the most probable states of the O-D matrix can be defined only in some confidence interval but not by point representation.

References

1. Ortuzar J. de D., Willumsen L. G., *Modelling transport*. 4th ed., John Wiley & Sons Ltd, Santiago, 2011.
2. Winston C., Small K. A., *The demand for transportation: Models and applications*, University of California, California, 1998.
3. Fratar T. J., *Vehicular Trip Distribution by Successive Approximation*, *Traffic Quarterly*, 1954, No. 8.
4. Quarmby D. A., *Choice of travel mode for the journey to work*, *Journal of Transport Economics and Policy*, 1967, Vol. 1, No. 3.
5. Barbier M., Merlin P., *Le futur réseau de transport en région de Paris*, Cahiers de l'I. A.U.R.P., Paris, 1966.
6. Jones I. S., *Gravity models and Generated Traffic*, *Journal of Transport Economics and Policy*, 1970, Vol. 6, No. 2.

7. Preston J., *Demand forecasting for new local rail stations and services*, *Journal of Transport Economics and Policy*, 1991, Vol. 25, No. 2.
8. Shuman H., Harding J., *Prejudice and the Norm of Rationality*, *Sociometry*, 1963, Vol. 27.
9. Brown L. A., Moore, E. G., *The intra-urban migration process. A perspective*, „General System”, 1970, Vol. 15.
10. Moore E. G., Brown L. A., *Urban acquaintance fields: an evolution of a spatial model*, „Environment and Planning”, 1970, No. 4.
11. Mokhtarian P. L., Bagley M. N., *Modeling Employees' Perceptions and Proportional Preferences of Work Locations: The Regular Workplace and Telecommuting Alternatives*, *Transportation Research*, 2000, Part A, No. 34.
12. Weise G., *Die Anwendung mathematisch-statistischer Berechnungsmethoden für die Bearbeitung der Prognose des individuellen Strassenverkehrs*, „Die Strasse”, 1965.
13. McCarthy T. R., Izraeli O., *Variations in travel distance, travel time and modal choice among SMSAs*, *Journal of Transport Economics and Policy*, 1985, Vol. 19, No. 2.
14. Shelejhovskij H., *The composition of the urban plan as the transport problem*, State Institute for Urban Design, Moscow, 1946.