BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES
DOI: 10.24425/bpasts.2025.154723
Special Session on Development Trends in Urban Systems

THIS IS AN EARLY ACCESS ARTICLE.
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

Synthetic Urban Agglomeration Modelling to Enable Big Data Applications in Transportation Systems

Michał ZAWODNY[1*], Piotr MACKIEWICZ[2]

[1] Faculty of Civil Engineering, Wrocław University of Science and Technology (Politechnika Wrocławska), 50-370 Wrocław, Poland
https://orcid.org/0000-0001-5709-3429
[2] Faculty of Civil Engineering, Wrocław University of Science and Technology (Politechnika Wrocławska), 50-370 Wrocław, Poland
https://orcid.org/0000-0002-3170-6415

Abstract. The paper proposes a new usage of Monte Carlo simulation in the field of transportation. The method allows to overcome problems connected to data availability in big data research, and make the research independent of biases connected to usage of existing cities and agglomerations. Urban development trends and emerging disruptive technologies, such as autonomous vehicles, can change the urban system. Simulations will be needed to ensure that urban agglomerations develop low-carbon emission transportation systems, by simulating non-existent characteristics of smart cities. In the paper, the Monte Carlo simulation was used to simulate the numbers of residents in each group in the city agglomeration. On the basis of that and the assumptions described in the paper, the OD (Origin-Destination) matrix of the simulated agglomeration was made. The simulation result is presented with the PTV Visum model and the simulated origin-destination matrix heatmap. However, the model and OD matrix presented is an example. The method allows to simulate city agglomeration of any size, depending on conducted research needs.

Key words: simulation, agglomeration, Monte Carlo

1. INTRODUCTION

For efficiently managing transport in urban systems, traffic models are made. The problem in research on future modes of transport and the use of new data sources in transport planning is the lack of agglomerations or cities with data resources or such modes of transport. To overcome this problem, the Monte Carlo simulation of agglomeration is proposed. The simulated agglomeration allows the research to be conducted in a hypothesis environment whenever full data is not accessible or costly to obtain. Based on the method, the PhD dissertation is about using numerous and different big data sources in transport modelling, as there is no city agglomeration that contains many sources of big data available to obtain. This method allowed to verify the hypothesis with promising results. The simulation presented in the paper was a basis for conducting research in the PhD dissertation. Without the simulation, research on numerous big data sources would not be possible without time- and cost-consuming data gathering.

The method is particularly needed as the current trend of urban development is deglomeration. Deglomeration is a challenge for urban transport systems, as new public transport connections are needed for efficient and low-carbon transportation in agglomerations. Simulations can lead to better planning in the development of agglomerations or can be used to simulate a future urban agglomeration.

In the literature, Monte Carlo-based simulations are done for simulating a variety of phenomena.

The sequential Monte Carlo method can be used to track travellers in indoor situations using Bluetooth low-energy beacons [1], when GPS is not available, for example, in metro stations. A similar approach using few sources of kinematic and positioning data was presented in [2]. In [3] the Monte Carlo simulation allowed to conduct a dissolved gas analysis with synthetic data. In [4] the SimSQL system is proposed, which allows SQL-based specification, simulation, and querying of database-valued Markov chains by using Monte Carlo simulation. Researchers in [5] also used the Markov Chain Monte Carlo (MCMC) simulation as in [4] , to model accident risk in marine transport. MCMC was also used to simulate financial data [6], and to create a forecast model for earthquakes [7]. In [8] the lack of "true" effective temperature measurements under different fuel rod conditions was solved using a Doppler effect database based on Monte Carlo simulation, covering various fuel temperature profiles and fuel rod conditions. The light propagation model in [9] was used

*e-mail: michal.zawodny@pwr.edu.pl

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

using Monte Carlo simulation to create databases. The Monte Carlo simulation database provided 300 models in the study [10] to evaluate the method for quick calculations of detector responses. In the case of dealing with uncertain data, a Monte Carlo-based system can be used [11].

Synthetic simulated databases are widely used in research. In [12] research shows how to build a synthetic database. Synthetic databases in can solve problems related to the limitation in access to data connected with privacy issue [13], the problem is common with Authors research field, as data about travellers can contain sensitive information, such as home address. Synthetic database was also used to train neural network in Captcha breaking [14], also with MCMC usage. Another example of synthetic database based on Monte Carlo methods was used in [15] where authors generated chemical structures and novel drug-like molecules. [16] used the Monte Carlo based method for computer-aided synthesis planning.

The Monte Carlo and synthetic databases are widely used in many fields. In the paper we would like to present a new usage in transportation. Usual use consists of cases such as Monte Carlo simulation to carry out the analysis of the transport system with a lack of historical data [17]. MCMC was used to impute data in ITS systems [18] [19]. Monte Carlo simulations were used along with machine learning methods for soil slope stability evaluation [20]. [21] used the Monte Carlo method to find an optimum solution to the fuzzy balance transportation problem. [22] uses Monte Carlo simulation to simulate different scenarios to solve the uncertainty of travel and dwell durations in order to allow the calculation of the choice of the shortest path under stochastic weather conditions. In [23] researchers tested the small-worldness of US rail and road networks. [24] used MCMC-based algorithm for the risk assessment of black swan on highways. The use of Monte Carlo simulation in [25] allowed to identify the vulnerable components of the virtual city network. In [26] a rational decision-making of urban transportation infrastructure investment was made with Monte Carlo simulation to verify the robustness of the optimisation solution and conduct multiple random experiments.

Simulations can enable researchers to work with future big data sources, such as IoT (internet of things), BIM and 3D GIS [27] from Digital Twins (DT) to better plan urban transport infrastructure. DT can also provide real-time data [28]. The Transportation DT are also proposed [29]. A review [30] summarizes wide use of DT in transport modelling.

