WARSAW UNIVERSITY OF TECHNOLOGY Index 351733
FACULTY OF CIVIL ENGINEERING
COMMITTEE FOR CIVIL AND WATER ENGINEERING
POLISH ACADEMY OF SCIENCES ISSN 1230-2945

DOI: 10.24425/ace.2021.136470

ARCHIVES OF CIVIL ENGINEERING

Vol. LXVII ISSUE 1 2021

POSSIBILITY OF BIG DATA APPLICATION FOR OD-MATRIX CALLIBRATION IN TRANSPORT DEMAND MODELS

A. BRZEZIŃSKI[1], T. DYBICZ[2]

Abstract. Modern IT and telecommunications technologies create new possibilities of data acquisition for the needs of traffic analyses and transport planning. At the same time, the current experience suggests that it is becoming increasingly difficult to obtain data on interurban travels of people in a traditional way (among others, in Poland there has been no comprehensive survey of drivers on the sections of non-urban roads since 2006). Within the framework of the INMOP 3 research project, an attempt was made to analyse the use of the Big Data application possibilities including data from SIM cards of the mobile telephony operator [1] and data from probe vehicle data (also known as "floating car data"), as data sources for carrying out the traffic analyses and modelling of travels by all means of transport in Poland. The article presents the manner, in which the data were used, as well as methodological recommendations for creating transport models at the national, regional and local levels. Especially the results of work can be applied for systematic passenger cars trip matrix update.

*Keywords***:** Big Data, traffic modelling, trip matrix, trip distribution, traffic forecast, vehicle probe data

[1] PhD., Eng., Warsaw University of Technology, Faculty of Civil Engineering, Al. Armii Ludowej 16, 00-637 Warsaw, Poland, e-mail: a.brzezinski@il.pw.edu.pl
[2] PhD., Eng., Warsaw University of Technology, Faculty of Civil Engineering, Al. Armii Ludowej 16, 00-637 Warsaw, Poland, e-mail: t.dybicz@il.pw.edu.pl.

1. INTRODUCTION[3]

The demand for the construction of an integrated Intermodal National Traffic Model in Poland covering various transport subsystems (road, railway, air) is likely to be met as a result of the "Principles of the traffic forecasting including other transport means" ["Zasady prognozowania ruchu drogowego z uwzględnieniem innych środków transportu"] research project, implemented in the period from 1 February 2016 to 30 April 2019, commissioned by the National Centre for Research and Development and the General Directorate for Roads and Motorways. The project (with the INMOP 3 acronym) was implemented by the consortium of the universities: The Institute of Roads and Bridges of the Warsaw University of Technology (leader) and Cracow University of Technology (partner) and assume the development of the hierarchical travel modelling and forecasting methodology (national, regional and local models) including the extended data access, i.e. Big Data in addition to data collected in a traditional manner, in measurements.

INMOP 3 also intends to solve a number of specific problems, such as taking into account the seasonality of traffic (changes in traffic in particular seasons and in relation to special periods, holidays), as well as the impact of bottlenecks in the transport network, especially during peak traffic periods (including traffic related to holiday and recreational travels). The project also assumes the preparation of the model updating procedures (database updating), which allow to periodically update the model and its individual elements. A method for updating the OD matrix of passenger cars with the use of the current data from the so-called probe vehicle data (also known as "floating car data"), was proposed. This article is dedicated to this issue.

2. BIG DATA

Modern IT and telecommunications technologies enable the collection and processing of a new type of data from many sources. It applies to: logging on to websites, questions entered into search engines, traffic on the Internet social networking sites, transactions and traffic generated in e-purchasing systems, sent e-mails, movements of mobile telephony users, GPS device movements, mapping services, etc. In relation to the data collected in a traditional way (e.g. on the basis of tests and

[3] Article funded as part of the research work of the Department of Transport Engineering and Surveying "Development trends in transport engineering and surveying" and a research grant for employees of the Warsaw University of Technology supporting scientific activities in the discipline of Civil Engineering and Transport

measurements), their feature is a multitude, and consequently, the need for special processing and analysis [2]. Furthermore, these data can be collected, analysed and made available in the past (collected backwards). In specific cases, the data can be obtained and analysed in real time (e.g. data on vehicle speeds obtained from probe vehicle data in traffic management centres). It is also possible to plan the periods, for which data providers will acquire them, taking into account special requirements, such as taking into consideration a selected special group of users. In 2001, in the Gartner company's report, the first systematisation of the concept of this type of data, describing them as Big Data, was implemented [3]. It was assumed that they are characterised by a three-dimensional structure, which should include: a large scope, fast collection time and a large variety of data types and sources. In 2012, the Big Data definition was extended by the requirement of using special analytical methods and technologies [4].

"Big Data" are used in many kinds of research papers, in medicine, education, public statistics, and for commercial purposes. They are also used in the analyses of movements/flow of people. Such an example includes the project entitled: Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics [5], implemented by the Eurostat consortium, Statistics Estonia and the private Positium company. The main objective of the project was to assess the possibility of using data from mobile networks in order to estimate movements/travels of people in and outside Estonia. The research showed a strong correlation between official data and the data obtained from the mobile networks – e.g. the same variability regarding the seasonality of movements was identified.

3. BIG DATA IN INMOP 3

Within the framework of the INMOP 3 research project, an attempt was made to use the data from the so-called probe vehicle data for the travel modelling. This concerns data about the users of the satellite navigation and GPS systems installed in vehicles, without personal information and actual vehicle identifiers. The data were obtained from over 80,000 fleet vehicles equipped with built-in GPS devices and 275,000 devices and applications for personal navigation. They were acquired and prepared in the form of a matrix of travels between predefined transport zones. The layout of zones adopted in the project was based on the administrative division of the country into counties. It also included road and railway border crossings, seaports and airports.

