WARSAW UNIVERSITY OF TECHNOLOGY Index 351733	DOI: 10.24425/ace.2022.141876
FACULTY OF CIVIL ENGINEERING COMMITTEE FOR CIVIL AND WATER ENGINEERING	ARCHIVES OF CIVIL ENGINEERING
POLISH ACADEMY OF SCIENCES ISSN 1230-2945	Vol. LXVIII ISSUE 3 2022

pp. 107 – 123

Research paper

Construction of the passenger rail traffic generation model

Andrzej Brzeziński[1], Andrzej Waltz[2]

Abstract: The article presents a new approach to building a passenger rail traffic generation model. It uses data on the number of passengers at stations and railway stops obtained from the databases of operators on the rail transport market through the Office of Rail Transport – market regulator – combined with data on the model of the area around the station built based on population, number of beds, individual motorization and gross domestic product (GDP). This enabled analyzing the potential of railway traffic generation at a very detailed level. The article presents a methodology for building a passenger rail traffic generation model and verification of this model based on limited variables describing railway stations and stops as well as traffic zones and available statistical data. The model takes into account three segments of the railway market: regional, interregional and inter-agglomeration transport. The results of these analyzes can be used to increase the accuracy and the reliability of rail traffic models used in the analysis of transport networks.

Keywords: railway traffic, traffic generation model, regression analysis, traffic forecasting, passenger traffic

[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, ORCID: 0000-0002-1256-6007

[2] PhD., Eng., independent consultant, e-mail: a.waltz@poczta.onet.pl, ORCID: 0000-0002-7566-3381

1. Introduction

The experience with the use of traffic models in Poland shows that they are most often built for the needs of road administration, large cities and potentially administrative regions. There is smaller experience related to the construction of models for the analysis of forecasted railway traffic. An exception is the model built for the needs of analyses related to the state investment project under the name of The Central Communication Port [9, 10]. Relatively little experience with rail traffic modelling translates into poor scientific achievements in articles and papers presented at conferences and seminars. Taking into account the implemented significant investment program aimed at the development of the railway infrastructure, strengthening the methodology of modeling rail travel is what we need at the moment.

Relatively little experience with railway models results, among others, from the restrictions on access to railway data, especially railway passenger surveys and ticket statistics, and restrictions on the availability of models built for railway companies. In the project "Principles of road traffic forecasting taking into account other means of transport" [2–4], financed under the project called Road Innovation Development by the National Center for Research and Development and the General Directorate for National Roads and Motorways, a methodology was created regarding integrated travel modeling and traffic forecasting on road and rail networks. The experience resulting from this project also confirms the limited availability of data necessary for the construction of railway models [15]. In practice, when building models, it is possible to use only public information from railway companies and the Polish Central Statistical Office (usually in the form of PDF files) and a map of the railway network, which is mainly useful when building a model of the railway network (table of objects on railway lines, regulations for assigning train paths, passenger train timetables, OBLIKO database). Such data can only be partially used for construction and calibration of the travel matrix (based on general data on transport work and medium-distance journeys).

The lack of access to data reduces the chances of building good quality railway models. Experience from other countries [6, 12, 16] shows that modeling rail traffic uses very detailed and real information from surveys and/or ticket sales databases.

In traffic generation models is taken into account that basic factors determining the number of passengers in rail transport are:

– transport offer in specific O-D relations,
– attractiveness) of the railway transport offer in relation to alternative modes of transport (bus, car).

The choice of the modal split, and thus the number of passengers, is determined by the relative utility of this modes of transport [8].

As a result, travel distribution models fall into two main types based on aggregated data, for example, ticket sales information and models based on survey results using data disaggregated at the level of interviews with system users. An example of the aggregated model is the Wardman model [13] which uses such variables as: population, GDP, motorization rate and total travel cost. In the Wardman model, the most important factor influencing the increase in the number of rail crossings is GDP. When disaggregated data (user surveys) [1] is available, it is possible to take into account personal characteristics (such as: age, gender, income, etc.), transport service characteristics and information on sources and destinations. Such models are more demanding at the data collecting and processing stage, hence, they are also more expensive.

In recent years, Poland started to publish data on passenger rail traffic which was not available at the time of the creation of the National Traffic Model developed as part of the research [5]. Currently, the data allows for the development of new variants of traffic generation models based on real data. Data on passenger rail traffic at the stops and stations level significantly improves the quality of the generation models describing this traffic.

Therefore, the passenger rail traffic generation model described in the article is the result of the continuous work on modeling rail traffic, undertaken in research [5, 17] using the available data on supply and demand in the rail market.

2. Estimating the number of rail passengers on the basis of data

In 2017, the first study [11] was prepared at the Railway Transport Office (Urząd Transportu Kolejowego, UTK), which estimated the number of passengers using stations and stops located on the entire railway network in Poland (Fig. 1. This data was then updated in 2018 and 2019. Passenger exchange data was collected from rail operators and mainly sourced from their ticketing systems. Only in the case of Szybka Kolej Miejska in Warsaw and Łódzka Kolej Aglomeracyjna, the source were the meters in vehicles. In fact, the exchange of passengers at stations and stops differs from the data from sales systems.