A part of existing Smart City mobility system, such as ITS are already used in transport modelling [31] for obtaining OD matrixes. The DT particularly for ITS systems is proposed in [32]. The Smart City mobility system can use autonomous vehicles traffic [33] [34], which are not existent in traffic in high level automation. Autonomous vehicles with V2I (vehicle-to-infrastructure) systems are another big data source for transportation modelling as V2I contains spatio-temporal data [35].

Data-driven policy is used with success in Great Britain, as part of government regulations [36], and Transport for London guidelines [37].

The authors want to further develop the usage of Monte Carlo simulation in the field of transportation systems. Due to the lack of a large number of big data sources from a specific urban agglomeration, it was decided to simulate a realistic urban agglomeration for the purpose of research on big data sources in the OD (Origin-Destination) matrix. The purpose of the simulation is also to avoid making a specific case study of an urban agglomeration, that is, due to the authors affiliation - Wroclaw on the basis of the 2018 and 2024 Comprehensive Traffic Research. It also allowed the use of nine sources of big data described in the one of authors PhD dissertation, at the time of publication, at the completion stage.

The paper is presented in the following order: Section 2 starts with agglomeration description according to transport zones, then describes elements of Monte Carlo simulation usage, followed by assumptions and calculations concerning traffic modelling, and visualisation of the results as a heatmap. Section 3 shows how the transportation model was made with its parameters, and results as traffic flows in the proposed transportation network model. Section 4 finally shows the conclusions of the research.

2. CALCULATIVE BASIS OF CITY AGGLOMERATION SIMULATION

It was decided to create a simplified agglomeration to 100 Transportation Analysis Zones (TAZs), which gave an adequate OD matrix of 100x100. 10 TAZ of the strict centre with the highest generation of trips were assumed. These are area numbers from 1 to 10. Then 40 TAZ of the central area of the agglomeration - numbers from 11 to 50 with an average number of trips. TAZ outside the main city of the agglomeration with the smallest number of trips, 40 TAZ with numbers 51 to 90. The last 10 TAZ with numbers 91 to 100 represent transit routes that cover trips to distant areas of the agglomeration and traffic outside the agglomeration with an impact on the flow of travellers in the agglomeration. To obtain a diverse number of residents in TAZ, it was decided to use a method of generating random numbers similar to that used in the Monte Carlo method. The pseudo-random number generator is MS Excel. Randomisation was performed using the =RANDOM() function, which returns a pseudorandom number in the range of real numbers from 0 to 1. To obtain the assumed number of trips, the function was multiplied and a constant was added. The function formulas for each type of TAZ are presented in Table 1. The formula for the number of residents of group X in the i-th region – Eq. (1) consists of the first number being the constant S plus the variable number Z multiplied by the pseudo-random number L generated from 0 to 1:

$$KX_i = S + Z * L \tag{1}$$

Eq. (1) copied to the MS Excel interface is:

$$KX_i = S + Z * RAND() \tag{2}$$

Table 1 is written in the layout of Eq. (2), but it can be read by comparing Eq. 2 to Eq. (1), where K0 are children up to 7 years of age, K1 are commuters, K2 are secondary school and university students, K3 are primary school students, and group K4 are retirees.

TABLE 1. Summary of Simulation of TAZ Residents' Groups

TAZ type/ Resident group	K0	K1	K2	K3	K4
1-10	600+600 *RAND()	2500+2500 *RAND()	1200+1200 *RAND()	1800+1800 *RAND()	1000+1000 *RAND()
11-50	250+250 *RAND()	1500+1500 *RAND()	500+500 *RAND()	750+750 *RAND()	400+400 *RAND()
51-90	100+100 *RAND()	700+700 *RAND()	200+200 *RAND()	300+300 *RAND()	150+150 *RAND()
91-100	600+600* RAND()	2500+250* RAND()	1900+190* RAND()	300+300* RAND()	150+150* RAND()

After generating the number of inhabitants for each TAZ, average trip frequencies per resident were assumed. For example, in the case of Comprehensive Traffic Researches, it would be more advisable to determine mobility for each TAZ separately. However, due to the simplification of formulas, it was decided to simplify mobility by assuming the same for each group, with differentiation by travel motivation. The assumptions were made to show differences between demographics. In the case, for example, commuters take an average of one trip to work daily, but not all of them come back to home after work. Therefore, 0,9 average trip from work to home is made and number of average other to home trips is increased in comparison to non home-baed tips to fill the missing 0,1 trip to home daily. Retirees make an average of one trip from home to other and other to home with high – 0,9 non home-based trips daily. Trip frequencies are listed in Table 2. In Table 2 HW stands for trips from Home to Work, HS – Home to School, HO – Home to Other, WH – Work to Home, SH – School to Work, OH – Other to Home, and NHB – None Home-Based trips.

TABLE 2. Trip Frequency

TAZ\Trip Purpose	HW	HS	HO	WH	SH	OH	NHB
K0	0	0	0	0	0	0	0
K1	1	0	0,2	0,9	0	0,3	0,2
K2	0	0,9	0,2	0	0,7	0,3	0,3
K3	0	0	1,1	0	0	1,1	1
K4	0	0	1	0	0	1	0,9

Due to the simulative properties of the model, it was decided to determine the attraction of individual TAZ for the given travel motivations in percentage form. Due to the fact that the numbers were not drawn randomly but rather defined specific ranges, the maximum and minimum values were presented in Tables 3 and 4. The numbers were assumed in a way that reflected the nature of the agglomerations. For example, schools were not established in each transport region, and the occurrence of universities was limited to the central area only.