The OD matrices were constructed on the basis of travels/movements of vehicles between individual counties in Poland, as well as international travels having its beginning or end in Poland and transit journeys through the territory of Poland. In case of international "inbound" travels to Poland, the border crossing or ferry terminal through which the vehicles entered the territory of Poland was treated as a starting point of the travel. It was similar in case of "outbound" travels from Poland, the border crossing or ferry terminal through which the vehicles leaving the territory of Poland was treated as an ending point of the travel.

The OD matrices were created in the division into two types of vehicles: cars and trucks, distinguished in the data collection system based on the supply voltage in the vehicle electrical system, and in the absence of this information, on the basis of the vehicle behaviour analysis or other available data. It was assumed that the group of passenger cars corresponds to passenger cars, and trucks correspond to delivery vans, trucks and lorries with trailers and semi-trailers.

The data from probe vehicle data can be collected in any time intervals. Due to the purpose of the project related to modelling of travels, a 60-minute interval was assumed as the source one. On this basis, it was possible to build matrices for any periods, such as daily, weekly, monthly, seasonal or annual ones. Due to budgetary constraints of the project, the data were collected for a total of 6 weeks of 2016. The data covered 42 days during which over 11 million travels/passing of passenger cars (Table 1) were recorded, which gives an average of 280,000 travels/passing of passenger cars for one day.

Table 1. The scope of data from the probe vehicle data of vehicles used in the INMOP 3 research project

Measurement week (7 days) in 2016	**Periods for which a travel database was obtained**	**Number of travels by passenger cars**
winter period	1 - 7 February	1,698,079
spring period	18 - 24 April	2,015,380
long weekend in May	28 April - 4 May	1,682,147
summer period	18 - 24 July	2,208,405
long weekend in August	11 - 17 August	1,919,324
autumn period	14 - 20 November	2,153,052
	In total	11,676,387

It was assumed that the adopted analysis periods will make it possible to test the travel characteristics in various periods of the year (in seasons). In total, taking into account the data on passenger cars and heavy vehicles, a database on 13 million of travels/passing was collected, 4 million of which are travels/passing made in two special weeks (with long May and August weekends).

These data can be aggregated to daily matrices [7]. Fig. 1 presents an example of the distribution of the potential of regions for selected hours during the day, on the selected spring day, for passenger cars. The data collected from probe vehicle data also included the vehicle speed data on individual sections of the national and regional road network.

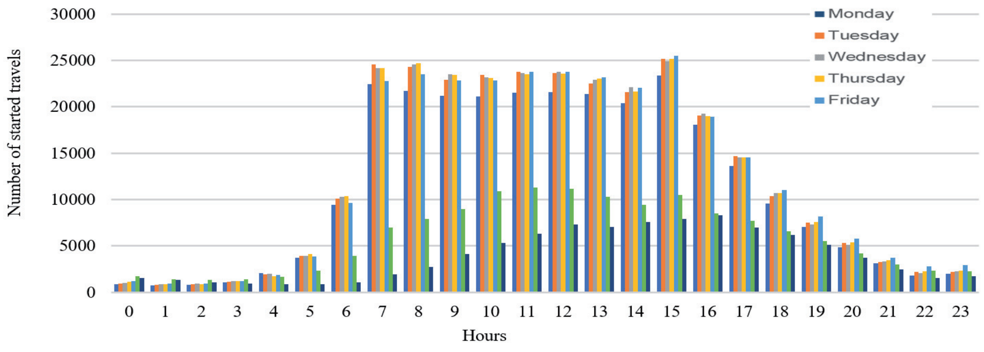

Fig. 1. Distribution of the number of travels (started) by passenger cars within 24 hours on a selected winter day in 2016. Source: analysis of data obtained from probe vehicle data

It also creates other opportunities of using data from probe vehicle data, e.g. in the context of determining the speed of vehicles on the transport network, or detection of overloads and location of bottlenecks [7]. An example of the test carried out in 2016 along the expressway (S2-A2) between the Pruszków tolling station and the Bemowo interchange was shown in Figure 2. The diagram shows the speed distribution of vehicles over time (on the vertical axis, subsequent hours from the morning to the afternoon) along the tested section. The green colour determines free traffic conditions, while the range and time of occurring disturbances are determined by the range from yellow to red. These data make it possible to analyse the impact and its extent of the bottleneck that is on this section of the Pruszków tolling station.

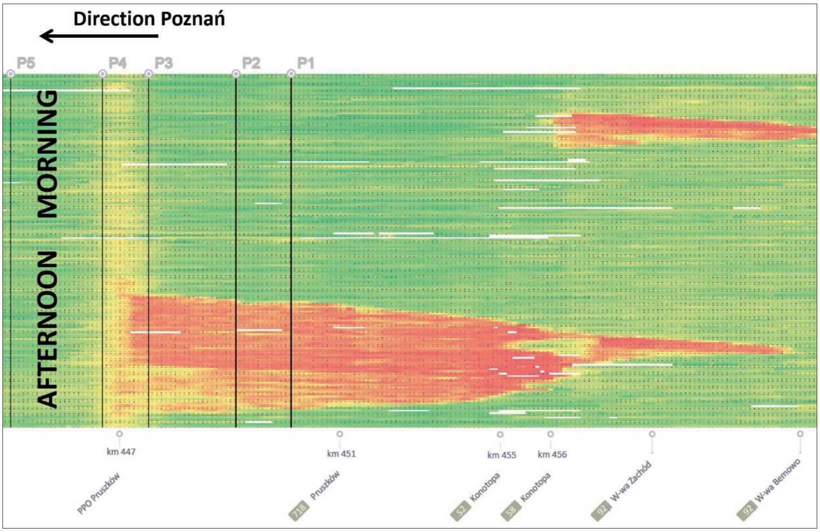

Fig. 2. Distribution of speed in time (vertical axis) and at the distance (section) for bottleneck associated with an inactive tolling station on the A2 motorway near the Pruszków area. Source: analysis of data obtained from probe vehicle data.