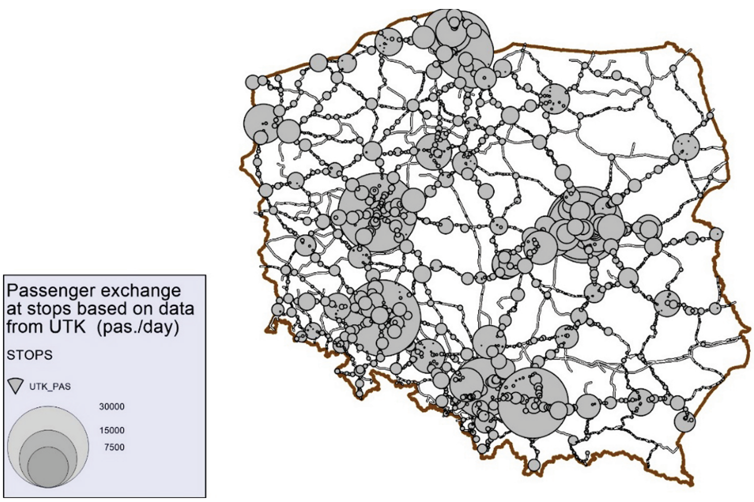

Fig. 1. The volume of passenger exchange at stations and stops based on data from carriers' ticketing systems (2017–2019). *Source:* UTK, VISUM traffic model

This is due to, inter alia, differences related to zone and section tickets used in metropolitan transport valid to the end station in a given zone. Therefore, zone offers and non-relational tickets make it difficult to properly identify passenger movement. It is then necessary to estimate the number of passengers in relation to number provided by carriers, e.g. on the basis of data from meters in vehicles, in proportion to the passenger exchange indicated on the basis of relational tickets.

On the basis of data from railway carriers in Poland, the amount of passenger exchange is obtained, determining the number of people getting on and off the trains, excluding people getting on and off outside Poland. In addition, in the case of stations or stops where trains of more than one carrier stop, with an average daily passenger exchange of more than 1000 passengers, general data was rounded to 100. In the case of stations with an average daily exchange of more than 1000 passengers where trains of one carrier stop or the share of one carrier is nearly 100% and stations with an average daily exchange of less than 1000 passengers, general data is provided in value ranges. For the purpose of building railway traffic generation models, the average value for the enre 2017–2020 period was used. It was taken as the base value for further analyzes. Data for 2020 was not used due to their unusual character due to the pandemic.

Data from ticket systems contain the total number of passengers. Determining the structure of passengers traveling in inter-agglomeration, interregional and regional traffic requires linking the ticket data with the number of trains of individual categories stopping at stations and stops. For this purpose, a simple model can be used to define the share of individual types of trips in the total trip using the formulas presented below. Coefficient 0.64 results from calibration at the level of conformity of the number of passengers for three types of trains with statistical data.

$$U_{\text{reg}} = T_{\text{reg}}/(T_{\text{reg}} + \text{POV}_{\text{mreg}} + 0.64 \cdot T_{\text{iagg}}) \tag{2.1}$$

$$U_{\text{ireg}} = T_{\text{ireg}}/(T_{\text{reg}} + T_{\text{ireg}} + 0.64 \cdot T_{\text{iagg}}) \tag{2.2}$$

$$U_{\text{iagg}} = 0.64 \cdot T_{\text{iagg}}/(T_{\text{reg}} + T_{\text{ireg}} + 0.64 \cdot T_{\text{iagg}}) \tag{2.3}$$

where: U_{reg} – the share of regional trips in the total trips generated at the station and stop or zone, U_{ireg} – the share of interregional trips in the total trips generated at a station and stop or zone, U_{iagg} – share of inter-agglomeration trips in total trips generated at a station and stop or zone, T_{reg} – number of regional trains stopping at the station or stop or in zone, T_{ireg} – number of interregional trains stopping at the station or in zone, T_{iagg}) – number of inter-agglomeration trains stopping at the station or in zone.

The correctness of the model of dividing trips into types can be checked using statistical data published by the Polish Central Statistical Office. The Polish Central Statistical Office provides data on passenger transport on an annual basis, in a summary form, broken down into types of trains:

– express (inter-agglomeration),
– fast – TLK, IR (interregional),
– passenger (regional).

The use of the trip division model to calculate the number of trips in the UTK statistics for stations and stops results in the number of trips in individual types of passenger transport, consistent with the statistics of the Central Statistical Office (Table 1).

Table 1. Number of passengers using the national railway network by train type in 2017–2019

	Inter-agglomeration		Inter-regional		Regional		Total
	million passengers / year	% share	million passengers / year	% share	million passengers / year	% share	million passengers / year
2017	6.2	2.0	35.8	11.9	259.6	86.1	301.6
2018	6.4	2.1	38.7	12.6	262.8	85.3	307.9
2019	6.8	2.1	41.0	12.3	285.5	85.6	333.3
Average from 2017–2019	6.5	2.1	38.5	12.2	269.3	85.7	314.3
Average based on UTK data for the years 2017–2019 by model	5.9	2.0	36.9	12.5	252.6	85.5	295.5

Source: [11, 14], own study.

The trip division model makes it possible to estimate for each stop the division of the total number of journeys into trips made by particular types of trains (Fig. 2).

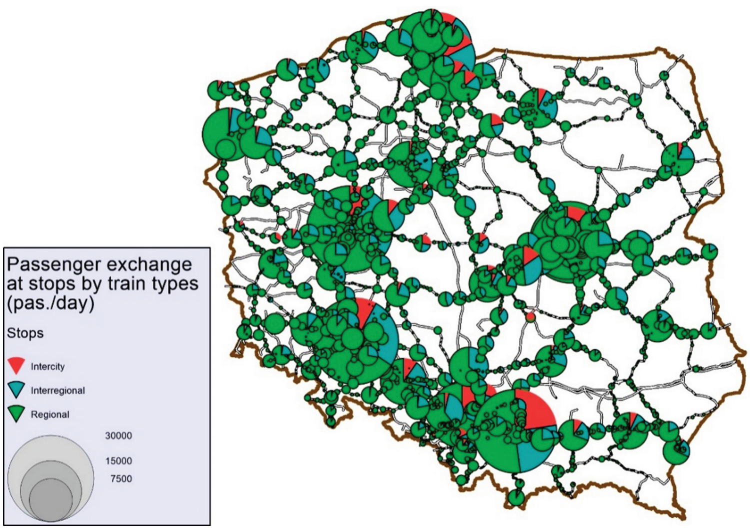

Fig. 2. The result of the application of the travel division model to estimate the volume of passenger traffic at stations and stops divided into train types (2017–2019). *Source*: UTK, VISUM traffic model

As the division of the area into zones is used at a later stage in the construction of the rail travel matrix, the data on the number of journeys at individual stations and stops should be aggregated. For each zone creating the area model, the number of journeys at stations and railway stops lying in the certain zone should be summed up (Fig. 3).

