

# DEMAND MODEL IN THE AGGLOMERATION USING SIM CARDS

**A. BRZEZIŃSKI<sup>1</sup>, T. DYBICZ<sup>2</sup>, Ł. SZYMAŃSKI<sup>3</sup>**

The road network development programme, as well as planning and design of transport systems of cities and agglomerations require complex analyses and traffic forecasts. It particularly applies to higher-class roads (motorways and expressways), which in urban areas, support different types of traffic. Usually there is a conflict between the needs of long-distance traffic, in the interest of which higher-class roads run through undeveloped areas, and the needs of bringing such road closer to potential destinations, cities [1]. By recognising the importance of this problem it is necessary to develop the research and methodology of traffic analysis, especially trip models. The current experience shows that agglomeration models are usually simplified in comparison to large city models, what results from misunderstanding of the significance of these movements for the entire model functioning, or the lack of input data. The article presents the INMOP 3 research project results, within the framework of which it was attempted to increase the accuracy of traffic generation in agglomeration model owing to the use of BigData – the mobile operator’s data on SIM card movements in the Warsaw agglomeration.

*Keywords:* transport system, trip modelling, demand model, trip generation, agglomeration models, BigData

## 1. INTRODUCTION

Trips within the agglomeration divides into four basic types:

- trips of the suburban area inhabitants carried out inside this area (within individual municipalities and between them),

<sup>1</sup> Warsaw University of Technology, Faculty of Civil Engineering, Institute of Roads and Bridges, ul. Lecha Kaczyńskiego 16, 00-637 Warsaw, mail: [a.brzezinski@il.pw.edu.pl](mailto:a.brzezinski@il.pw.edu.pl)

<sup>2</sup> Warsaw University of Technology, Faculty of Civil Engineering, Institute of Roads and Bridges, ul. Lecha Kaczyńskiego 16, 00-637 Warsaw, mail: [t.dybicz@il.pw.edu.pl](mailto:t.dybicz@il.pw.edu.pl)

<sup>3</sup> TransEko Design-Consulting Office, [l.szymanski@transeko.pl](mailto:l.szymanski@transeko.pl)

- trips of the suburban area inhabitants connected with the main city of the agglomeration,
- trips of inhabitants of the main city of the agglomeration related to the suburban area,
- trips of inhabitants of the main city of the agglomeration, inside this city.

In models developed for the agglomerations, the trips of inhabitants of the main city are better described because they are usually subjected to transport research. The research is less frequently carried out in the suburban areas, and if so, the trips related to the main city are examined more often than other trips (between suburban cities/municipalities). It usually results from a misunderstanding of the suburban area significance for the transport system analyses of a large city and willingness to reduce the research costs. Insufficient information about the trips in the area around the main agglomeration city force to use simplified demand models, e.g. coming down to a description of all types of trips with two models:

- a model of the trips of inhabitants of the main city, mapping both trips inside this city and outbound trips from the city (if possible, based on the results of a comprehensive traffic survey),
- a model of trips of the suburban area inhabitants, mapping trips inside individual municipalities, between municipalities and trips connected with the main city of the agglomeration.

However, in case of the trips carried out by the suburban area inhabitants, such an approach is too generalised. The application of the same mobility parameters, resistance function, modal split, etc. for trips inside the suburban area and for trips from the area to the main city does not allow to determine the number of generated trips and their spatial distribution (trip matrix) precisely enough. At the stage of the calibration of models, it often results in the necessity of interfering in the resulting matrices, their reduction or increase. The suburban area is very heterogeneous from the perspective of travel behaviours. The municipalities are characterised by different trip potentials generated to the suburban area and to the main city. It requires an individual approach at the traffic generation stage.

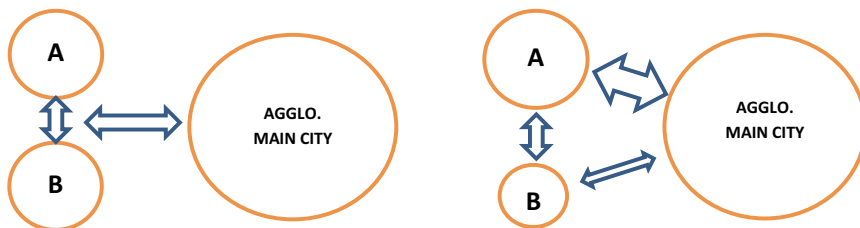


Fig. 1. Diagrams of the trip models in the agglomeration – simplified (left side) and with an individual approach to municipalities (right side): A, B – suburban area municipalities

