RAILWAY PASSENGER TRAFFIC MODEL

Andrzej Brzeziński¹, Tomasz Dybcz¹, Andrzej Waltz²

¹Wydział Inżynierii Lądowej, Politechnika Warszawska, Warszawa
²independent expert

ABSTRACT
The paper addresses the issues of railway trip modelling in conditions of restricted access to railway data, and especially, the results of surveys among railway passengers and databases from railway systems. It presents the gathered experience associated with the development of a passenger railway transport model in the research project “Principles of forecasting road traffic, taking into account other means of transport”, implemented within the project Rozwój Innowacji Drogowych (Road Innovation Development), organized by the National Centre for Research and Development and the General Directorate for National Roads and Motorways. The article also presents a proposed methodology for developing a railway trip generation and distribution model based on limited variables describing transport zones (and stations) and the available statistical data on the number of trips, transport tasks and the average trip distance divided into three railway market segments: regional, interregional and interagglomeration transport. It also discusses the results of applying the developed model, including the obtained railway transport structures for transport types, and the model calibration results.

Key words: railway, trip generation models, trip distribution models, trip matrix, railway model

INTRODUCTION
A passenger railway transport system is treated as a priority by the EU and in Poland. It results from the potential of this branch of transport and its competitiveness relative to other means of transport, in terms of trip time and comfort, the safety of transport, and the lower environmental impact that other means of transport. This priority translates to railway development programs, including the ones in the field of infrastructural investments. It is expressed by the National Railway Program (KPK), which, in a long-term perspective, includes railway investments worth PLN 66 billion over 220 projects for 9,000 km of tracks. The strategic objectives of the program include: improving railway passenger transport attractiveness in agglomera-

© Copyright by Wydawnictwo SGGW
plans). There is much less experience associated with developing models for the analysis and forecasting railway trip. Therefore, the need of the moment, also due to the ongoing investment process in Poland, is strengthening the rail trip modelling methodology and increasing the functionality of models, through including trip intermodality.

A relatively minor experience associated with railway traffic models results from, i.a., restrictions in the access to railway data, especially passenger studies and ticket statistics, and limitations in the access to models developed for railway companies. They treat models as tools used for internal analysis, containing confidential data impossible for general sharing. The confirmation of this thesis is, for once, the small number of articles and conference papers addressing this issue in Poland. For example, in the research-technical bulletin of the SiTK (Association for Communication Engineers and Technicians) Kraków branch, published in 2014 after a specialist conference “Trip Modelling and Traffic Forecasting”, only two out of 32 articles were related to railway issues and one to trip modelling.

In the light of the above, the objective of this article is to present the possibilities of estimating the models of generation and distribution of trips between train stations, taking into account the available data on supply and demand in the railway market.

RAILWAY TRAFFIC MODELS

The issue of modelling trips associated with railway was approach, i.a., within the research projects of the Institute of Roads and Bridges at the Warsaw University of Technology (Brzeziński and Jesionkiewicz-Niedzińska, 2016). The results of tests performed within the Warsaw agglomeration show that most trips to a train station is performed on foot – 68%, with a minority by means of transport (car, bicycle, bus) – 32%. An average trip distance just exceeds 3 km, while 93% of the trips on foot falls within a range of 0.3–1.6 km, and 90% of trips by car do not exceed 3.1 km.

The project “Principles for forecasting road traffic, taking into account other means of transport”, within the Joint Undertaking named the Road Innovation Development, organized by the National Centre for Research and Development and the General Directorate for National Roads and Motorways¹, involves developing a methodology for trip modelling and road traffic forecasting, taking into account other means of transport, also railway, and trip models (road, rail, air). Previous experience resulting from the project confirms limited availability of data necessary to develop railway traffic models (Waltz, 2014). By principle, it is only possible to use commonly available information of railway companies and GUS (Central Office of Statistics) – usually in the form of PDF files, and a railway network map, which enables constructing a transport network model (table of objects along railway line, train route assignment rules, passenger train timetable, OBLIKO database) and partially, the construction and calibration of a trip matrix (based on general data on pass.-km and average trip distances). However, there is no possibility to use the transportation network models developed by PKP PLK or ticket system databases (e.g. PKP Intercity). Data of such type are in the possession of railway companies but is unavailable.