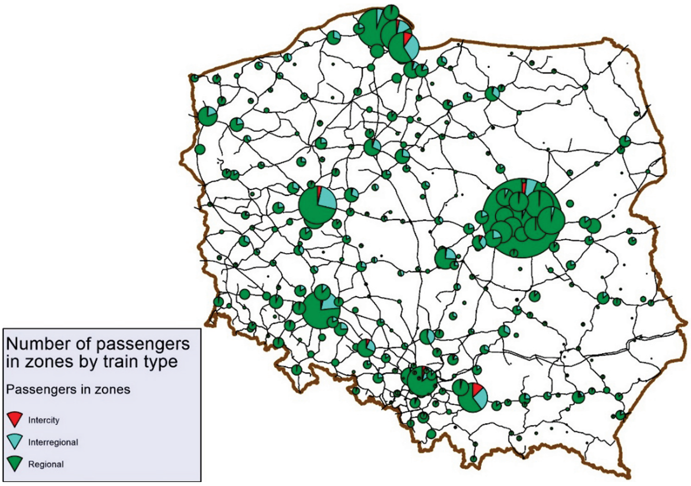

Fig. 3. Example of passenger traffic aggregation at railway stations and stops to the level of zones by train type 2017–2019). *Source*: UTK, VISUM traffic model

3. Socio-economic data in the area model

Building the railway traffic generation model requires socioeconomic data for the current state and, in the case of forecasting models, for the forecast period. Due to the way in which the travel model is built (based on the area model), it is necessary to have statistical data taking into account the division into administrative units (communes, poviats, subregions, administrative regions). Depending on the needs (the purpose of the analysis or the way of presenting the results), subsequent divisions may be aggregated up to the country level.

In the case of transport analyzes on the national scale, the most common division of the area model to zones, the boundaries of which coincide with poviats (379 regions). Then all statistical data for the existing state or prognostic data is reduced to this mapping scale. It is also convenient because the distinguishing feature of each administrative unit is its unique statistical number, thanks to which it is possible to assign to them any sets of statistical data in which such numbers appear.

In the area model, the regions inside the country are complemented by border crossings, which are points generating foreign traffic. Border crossings are excluded from the UTK statistics.

The main source of statistical data enabling the creation of the area model is the Local Data Bank provided by the Polish Central Statistical Office. Data from this bank is used in very different social and economic fields, including transport models. The website of the Polish Central Statistical Office is a good and convenient source for downloading this information. By specifying the territorial breakdown level, year and dataset, you can get a table with the data in the EXCEL spreadsheet format, and by using the territorial unit statistical number, you can then enter this data into the area model. The database of the Polish Central Statistical Office contains mainly historical data, most of which is published during the year after the end of a given calendar year. Some data also concern the forecast state. From the point of view of the travel generation model, the most important data are population data, including a 30-year forecast divided into poviats. Other information assigned to administrative units available at the Polish Central Statistical Office and were used in the traffic generation model:

– tourist potential expressed in the number of accommodation places,
– individual motorization index,
– level of the gross domestic product (GDP).

In building the forecast traffic generation model, it is also necessary to have GDP growth forecasts, on which the calculations of transport traffic forecasts are generally based. The GDP growth rate, through the traffic generation models, has a clear impact on the growth rate of passenger transport. It is necessary not only to have a general knowledge of the GDP growth rate for the entire country, but also to define a forecast of GDP growth in its individual regions. In the case of network analyzes, where matrixes of travel between zones are used, it is advisable to determine the differentiation of GDP at the poviat level because, at this level of detail, the matrixes of movements of people and goods are built for the current state and for the time period considered in the forecast.

As traffic forecasts for investments in transport should be made for the period of 30 years, data on GDP should also refer to distant periods, currently at least until 2045–2050. The rules for using data on macroeconomic indicators for Poland are set out in the document [13] which specifies the GDP growth rate for Poland in the next 30 years (Table 2).

Table 2. Assumptions for the GDP growth rate for 2020–2035

Year	GDP growth [%]	YEAR	GDP growth [%]
2020	–2.70	2028	3.00
2021	4.00	2029	2.80
2022	3.40	2030	2.80
2023	3.00	2031	2.70
2024	3.00	2032	2.60
2025	3.00	2033	2.50
2026	3.10	2034	2.40
2027	3.10	2035	2.30

Source: [7].

The source of data on forecasts of GDP and GDP per capita for the needs of transport planning can also be a research [5] which projected the size of GDP per capita until 2045 in the system of 72 subregions using the linear regression analysis. The projection takes into account the demographic forecast for each sub-region and the GDP forecast for the whole country.

4. Assumptions for the model of passenger rail traffic generation

In the preparation of traffic generation models, regression models are used in which the explained variable is the number of passengers at stops, and the explanatory variables are information about the area surrounding the stops and the transport offer expressed in the number of trains stopping at these stops. The number of trains stopping at a station determines the level of supply and availability of the railway and has a specific impact on the number of passengers generated. Explanatory variables should equally available and reliable at the level of the analysis of the current state and in the forecasting periods.

Thus, the model of the transport network should contain historical data specifying the number of passengers getting on / off at stations and stops. On the other hand, the area model should contain data characterizing the area in the vicinity of the station (e.g. divided into communes). The railway traffic generation potential is determined by the accessibility of the station or railway stop to users, mostly those living in its vicinity. Determining this potential requires examining the impact of the spatial development in the vicinity of the station to the volume of traffic at that station. Based on the experience and the application of regression analysis for various ranges of station influence, it was found that the most reliable information was the information about the station's vicinity within a radius of 2 and 5 kilometers. Within the radius of up to 2 kilometers, the influence of these variables is significant from the point of view of the generation potential of agglomeration and regional traffic, as it is the range of walking distance to the stop. For interregional and inter-agglomeration traffic, the variables describing the stop vicinity within the radius of 5 kilometers better describe the number of passengers than the variables from a larger radius (a better correlation was obtained). This allows to determine the relationship between the number of passengers at the station and the explanatory variables characterizing the environment in which these stations are located.