Within the framework of the INMOP 3<sup>1</sup> research project [2] [3], it was attempted to enrich the methodology of building local trip models thanks to the use of the mobile operators' data on the SIM card movements. This type of data makes it possible to build a trip matrix for any hour, for any day of the week and any period of the year. They can also be aggregated in any way, e.g. for daily matrices [4]. The data were obtained for Warsaw (divided into districts) and for 60 municipalities of the suburban area (Warsaw agglomeration). For the purpose of analyses, a period of 15-17 November 2016 (Tuesday-Thursday) was used as representing typical behaviours on weekdays in the autumn.

## 2. TRIP ANALYSIS IN THE AGGLOMERATION BASED ON SIM DATA

As part of INMOP 3, the data on movements of SIM cards users which moved between the assumed individual suburban area municipalities, between districts in Warsaw and between the suburban area municipalities and Warsaw districts, were used. Internal trips (within municipalities and districts) were omitted. The trip matrices were developed on the basis of data from the SS7 signalling network of the T-Mobile Poland telecommunications operator. These data contain detailed information (among others, call time, and IMSI – the unique number assigned to each SIM card) for each mobile phone connection with base stations called BTS (within the TMPL -T-mobile Poland network), but without distinction of the means of transport (data from SIM cards do not allow this). From the entire collection of SIM cards at the operator's disposal, those whose users gave marketing consents to the data analysis were taken. In total, 0.47 million trips were analysed in the area of Warsaw agglomeration. By virtue of the specificity of the analysed data in the trip definition, the assumption of the maximum 3 hours stop time was adopted. It means that the SIM card user movement from A municipality to B municipality and stopping for time longer than 3hours in C municipality is respectively treated as two trips A to C & C to B. If the stop time is shorter than 3 hours it is treated as a one trip.

The necessity of data anonymisation forced the elimination of some trips, e.g. in the data aggregation in case of the small number of trips (resulting from a high level of data details, i.e. data presented in an hour, not during the day). When aggregating data, it was assumed that when travelling, the number on a given relation is equal to or less than 5, the average value of all these trips (whose value was less

---

<sup>1</sup> The project "Principles of road traffic forecasting including other means of transport" ["Zasady prognozowania ruchu drogowego z uwzględnieniem innych środków transportu"] implemented in the period from 1 February 2016 – 30 April 2019, commissioned by the National Centre for Research and Development (NCBiR) and the General Directorate for Roads and Motorways (GDDKiA) by the consortium of the consortium of technical universities: the Institute of Roads and Bridges at the Warsaw University of Technology (leader) and the Cracow University of Technology (partner)

than or equal to 5) is used. As a result of the analysis, the data on movements between the municipalities/districts and the number of trips generated by the particular municipalities/districts, and the number of trips per one inhabitant in the division into all trips and only trips related to Warsaw, were obtained.

It occurred that the greatest potential of traffic generation in the suburban area – over 5%, has Pruszków, the city near Warsaw. Over 4% of trips generates Lesznowola and Piaseczno cities. In case of trips only to Warsaw, the greatest generators are: Ząbki, Lesznowola and Marki (approx. 5% of trips between suburban area nad Warsaw).

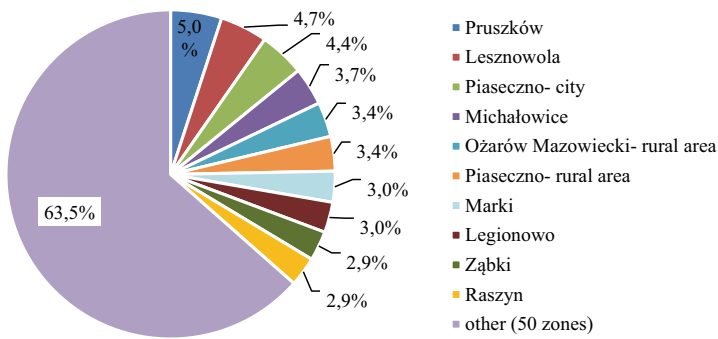


Fig. 2 The most important 10 trip generators in the Warsaw suburban area [%], based on the SIM cards data