TABLE 3. Maximum percentage values of attraction in TAZ

TAZ\Attraction	Work	Primary School	Secondary School	University	Other
1-10	3,10%	4,50%	6,50%	25,40%	4,50%
11-50	1,70%	3,90%	6,10%	0%	1,40%
51-90	1,10%	1,20%	0%	0%	0,70%
91-100	3,20%	0%	0%	0%	2%

TABLE 4. Minimum percentage values of attraction in TAZ

TAZ\Attraction	Work	Primary School	Secondary School	University	Other
1-10	1,50%	2,50%	3,20%	10,80%	2,50%
11-50	1,30%	0,20%	2,20%	0%	0,50%
51-90	0,10%	0,20%	0%	0%	0,20%
91-100	1,40%	0%	0%	0%	0,90%

Using the above data, we proceeded to generate a travel matrix. The trip generation of the regions (Y) was calculated using Eq. (3), where R stands for liveliness coefficient with a subscript showing the abbreviation of the travel motivation, e.g. OH as Other-Home, %A stands for the attraction of a particular type of attraction, e.g., %AWj is the % attractiveness of the j-th region for the work motivation, %ASPr is the percentage of attractiveness for the primary school learning motivation, similarly, the subscript S and U stands for secondary schools and universities. %AOj denotes the attractiveness of the j-th region for travel other than, i.e., not related to work and home.

$$
\begin{aligned}
Y_{ij} &= R_{HW1} * K1_i * \%AW_j + R_{HS2} * K2_i * \\
&* \left(\frac{\%AS_j}{\sum AS_j} + \frac{\%AU_j}{\sum AU_j}\right) + R_{HS3} * K3_i * \%ASPr_j + \qquad (3)\\
&+ R_{OH1} * K1_i * \%AO_j * 0{,}8 + R_{OH2} * K2_i * \%AO_j * 0{,}8 + \\
&+ R_{OH3} \ * K3_i * \%AO_j * 0{,}8 + R_{DH3} * K4_i * \%AO_j * 0{,}8 + \\
&+ R_{NHB1} * K1_i * \%AO_j * 0{,}8 + R_{NHB2} * K2_i * \%AO_j * 0{,}8 + \\
&+ R_{NHB3} * K3_i * \%AO_j * 0{,}8 + R_{NHB4} * K4_i * \%AO_j * 0{,}8
\end{aligned}
$$

Using Eq. (3) the OD matrix was made. Shown as heatmap on Fig. 1. The higher number of trips is shown in magnitude of color. On the top side of figure the legend is shown with range of numbers.

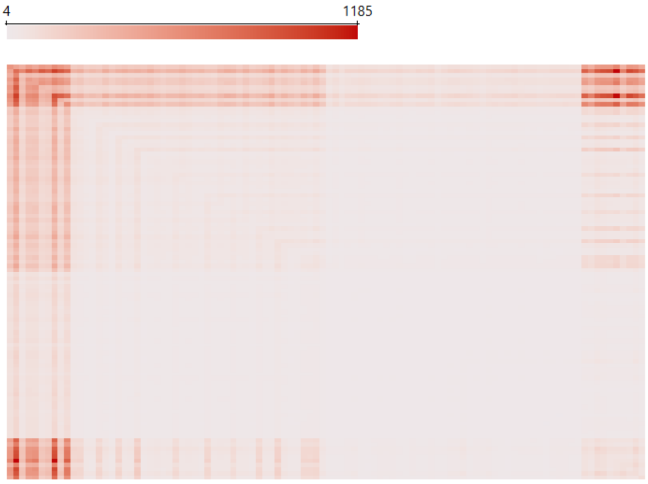

Fig.1. Simulated OD Matrix

The simulation process is shown step by step as flowchart in Fig. 2.

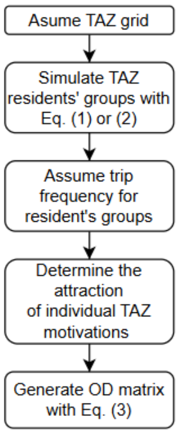

Fig.2. Simulation flowchart

3. AGGLOMERATION MODEL

The method only concerns the OD matrix, therefore, the network is simplified. For example, the modal split was not taken into account or the basic parameters of the PTV Visum programme were not changed.

First, a TAZ grid was placed. The grid has 100 of them in a simulated agglomeration of about half a million inhabitants. To create an accurate transport model, it would be necessary to create a larger number of TAZ, but this goes beyond the scope of this paper, so it was decided not to divide the simulated TAZ into smaller ones. The layout started from the strict city centre, zones 1 to 10. The assumption is an area of dense development with a high population density with high magnitude of trip generation, from and to agglomeration centre. This area is surrounded by zones 11 to 50, which is an urban development area with a smaller number of inhabitants. The role of the zones is realistic zone layout in the agglomeration, and trip generation not only from outer zones, but traffic flow inside the city. Both types of zones are shown in the screenshot from the PTV Visum programme in Fig. 3.