4. CONDITION FOR USING BIG DATA

At the stage of obtaining probe vehicle data and after their initial analysis, a number of problems related to the use of Big Data was identified. The most important of them are related to: the necessity of data anonymisation, the criteria for defining travels, the lack of the possibility to determine the travel motivation and the method of extending data to the entire population.

According to the applicable law, the obtained data must be anonymised. In case of data from GPS, it eliminates from the database those records that relate to the travels during which the fleet operator changed the vehicle identifier and the route with a small number of travels (less than three travels in a given time interval were treated as impossible to be anonymised and introduced to the obtained database). These problems were partly solved by using longer time intervals.

The use of data from probe vehicle data poses a problem with clarifying the definition of travel. In this case, it is not possible to clearly determine the source and purpose, e.g. on the basis of a direct conversation with the user of the transport system. It is necessary to interpret a fairly complex numerical record of the travel, which may contain a number of movement and stoppage sections. By analysing the available database from GPS devices, the following definition of travel was adopted:

a) minimum travel distance: 1,000 metres,

b) the maximum stoppage time not interrupting the travel is a variable value which increases with the travel time; limit values: 3 min for t = 0 min, 60 min for t >= 4 h, where t is the travel time from the beginning of the travel proportionally reduced to the used stoppage time,

c) in case of the stoppage in a car park having the nature of a rest area near the motorway, expressway or important national road, the permitted stoppage time was extended to 5 hours.

The database analysis also revealed a number of phenomena requiring a special interpretation. For example, the occurrence of the travel from the x region and to the x region, in the course of which the departure from the x region to the y region was recorded. In such situations, such a travel can be treated as internal (in the x region) or as two travels x-y and y-x. The second interpretation was adopted in the research work.

Regardless of the multitude of data obtained as Big Data, they represent only a part of the population. The problems with the method for the test extension are known and described in the literature, e.g. [6] and [Daas et al. 2015]. In the Netherlands, (bulk data obtained from induction loops installed in surfaces) in order to ensure the representativeness, the data have been overweighed taking into account information about the location of detectors, road construction (distribution of connections) and the length of sections between the successive measurements. As part of INMOP 3, it was decided to use the current data stored in the OD matrices of the General Directorate for Roads and Motorways, as described below.

5. MATRIX UPDATE METHOD

Within the framework of the INMOP 3 project, a method for updating the OD matrix of passenger cars, which is used in the National Traffic Model applied so far by the General Directorate for Roads and Motorways, was proposed. Originally this matrix was created in 2007 by submitting information on the real relations of the traffic obtained from the roadside origin-destination surveys supported with the theoretical traffic structure developed with a gravity model.

In subsequent years, the surveys for the needs of the national OD matrix ceased to be carried out, and the matrix was verified and updated based only on the available results of traffic volume measurements, especially the results of the General Traffic Census (GPR2010 and GPR2015). By taking into account the changes occurring in the land use, transport network and transport behaviours, the reliability of this method decreased from year to year, in particular in terms of the ability to map the spatial distribution of travels and their distance.

The chance of at least partial replacing of the traffic survey results to update the OD matrix is created by Big Data, in particular probe vehicle data. By analysing the data obtained within the framework of the INMOP 3 project, it was noted that in a large test, they allow for an in-depth analysis of travels by passenger cars across the country, between any defined zones (e.g. in the layout of districts, communes and counties). This allows to analyse the directions and travel distances, taking into account the variability of behaviours depending on the days of the week or season.

From the combination of seasonal passenger car OD/passing matrices obtained from probe vehicle data (Figure 3), the relationship that takes into account the number of weeks of the occurrence of a given season among all weeks of the year was developed.

The following relationship was used to connect seasonal matrices:

(5.1) M_SDR_SO_PVD = (M_Winter_SO_PVD*12/48 + M_Spring_SO_PVD*13/48 + M_Mayweekend_SO_PVD*1/48 + M_Summer_SO*7/48 + M_Augustweekend_SO_PVD*1/48 + M_Autumn_SO_PVD*14/48)/7

where:

M_Winter_SO_probe vehicle data – weekly matrix of travels by passenger cars obtained from probe vehicle data for the winter period,

M_Spring_SO_probe vehicle data – weekly matrix of travels by passenger cars obtained from probe vehicle data for the spring period,

M_Mayweekend_SO_PVD – weekly matrix of travels by passenger cars obtained from probe vehicle data for the week in which the long May weekend occurred,

M_Summer_SO_PVD – weekly matrix of travels by passenger cars obtained from probe vehicle data for the summer period,

M_Augustweekend_SO_PVD – weekly matrix of travels by passenger cars obtained from probe vehicle data for the week in which the long August weekend occurred,

M_Autumn_SO_PVD – weekly matrix of travels by passenger cars obtained from probe vehicle data for the autumn period.