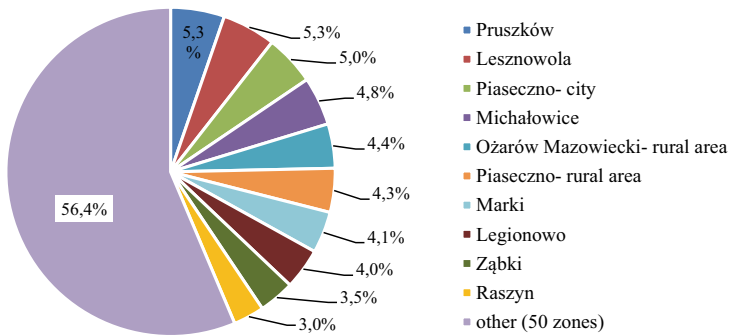


Fig. 3 The most important 10 trip generators to Warsaw [%], based on SIM cards data

The analysis of SIM cards revealed that there is no simple relationship between the trip generation potential in a given municipality and the share of trips to Warsaw. It means that there are groups of municipalities that are characterised by a stronger connection with Warsaw and those for whom there are other equally strong connections not related to Warsaw. It confirmed the previously described



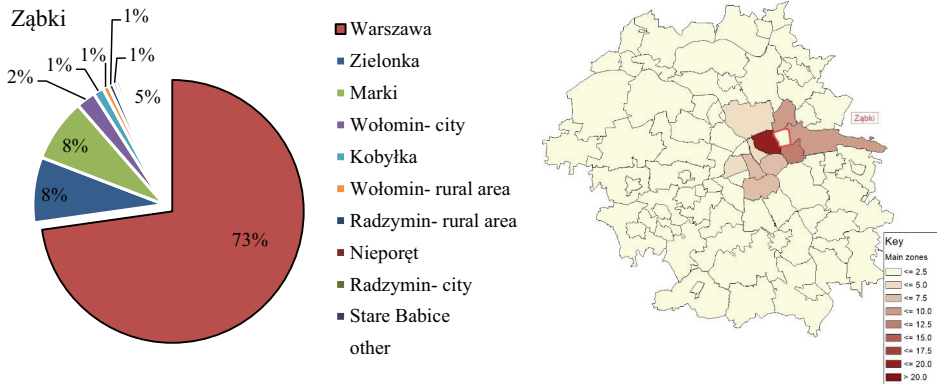


Fig. 5. City of Żąbki – percentage structure of external trips to municipalities/districts of the Warsaw agglomeration and spatial traffic distribution – based on SIM cards

An example of the generator, where the trips to Warsaw are not dominating, is city of Błonie. In this case, as much as 55% of the trips takes place to the neighbouring Błonie rural area, Ożarów Maz. rural area (7%) and to Grodzisk Mazowiecki rural area and Leszno (4%).

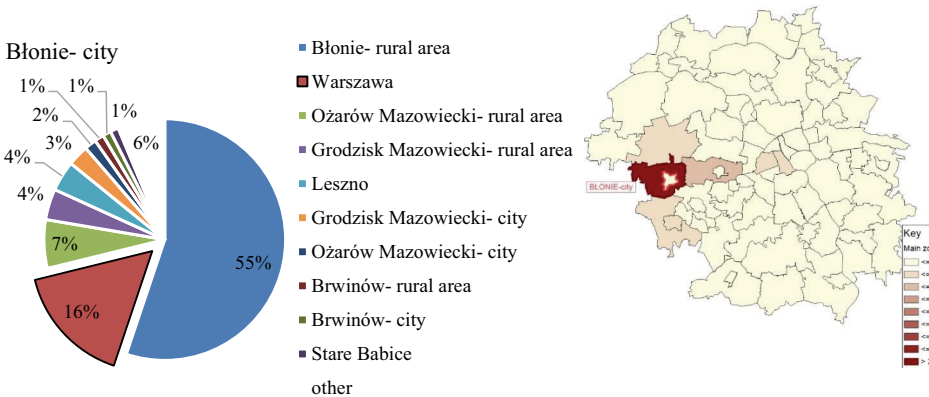


Fig. 7. City of Błonie – percentage structure of external trips to the municipalities/districts of the Warsaw agglomeration and traffic spatial distribution – based on SIM cards

### 3. TRIP GENERATION

In the traffic generation model, the variables describing the characteristics of travellers to and from particular transport zones are used. They usually come from surveys or statistical databases. Depending on the trip purpose, the used means of transport and trip periods, it is possible to apply different variables describing trips demand [5]. In case of the agglomeration model, apart from the

movements inside Warsaw, 11 explanatory variables were used: the number of inhabitants in the suburban area (LM-S), the number of inhabitants in Warsaw (LM-W), the number of workplaces in the suburban area (LMP-S), the total number of workplaces in Warsaw (LMP-W), the number of workplaces in services in the suburban area (LMPU-S), the number of workplaces in services in Warsaw (LMPU-W), the number of places in schools in the suburban area (LMS-S), the number of places in schools in Warsaw (LMS-W), the number of places at universities in Warsaw (LMU), the area of large shopping centres [ $m^2$  PWOH], share of inhabitants aged 6 and more (6plus).