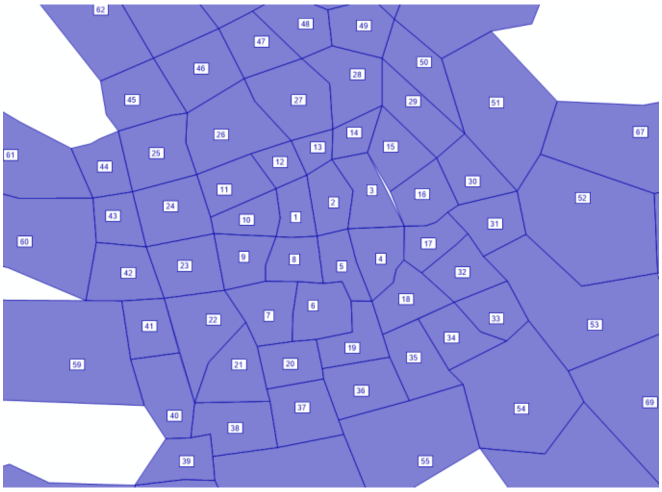

Fig.3. TAZ grid in the centre of the agglomeration

It was found that the layout of the urban network is not to be a perfect grid of squares, despite the fact that the dimension of the OD matrix of 100x100 allowed for this. It was also decided that it would not be a city with a typical American layout with a latitudinal network of connections. The agglomeration has a typical shape of European cities and spreads out with branches. It took the shape of an "octopus" with "tentacles" in the form of branches. Areas from 51 to 90 present the shape of such a layout. Some of them are so-called villages adjacent to the agglomeration, tightly wrapping the grid of areas 1-50. The rest represent branches of the agglomeration, being towns concentrated around access roads. “Tentacle” zones represents characteristic of deglomeration phenomena in the simulated agglomeration, with low population density and large area around the agglomeration centre. Areas 91-100 present the influence of distant towns in relation to the agglomeration. This simplification was decided for the sake of the readability of this paper. A very large model would make it impossible to show most of the numerical values on the connections. These TAZ are symbolically shown as square zones. The layout of the remaining TAZ and thus the entire layout of TAZ in the simulated agglomeration is shown in Fig. 4. With the TAZ grid ready, a network of main connections in the agglomeration was planned. Due to the research focused on OD Matrix only, it was decided not to do it in full accuracy and limited itself to creating a network of main urban arteries, without residential roads. The dissertation also does not include a division into means of transport, so a general network was created in the software based on a high-capacity road network to visualise only the flows of travellers between areas under relaxed conditions.

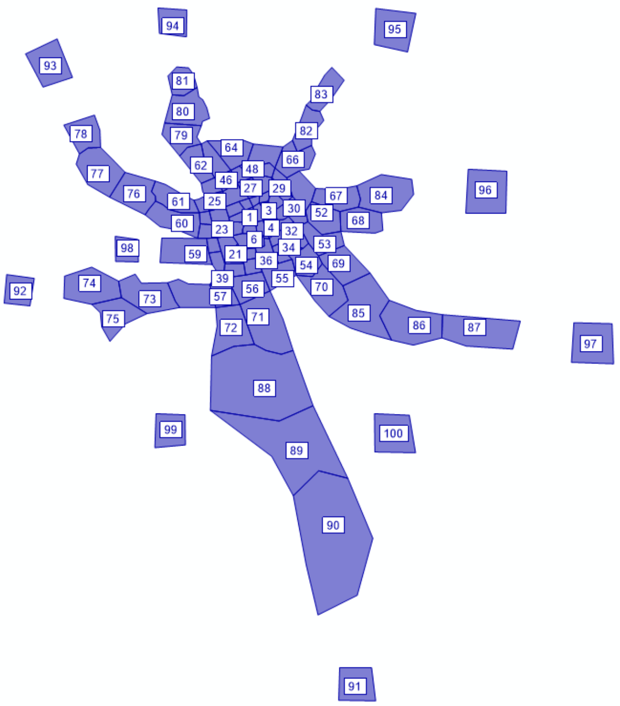

Fig.4. TAZ grid in agglomeration

Fig. 5 shows the system of nodes throughout the agglomeration. The nodes have the basic parameters from PTV Visum, to lower affect of control systems, all of them were set to "unknown" control type, with all types of turns on intersections possible. Also, nodes have no capacity limits, as for macroscopic characteristics of the model, the intersections do not have set particular geometry.

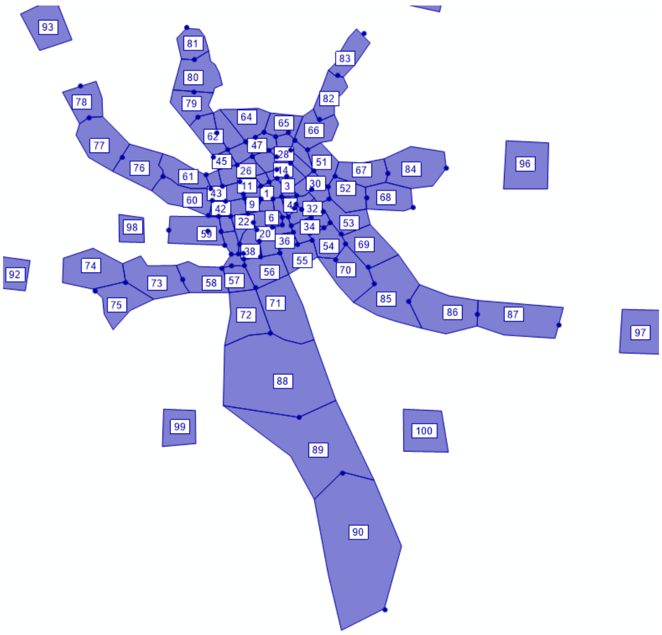

Fig.5. Nodes grid in agglomeration

Based on the nodes in Fig. 5, a road network was created. Fig. 6 shows the layout of connections in the centre of the agglomeration and Fig. 7 shows a distant view of the entire agglomeration model. Figure 8 shows the network without the zones layer enabled.