This is a significant restriction of the possibilities to construct railway traffic models. Experience from projects executed in other countries (Peterson, 2007), (Nielsen & Hansen, 2007) and (Cordera, Samundo, Dell’Olio & Ibeas, 1998) show that modelling utilizes detailed, actual information coming from surveys and/ or ticket sales databases. As a result, trips distribution models generally fall into two types, based on aggregated data, which utilize, for example, information regarding ticket sales and models based on surveys, utilizing data disaggregated at the level of interviews with system users. An example of an aggregated model based on railway ticket sales data is the Wardman’s model (Wardman, 1998), which proposed a production and distribution model in the UK, using such variables as: population, GDP, motorization rates and the total trip cost. He concluded that the most important factor

¹ The INMOP 3 project, implemented by a consortium between the Warsaw University of Technology (Institute of Roads and Bridges, consortium leader) and the Tadeusz Kościusko Kraków University of Technology over a period from 15.02.2016 to 30.04.2019.
impacting the number of railway trips was the GDP. If disaggregated data is unavailable (user surveys) (Ben-Akiva and Lerman, 1985), it is possible to take into account personal features (age, sex, income, etc.), the specification of transport services and information on trip origins and destinations. Models of such type are more demanding at the data collecting and processing stage, hence, are also more expensive.

In Polish conditions, taking into account the lack of transport system (including railway) user tests (disaggregated model) and the unavailability of data from ticket systems (aggregated model), the construction of commonly available railway trip models induces the need to search for other matrix development and calibration methods, omitting the unavailable data.

TRIP GENERATION MODEL

A railway traffic model development, trip matrix especially, is strongly dependant on data availability. Apparently in the case of railway passenger transport, the trip matrices could be developed using ticket sales statistics, which determine the distribution of trip origins and destinations, and their potentials. However, free access to railway data is usually strongly restricted, and ticket statistics are treated as trade secrets of transportation companies. It is sometimes possible to use historic data (e.g. ticket statistics from previous years), by calibrating the matrices based on socio-economic data and parameters determining the availability of means of transport.

The first analyses of this type were conducted when developing railway traffic forecasts for the elaboration of “Master plan for railway traffic in Poland until 2030”. The elaboration was adopted in 2008 by the contemporary Ministry of Infrastructure as the fundamental document stipulating the direction of development for the railway network in Poland. All data regarding the railway network and the data on transport owned by railway carriers had been made available for the purposes of this project. The data included the number of passengers per station and the number of trips between stations. It enabled the execution of comprehensive traffic analyses throughout the entire network. In consecutive years, as far as data access allowed, the analyses were updated.

The INMOP 3 project involved the development of railway trip generation model, where the basic variable from the demand side is the number of inhabitants within the surroundings of train stations, within a variable access radius, depending on the traffic character. The further variables, improving model quality, include: economic level expressed by PKB/capita and motorization degree of these inhabitants, defined by the cars/capita number. In the case of interregional and inter-agglomeration trips, where the share of business and tourism trips is higher, additional variables were used: the number of companies and the size of the tourist-accommodation base. The most important explanatory variable from the demand side was assumed to be the number of trains per day, available to these inhabitants.

