City railways – who are their passengers?

Oleg Pokusaev¹, Alexander Chekmarev¹ and Dmitry Namiot²

¹ RUT (MIIT), 127994, 9b9 Obrazcova Street, Moscow, Russia
² Lomonosov Moscow State University, 119991, Leninskie Gory Moscow, Russia

e-mail: dnamiot@gmail.com

Abstract. In this paper, the approaches to forecasting of passenger flows for new lines of the city railway are analyzed. The article analyzes the experience gained in designing of Moscow central diameters, in the modeling and analysis of their integration with the Moscow Metro and ground urban passenger transport. The main attention is paid to how modern telecommunication technologies make it possible to measure various movements of mobile subscribers, who are passengers, quite accurately. It is the development of telecommunication technologies that is changing the approaches to solving transport problems. What used to be seen as a goal for most transport tasks and was a forecasting goal has now become raw data and can be accurately measured. The article considers what the analysis of transport flows in a modern city now looks like.

Keywords: Smart City; smart mobility; urban railways.

1. Introduction

The article describes the experience gained at RUT (MIIT) in forecasting traffic flows for city railway lines. This urban railway is the so-called Moscow Central Diameters (MCD). The lines of this railway are radial lines of suburban railways, which are connected in this project through the city center and have multiple points of intersection with the metro. MCD is the largest transport project in Moscow and the Moscow region. In practice, it is a new land-based metro, which combines the format of suburban trains and the traditional metro in the capital (Fig. 1).

The project description notes that the length of the main railway tracks within the Old Moscow exceeds 250 km. There are 9 railway terminals in the city that accept electric trains from 11 directions. At the same time, historically, the overwhelming number of passengers use the railroad only as a delivery vehicle, changing later on to the subway. The use of the railway for intra-city travel was extremely low.

At the same time, it was noted in the RUT studies (MIIT) that the railway lines were completely isolated. This gives independence to the traffic situation in the city. As it was repeatedly noted in international studies, today the railway is the only type of transport that can provide mass transportation without traffic jams. At the same time, rail traffic has huge development potential. The rolling stock on the railway has a large size, and trains have a longer length compared to the subway. It is possible to construct additional tracks in almost all directions, which will significantly increase the capacity.
Figure 1. The Moscow Central Diameters [1]

Justification (evaluation of socio-economic effects of the project) includes significant saving of passenger time, new parameters of transport accessibility of different districts of the city and region, improvement of ecology, creation of new jobs, etc. Initial assessments showed the possibility to reduce traffic by 10-12%. More details can be found, for example, in the paper [2].

The rest of the paper is organized as follows. In Section 2, we discuss the source data for our prognosis. In Section 3, we provide our analysis. And Section 4 is devoted to the conclusion.

2. On source data for prognosis

Naturally, this is the most important issue for any forecasting tasks (not only transport). For the railway in the Moscow region, a passenger shows (scans) his ticket twice – at the entrance and at the exit. This is anonymous data, there is no information about the passenger, there is time, ticket type and, accordingly, the starting and ending station. In terms of social networks, there are two check-in/check-out marks [3]. In transport analysis, it is called the correspondence matrix (in Russian papers) or OD matrix (origin – destination) in English papers [4]. In other words, for the railway, we have a complete correspondence matrix. This is a very important point, because in many transport – related papers such information is a goal. The construction of the correspondence matrix is the final task. Here, in our case, the matrix of correspondences becomes the initial data. This information can be analyzed by itself, only instead of forecast tasks (there is no need to predict what is accurately measured), for example, there are tasks of behavior analysis. There are examples of such studies in [5, 6].

When moving to the metro in Moscow, passengers show their travel documents (pay for travel) once, at the entrance. The final point of the route can be assessed using heuristic methods, which are applicable in case of electronic payment. In Moscow, such a tool is a Troika payment card. The meaning of heuristic is as follows. If the use of the card with the number $N$ at station $A$ at time $t_1$ is
marked, it is necessary to determine its next use at some station \( B \) [7] at time \( t_2 \). It is very likely that the route was \( A \rightarrow B \), after which the passenger was at home or at work (depending on the time of a day). Then the time interval \( t_2-t_1 \) includes the travel time, which for the subway can be accurately calculated and the passenger's time outside the transport system.