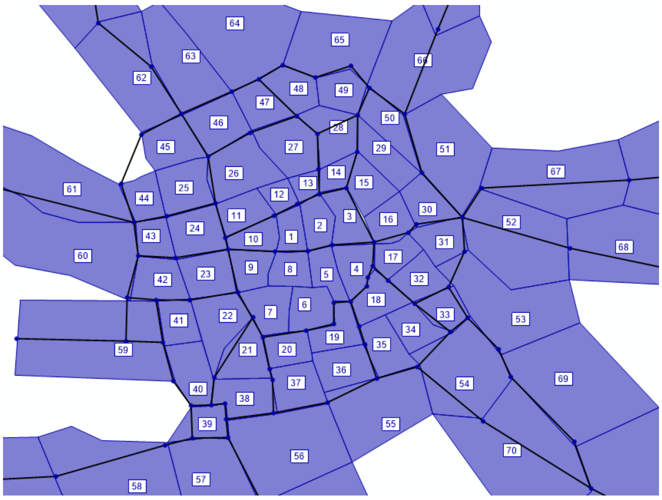

Fig.6. Connections in agglomeration centre

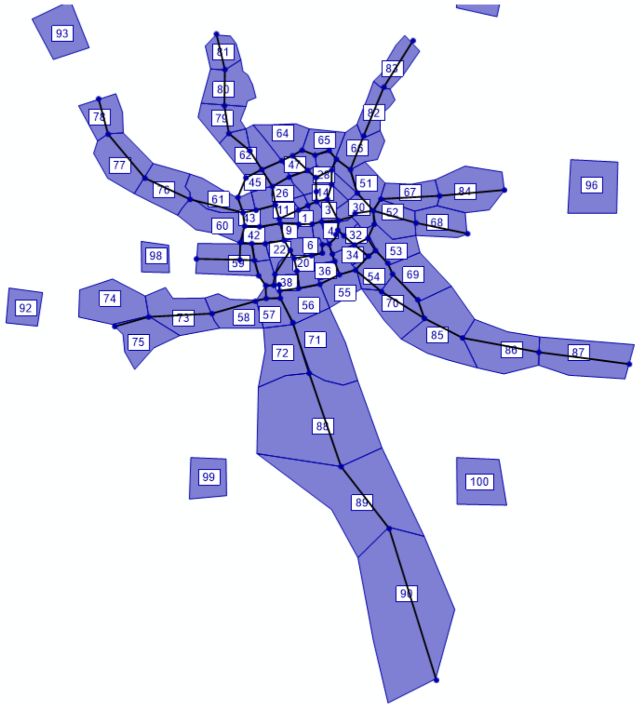

Fig.7. Links in Agglomeration

To show traffic flows without infrastructure constraints, each of the connections has the same parameters. These are simply the basic parameters suggested by PTV Visum. Capacity – CapPrT is 99999 and V0PrT is 50 km/h. There is also no modal split between modes of transport, so the model is not affected by the lack of parameters for public transport.

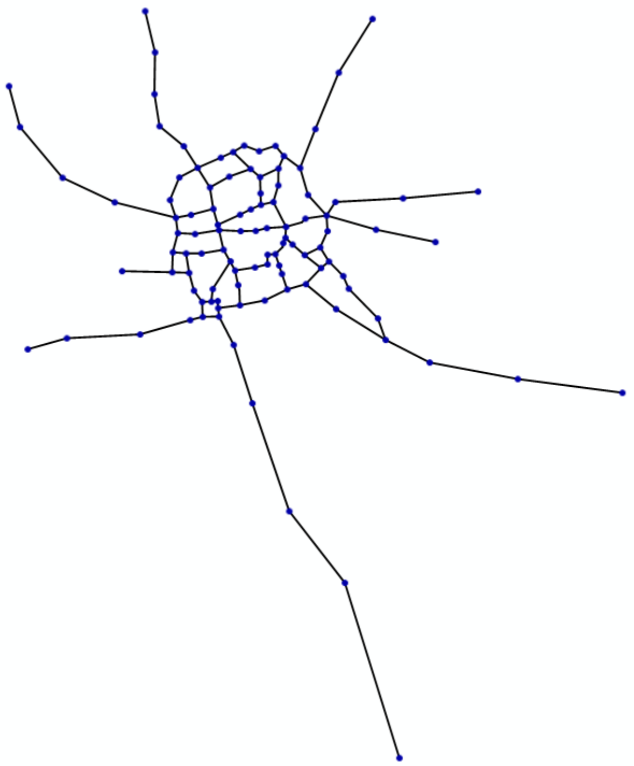

Fig.8. Links in agglomeration without zones shown

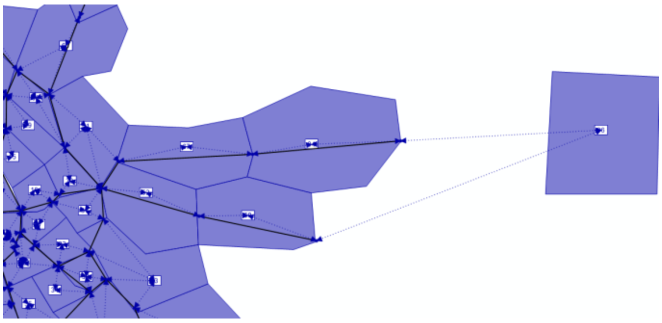

Fig.9. Part of the model showing how zones are connected to links.