On the basis of the results of the comprehensive traffic surveys in Warsaw [6], the diversification was introduced into the 18 most-represented trip purposes, 6 within the suburban area: Home-Work, Work-Home, Home-School, School-Home, Home-Other, Other-Home; 8 in relation from the suburban area to Warsaw: Home-Work, Work-Home, Home-School, School-Home, Home-University, University-Home, Home-Other, Other-Home and 4 common ones for trips across the suburban area and to Warsaw: Home-Large Shopping Centres, Large Shopping Centres-Home, Home-Other, Other-Work.

The SIM card movement analysis results were used to introduce calibration factors, individually for each municipality in the suburban area, separately for trips in the suburban area, for trips related to Warsaw and for the trips of Warsaw inhabitants connected with the suburban area:

- the trip generation coefficient in the suburban area, which changes the mobility in a given transport zone, affect the matrix size change (Gen-S),
- the coefficient generating the trips related to Warsaw, which changes the mobility of trips related to Warsaw in a given transport zone, affects the matrix size change (Gen-W),
- the absorption coefficient of trips in the suburban area, which changes the transport zone attraction, does not affect the matrix size, and it has an impact on the spatial distribution change and it concerns variables describing the workplaces, i.e. data burdened with high uncertainty (Abs-S),
- the absorption coefficient of the Warsaw inhabitants' trips connected with the suburban area, which changes the transport zone attraction, does not affect the matrix size, and it has an impact on the spatial distribution change; Abs-W - it adopts the values lower or equal to 1,
- the attraction of large shopping centres coefficient (WOH).

It allowed to create a separate description of trip behaviours to and outside Warsaw. Therefore, the possibility of mapping the local conditions, which affect the traffic generation and absorption values, was introduced into the agglomeration scale. The rules of determining the trips potentials of the suburban area inhabitants are presented in Table 1.





0-3	0.438	0.429	0.015	0.001	0.554	0.528	0.018	0.001
3-6	0.302	0.256	0.063	0.017	0.308	0.266	0.049	0.016
6-9	0.118	0.107	0.056	0.044	0.075	0.077	0.067	0.070
9-12	0.067	0.073	0.079	0.092	0.015	0.055	0.078	0.137
12-15	0.033	0.045	0.123	0.136	0.024	0.028	0.166	0.169
15-18	0.003	0.028	0.131	0.148	0.003	0.017	0.187	0.163
18-21	0.015	0.021	0.161	0.140	0.005	0.010	0.171	0.137
21-24	0.001	0.012	0.136	0.132	0.001	0.006	0.083	0.114
24-27	0.004	0.008	0.139	0.102	0.002	0.004	0.115	0.081
27-30	0.005	0.006	0.065	0.074	0.000	0.003	0.043	0.051
30-33	0.006	0.005	0.024	0.051	0.005	0.002	0.000	0.032
33-36	0.000	0.004	0.005	0.029	0.000	0.001	0.025	0.016
36-39	0.000	0.003	0.004	0.016	0.004	0.001	0.000	0.007
39-42	0.001	0.002	0.000	0.009	0.001	0.001	0.000	0.004
42-45	0.007	0.001	0.000	0.004	0.003	0.001	0.000	0.002
>45	0.000	0.002	0.000	0.004	0.000	0.001	0.000	0.001
Compatibility R <sup>2</sup>	0.99		0.87		0.99		0.85	
Number of relationships (surveys)	389		593		220		108	

Table 3. Comparison of the share of trips related to Warsaw on the basis of SIM card movements, of the updated and base model