The computation procedure adopted the following sequence of actions:

- Assigning stations to zones. Stations handling individual train categories (regional, interregional, inter-agglomeration) were defined for each transport zone, based on a railway network model and valid information regarding the timetable.
- Determining the number of inhabitants within a station range. The numbers of inhabitants divided into the inhabitants of rural and urban areas are calculated per each station (within a radius of 2, 3 and 10 km). Demographic data are provided by GUS in the Local Data Bank2 (BDL), broken down into municipalities, which in turn, are divided into rural and urban, which enables obtaining information of such type. The population forecast is also determined by GUS. Current forecasts come from 2014, but their new version, based on recent trends and new methodology is under development. The vicinity of neighbouring stations of the same time (handling the same train types) is taken into account. Should a tested station impact

---

2 The GUS Bank of Local Data is the biggest data base on economy, society and the environment in Poland. It offers over 40,000 statistical features grouped by topics. It is available in the NUTS 2016 (macroregions, regions and subregions) system of statistical units.
radius include also other stations, its impact area is appropriately corrected (reduced). It automatically affects the change in the number of inhabitants within an impact area.

- Determining the economic potential. The economic potential of potential passengers, which is the product of the number of inhabitants within a station impact area and the GDP/capita for this zone, is calculated for each zone. The GDP per capita, both for the existing state, as well as the forecasts, is stated at a national subregion level. The existing state and historic data are in the BDL, while the data regarding the forecast were developed within the INMOP 3 project, based on a forecast by the Ministry of Development and Finances. The GDP per capita level in a poviat is adopted depending on the subregion it is located in.

- Determining the motorization potential. The number of cars owned by potential railway passengers, which is the product of the number of inhabitants within a station impact area and the motorization index for this region is calculated for each zone. The motorization index for the existing state is given by GUS for each poviat. Also, at the same division level, within the INMOP 3 project, the motorization index forecast was developed.

- Determining the attractiveness of a zone. Each zone has an assigned number of companies and accommodation beds within a station impact area. The data from the GUS Local Data Banks can be used. This data is given for the division into poviat and municipalities.

- Determining availability. The total number of trains stopping daily at stations in a given zone, divided into transport types (based on commonly available electronic versions of train timetables) is calculated for each zone.

- Determining the potential number of railway passengers. They are calculated for each zone, based on the regression equations determined below, the total number of potential railway passengers and within the impact areas of stations defined for this zone (Brzeziński, Dybicz & Jesionkiewicz-Niedzińska, 2013, Brzeziński, Dybicz & Rezwowski-Mosakowska 2015).

Based on the data regarding the number of passengers using the stations, the impact of the aforementioned variables characterizing station surroundings on the size of station-generated traffic was studied. In the course of regression analysis, the impact of these variables calculated for various station impact radii was studied. During the analysis, it was concluded that the most reliable information were the ones regarding the station surroundings within a 2, 5 and 10 km radius. Smaller radii were reliable for short, regional trips, while bigger for trans-regional trips. An approach taking the radial availability of railway stations is entitled. The potential number of passengers is calculated

![Fig. 1. The number of inhabitants within the impact area of a station in traffic zones](architectura.actapol.net)
taking into consideration population around the station, on the basis of the population density. Regression analysis confirmed that this variable is sufficient to estimate the travel potential. An advantage of such approach concerns data availability also for forecast periods, possibility to apply them for more than 2,000 stations and the whole process automatization.

Program VISUM commonly used to model railway traffic, for example AECOM (2014), enable calculating the totals of these variables for the surroundings of each station, within a set distance radius.

The regression analysis procedure involved introducing different variables describing zones (poviats), available in the BDL. These are, apart from the ones referred to above, the data on the number of learners (students and secondary school pupils), manufacturing volume, municipality income, average income, number of large format commercial facilities, etc. After analysing the impact of these variables on the traffic volume and the possibility of achieving them not only for the current state but also for the forecast period, the traffic generation model was based on 6 variables: the number of inhabitants, the number of cars owned by the inhabitants within a station radius, GPD per capita within a 2 and 5 km station radius, as well as the number of registered companies and hotel beds within a 10 km radius from a station. Using these variables, the regression analysis procedure involved developing formulas determining the number of potential railway trips within a region, for which the calculation results is most similar to the actual trip number values obtained from ticket statistics.