Fig. 9 shows how the regions were connected to the transport network. Local roads were not modelled into the zones due to the readability of the model. This was not necessary at the level of accuracy of the flows based on the OD matrix. In this case, an alternative option for presenting the results could be a classical flow chart, but for descriptive purposes, it was decided to show the result on a simplified model. After the simulated OD matrix was introduced, the model was recalculated. The result in the form of traffic volumes on the links is presented in Fig. 10 and Fig. 11. Fig. 10 shows the central area of the agglomeration. Fig. 11 shows a view of the entire agglomeration area. Due to the lack of an external bypass of the agglomeration, large flows of travellers pass through the centre, reflecting travel between transit roads. Similarly, the area of the strict centre (TAZs 1 to 10), which also gathers a large number of residents, "draws" large flows of traffic. A unique branch of the agglomeration is the southern branch, where three TAZs representing external roads converge, because it causes a significantly higher number of trips. Fig. 1 shows the OD matrix for better readability in the form of a heat map.

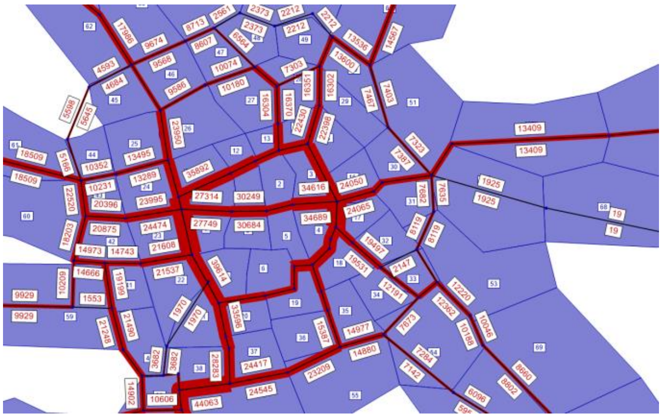

Fig.10. Traffic flows in simulated agglomeration centre

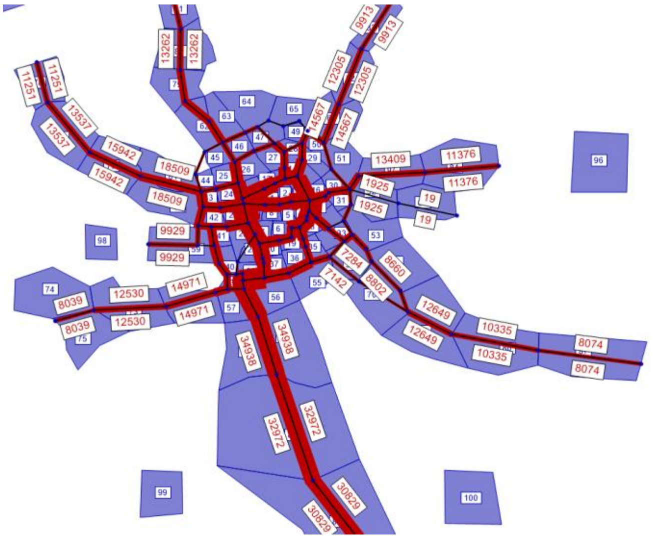

Fig.11. Traffic flows in simulated agglomeration – agglomeration view

Figs. 10 and 11 show that some characteristics of agglomerations are simulated correctly. Each type of TAZ shows the expected travel magnitude. The high density of traffic flows shows the popular deglomeration pattern in recent agglomerations, as the high traffic flow comes from outside of the simulated city into it. The links with varying traffic levels, allows to check how the model will behave with other data input, such as big data based OD matrixes.

4. CONCLUSIONS

In the paper, simulated agglomeration was presented using the Monte Carlo method. A new usage of the method in the transportation field is proposed to work with future big data sources and using not available and costly sources to obtain existing sources. Simulated agglomeration allows to reduce costs of research, and avoid biases of particular cities agglomeration. The method also allows to make a research on future agglomeration with smart city characteristics, that does not exist yet. Smart cities will have numerous sources of big data. IoT usage such as V2I (vehicle to infrastructure) communication, or data from digital twins network will provide rich data sources. Also, the simulated agglomeration does not contain biases from existing cities, as was possible in the research conducted only in a single city case study. The PTV Visum model, and the OD Matrix represent 24-hour period of travels, therefore, matrix is symmetric. Traffic flows have realistic value to around of approximately 500 000 residents agglomeration simulated in the case. The maximum average daily traffic in the main links in the agglomeration is around 100 000 trips in both directions. Less important links show values of a few thousand trips. Without transit trips, the values are realistic, as local roads are not present in the model. The assumptions made to the model, however, can lead to bigger numbers on the shorter links between TAZs with greater trip generation and attraction, as the capacity limit of links and nodes is set high, as mentioned in Section 3. With the capacity limit set lower, the route choice algorithm would redirect traffic into alternate routes. The model, however, represents flows in OD matrix and does not focus on route choice, as the use presented in the paper focusses on OD matrix, not the classical four-step travel model. The method allows to simulate agglomerations of different sizes by using other number of TAZ. In addition, the model can contain more details depending on the research needs. Such models can be helpful to city planners. The steps shown in Fig. 2 can also help sustainability initiatives simulate missing parts of information in transportation models. For example, in Wroclaw Comprehensive Traffic Research 2018, OD matrices were available from Big Data sources, however, information about the real OD matrix of the city was not available publicly. With trip frequencies available publicly, it is possible to use the method for completing the OD matrix that was not available or existent at the time. City planners can use simulation steps in case of missing data for models.

In the future research authors plan to publish usage of the model with big data sources implementing the method into a particular research problem. The second plan for future research will be model validation by acquiring an existing agglomeration transport model and comparing it with a similar simulated model.

The authors' scientific contributions: conceptualization, M.Z. and P.M.; methodology, M.Z.; software, M.Z.; validation, M.Z. and P.M.; formal analysis, M.Z. and P.M.; investigation, M.Z. and P.M.; resources, M.Z.; data curation, M.Z.; writing—original draft preparation, M.Z. and P.M.; writing—review and editing, M.Z.