5. Process of data preparation for the passenger rail traffic generation model

After analyzing the data from the demand side, its availability for the needs of model execution and the impact on the accuracy of the regression model, the following explanatory variables were selected for detailed analyzes:

- number of inhabitants,
- economic potential of the inhabitants,

– number of passenger cars,
– number of beds.

The procedure of data preparation for the model of passenger rail passenger traffic generation requires the following steps:

– For each zone, on the basis of the railway network model, defining (indicating) stations and stops serving trains of particular categories (inter-agglomeration, inter-regional, regional). For this purpose, information from the timetable can be used to determine what type of trains stop at a given stop.
– For each station or stop, calculation of the number of inhabitants within the radius of 2 and 5 kilometers, divided into rural and urban inhabitants. The vicinity of other stations of the same type should be considered, taking into account that some residents have a choice of two or more stations.
– For each zone, the calculation of the sum of inhabitants around the stations or stops defined for (assigned to) a given zone.
– For each station or stop and zone, the calculation of the economic potential being the product of the number of inhabitants around the station or stop within a radius of 2 to 5 kilometers and the GDP per capita assigned to this zone.
– For each station or stop and zone, the calculation of the number of passenger cars being the product of the number of inhabitants around the station within the radius of 2 and 5 kilometers and the individual motorization index assigned to this zone.
– For each station or stop and zone, the calculation of the number of beds around the station within the radius of 5 kilometers.
– For each station or stop and zone, the calculation of the sum of all trains stopping during the day.

Fig. 4 shows an example of the result of the analysis concerning the number of inhabitants within the radius of 5 km around stations and stops serving the traffic of regional trains.

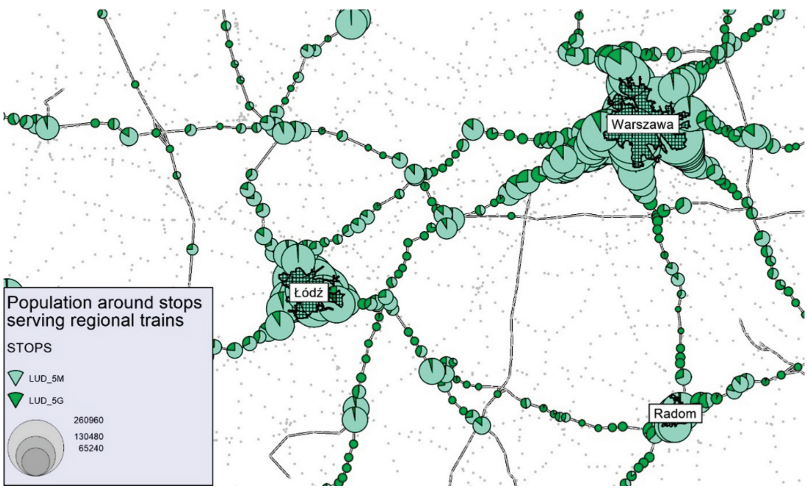

Fig. 4. Rural and urban population around stations and stops serving the regional train traffic).
Source: own study

6. Regression analysis

Regression analysis was used to examine the influence of the explanatory variables on the compliance of the results obtained from the travel model with the control statistical data. When selecting the explanatory variables, their correlation was taken into account so as to limit their number and introduce those that are available at the stage of model construction for the current state and forecasting periods and have a stronger impact on the correct calculation result. The analysis was carried out at the level of traffic at railway stations or stops and for aggregate traffic to zones (e.g. several stations or stops in one zone). The analysis was performed for all stops in the national network for which there were data on the number of passengers.

As a result of the analyzes, regression equations were obtained, the form of which is presented below. They define the sum of rail trips generated in individual regions and at individual railway stations or stops.

Trips generated in zones

$$P_{\text{rej}} = a \cdot T_{\text{reg}} + b \cdot T_{\text{ireg}} + c \cdot T_{\text{iagg}} + d \cdot (\text{POP}_{u\_\text{REG}\,2\,\text{km}} \cdot \text{GDPM}) + e \cdot (\text{POP}_{u\_\text{IAGG}\,5\,\text{km}} \cdot \text{GDPM}) + f \cdot (\text{POP}_{r\_\text{IREG}\,5\,\text{km}} \cdot \text{MOT}_{\text{car}}) + g \cdot (\text{POP}_{r\_\text{IREG}\,5\,\text{km}} \cdot \text{GDPM}) + h \cdot \text{MHOTEL}\,5\,\text{km} \tag{6.1}$$

where: P_{rej} – number of trips in the zone generated at stations or stops in the zone, POC_{reg} – number of regional trains stopping at stations or stops with regional traffic in the zone, POC_{mreg} – number of interregional trains stopping at stations with interregional traffic in the zone, POC_{mag}) – number of inter-agglomeration trains stopping at stations with inter-agglomeration traffic in the zone, $\text{POP}_{u\_\text{REG}\,2\,\text{km}}$ – number of urban residents within 2 km from stations serving regional traffic in the zone, $\text{POP}_{u\_\text{IAGG}\,5\,\text{km}}$ – number of urban residents within a radius of 5 km from the stations serving inter-agglomeration traffic in the zone, $\text{POP}_{r\_\text{IREG}\,5\,\text{km}}$ – number of rural inhabitants within a 5 km radius from stations serving interregional traffic in the zone, $\text{POP}_{r\_\text{IREG}\,5\,\text{km}}$ – number of rural inhabitants within a 5 km radius from stations serving interregional traffic in the zone MHOTEL 5 km – number of beds within a 5 km radius around stations in the zone, MOT_{car} – motorization index per capita in the region (capacity / 1000 inhabitants), GDPM – GDP per capita in the region (thousand PLN/inhabitant). Regression coefficients are shown in Table 3.