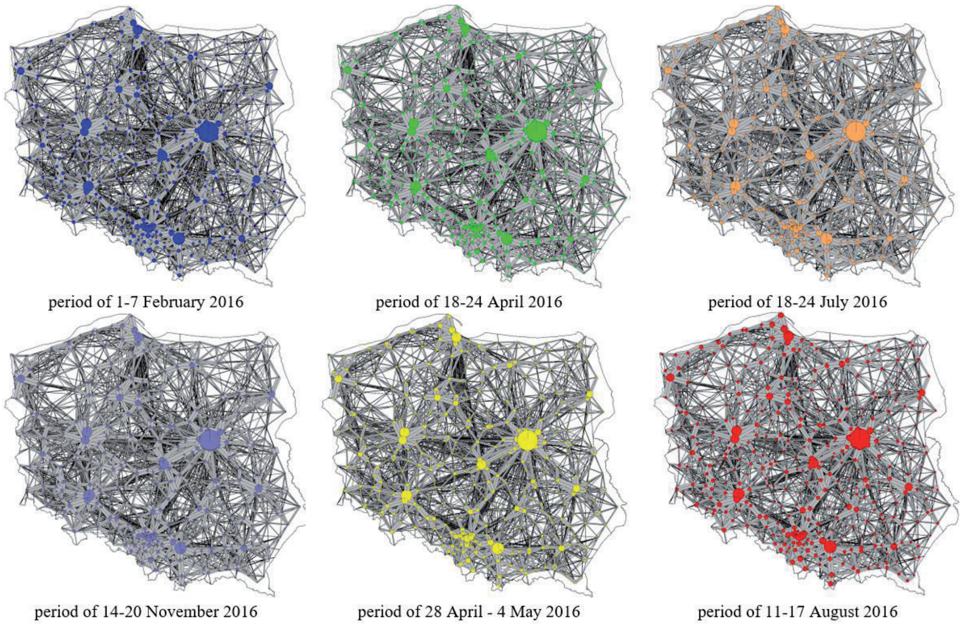

Fig. 3. The structures of travels by passenger cars in the periods of 2016, Source: analysis of data obtained from probe vehicle data.

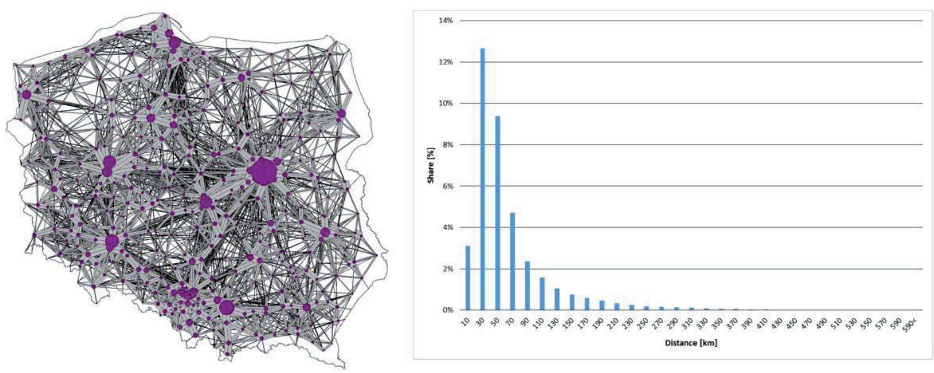

Fig. 4. Distribution of travels by passenger cars on the average day of the year and the distribution of the travel distances for passenger cars developed on the basis of the combination of seasonal structures of 2016, Source: analysis of data obtained from probe vehicle data.

However, probe vehicle data can be used as one of the elements of the matrix update process. Similarly, as in case of the matrix built in a traditional way, based on the origin-destination survey, using the data from probe vehicle data, it is not possible to build a complete interzone (county) matrix. With fairly large fragmentation of zones, not all of travels are represented with the sufficient number in the database including movement of vehicles data. Therefore, at the outset, a reasonable approach is to aggregate the zones into macro-zones and to determine the macro-zones potentials of traffic generation and absorption. Due to the number of movements in the network on one day, the differentiation of individual zones (counties), in terms of the potentials of traffic generation and the differentiation of traffic spatial distribution, the zones were connected in their corresponding 73 macro-zones (subregions) [8].

Then, in the system of macro-zones, using the data from probe vehicle data, 6 seasonal OD matrices corresponding to particular periods of the year were built, and on their basis, the average daily annual matrix of the movement of passenger cars was created. It resulted in the spatial distribution of travels by passenger cars between subregions of Poland and a reliable, actual histogram of the travel distance, as well as the determination of the average travel distance.

In this way, two matrices of travels by passenger cars, aggregated into a system of 73 macro-zones – subregions (matrix size 73x73), were obtained:

a) aggregated full matrix of the General Directorate of Roads and Motorways calibrated to the General Traffic Measurement 2015, but with the current spatial distribution of the travel recorded in it (Figure 5),

b) aggregated matrix developed on the basis of data from probe vehicle data, which maps the current spatial distribution of the travel (of 2016) and the travel distance, but built on a travel test (Figure 6).