REFERENCES

[1] F. Daniş and A. Cemgil, “Model-Based Localization and Tracking Using Bluetooth Low-Energy Beacons,” *Sensors*, vol. 17, no. 11, p. 2484, Oct. 2017, doi: 10.3390/s17112484.

[2] P. Baranski, M. Polanczyk, and P. Strumillo, “Fusion of Data from Inertial Sensors, Raster Maps and GPS for Estimation of Pedestrian Geographic Location in Urban Terrain,” *Metrol. Meas. Syst.*, vol. 18, no. 1, pp. 145–158, Jan. 2011, doi: 10.2478/v10178-011-0014-3.

[3] E. K. Babu, I. Bashir, G. Pillai, and K. C. Jyothi, “Generation of synthetic datasets for transformer’s dissolved gas analysis using Monte-Carlo Simulation,” in *2021 56th International Universities Power Engineering Conference (UPEC)*, Middlesbrough, United Kingdom: IEEE, Aug. 2021, pp. 1–4. doi: 10.1109/UPEC50034.2021.9548242.

[4] Z. Cai, Z. Vagena, L. Perez, S. Arumugam, P. J. Haas, and C. Jermaine, “Simulation of database-valued markov chains using SimSQL,” in *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, New York New York USA: ACM, Jun. 2013, pp. 637–648. doi: 10.1145/2463676.2465283.

[5] S. Faghih-Roohi, M. Xie, and K. M. Ng, “Accident risk assessment in marine transportation via Markov modelling and Markov Chain Monte Carlo simulation,” *Ocean Eng.*, vol. 91, pp. 363–370, Nov. 2014, doi: 10.1016/j.oceaneng.2014.09.029.

[6] F. Bruni Prenestino, E. Barbierato, and A. Gatti, “Robust Synthetic Data Generation for Sequential Financial Models Using Hybrid Variational Autoencoder–Markov Chain Monte Carlo Architectures,” *Future Internet*, vol. 17, no. 2, p. 95, Feb. 2025, doi: 10.3390/fi17020095.

[7] H. Karaca, “Generation of synthetic catalog by using Markov chain Monte Carlo simulation and inverse Poisson distribution,” *J. Seismol.*, vol. 25, no. 4, pp. 1103–1114, Aug. 2021, doi: 10.1007/s10950-021-10018-z.

[8] G. K. Delipei, P. Rouxelin, M. Avramova, E. Ivanov, and K. Ivanov, “Doppler effective temperature database based on Monte Carlo simulations,” *Best Estim. Plus Uncertain. Int. Conf. BEPU 2024*, May 2024.

[9] B. Rezaeifar *et al.*, “A deep learning and Monte Carlo based framework for bioluminescence imaging center of mass-guided glioblastoma targeting,” *Phys. Med. Biol.*, vol. 67, no. 14, p. 144003, Jul. 2022, doi: 10.1088/1361-6560/ac79f8.

[10] L. Lin, Y. Wang, Q. Zhang, and Y. Li, “A Monte Carlo-based adaptive weighted reduced order model for gamma density measurement,” *Ann. Nucl. Energy*, vol. 178, p. 109341, Dec. 2022, doi: 10.1016/j.anucene.2022.109341.

[11] R. Jampani, F. Xu, M. Wu, L. L. Perez, C. Jermaine, and P. J. Haas, “MCDB: a monte carlo approach to managing uncertain data,” in *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, Vancouver Canada: ACM, Jun. 2008, pp. 687–700. doi: 10.1145/1376616.1376686.

[12] D. Libes, D. Lechevalier, and S. Jain, “Issues in synthetic data generation for advanced manufacturing,” in *2017 IEEE International Conference on Big Data (Big Data)*, Boston, MA: IEEE, Dec. 2017, pp. 1746–1754. doi: 10.1109/BigData.2017.8258117.

[13] T. E. Raghunathan, “Synthetic Data,” *Annu. Rev. Stat. Its Appl.*, vol. 8, no. 1, pp. 129–140, Mar. 2021, doi: 10.1146/annurev-statistics-040720-031848.

[14] T. A. Le, A. G. Baydin, R. Zinkov, and F. Wood, “Using synthetic data to train neural networks is model-based reasoning,” in *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA: IEEE, May 2017, pp. 3514–3521. doi: 10.1109/IJCNN.2017.7966298.

[15] M. Roucairol, A. Georgiou, T. Cazenave, F. Prischi, and O. E. Pardo, “DrugSynthMC: An Atom-Based Generation of Drug-like Molecules with Monte Carlo Search,” *J. Chem. Inf. Model.*, vol. 64, no. 18, pp. 7097–7107, Sep. 2024, doi: 10.1021/acs.jcim.4c01451.

[16] X. Wang *et al.*, “Towards efficient discovery of green synthetic pathways with Monte Carlo tree search and reinforcement learning,” *Chem. Sci.*, vol. 11, no. 40, pp. 10959–10972, 2020, doi: 10.1039/D0SC04184J.

[17] V. D. A. Guimarães, I. C. Leal Junior, and M. A. V. Da Silva, “Evaluating the sustainability of urban passenger transportation by Monte Carlo simulation,” *Renew. Sustain. Energy Rev.*, vol. 93, pp. 732–752, Oct. 2018, doi: 10.1016/j.rser.2018.05.015.

[18] D. Ni and J. D. Leonard, "Markov Chain Monte Carlo Multiple Imputation Using Bayesian Networks for Incomplete Intelligent Transportation Systems Data," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 1935, no. 1, pp. 57–67, Jan. 2005, doi: 10.1177/0361198105193500107.