Community	Trip share in related to Warsaw			Community	Trip share in related to Warsaw		
	SIM	Updated model	Base model		SIM	Updated model	Base model
Halinów	50%	51%	31%	Otwock	41%	48%	38%
Izabelin	66%	67%	68%	Ożarów Maz. rural area	41%	44%	42%
Jabłonna	36%	42%	51%	Ożarów Maz.- city	37%	38%	20%
Józefów	59%	52%	54%	Piaseczno – city	39%	41%	39%
Karczew	18%	19%	29%	Piaseczno - rural area	36%	37%	32%
Kobyłka	36%	37%	31%	Piastów	49%	51%	41%
Konstancin-city	49%	51%	27%	Pruszków	38%	40%	42%
Konstancin-rural area	40%	44%	59%	Raszyn	56%	58%	37%
Legionowo	38%	36%	59%	Stare Babice	64%	62%	55%
Lesznowola	45%	44%	34%	Sulejówek	69%	68%	55%
Łomianki - city	58%	60%	67%	Wiązowna	50%	52%	32%
Łomianki - rural area	51%	48%	49%	Wołomin -city	35%	37%	29%

Marki	65%	66%	51%	Wołomin - rural area	27%	28%	28%
Michałowice	43%	46%	32%	Ząbki	73%	74%	60%
Nieporęt	52%	55%	51%	Zielonka	54%	57%	36%
Compatibility R <sup>2</sup>						0.96	0.37

The results obtained from the updated model were compared to data on the SIM card movements. The generation shares of individual municipalities were compared in the division into trips within the suburban area and related to Warsaw. For this purpose, the matrix calculated according to the transport model was transformed into a matrix that could be compared to the one containing information on the SIM card movements. The daily matrix was prepared for all the trip purposes calculated in the traffic model aggregated to the municipalities, excluding the internal trips. The comparison of the most important agglomerational municipalities from the perspective of the trip generation share calculated on the basis of the SIM card movements with the results of the adjusted transport model is presented in Table 3. The comparison of the trip distance distribution with the results of the adjusted model is shown in Fig. 8.

Table 4. Comparison of the selected municipalities in terms of the external trip generation share – based on SIM cards and the updated agglomeration transport model

Municipality	All trips		Trips within the suburban area		Trips from the sub. area to Warsaw		The share of the trip to Warsaw	
	SIM	MODEL	SIM	MODEL	SIM	MODEL	SIM	MODEL
Pruszków	5.0%	5.0%	5.2%	5.1%	4.8%	4.9%	38%	40%
Lesznowola	4.7%	4.7%	4.3%	4.5%	5.3%	5.1%	45%	44%
Piaseczno	4.4%	4.8%	4.5%	4.8%	4.4%	4.7%	39%	41%
Michałowice	3.7%	3.3%	3.5%	3.0%	4.0%	3.7%	43%	46%
Ożarów Mazowiecki	3.4%	3.3%	3.3%	3.1%	3.5%	3.5%	41%	44%
Piaseczno	3.4%	3.6%	3.6%	3.8%	3.0%	3.2%	36%	37%
Marki	3.0%	3.1%	1.8%	1.8%	5.0%	5.0%	65%	66%
Legionowo	3.0%	3.5%	3.1%	3.8%	2.9%	3.0%	38%	36%
Ząbki	2.9%	2.8%	1.3%	1.2%	5.3%	5.1%	73%	74%
Raszyn	2.9%	2.8%	2.1%	2.0%	4.1%	3.9%	56%	58%
Other	63.5%	63.2%	67.2%	67.0%	57.8%	57.7%	-	-
All municipalities – compliance R <sup>2</sup>	0.96		0.92		0.99		0.98	

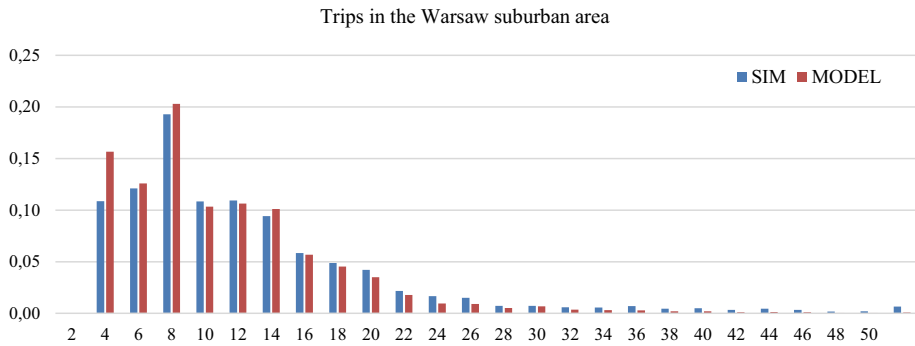


Fig. 8 Comparison of the trip distance distribution [km] in the suburban area based on SIM card movements and the updated transport model