Table 3. Regression coefficients – trips generated in zones

A	*B*	*C*	*d*	*e*	*f*	*g*	*H*
6.2156	38.9921	6.6446	0.1981	0.1761	–0.3608	2.324	0.0141

Source: own study.

It should be noted that, in the case of the number of passenger cars in rural areas, the indicator has the "MINUS" sign, which means that it reduces the number of rail journeys due to the greater availability of the car resulting from the higher motorization index.

The following model compliance parameters were obtained:
- R-square = 95.1293 percent,
- Standard estimation error = 2466.4,
- Mean absolute error = 1262.91.

As a result of the regression analysis for the travel model generated in the regions, a relatively high correlation coefficient of 0.95 was obtained (Fig. 5). However, it should be remembered that the input data on the number of passengers is largely derived from various types of conversion factors used to determine the number of tickets sold by carriers, especially when it comes to journeys based on season tickets. The input data for the analysis may therefore be subject to an error, hence, in some regions, there may be large differences between the values calculated using the model and those provided by carriers. This does not necessarily mean that the model is imperfect.

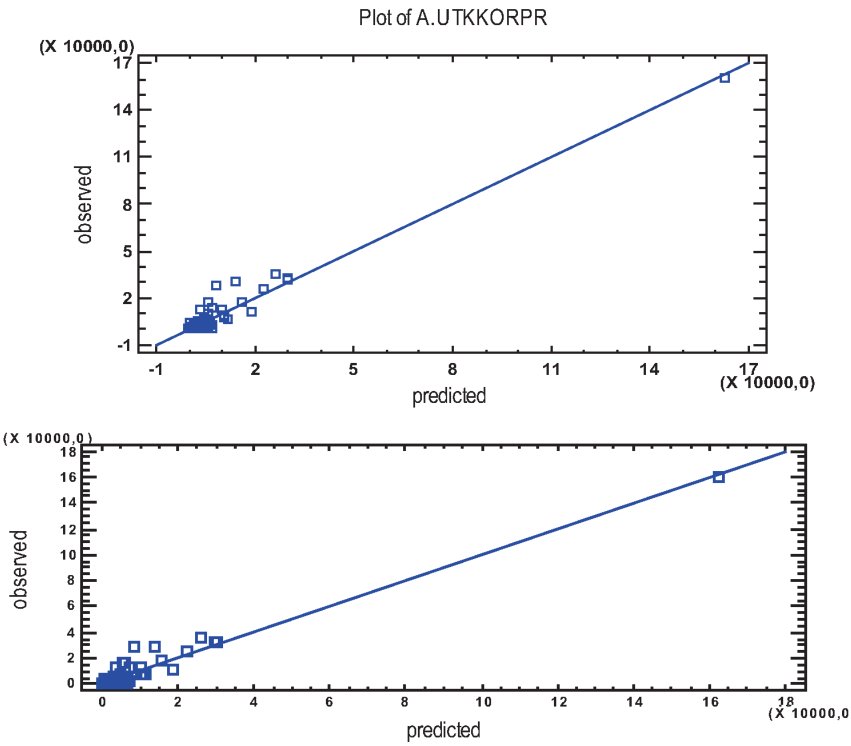

Fig. 5. Compatibility of the values calculated using the model with the observed values for zones. *Source*: own study

Trips generated at stations and stops

$$P_{st} = A \cdot T_{reg} + b \cdot T_{ireg} + C \cdot T_{iagg} + d \cdot (\mathrm{POP}_{u\_\mathrm{REG}\,2\,\mathrm{km}} \cdot \mathrm{GDPM}) + e \cdot (\mathrm{POP}_{u\_\mathrm{IAGG}\,5\,\mathrm{km}} \cdot \mathrm{GDPM}) + f \cdot (\mathrm{POP}_{r\_\mathrm{IREG}\,5\,\mathrm{km}} \cdot \mathrm{MOT}_{car}) + g \cdot (\mathrm{POP}_{r\_\mathrm{REG}\,5\,\mathrm{km}} \cdot \mathrm{GDPM}) + h \cdot \mathrm{MHOTEL}\,5\,\mathrm{km} \tag{6.2}$$

where: P_{st} – number of trips generated at the station or stop, T_{reg} – number of regional trains stopping at the station or stop serving regional traffic, T_{ireg} – number of interregional trains stopping at the station serving interregional traffic, T_{iagg} – number of inter-agglomeration trains stopping at the station serving inter-agglomeration traffic, $POP_{u\_REG\,2\,km}$ – number of urban residents within 2 km around the station or stop serving regional traffic, $POP_{u\_IAGG\,5\,km}$ – number of urban residents within a 5 km radius around the station serving inter-agglomeration traffic, $POP_{r\_IREG\,5\,km}$ – number of rural residents within a 5 km radius around the station or stop handling inter-agglomeration traffic $POP_{r\_REG\,5\,km}$ – number of rural residents within a 5 km radius around the station or stop handling regional traffic, MHOTEL 5 km – number of beds within a 5 km radius around the station or stop, MOT_{car} – motorization index per capita in the zone where the station or stop is located (capacity/1000 inhabitants), GDPM – GDP per capita in the zone where the station or stop is located (thousand PLN/inhabitant). Regression coefficients are shown in Table 4.

Table 4. Regression coefficients and model compatibility parameters – trips generated at stations and stops

A	*B*	*C*	*d*	*e*	*f*	*g*	*H*
10.6277	39.3989	18.0322	0.0869	0.2500	–0.1362	0.0181	0.0873
R-square = 63.2564 percent				Standard estimation error = 1805.44			
Mean absolute error = 739.483							

Source: own study.