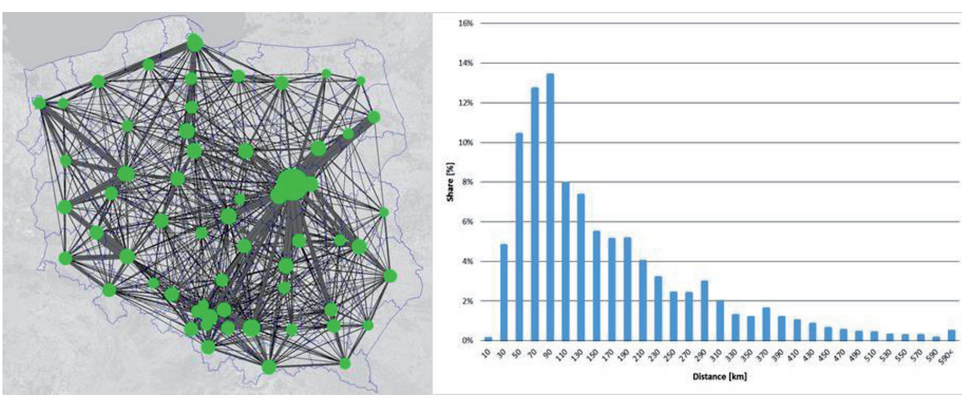

Fig. 5. Distribution of travels by passenger cars aggregated into subregions based on the matrix of the General Directorate of Roads and Motorways, Source: analysis of data obtained from the General Directorate of Roads and Motorways.

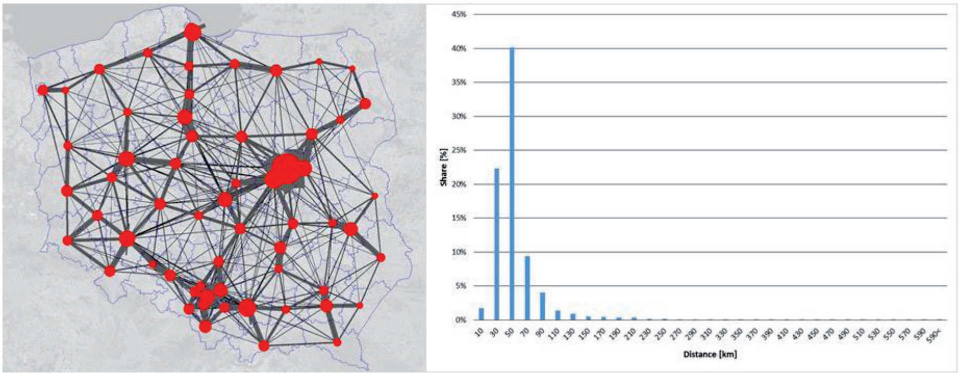

Fig. 6. Distribution of travels by passenger cars aggregated into subregions based on probe vehicle data, Source: analysis of data obtained from probe vehicle data.

The assumption was made that at this stage, the current matrix of the General Directorate of Roads and Motorways will be used in order to determine:

- macro-zone potentials (subregions) in terms of the volume of generated traffic of passenger cars,
- differentiation indices of the potentials of zones within macro-zones (subregions);

and thus to expand the matrix from probe vehicle data (Figure 7), and then to disaggregate it to the system of zones.

The matrix constructed in this way requires calibration. The method of successive iterations should be led to the compatibility of the histogram of the travel distance of the resulting matrix with the histogram of the travel distance developed on the basis of data from probe vehicle data. It may be related to the use of correction factors, individual zones potentials, and thus the introduction of the correction of potentials recorded in the current matrix of the General Directorate of Roads and Motorways.

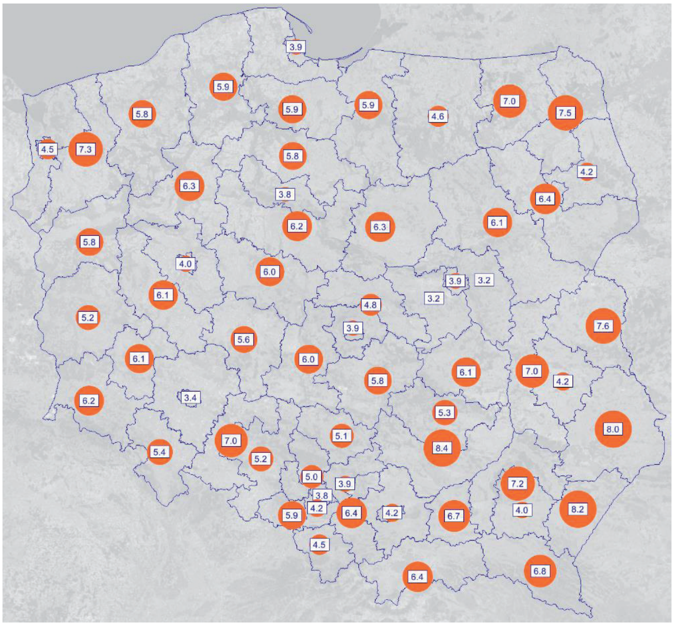

Fig. 7. Expanding rates of matrix obtained from probe vehicle data for particular subregions, Source: analysis of data obtained from the General Directorate of Roads and Motorways.

In the next steps, just after inclusion of other matrices of traffic of passenger cars (e.g. for international traffic), another calibration step is possible, i.e. comparison of total transport works expressed in vehicle kilometres, which are performed on the road network by passenger cars with statistical data. Since the network model contains only national and regional roads, then, the checksums must include transport works implemented on these roads. The sources of data on the transport work on the road network are [9] and [10].

The results obtained from these two sources are different. The data from GPR2015 only cover those sections on which the traffic measurement was performed. It is not performed on a part of the roads in urban areas. However, the Central Statistical Office report includes all the roads for which the average network load values resulting from the traffic measurement are available. In the first case, it is the sum of the products of the traffic flow at the section and section length, and in the second one, the product of the average traffic flow for a given category of roads and the total length of these roads according to the statistics of the Central Statistical Office. The values that should be obtained in the model should be between these two values. It should also be noted that the OD matrices obtained at this stage should not be considered as final. Their checking and possible correction must take place at the stage of the network analyses in the traffic distribution procedure on the road network. Therefore, in the matrix calibration process, one should strive for obtaining the results of the control values close to those available from the measurements and statistics, but not necessarily identical ones.