The form of the formulas is following:

Regional trip:
\[ LP_{reg} = a \cdot POC_{reg} + b \cdot LUD_{5w} + c \cdot LUD_{2m} + d \cdot SO_{5m} + e \cdot PKB_{5m} \]

Interregional trip:
\[ LP_{mreg} = a \cdot POC_{mreg} + b \cdot LUD_{5m} - c \cdot SO_{5m} + e \cdot PKB_{5m} + e \cdot FIRMY_{10} \]

Interagglomeration trip:
\[ LP_{mag} = a \cdot POC_{mag} - c \cdot SO_{5m} + d \cdot PKB_{5m} + e \cdot MHOTEL_{10} \]

where:
- \( LP_{reg} \) – the number of trips commenced in regional trains within a given zone,
- \( LP_{mreg} \) – the number of trips commenced in interzonal trains within a given zone,
- \( LP_{mag} \) – the number of trips commenced in interagglomeration trains within a given zone,
- \( POC_{reg} \) – the number of zoneal trains stopping at the stations in the zone,
- \( POC_{mreg} \) – the number of interzonal trains stopping at the stations in the zone,
- \( POC_{mag} \) – the number of interagglomeration trains stopping at the stations in the zone.
LUD5w  – number of inhabitants within the station range – 5 km radius,
LUD2m  – number of inhabitants within the station range – 2 km radius,
SO5m  – the number of cars belonging to inhabitants within a radius of 5 km,
PKB5m  – GDP within a 5 km radius around a station
FIRMY10  – the number of companies within a station impact area – 10 km,
MHOTEL10  – the number of accommodation beds within a station impact area – 10 km,
a, b, c, d, e  – function coefficients.

TRIP DISTRIBUTION MODEL

The next step involves using the obtained regional traffic generation potentials for developing a trip distribution model. In the case of railway trip, the division of trip into three segments was suggested (regional, interregional, interagglomeration) which correspond to the existing types of trains.

The task of a distribution model is to link defined number of trips commencing and terminating in a given traffic zone (traffic generation and absorption potentials) with other zones, thus forming a displacement matrix, mapping the spatial distribution of trips.

A traditional gravity model was used for this purpose (Brzeziński, 2016, Cordera, 2018), which takes into account that trip demand between each pair of zones is directly proportional to the product of the number of trips commencing in the initial zone and the number of trips terminating in the target zone, weighed by the spatial resistance function linking a pair of zones and characterized by, e.g., trip time, distance or cost.

Therefore, the gravity model has the form:

\[ T_{ij} = \frac{P_i \cdot A_j \cdot F_{ij}}{\sum A_j \cdot F_{ij}} \]

where:

\( T_{ij} \)  – the number of trips between a given pair of zones,
\( P_i \)  – generating potential of a zone,
\( A_j \)  – absorption potential of a zone,
\( F_{ij} \)  – spatial resistance function.

In turn, the spatial resistance function has the form:

\[ F_{ij} = a \cdot l_{ij}^b \cdot e^{c\cdot i} \]

where:

\( l_{ij} \)  – distance between zones,
a, b, c  – resistance function parameters.

In the case of the discussed railway model, it was assumed that the basis for trip distribution would be the volumes of traffic generated in the zones and trips times between the zones. Time best reflects the technical condition of infrastructure, and simultaneously, is the most important element impacting a decision to trip by train. Trip cost is the second factor, but due to a comprehensive ticket price determination system, varying depending on the carrier, it was impossible to obtain a sufficiently reliable model utilizing this variable. The second factor is the difficulty in determining future ticket prices within the forecast period. The function adopted as a result of model calibration for three analysed trip segments are shown in the figure below.

There are many ways to calibrate the parameters occurring in the trip distribution procedure. It is usually used in the iteration process, until a matrix is sufficiently consistent with the possessed measurement data. Most of these methods is of heuristic nature. The function calibration process was developed with a standard calibration procedure, based on trip histogram compliance analysis obtained from actual trip matrices and obtained as a result of the calculations.