[19] E. S. Park, L. R. Rilett, and C. H. Spiegelman, "A Markov Chain Monte Carlo-Based Origin Destination Matrix Estimator that is Robust to Imperfect Intelligent Transportation Systems Data," *J. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 139–155, Aug. 2008, doi: 10.1080/15472450802262364.

[20] F. Ahmad, P. Samui, and S. S. Mishra, "Machine learning-enhanced Monte Carlo and subset simulations for advanced risk assessment in transportation infrastructure," *J. Mt. Sci.*, vol. 21, no. 2, pp. 690–717, Feb. 2024, doi: 10.1007/s11629-023-8388-8.

[21] A. S. Mhaske and K. L. Bondar, "Fuzzy Transportation by Using Monte Carlo method," *Adv. Fuzzy Math.*, vol. 12, no. Number 1, pp. 111–127, 2017.

[22] Y. Peng, Z. Mo, and S. Liu, "Passenger's routes planning in stochastic common-lines' multi-modal transportation network through integrating Genetic Algorithm and Monte Carlo simulation," *Arch. Transp.*, vol. 59, no. 3, pp. 73–92, Sep. 2021, doi: 10.5604/01.3001.0015.0123.

[23] P. R. Aldrich *et al.*, "Monte Carlo tests of small-world architecture for coarse-grained networks of the United States railroad and highway transportation systems," *Phys. Stat. Mech. Its Appl.*, vol. 438, pp. 32–39, Nov. 2015, doi: 10.1016/j.physa.2015.06.013.

[24] A. Z. Deriba and D. Y. Yang, "Performance-Based Risk Assessment for Large-Scale Transportation Networks Using the Transitional Markov Chain Monte Carlo Method," *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part Civ. Eng.*, vol. 11, no. 1, p. 04024090, Mar. 2025, doi: 10.1061/AJRUA6.RUENG-1402.

[25] O. Kammouh, A. Cardoni, S. Marasco, and G. P. Cimellaro, "Resilience Assessment of City-Scale Transportation Networks Using Monte Carlo Simulation," *Proc. Ninth Int. Conf. Bridge Maint. Saf. Manag.*, pp. 1902–1910, Jul. 2018.

[26] M. Wang and S. Yu, "Application of Bayesian network and genetic algorithm enhanced Monte Carlo method in urban transportation infrastructure investment," *J. Comput. Methods Sci. Eng.*, p. 14727978241313262, Jan. 2025, doi: 10.1177/14727978241313262.

[27] Y. Gao, S. Qian, Z. Li, P. Wang, F. Wang, and Q. He, "Digital Twin and Its Application in Transportation Infrastructure," in *2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI)*, Beijing, China: IEEE, Jul. 2021, pp. 298–301. doi: 10.1109/DTPI52967.2021.9540108.

[28] K. Kušić, R. Schumann, and E. Ivanjko, "A digital twin in transportation: Real-time synergy of traffic data streams and simulation for virtualizing motorway dynamics," *Adv. Eng. Inform.*, vol. 55, p. 101858, Jan. 2023, doi: 10.1016/j.aei.2022.101858.

[29] M. S. Irfan, S. Dasgupta, and M. Rahman, "Toward Transportation Digital Twin Systems for Traffic Safety and Mobility: A Review," *IEEE Internet Things J.*, vol. 11, no. 14, pp. 24581–24603, Jul. 2024, doi: 10.1109/JIOT.2024.3395186.

[30] D. Nag, F. Brandel-Tanis, Z. A. Pramestri, K. Pitera, and Y. K. Frøyen, "Exploring digital twins for transport planning: a review," *Eur. Transp. Res. Rev.*, vol. 17, no. 1, p. 15, Mar. 2025, doi: 10.1186/s12544-025-00713-0.

[31] S. Afandizadeh Zargari, A. Memarnejad, and H. Mirzahossein, "Hourly Origin–Destination Matrix Estimation Using Intelligent Transportation Systems Data and Deep Learning," *Sensors*, vol. 21, no. 21, p. 7080, Oct. 2021, doi: 10.3390/s21217080.

[32] L. Bao, Q. Wang, and Y. Jiang, "Review of Digital twin for intelligent transportation system," in *2021 International Conference on Information Control, Electrical Engineering and Rail Transit (ICEERT)*, Lanzhou, China: IEEE, Oct. 2021, pp. 309–315. doi: 10.1109/ICEERT53919.2021.00064.

[33] P. Ribeiro, G. Dias, and P. Pereira, "Transport Systems and Mobility for Smart Cities," *Appl. Syst. Innov.*, vol. 4, no. 3, p. 61, Sep. 2021, doi: 10.3390/asi4030061.

[34] M. A. Richter, M. Hagenmaier, O. Bandte, V. Parida, and J. Wincent, "Smart cities, urban mobility and autonomous vehicles: How different cities needs different sustainable investment strategies," *Technol. Forecast. Soc. Change*, vol. 184, p. 121857, Nov. 2022, doi: 10.1016/j.techfore.2022.121857.

[35] C. Xu, H. Luo, H. Bao, and P. Wang, "STEIM: A Spatiotemporal Event Interaction Model in V2X Systems Based on a Time Period and a Raster Map," *Mob. Inf. Syst.*, vol. 2020, pp. 1–20, Feb. 2020, doi: 10.1155/2020/1375426.

[36] "Transport analysis guidance (TAG)." Transport Appraisal and Strategic Modelling (TASM) Division, Zone 2/25 Great Minster House, 33 Horseferry Road, London, Nov. 29, 2018.

[37] L. Beeston, R. Blewitt, S. Bulmer, and J. Wilson, "Traffic Modelling Guidelines Version 4.0." Transport for London, Sep. 2021.