The figure below shows the resulting interregional matrix of travels by passenger cars and the resulting histogram of the travel distance from the model OD matrix of passenger cars (Figure 8).

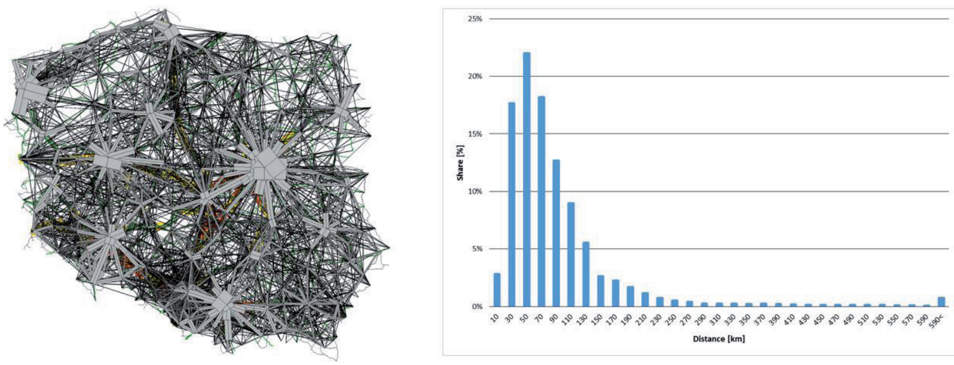

Fig. 8. The resulting inter zone distribution of travels by passenger cars and the resulting histogram of the travel distance in 2016, Source: analysis of data obtained from probe vehicle data.

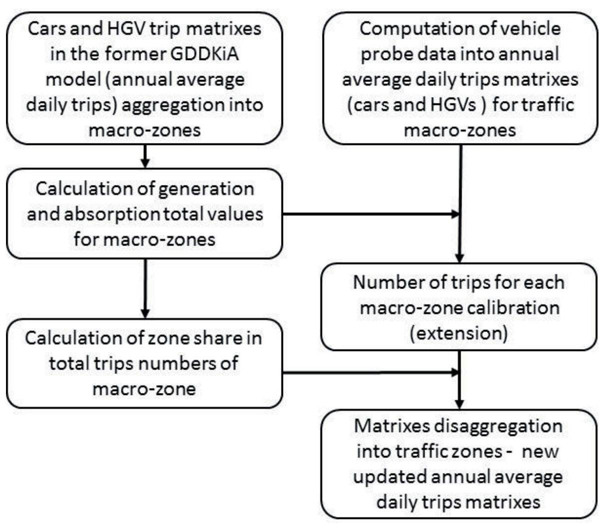

Fig. 9. Procedure of cars and HGVs trip matrixes updating.

6. CONCLUSIONS

The previous experience of the INMOP 3 research project and the carried out analyses of the possibility of using Big Data in modelling the travels are proved to be useful. The article presents the results of the possibilities of using probe vehicle data to update the OD matrix. In a situation where no reliable research of transport behaviours on the road network has been carried out for years, the possibility of using new data sources to improve the quality of the used models has been confirmed. In the construction process, and then in the calibration of the OD matrix, it was managed to obtain new actual spatial distribution of travels by passenger cars, with very high compatibility of the travel distance histograms.

An unquestionable advantage of using this type of data is the automation of information collection and thanks to that limitation of mistakes made by a human (e.g. an interviewer), the multitude of available data, which allows to reliably reproduce the spatial distribution of travels, travel distances and speeds, taking into account the seasonality of traffic and changes in the transport behaviours. These aspects in the current models were usually omitted due to the time of obtaining the measurements (usually in spring or early autumn). The possibility of carrying out the matrix changes systematically, e.g. once per year, and thanks to that limiting one of the basic disadvantages of the currently used model, its out-of-date is also invaluable.

A full transition in the matrix construction to Big Data also requires overcoming the problem resulting from the simplifications associated with the impossibility of assigning motivation to travels. It raises the problem at the stage of the travel distribution to the road network, when choosing the travel paths. The solution may include the use of separate gravity models for the travel groups in the travel distance groups, while assigning various travel motivations to these groups. However, it is also possible that one cannot lose the sight of the possibility of carrying out periodic transport tests (surveys on the transport network and in households), e.g. every 5 years, treating their results as base data for building the matrices (including the travel motivation), and Big Data (in this case from probe vehicle data) to update them in shorter periods (e.g. once per year).

REFERENCES

1. Brzeziński A., Dybicz T., Szymański Ł.: Demand Model in the Agglomeration using Sim Cards, w: Archives of Civil Engineering, vol. 65, nr 1, 2019, ss. 143-156, DOI:10.2478/ace-2019-0010
2. M. Beręsewicz, M. Szymkowiak: Big data w statystyce publicznej- nadzieje, osiągnięcia, wyzwania i zagrożenia, Ekonometria 2(**48**) : 9-22, 2015.
3. L. Douglas, "3D Data Management: Controlling Data Volume, Velocity and Variety". Gartner. February (2001)