## 5. CONCLUSIONS

In case of the local models developed for large cities, it is very important to include not only the movements within these cities, but also in their related functional areas – in the entire agglomeration area. The movements occurring in the area around the main city and those between the suburban area and this city should be mapped. The BigData (SIM card database) analysis carried out as part of the INMOP 3 project proved the occurrence of strong differentiation in particular agglomeration areas (municipalities) in terms of: total generated traffic volume, structure of performed trips (between the municipalities of the suburban area, to the main city) and spatial trip distribution (connections between particular zones). It indicates the need to build more accurate models of the trip generation in the agglomeration areas, and consequently, more accurate trip models. The use of simplifications, which jointly treat the trip generation within the main city of the agglomeration and the trip generation from this city to the suburban area, as well as jointly the trip generation within the suburban area and between the area and the main city, is insufficient. The studies shows it on the example of the Warsaw agglomeration. The strong differences between the municipalities in the agglomeration indicate that it is justified to build separate models of the trip generation for the trip within the suburban area, between the suburban area and the main city, between the main city and the suburban area as well as within the main city.

This approach raises problems related to access to data. They are often limited only or almost exclusively (e.g. in connection with small statistical samples of surveys) to the main city of the agglomeration. However, as shown in the article, it is possible to use unconventional data sources, e.g. the mobile operators' data on the SIM card movements. These data can be used for detailed

examination of particular areas (municipalities, districts and even smaller territorial units in the agglomeration) in terms of the generated traffic potential, the structure of this movement and the relationships between the areas. As a result, it gives the possibility to increase the accuracy of the traffic generation models, e.g. by introducing calibration coefficients that differentiate individual areas between each other. The unquestionable benefit resulting from the increased accuracy of the agglomeration model is an increase of its usefulness in traffic analyses, especially on the access roads to the agglomeration centre and between the other agglomeration municipalities.

## REFERENCES

1. Rudnicki A., Sarna S.: *Problemy planowania autostrad w dużych miastach*. Mat. XLII Konferencji Naukowej Problemy Naukowo-Badawcze Budownictwa, Krynica 1996.
2. Brzeziński A., Dybicz T., Szymański Ł. Wykorzystanie KMR do modelowania ruchu w miastach na przykładzie Warszawy. Konferencja Naukowo-Techniczna Miasto i Transport 2018, Politechnika Warszawska.
3. Brzeziński A., Dybicz T., Szymański Ł., Włodarek P. Zastosowanie Krajowego Modelu Ruchu w analizie przebiegu dróg szybkiego ruchu przez Warszawę. Problemy komunikacyjne miast warunkach zatłoczenia motoryzacyjnego. *Annały inżynierii ruchu i badań transportowych*, 2017, t.1 (XI)
4. Brzeziński A., Dybicz T., Jesionkiewicz-Niedzińska K., Szymański Ł. Zastosowanie "Big Data do budowy "sezonowych macierzy podróży w Krajowym Modelu Ruchu". *Problemy komunikacyjne miast w warunkach zatłoczenia motoryzacyjnego. Annały inżynierii ruchu i badań transportowych*, 2017, t.1 (XI)
5. Peterson A. The Origin-Destination Matrix Estimation Problem – Analysis and Computations. Linköping Studies in Science and Technology. Dissertations No. 1102. Norrköping 2007 r.
6. *Warszawskie Badanie Ruchu 2015 wraz z opracowaniem modelu ruchu. Raport z etapu IV. Model ruchu*. Warszawa, marzec 2017 r.. PBS sp. z o.o, Politechnika Krakowska, Politechnika Warszawska.

## LIST OF FIGURES AND TABLES:

Fig. 1. Diagrams of the trip models in the agglomeration – simplified (left side) and with an individual approach to municipalities (right side): A, B – suburban area municipalities

Rys. 1. Schematy modeli podróży w aglomeracji – uproszczony (strona lewa) i z indywidualnym podejściem do gmin (strona prawa).