As a result of the regression analysis for the travel model generated at the stations, the correlation coefficient of 0.63 was obtained (Fig. 6). The results obtained from the model are therefore burdened with a relatively large error. As in the case of the analysis for zones,

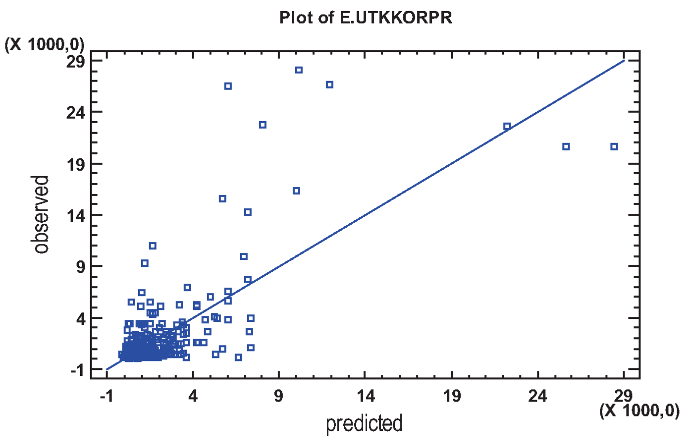

Fig. 6. Compatibility of the values calculated using the model with the observed values for stations and stops. *Source*: own study

it should be remembered that the input data on the number of passengers is also subject to an error that may cause differences.

7. Sensitivity analysis

For the model of railway traffic generation created in this way, a test of the sensitivity of the model to changes in the level of explanatory variables was carried out. The following shows the size of the growth rates for aggregate traffic and for individual types of travel, depending on the changes:

- GDP,
- population,
- motorization index,
- number of beds,
- number of trains serving the zone.

It should be noted that the dependence of the number of passengers on GDP is used practically in all models, because it is an obvious dependence resulting from the available data. The problem may concern use of an inadequate flexibility indicator that gives inflated forecast results, rather than just taking advantage of the dependence of traffic volume on GDP.

During the analysis, changes in the level of subsequent explanatory variables were introduced, thus, in the growth rates in the following lines (Table 5), all previous changes

Table 5. Sensitivity of the railway traffic generation model to changes in explanatory variables

	Travel growth rates compared to the base year			
	Total trips	Inter-agglomeration trains	Inter-regional trains	Regional trains
Base year	1.000	1.000	1.000	1.000
50% GDP growth	1.214	1.201	1.222	1.214
100% increase in the number of beds	1.251	1.252	1.250	1.252
20% increase in the motorization index	1.213	1.235	1.209	1.213
Population decline by 10%	1.172	1.185	1.167	1.172
50% increase in the number of inter-agglomeration trains	1.177	1.751	1.158	1.167
50% increase in the number of interregional trains	1.321	1.831	1.860	1.231
50% increase in the number of regional trains	1.534	1.500	1.534	1.535

Source: own study.

are included. The last line shows the level of traffic generation growth taking into account all the analyzed increases in explanatory variables.

8. Conclusions

The basic factors determining the number of passengers in rail transport are transport offer and attractiveness of the railway transport. This influences the modal wich is one of the most important parameters indicating the legitimacy of railway investments. The literature shows that the choice of the modal split, and thus the number of passengers, is determined by the relative utility of this modes of transport.

The article presents the methodology of passenger rail traffic generation, based on the available data sources. As described in the introduction, the sources are the publications of the Office of Rail Transport, in which information has been provided on the passenger traffic load since 2017 regarding all stations and railway stops in Poland. As the description of the presented method shows, these are the values estimated on the basis of various sources. The received data is given in the ranges "from – to", so the data itself contains a certain range of uncertainty as to the specific number of passengers. As a rule, the reported upper end of the range is 30–50% greater than the lower end, so the uncertainty range is significant. Also, a comparative analysis of data from 3 years shows that, in many cases, the differences in the number of passengers between consecutive years are large, and it is very difficult to interpret the reasons for such differences.

For the work on the passenger rail traffic generation model, average values from the given passenger number ranges for 3 years were adopted as the base data, assuming that this gives the picture that is closest to the current state. Nevertheless, the obtained results, especially for trips calculated at the stop level, sometimes differ significantly from the values entered as the base data. At the same time, the model gives correct results for the number of trips summed up at the country level and, in many cases, satisfactory results for calculations at the level of traffic zones. Errors in estimating traffic at stops are eliminated when data is aggregated to the level of zones.

The model can be used to calculate the forecast passenger rail traffic in the Intermodal National Traffic Model, taking into account all the above-mentioned reservations. Additionally, it may be used in two ways.

The first option is to estimate the size of the potential rail travel market if the location of a new stop is proposed. By introducing data describing the surroundings of this stop into the regression equation, it is possible to obtain information about the possible number of passengers that may use this stop after its launch.

The second possibility is to use the model to calculate the growth rates for the forecast period by introducing the forecast values of the explanatory variables (population, automotive, GDP, etc.). The obtained growth rates can be applied directly to the base year data obtained in any way deemed reliable. This may be data from the UTK collections relating to specific stations, but also information obtained directly from carriers or obtained through own measurements. The method of obtaining such indicators from the model is presented synthetically in the sensitivity analysis.

References

[1] M.E Ben-Akiva, S.R. Lerman, *Choice Analysis: Theory and Application to Travel Demand*. Cambridge: MIT Press, 1985.

[2] A. Brzeziński, T. Dybicz, "Possibility of Big Data Application for OD-matrix Callibration in Transport Demand Models", Archives of Civil Engineering, 2021, vol. 67, no. 1, pp. 215–232.

[3] A. Brzeziński, T. Dybicz, "Modelowanie ruchu z wykorzystaniem innowacyjnych źródełdanych", in *Badania naukowe w Instytucie Dróg i Mostów – monografia jubileuszowa*, P. Olszewski, Ed. Oficyna Wydawnicza Politechniki Warszawskiej, 2019, pp. 209–228.

[4] A. Brzeziński, T. Dybicz, A. Waltz, "Railway passenger traffic model ", *Acta Scientiarum Polonorum. Series: Architectura*, 2018, vol. 17, no 4, pp. 5–17, DOI: 10.22630/ASPA.2018.17.4.36.