In the discussed case, trip matrices required additional calibration, so that trips sums and the executed pass.-km were consistent with statistical data provided by GUS. A table with reference totals used in the calibration process and the ones obtained as a result of model calculations can be found below.

A high level of compliance was achieved also in the case of trip length distribution. The table below shows a comparison of average rail trip distances as per GUS statistics and the developed model. Trip

Fig. 3. Spatial resistance function for three types of transport in a railway model

Table 1. A comparison of railway passenger transport volume as per GUS and the model

<table>
<thead>
<tr>
<th>Types of transport</th>
<th>GUS statistics – railway transport</th>
<th>Model statistics from the passenger trips matrix</th>
<th>Calibration coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pass.·day⁻¹</td>
<td>pass.-km·day⁻¹</td>
<td>pass.·day⁻¹</td>
</tr>
<tr>
<td>Regional</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>674 117</td>
<td>26 230 136</td>
<td>673 299</td>
</tr>
<tr>
<td>Interregional</td>
<td>65 753</td>
<td>14 846 301</td>
<td>65 928</td>
</tr>
<tr>
<td>Interaglomeration</td>
<td>16 712</td>
<td>5 664 110</td>
<td>16 652</td>
</tr>
<tr>
<td>Total</td>
<td>756 712</td>
<td>46 740 548</td>
<td>755 879</td>
</tr>
</tbody>
</table>

Source: GUS (2014) and own elaboration.

Table 2. A comparison of average trip distances in rail passenger transport as per the GUS and the model

<table>
<thead>
<tr>
<th>Types of transport</th>
<th>GUS statistic</th>
<th>Model statistics</th>
<th>Calibration coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regional</td>
<td>38.9</td>
<td>38.7</td>
<td>1.01</td>
</tr>
<tr>
<td>Interregional</td>
<td>225.79</td>
<td>220.1</td>
<td>1.03</td>
</tr>
<tr>
<td>Interagglomeriation</td>
<td>338.92</td>
<td>303.8</td>
<td>1.12</td>
</tr>
<tr>
<td>Mean value</td>
<td>61.77</td>
<td>61.4</td>
<td>1.01</td>
</tr>
</tbody>
</table>

Source: GUS (2014) and own elaboration.

distance percentage distributions for the analysed market segments are shown in the figures below.

Resultant matrices for regional, interregional and inter-agglomeration trips, as well as a joint matrix were obtained as a result of the suggested railway passenger trip generation and distribution process, and the calibration process. Graphic illustrations of calculated matrices (trip structures) are presented in the figures below.
Fig. 4. Trip distance distribution in rail passenger transport as per the model. Regional transport, $L_{avg} = 38.7$ km

Fig. 5. Trip distance distribution in rail passenger transport as per the model. Intereregional transport
Fig. 6. Trip distance distribution in rail passenger transport as per the model. Inter-agglomeration transport, $L_{\text{avg}} = 303.8$ km

Fig. 7. Trip distance distribution in rail passenger transport as per the model. Total passenger transport, $L_{\text{avg}} = 61.4$ km
Fig. 8. Passenger trip matrices as per transport type – regional. Year 2015

Fig. 9. Passenger trip matrices as per transport type – interregional. Year 2015

**Fig. 10.** Passenger trip matrices as per transport type – inter-agglomeration. Year 2015

**Fig. 11.** Passenger trip matrices as per transport type – total. Year 2015
The sum of the matrices calculated this way gives a matrix of all railway trips, and the obtained average trip distance is consistent with the one shown in the statistics.

CONCLUSIONS

Limited availability of railway data hinders the work on railway traffic models in Poland. By principle, it is possible to use only commonly available maps and data published by railway companies. Comprehensive studies of transport system users (basis for developing a disaggregated model) are unavailable, except for certain local or regional studies, and ticket system data (aggregated model), except for certain local carriers. It especially hinders modelling trip at national and inter-regional levels.