Telecom operators' data can be used for accurate travel measurements. Mobile communication support in the Moscow Metro means having base stations directly in the metro. This, in turn, means that a mobile operator can determine exactly when a passenger's mobile phone switched to work with a base station in the metro (a passenger entered the metro) and when the same phone switched to work with a base station in the city. The location of the base stations is known, which allows determining the exact entrance station and exit station. Thus, it turns out that for the metro all movements are also fully known. If we consider the metro itself, for any moment of daytime the number of mobile users (passengers) moved between any two pairs of stations is known. Therefore, there is also no need to make a forecast, because we know exactly the traffic between the stations. It should also be noted that the correspondence matrix, in this case, works as some kind of measuring tool, which reflects the events taking place in the city. For example, a new bus route from the suburbs has appeared, which caused a new flow of passengers at the entrance in the morning, the bus station has been moved from one station to another – these events cannot be predicted by the metro correspondence matrix. It is only possible to see some changes in the movement patterns and suggest looking for explanations for them. It means that different methods of correspondence matrix construction, which are considered as the final task of transport flow analysis [8], are not relevant (again, because there is no need to predict accurately measured). The methods of detecting flow characteristics again come to the fore [9].

Telecom operators can evaluate subscriber movements and are simply linked to certain geographic areas [10]. The minimum size of such an area is determined by the density of equipment placement. It is also necessary to take into account that in order to maintain privacy (in this case – the inability to track individual movements) the data will be grouped. This approach is called k-anonymity [11]. The real size for the Moscow region (data, available for our research) is a square of 500×500 meters. The correspondence matrix describes the movements from one such square to another within a selected time interval.

An important point in this reasoning is that this dataset means that all the movement flows are actually known. They do not need to be predicted, they are precisely measured. In the case of geographical squares, certainly, they should be mapped somehow on transport possibilities. Moving a known number of mobile subscribers between two squares must be done somehow. It has to be in line with the existing transport possibilities. In other words, the shortest route from the center of one square to another must be taken into account and well as the load of the transport involved in this route. Someone will "pick up" the subway, someone ground transport, the load of which at landings is taken into account using the same Troika payment card, the rest are those who moved by car. Thus, if we talk about the forecast, it is only in terms of determining the number of those who moved by car. More precisely – those who are not covered by public transport. But these will also be sufficiently determinate calculations.

3. **On data analysis**

Obviously, after the launch of the city railway, the "old" passengers of the railway will continue to use it. For them, this launch brings the only convenience in the form of interval traffic and comfortable trains.

Changing centers (connection to the subway) means integration of railway stations into the subway system. Accordingly, users moving between two stations can follow part of the track on the railway. All the pairs of metro stations for which the use of the city railway (as a part of the track or even the whole track) are known. They can be collected from a combined metro and MCD scheme. An example of such a scheme is shown in Fig. 2. This is static information, it is known before the movement and depends only on the topology.
Assuming that passengers will always choose a shorter route (it is necessary to take into account, of course, transfer time), it is possible to determine the movements to be used by MCD. Flows at metro stations are known by direction (see the previous section), existing flows at railway stations are also known. Accordingly, new flows can be identified. This also results, to a large extent, in a deterministic calculation. Thus, the calculation of possible switching of metro passengers to the use of the railway does not cause technical difficulties, it was carried out in the early stages of the project justification at the RUT Center for High Speed Transport Systems [12].

![Combined metro map](image)

*Figure 2. The combined metro map [13].*

More interesting is the question of the possible number of new (attracted) passengers. In the literature, it was noted that the city railways were the only means of transport that could encourage some drivers to refuse to travel in their car. And this is one of the urgent tasks for a smart city – to limit the use of cars [2].
Here we can offer the following approach. It is necessary to determine the total number of potential passengers who move (according to operators) from the area of station availability. By European standards, it is, for example, about 3 km from the station. The total number of people leaving the area during the day is the potential maximum for the respective railway station. The number of passengers entering the station (and their distribution by time) is known. Accordingly, it can be understood what percentage of all possible passengers is currently served by rail transport. This is the second point in the analysis. It is illustrated in Fig. 3.

**Figure 3. Area to area movement**

In general, there is no reason for this percentage to vary greatly by stations. Accordingly, it is likely to be leveled off. Stations with low percentages are likely to experience a higher influx of passengers (their coverage percentage is likely to increase).