[5] A. Brzeziński, "Zasady prognozowania ruchu drogowego z uwzględnieniem innych środków transportu", presented at Projekt INMOP3 – założenia, cele, innowacyjność, dostępne dane. Międzynarodowa konferencja naukowo-techniczna: Innowacyjne metody prognozowania ruchu w ramach projektu RID_I_62, 28 maja 2019 r.

[6] R. Cordera, R. Samnudo, L. Dell'Olio, A.Ibeas, "Trip distribution model for regional railway servies considering spatial effects between stations", *Transport Policy*, 2018, vol. 67, pp. 77–84, DOI: 10.1016/j.tranpol.2018.01.016.

[7] *Guidelines on issues related to the preparation of investment projects, including income-generating projects and hybrid projects for the years 2014–2020 of 10 January 2019*. GUS, 2020.

[8] X. Li, H. Huang, J. Li, H. Xu, "A research on rail traffic generation forecasting based on BaiduMap API", *Shandong Science*, 2017, vol. 30, no. 1, pp. 82–88, DOI: 10.3976/j.issn.1002-4026.2017.01.013.

[9] *Pasażerski Model Transportowy, Skrócony Raport Techniczny*. Warszawa: Biuro Planowania i Rozwoju Kolei, CPK, 2020.

[10] *Pasażerski Model Transportowy i sieciowa prognoza ruchu, część szczegółowa*. Warszawa: Biuro Planowania i Rozwoju Kolei, CPK, 2020.

[11] *Passenger exchange at stations in Poland in 2017, 2018, 2019*. Urząd Transportu Kolejowego.

[12] A. Peterson, "The Origin-Destination Matrix Estimation Problem – Analysis and Computations", *Linkoping Studies in Science and Technology. Dissertations No. 1102*. Norrkoping, Sweden, 2007.

[13] O. Nielsen, C. Hansen, "Updating trip matrices for Copenhagen using multiple data sources", *Trafikdage på Aalborg Universitet*, 2007.

[14] Statistical yearbooks of the Central Statistical Office "Transport – activity results" 2017–2019.

[15] A. Waltz, "Krajowy model transportowy i jego zastosowanie w prognozowaniu ruchu dla potrzeb PKP Intercity S.A.", *Zeszyty naukowo-techniczne Stowarzyszenia Inżynierów i Techników Komunikacji Rzeczpospolitej Polskiej, Oddziałw Krakowie*, 2014, vol. 103, no. 1, pp. 393–417.

[16] M. Wardman, "The value of travel time: A review of British evidence", *Journal Transport Economics and Policy*. 1998, vol. 32, no. 3, pp. 285–316.

[17] *Zasady prognozowania ruchu drogowego z uwzględnieniem innych środków transportu DZP/RID-I-62/11/NCBR/2016*.

Budowa modelu generacji ruchu kolejowego

Słowa kluczowe: ruch kolejowy, model generacji ruchu, analiza regresyjna, prognozowanie ruchu, ruch pasażerski

Streszczenie:

W artykule przedstawiono nowe podejście do budowy modelu generacji pasażerskiego ruchu kolejowego. Przedstawiono propozycję metodyki budowy modelu generacji pasażerskiego ruchu kolejowego oraz weryfikacji tego modelu w oparciu o ograniczone zmienne opisujące stacje i przy-

stanki kolejowe oraz rejony komunikacyjne i dostępne dane statystyczne Wykorzystano w nim dane o liczbie pasażerów na stacjach i przystankach kolejowych w połączeniu z danymi opisującymi model obszaru wokół stacji budowany w oparciu o liczbę ludności, liczbę miejsc noclegowych, motoryzację indywidualną i produkt krajowy brutto (PKB). Wykorzystanie tych danych umożliwiło to przeprowadzenie analizy potencjału generacji ruchu kolejowego na dużym poziomie szczegółowości. W metodzie uwzględniono trzy segmenty rynku kolejowego: przewozy regionalne, międzyregionalne i międzyaglomeracyjne.

Doświadczenia z wykorzystywania modeli ruchu w Polsce pokazują, że najczęściej są one budowane na potrzeby administracji drogowej oraz na potrzeby dużych miast i ew. województw. Znaczenie mniej jest doświadczeń związanych z budową modeli służących analizie prognozowanego ruchu kolejowego. Stosunkowo niewielkie doświadczenia związane z modelami kolejowymi wynikają m.in. z wieloletnich ograniczeń w dostępie do danych kolejowych, zwłaszcza badań pasażerów kolei i statystyk biletowych oraz ograniczeń w dostępności modeli budowanych na rzecz spółek kolejowych. Brak dostępu do danych zmniejsza szansę na dobrą jakość budowanych modeli kolejowych. Doświadczenia z innych krajów wskazują, że w modelowaniu ruchu kolejowego korzysta się z bardzo szczegółowych i rzeczywistych informacji pochodzących z badań ankietowych i/lub baz danych o sprzedaży biletów.

W Polsce w ostatnich latach zaczęto publikować dane o pasażerskim ruchu kolejowym. Pozwala to na opracowanie nowych wariantów modeli generacji ruchu, opartych o rzeczywiste dane i w ten sposób istotnie zwiększyć jakość i dokładność prognozowania ruchu. Od roku 2017 Urząd Transportu Kolejowego (UTK) rozpoczął publikowanie sumarycznych informacji o liczbach pasażerów korzystających ze stacji i przystanków położonych na całej sieci kolejowej w Polsce. Wykorzystanie tych danych do budowy modeli ruchu wymaga ustalenia struktury pasażerów podróżujących różnymi typami pociągów, tj. w ruchu międzyaglomeracyjnym, międzyregionalnym oraz regionalnym. W artykule przedstawiono metodę określania, za pomocą wzorów, udziałów podróży poszczególnych typów w sumie podróży kolejowych. Poprawność zaproponowanej metody podziału podróży na rodzaje sprawdzono wykorzystując dane statystyczne publikowane przez Główny Urząd Statystyczny.