Fig. 2. 10 most important trip generators in the Warsaw suburban area [%], based on the SIM cards data

Rys. 2. Zestawienie 10 najistotniejszych rejonów pod względem udziału w generacji podróży w strefie podwarszawskiej [%] – analiza na podstawie danych z kart SIM

Fig. 3. 10 most important trip generators to Warsaw [%], based on SIM cards data

Rys. 3. Zestawienie 10 najistotniejszych rejonów pod względem udziału w generacji podróży ze strefy do Warszawy [%] – analiza na podstawie danych z kart SIM

Fig. 4. Warsaw suburban area municipalities – share of trips to Warsaw in relation to all the generated trips

Rys. 4. Gminy strefy podwarszawskiej - udział podróży do Warszawy w stosunku do wszystkich generowanych podróży

Fig. 5. Żąbki – percentage structure of external trips to municipalities/districts of the Warsaw agglomeration and spatial traffic distribution – based on SIM cards

Rys. 5. Żąbki – struktura procentowa podróży zewnętrznych do gmin/dzielnic aglomeracji warszawskiej oraz rozkład przestrzenny ruchu – na podstawie kart SIM warszawskiej oraz rozkład przestrzenny ruchu – na podstawie kart SIM

Fig. 7. Błonie – percentage structure of external trips to the municipalities/districts of the Warsaw agglomeration and traffic spatial distribution – based on SIM cards

Rys. 7. Błonie - struktura procentowa podróży zewnętrznych do gmin/dzielnic aglomeracji warszawskiej oraz rozkład przestrzenny ruchu – na podstawie kart SIM

Fig. 8. Comparison of the trip distance distribution in the suburban zone based on SIM card movements and the adjusted model

Rys. 8. Porównanie rozkładów długości podróży w strefie podmiejskiej na podstawie przemieszczeń kart SIM i skorygowanego modelu.

Tab. 1. The rules of determining trips potentials of the suburban area inhabitants

Tab. 1. Zasady wyznaczania dobowych potencjałów ruchotwórczych podróży mieszkańców aglomeracji warszawskiej

Tab. 2. Comparison of the trip distance distribution – suburban area inhabitants

Tab. 2. Porównanie rozkładów długości podróży mieszkańców strefy podwarszawskiej

Tab. 3. Comparison of the share of trips related to Warsaw on the basis of SIM card movements, of the updated and base model (MTAW 2016)

Tab. 3. Porównanie udziału podróży związanych z Warszawą na podstawie przemieszczeń kart SIM, zaktualizowanego i bazowego modelu

Tab. 4. Comparison of the selected municipalities in terms of the external trip generation share – based on SIM cards and the updated agglomeration transport model

Tab. 4. Porównanie wybranych gmin pod względem udziału generacji podróży zewnętrznych – na podstawie kart SIM oraz skorygowanego modelu transportowego

## Model popytu w aglomeracji z wykorzystaniem kart SIM

*Słowa kluczowe: system transportowy, modelowanie podróży, model popytu, generacja podróży, BigData*

### STRESZCZENIE

Program rozwojowy sieci drogowej jak i planowanie i projektowanie układów komunikacyjnych miast i aglomeracji wymaga wykonywania złożonych analiz i prognoz ruchu. Dotyczy to zwłaszcza dróg wyższych klas, także autostrad i dróg ekspresowych, które w obszarach zurbanizowanych obsługują ruch docelowy i

tranzytowy. Istnieje konflikt między potrzebami ruchu na duże odległości w interesie którego leży, aby autostrada przebiegała przez tereny niezabudowane i potrzebami zbliżenia autostrady do potencjalnych celów podróży, których największymi koncentracjami są miasta [1].

Dostrzegając wagę problemu niezbędne jest rozwijanie metodyki badania i analizowania ruchu, a zwłaszcza budowy modeli podróży. Dotychczasowe doświadczenia wskazują, że zwłaszcza w modelach dla aglomeracji stosowane są uproszczone odwzorowania przemieszczeń w strefie aglomeracyjnej (poza głównym miastem) co wynika z niezrozumienia znaczenia tych przemieszczeń dla funkcjonowania całości modelu, bądź wynika z braku danych wejściowych (z badań) pozwalających na zbudowanie wiarygodnych modeli. Artykuł przedstawia wyniki projektu badawczego INMOP 3 w ramach którego podjęto próbę zwiększenia dokładności modelu aglomeracyjnego w części dotyczącej modelu generacji ruchu, dzięki wykorzystaniu BigData – danych operatora telefonii komórkowej o przemieszczeniach kart SIM na obszarze aglomeracji warszawskiej.