Since different cities in the world already have experience in putting urban railways into operation, there are also calculated a posteriori data on the number of car passengers who have switched to using the railroad. This figure ranges from 15-20%. Accordingly, for the residents of the suburbs, we can get two figures for each station. First, it is the maximum possible number of passengers on the railway based on the proportion of trips not covered by the railway. This would be the upper estimate. Secondly, we can get a lower estimate as 20% of the number of trips not covered by the railway.
The suburban railway to MCD operated with a daytime technological break. Accordingly, there are no entrances and exits in the statistics during the daytime. As one of the demand criteria, it is possible to use the analysis of inputs immediately after the break. To what extent do these figures exceed (if they exceed) the average values? Large excesses are likely to be an indicator of "deferred" demand, which was interrupted by the technological break, and they will indicate in favor of the overall growth in the use of this station.

A special case is formed by the end-stations. They can be the point of entry into the transport system of Moscow for residents of adjacent districts who can get to these stations, for example, by bus or by car. An example is the Lobnya station as the end-station of MCD1 and the Dmitrovsky district lying to the north. Here, potential passengers should be assessed on a larger radius than 3 km.

Technically, movement data (correspondence matrix) is CSV files where geographical areas are defined by references to their representation in QGIS. QGIS is an open source geo-information system [14]. It is a good choice in comparison with the really expensive systems like ArcGIS [15]. Using QGIS as a base, it is possible to create semantic layers and describe movements not in terms of 500x500 meters geographical squares but in terms of objects (in our case, object of transport infrastructure). For example, displacements from the area (as a set of squares) around a particular railway station $N$ to an area (as another set of squares) near some station $M$.

It is the analysis of the possible proportion of those who move from the vicinity of the station that serves as the traffic estimation.

4. Conclusion
In this paper, we describe the approaches to the evaluation of the traffic of the city railway. First of all, the results obtained during the studies carried out under the auspices of the Centre for High-speed Transport Systems were considered. The technical basis for making the forecast is the data of mobile operators, which show information on the movements of mobile subscribers in the city and its suburbs. The data are currently provided as CSV files linked to the QGIS system. The semantic level added to the system allows constructing such movements already in terms of transport objects. The main elements of the semantic level are the transport infrastructure elements. As a result, information about movements between areas tied to railway stations and, for example, subway stations appears.

References
[1] Invest Moscow http://en.investmoscow.ru Retrieved: Apr, 2020
[2] Namiot D et al 2018 On the assessment of socio-economic effects of the city railway International Journal of Open Information Technologies 6 (1) pp 92–103.
[3] Namiot D and Sneps-Sneppe M 2015 The physical web in smart cities Advances in Wireless and Optical Communications (RTUWO) pp 46–49.
[4] Li X M, Huang G Z and Tang J 2008 Passenger flow forecasting based on OD-matrix estimation model Journal of the China Railway Society 30 (6) pp 7–12.
[5] Namiot D, Pokusaev O and Kupriyanovsky V 2019 On railway stations statistics in Smart Cities International Journal of Open Information Technologies 7 (4) pp 19–24.
[6] Misharin A et al 2018 On Passenger Flow Estimation for new Urban Railways IOP Conference Series: Earth and Environmental Science 177 (1) p 012012
[7] Namiot D and Sneps-Sneppe M 2017 A Survey of Smart Cards Data Mining AIST (Supplement) pp 314–325.
[8] Jianfeng L, Jinhai L, Feng C and Yanqiu Z 2010 Review on station-to-station OD matrix estimation model and algorithm for urban rail transit 2010 second international conference on computer modeling and simulation 3 pp 149–153
[9] Nekraplonna M et al 2019 Metro correspondence matrix analysis International Journal of Open Information Technologies 7 (7) pp 68–80.
[10] Steenbruggen J, Tranos E and Nijkamp P 2015 Data from mobile phone operators: A tool for smarter cities? Telecommunications Policy 39(3-4) pp 335–346.
[11] El Emam K and Dankar F K 2008 Protecting privacy using k-anonymity Journal of the American Medical Informatics Association 15(5) pp 627–637.
[12] The RUT Center for High Speed Transport Systems http://rut.digital Retrieved: Apr, 2020
[13] An interactive metro map https://www.moscowmap.ru/metro.html Retrieved: Apr, 2020
[14] QGIS https://qgis.org/en/site/ Retrieved: Apr, 2020
[15] ArcGIS https://www.arcgis.com/index.html Retrieved: Apr, 2020

Published under licence in Journal IOP Conference Series Earth and Environmental Science