4. A. De Mauro, M. Greco, M. Grimaldi, "A Formal definition of Big Data based on its essential Features". Library Review. **65**: 122–135. doi:10.1108/LR-06-2015-0061, (2016)
5. Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics Consolidated Report Eurostat Contract No 30501.2012.001-2012.452; Eurostat (2014)
6. A. Brzeziński, T. Dybicz, K. Jesionkiewicz-Niedzińska, Ł. Szymański, 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, t.1 (**XI**), (2017)
7. A. Brzeziński, T. Dybicz, Ł, Szymański, P. Włodarek, Innowacyjne podejście do budowy Krajowego Modelu Ruchu". Problemy komunikacyjne miast w warunkach zatłoczenia motoryzacyjnego. Annały inżynierii ruchu i badań transportowych, t.1 (XI), (2017)
8. Klasyfikacja NUTS3 wg. http://stat.gov.pl/statystyka-regionalna/jednostki-terytorialne/klasyfikacja-nuts/klasyfikacja-nuts-w-polsce/, stan wg 1 stycznia 2018 r.
9. Generalny Pomiar Ruchu 2015 w którym dla odcinków sieci podany jest SDRD, a dla każdego odcinka pomiarowego podana jest jego długość. Zbiór taki jest dostępny dla obydwu typów dróg i w prosty sposób można uzyskać sumaryczną pracę przewozową.
10. Raport GUS o wielkości pracy przewozowej na sieci dróg z roku 2015.

LIST OF FIGURES AND TABLES

Rys. 5. Rozkład przestrzenny podróży odbywanych samochodami osobowymi zagregowany do podregionów opracowany na podstawie macierzy podróży z Generalnej Dyrekcji Dróg Krajowych i Autostrad. Zródło: opracowanie własne na podstawie danych z Generalnej Dyrekcji Dróg Krajowych i Autostrad.

Fig. 6. Distribution of travels by passenger cars aggregated into subregions based on probe vehicle data, Source: own development based on data obtained from probe vehicle data. Source: own development based on data obtained from probe vehicle data.

Rys. 6. Rozkład przestrzenny podróży odbywanych samochodami osobowymi zagregowany do podregionów opracowany na podstawie danych z sondowania pojazdow. Źródło: analizy własne oparte o dane z sondowania pojazdów.

Fig. 7. Expanding rates of matrix obtained from probe vehicle data for particular subregions, Source: own development based on data obtained from the General Directorate of Roads and Motorways.

Rys. 7. Wskaźniki rozszerzenia macierzy opracowanej na podstawie danych z sondowania pojazdów w układzie podregionów. Żódło: opracowanie własne na podstawie danych Geenralnej Dyrekcji Dróg Krajowych i Autostrad.

Fig. 8. The resulting inter zone distribution of travels by passenger cars and the resulting histogram of the travel distance in 2016, Source: own development based on data obtained from probe vehicle data.

Rys. 8. Wynikowy rozkład przestrzenny podróży odbywanych samochodami osobowymi pomiędzy rejonami komunikacyjnymi I wynikowy histogram długości podróży w roku 2016. Żródło: opracowanie własne na podstawie danych z sondowania pojazdów.

Fig. 9. Procedure of cars and HGVs trip matrixes updating.

Rys. 9. Procedura aktualizacji macierzy samochodów osobowych i towarowych

Table 1. The scope of data from the probe vehicle data of vehicles used in the INMOP 3 research project.

Tabela 1. Zakres danych z sondowania pojazdów wykorzystanych w projekcie badawczym INMOP 3.

MOŻLIWOŚĆ ZASTOSOWANIA BIG DATA DO KALIBRACJI MACIERZY ŹRÓDŁO-CEL W MODELACH PODRÓŻY

Słowa kluczowe: Big Data, modelowanie podróży, macierze podróży, rozkład przestrzenny ruchu, prognozy ruchu, sondowanie pojazdów.

STRESZCZENIE:

Współczesne technologie informatyczne oraz telekomunikacyjne tworzą nowe możliwości pozyskiwania danych na potrzeby analiz ruchu i modelowania systemów transportu. Równocześnie dotychczasowe doświadczenia wskazują, że coraz trudniej jest pozyskiwać dane o międzymiastowych podróżach osób w sposób tradycyjny (między innymi od 2006r. nie odbyło się żadne kompleksowe badanie ankietowe kierowców na odcinkach dróg zamiejskich). Prowadzone są jedynie wyrywkowe badania w małej skali do tego w sposób niesystematyczny. Tworzy to problemy lub wręcz uniemożliwia wykonywanie systematycznej aktualizacji modeli systemów transportowych, które są stosowane do analiz i prognozowania ruchu. Dotyczy to także Krajowego Modelu Ruchu Generalnej Dyrekcji Dróg Krajowych i Autostrad

stosowanego praktycznie we wszystkich największych projektach drogowych w Polsce, zwłaszcza w budowie autostrad i dróg ekspresowych.

Dane typu Big Data w stosunku do danych gromadzonych w sposób tradycyjny (np. na podstawie badań i pomiarów) zbierane są na skalę hurtową. Mogą być gromadzone, analizowane i udostępniane w czasie przeszłym (pozyskane wstecz). W szczególnych przypadkach mogą być pozyskiwane i analizowane w czasie rzeczywistym (np. dane o prędkościach pojazdów pozyskiwane z sondowania pojazdów w centrach zarządzania ruchem). Możliwe jest również zaplanowanie okresów dla których dostawcy danych będą je pozyskiwać z uwzględnieniem specjalnych wymogów, jak np. uwzględnienie wybranej szczególnej grupy użytkowników

W ramach projektu badawczego INMOP 3 podjęto próbę wykorzystania do modelowania podróży danych z tzw. sondowania pojazdów, czyli danych o użytkownikach nawigacji satelitarnej i systemów GPS zamontowanych w pojazdach, bez informacji osobowych i rzeczywistych identyfikatorów pojazdów. Pozyskano dane z ponad 80 tys. pojazdów flotowych wyposażonych we wbudowane urządzenia GPS oraz 275 tys. urządzeń i aplikacji do nawigacji osobistej. Ujęto je w formie macierzy podróży odbywanych pomiędzy rejonami komunikacyjnymi. Przyjęty w projekcie układ rejonów komunikacyjnych opiera się o administracyjny podział kraju na powiaty. Uwzględniono w nim również przejścia graniczne drogowe i kolejowe, porty morskie i lotniska.