Within the project the “Principles for forecasting road traffic, taking into account other means of transport” a new approach was proposed how to build a trip generation and distribution model based on a division of the area into povias, utilizing:

- for the traffic generation, the railway station surroundings analysis (in a 2 and 5 km radius) and such variables as: population, GDP, motorization factor, number of companies, number of beds, number of trains,
- for trip distribution: volume of traffic generated within zones and the trip time between zones with spatial resistance functions and for three types of transport (regional, interregional and inter-agglomeration),
- for matrix calibration: number of trips/day, number of passenger km/day and average trip distance for three types of transport.

The suggested trip generation-distribution model was tested by developing a national railway passenger trip matrix, which provided good coefficients of conformity with reference parameters. In the case of medium-distance trips, at a level of 90–99.5%, depending on the type of service.

The new approach also concerns trip distribution based on the generalized cost of travel, which is a component of the cost of travel and the different means of transport, taking account of their participation in the overall transport tasks. Classical approach is based only on the lengths or times of travel.

According to the authors, the application of this method provides good results and it may be successfully used for modelling railway trips. It can also be enhanced through developing the studies of transport behaviour of persons using railway stations and stops. Nonetheless, we cannot lose sight of models based on comprehensive railway passenger studies and ticket system databases. Access to data of such type, for the needs of trip modelling, should not be restricted. This is because modelling results facilitate better planning and designing financial investment from public funds.

REFERENCES

AECOM & ESRI (2014). National Transport Model, Variable Demand Model.


MODEL KOLEJOWEGO RUCHU OSOBOWEGO

STRESZCZENIE

Budowa modeli ruchu kolejowego, a zwłaszcza macierzy podróży jest silnie uwarunkowana dostępnością danych. Pozornie w przypadku kolejowych przewozów pasażerskich macierze podróży mogłyby być budowane na podstawie statystyk sprzedanych biletów, określających rozmieszczenie źródeł oraz cele podróży i ich potencjały. Swobodny dostęp do danych klęowych jest najczęściej silnie ograniczony, a statystyki bilete są traktowane jako tajemnice handlowe spółek przewozowych. Ograniczona dostępność danych kolejowych utrudnia prace nad modelami kolejowymi, a zwłaszcza modelowanie podróży na krajowym i międzyregionalnym poziomie. W artykule przedstawiono propozycję metody budowy modelu kolejowego z wykorzystaniem ogólnodostępnych danych publikowanych przez spółki kolejowe.

W artykule przedstawiono wyniki uzyskane w ramach projektu „Zasady prognozowania ruchu drogowego z uwzględnieniem innych środków transportu” (INMOP 3), w tym opisano (1) metodykę budowy modelu generacji podróży kolejowych dla trzech typów pociągów (regionalnych, międzyregionalnych i międzyaglomeracyjnych), uwzględniającą ogólnodostępne zmienne, takie jak: liczba mieszkańców w otoczeniu stacji kolejowych, poziom ekonomiczny wyrażony wielkością PKB na mieszkańca, stopień zmotoryzowania mieszkańców, liczba firm, wielkość bazy turystyczno-noclegowej oraz liczba pociągów na dobę; (2) metodykę dystrybucji podróży odpowiadającą w. rodzajom pociągów, bazującą na wielkości ruchu generowanego w rejonach oraz czasie przejazdu pomiędzy rejonami oraz (3) proces kalibracji wykonany standardową procedurą z wykorzystaniem analizy zgodności histogramów podróży otrzymanych z macierzy podróży rzeczywistych oraz otrzymanych w wyniku obliczeń.

W rezultacie zaproponowanego procesu generacji i dystrybucji podróży kolejowych osób oraz procesu kalibracji przedstawiono macierze wynikowe dla podróży regionalnych, międzyregionalnych i międzyaglomeracyjnych oraz macierz łączną.

Słowa kluczowe: kolej, modele generacji podróży, modele dystrybucji podróży, macierze podróży osób, model kolejowy