Zbudowanie modelu generacji ruchu kolejowego wymaga dysponowania także danymi socjoekonomicznymi opisującymi stanu istniejący, a w przypadku modeli prognostycznych okresy prognozy. Ze względu na sposób budowy modelu podróży (w oparciu o model obszaru) niezbędne jest dysponowanie danymi statystycznymi uwzględniającymi podział na jednostki administracyjne (gminy, powiaty, podregiony, województwa). W zależności od potrzeb (celu analizy czy też sposobu prezentacji wyników) kolejne podziały mogą być agregowane, aż do poziomu kraju. W przypadku modelu generacji ruchu opisanego w artykule zastosowano podział obszaru na rejony komunikacyjne, których granice pokrywają się z powiatami (379 rejonów).

Głównym źródłem danych statystycznych pozwalającym na stworzenie modelu obszaru jest Bank Danych Lokalnych udostępniany przez Główny Urząd Statystyczny. W bazie GUS znajdują się głównie dane historyczne, a niektóre dane dotyczą również stanu prognozowanego. Najistotniejsze z punktu widzenia modelu generacji podróży są dane ludnościowe, w tym prognoza opracowana na 30 lat w podziale na powiaty. Wśród innych informacji przypisanych do jednostek administracyjnych, które są dostępne w GUS i które zastosowano w opisywanym modelu generacji ruchu są: potencjał turystyczny wyrażony liczbą miejsc noclegowych, wskaźnik motoryzacji indywidualnej i poziom produktu krajowego brutto (PKB).

Budowa modelu generacji ruchu prognozowanego wymaga także dysponowania prognozami wzrostu PKB, na których generalnie opierają się obliczenia prognoz ruchu w transporcie. Tempo wzrostu PKB, poprzez modele generacji ruchu, wpływa w sposób jednoznaczny na tempo wzrostu przewozów pasażerskich. Konieczna jest nie tylko ogólna wiedza o tempie wzrostu PKB dla całego kraju, ale również określenie prognozy wzrostu PKB w poszczególnych jego rejonach. W przypadku analiz sieciowych, gdzie operuje się macierzami podróży pomiędzy rejonami komunikacyjnymi, wskazane jest określenie zróżnicowania PKB na poziomie powiatu, ponieważ na takim poziomie szczegółowości budowane są macierze przemieszczeń osób i towarów dla stanu istniejącego oraz dla horyzontów czasowych rozpatrywanych w prognozie. W artykule wskazano źródła danych dot. prognozowanego PKB.

W opisywanym modelu generacji ruchu pasażerskiego kolejowego wykorzystano modele regresyjne w których zmienną objaśnianą jest liczba pasażerów na przystankach, a zmiennymi objaśniającymi informacje o obszarze otaczającym przystanki oraz oferta przewozowa wyrażona liczbą pociągów zatrzymujących się na tych przystankach. Tak więc w modelu sieci transportowej korzystano z danych określających liczbę wsiadających/wysiadających na stacjach i przystankach, a w modelu obszaru z danych charakteryzujących obszar w otoczeniu stacji. O potencjale generacji ruchu kolejowego decyduje dostępność stacji lub przystanku kolejowego dla użytkowników, w większości dla mieszkających w jego otoczeniu. Ustalenie tego potencjału wymaga zbadania wpływu zagospodarowania przestrzennego w otoczeniu stacji na wielkość ruchu na tej stacji. Na podstawie doświadczeń i zastosowania analizy regresyjnej dla różnych zasięgów oddziaływania stacji stwierdzono, że najbardziej miarodajne okazują się informacje o otoczeniu stacji w promieniu do 2 i do 5 kilometrów. W promieniu do 2 kilometrów istotny jest wpływ tych zmiennych z punktu widzenia potencjału generacji ruchu aglomeracyjnego i regionalnego, ponieważ jest to zasięg dojścia pieszego do przystanku. Dla ruchu międzyregionalnego i międzyaglomeracyjnego zmienne opisujące otoczenie przystanku w promieniu do 5 kilometrów lepiej opisują liczbę pasażerów niż zmienne z większego promienia (uzyskano lepszą korelację). Pozwala to określić zależności pomiędzy liczbą pasażerów na stacji a zmiennymi objaśniającymi charakteryzującymi otoczenie w jakim te stacje się znajdują.

Do zbadania wpływu zmiennych objaśniających na zgodność wyników otrzymanych z modelu podróży z kontrolnymi danymi statystycznymi zastosowano analizę regresyjną. Przy wyborze zmiennych objaśniających brano pod uwagę ich korelację tak, aby ograniczyć ich liczbę i wprowadzać te które są dostępne na etapie budowy modelu dla stanu istniejącego i okresów prognostycznych oraz silniej wpływają na prawidłowy wynik obliczeń. Analizę przeprowadzono na poziomie ruchu na stacjach lub przystankach kolejowych oraz dla ruchu zagregowanego do rejonów komunikacyjnych (np. kilka stacji lub przystanków w jednym rejonie).

W wyniku analiz otrzymano równania regresyjne, określające sumę podróży kolejowych generowanych w poszczególnych rejonach oraz na poszczególnych stacjach lub przystankach kolejowych. Opracowany model można wykorzystać do liczenia prognozowanego pasażerskiego ruchu kolejowego np. na potrzeby prowadzonych inwestycji kolejowych. Dodatkowo zaproponować można jego wykorzystanie na dwa sposoby. Pierwsza możliwość to oszacowanie wielkości potencjalnego rynku podróży kolejowych w przypadku gdy proponowana jest lokalizacja nowego przystanku, a druga, to wykorzystanie modelu do liczenia wskaźników wzrostu dla okresu prognozowanego, poprzez wprowadzanie prognozowanych wartości zmiennych objaśniających (ludność, motoryzacja, PKB, itd.).

Received: 05.10.2021, Revised: 28.12.2021