Analizując pozyskane dane zauważono, że umożliwiają one na dużej próbie, pogłębioną analizę podróży odbywanych samochodami osobowymi w skali całego kraju, pomiędzy dowolnie określonymi rejonami komunikacyjnymi (np. w układzie dzielnic, gmin, powiatów), a zatem umożliwiają analizę kierunków i długości podróży, z uwzględnieniem zmienności zachowań w zależności od dni tygodnia czy pory roku (tzw. sezonowość ruchu).

Jednak dane z sondowania pojazdów mogą być wykorzystane tylko jako jeden z elementów procesu aktualizowania macierzy. Podobnie jak w przypadku macierzy budowanej w sposób tradycyjny, na podstawie badań źródło-cel, wykorzystując dane z sondowania pojazdów nie jest możliwe zbudowanie kompletnej macierzy międzyrejonowej (międzypowiatowej). Przy dość dużym rozdrobnieniu rejonów, nie wszystkie z nich są reprezentowane z wystarczającą liczebnością w bazie danych o przemieszczeniach pojazdów. Dlatego też na wstępie rozsądnym podejściem jest agregacja rejonów do stref i wyznaczenie strefowych potencjałów generacji i absorpcji ruchu (np. w układzie 73 podregionów statystycznych).

Następnie już w układzie stref, wykorzystując dane z sondowania pojazdów, zbudowano 6 sezonowych macierzy podróży odpowiadających poszczególnym okresom roku, a na ich podstawie średniodobową roczną macierz ruchu samochodów osobowych. Uzyskano w ten sposób aktualny rozkład przestrzenny podróży samochodami osobowymi pomiędzy podregionami Polski oraz miarodajny, rzeczywisty histogram długości podróży oraz średnie długości podróży. Powstały w ten sposób dwie macierze podróży samochodami osobowymi, zagregowane do układu 73 stref - podregionów (rozmiar macierzy 73x73):

- zagregowana pełna macierz GDDKiA skalibrowana do GPR 2015, ale z utrwalonym w niej dotychczasowym rozkładem przestrzennym podróży,
- zagregowana macierz opracowana na podstawie danych z sondowania pojazdów, odwzorowująca aktualny rozkład przestrzenny podróży (z roku 2016) oraz długości podróży, ale zbudowaną na próbie podróży.

Przyjęto założenie, że na tym etapie dotychczasowa macierz GDDKiA zostanie wykorzystana do określenia:

- potencjałów stref (podregionów) w zakresie wielkości generowanego ruchu samochodów osobowych,
- wskaźników zróżnicowania potencjałów rejonów w obrębie stref (podregionów).

i do rozszerzenia macierzy z sondowania pojazdów, a następnie jej dezagregacji do układu rejonów.

Tak zbudowana macierz wymaga kalibracji. Metodą kolejnych iteracji należy doprowadzić do zgodności histogramu długości podróży macierzy wynikowej z histogramem długości podróży opracowanym na podstawie danych z sondowania pojazdów. Może wiązać się to ze stosowaniem współczynników korygujących poszczególne potencjały rejonowe, a tym samym wprowadzaniem korekty potencjałów zapisanych w dotychczasowej macierzy GDDKiA.

W kolejnych krokach, już po dołączeniu innych macierzy ruchu samochodów osobowych (np. dla ruchu międzynarodowego) możliwy jest kolejny krok kalibracyjny, tj. porównanie sumarycznych prac przewozowych wyrażonych w pojazdokilometrach, jakie są wykonywane na sieci drogowej przez pojazdy osobowe z danymi statystycznymi. Ponieważ model sieci zawiera tylko drogi krajowe i wojewódzkie, sumami kontrolnymi muszą być wówczas prace przewozowe wykonane na tych drogach.

Dotychczasowe doświadczenia z realizacji projektu badawczego INMOP 3 i przeprowadzone analizy możliwości wykorzystania Big Data w modelowaniu podróży okazują się obiecujące. Niewątpliwą zaletą wykorzystywania tego typu danych jest automatyzacja zbierania informacji równoznaczna z ograniczeniem błędów popełnianych przez człowieka (np. ankietera), masowość dostępnych danych, pozwalająca wiarygodnie odtworzyć rozkład przestrzenny podróży, odległości i prędkości podróży oraz uwzględnić sezonowość ruchu i zmiany zachowań komunikacyjnych. Te aspekty w dotychczasowych modelach były zwykle pomijane, ze względu na czas sprowadzenia pomiarów (zwykle wiosną lub wczesną jesienią. Nie do przecenienia jest także możliwość przeprowadzania zmian macierzy systematycznie, np. raz w roku, ograniczając w ten sposób jeden z podstawowych mankamentów dotychczas stosowanego modelu, jego nieaktualność.

Pełne przejście w budowie macierzy na dane typu Big Data wymaga także pokonania problemu wynikającego z uproszczeniami związanymi z brakiem możliwości przypisania motywacji do odbywanych podróży. Wywołuje to problem na etapie dystrybucji podróży na sieć transportową, przy wyborze ścieżek podróży. Rozwiązaniem może być stosowanie odrębnych modeli grawitacyjnych dla grup podróży w grupach odległości podróży, przypisując niejaki tym grupom różne motywacje podróży.

Received: 02.07.2020, Revised: 17.10.2020