On Passenger Flow Estimation for new Urban Railways

A Misharin¹, D Namiot² and O Pokusaev³
¹ RUT (MIIT), 9b9 Obrazcova str., Moscow, 127994, Russia
² Lomonosov Moscow State University, Leninskie Gory Moscow, 119991, Russia
³ Center for High-Speed Digital Transport RUT (MIIT), 9b9 Obrazcova str., Moscow, 127994, Russia

e-mail: dnamiot@gmail.com

Abstract. In this paper, we discuss the questions, associated with the forecast for passenger traffic for urban railways. The aim of the study is to select and verify the model for predicting passenger traffic of new urban railways. The article is based on the practical tasks implemented during the project phase for new urban railways in Moscow, Russia. We are considering data sources for building the forecast, as well as practical models that can be used to obtain numerical estimates. Among the discussed data sources, we target migration data that can be collected with the help of telecommunications operators, and information on the use of public transport, obtained from the validation of transport cards. Also, in this paper, we investigate the metrics for traffic along the new city rail line, which can be determined on the basis of the projected passenger traffic. The result of the work was the constructed model of the transport behavior of passengers, taking into account the availability of new urban railways and a set of metrics for assessing the functioning of this transport tool.

Keywords: Smart City; smart mobility; urban railways.

1. Introduction
It is a well-known fact that mobility (smart mobility) is one of the key elements for Smart Cities. Urban railways recently attracted much attention, because they are almost the only transport in the city, which can move without traffic jams. This article discusses the issues of modeling of passenger flow for urban railways. City railways, similar to other transport projects in cities, are quite expensive enterprises, which require comprehensive calculations and justifications. An expected traffic is, of course, one of the key parameters. This traffic is the basis for calculating social and economic effects from the railroad, many of which are based on the expected traffic [1]. For example, the predicted traffic allows us to estimate the revenue from the future passenger transportation. Also, it allows us to estimate saving time by passengers from the introduction of new routes, it lets us estimate a decrease in transport fatigue, reachability of new areas, etc. In this connection, we could mention once again the paper [1], where we, in fact, considered various tasks assessment of social and economic effects.

The objectives of this study were the selection and verification of a model suitable for estimating (predicting) the traffic of new urban rail lines, selecting initial data for modeling, selecting metrics for assessing the functioning of the constructed transport lines. This is the first such kind of research for Moscow.

The rest of the paper is organized as follows. In Section 2, we discuss research methods. In Section 3, we provide our analysis. And Section 4 is devoted to the conclusion.
2. On passenger flow models

This paper based on some work performed during early stages of Moscow radial urban railways project (so-called Moscow diameters). The base element for passenger flows prediction is data on migration within the city. Naturally, passengers are traveling residents or visitors, some of whom (and it is what will be evaluated in our model) could take advantage of a new mode of transport (a new railway line in this case).

Two basic approaches to assessing migration data could be mentioned here. As per the US, for example, the good sources of migration data are surveys (census), which the statistical service collects in the US. Answers to polls are required by law. These surveys include data on migration too [2]. As a good example of such data, we could mention paper [3] with collected statistics for San Francisco. The paper [4] presents detailed information on surveys in the Bay Area (California). At the same time, the accuracy of the information is quite high. Information is collected (monitored) at the level of city blocks (there are about 4,000 blocks in San Francisco). Accordingly, the correspondence matrix (it is a Russian terminology, US scientists use origin-destination or OD matrix abbreviation) marks the movement from block to block. It is illustrated in Figure 1.

![Figure 1. Correspondence (OD) matrix [2]](image)

We should have a similar approach in mind, perhaps we should strive to create something like this in Russia, but so far we have nothing like it. Actually, the survey data is available even in the US not in all places. In such case, we need some automatic methods collection of information on movements. Unconditionally, the leader here is the data of mobile operators. The typical examples of this kind of research are presented in papers [5–7]. Note that due to telecommunication standards, the technical possibilities in this part in Russia are no different from what is available in the US or Europe. Data that are collected by telecommunication operators, technically, can describe different aspects of behavior for mobile subscribers that could be used for creating the same correspondence (or OD) matrix.

One of the first deployment models for telecom data is linked with the analysis of user activity, measured in Erlangs (1 Erlang is the use by a subscriber of a network during 1 hour). With this model it possible to determine areas of activity for users in the city (it is so-called social dynamics). The idea of use is very simple – if we register, for example, calls from a specific phone from the same place in the morning and in the evening, and then in the daytime from another, we can tell how the route looks in the morning from home to work for this phone (for its owner). For example, in [2] algorithm analyzed places for phone calls and it was assumed that time from 21:00 to 6:00 corresponds to a call
from home, and time 9:00 to 17:00 corresponds to a call from work. It, accordingly, gave one route. In the paper [8], authors use a slightly different approach. For example, their algorithm assumes that the location from which the user leaves in the morning and to which he returns at night is his home. It is also assumed that location of the longest repeating stay on weekdays is a workplace of this mobile subscriber. This algorithm describes the construction of a general model of mobility in the city of Boston, where according to the operators’ data house-work pairs have been restored for 75% of subscribers. So, such model can predict user moves from a 10-minute interval.

Other ways of collecting parameters (for example, Location update, Handover [9]) allow come, as a result, to the same matrix of correspondence (starting and ending points in mobile phone movements). Note, that the end-point for the particular route can be defined empirically, as a certain time interval during which the mobile device remained in place. Basically, it is a complete analogy with the concept of a web session. A visit to a website by a specific user is also determined (completed) by time analysis for inactivity of the user [10].

The next important thing is to get the metrics for the passenger flow on the railway. The idea was borrowed from [2] – it is the distribution of travels by duration (Fig. 2).

As the next characteristics, we used the centrality [12]. The idea is that the structural properties of the network determine the flow through the network. The centrality for node \(v\) could be defined as:

\[
b_c(v) = \sum_{s,t \in V, s \neq t} \frac{\sigma(s \rightarrow t | v)}{\sigma(s \rightarrow t)}
\]

where \(\sigma(s \rightarrow t | v)\) is a number of shortest paths from station \(s\) to station \(t\), passed through \(v\), and \(\sigma(s \rightarrow t)\) is a total amount of shortest paths from \(s\) to \(t\).

And according to the research, the traffic passing through the station correlates with its centrality. The more shortest paths pass through the station, the more passengers travel through it. That makes it possible to predict the volumes of passengers, since the characteristics of the nodes can be calculated statically (based only on the network topology).

![Figure 2. Distribution of travels by duration [2]](image)

The centrality metric has also been used to analyze the resilience of a network [2]. By removing key nodes or edges between key nodes in the network (in the graph), it can be observed that the flow patterns through different paths will be highly affected. This remove operation will simulate the effect of disruption in real physical transportation networks (for railways it can simulate train breakdowns, for example). So, it could help to city administration (to transport department) identify whether
transport system incorporates sufficient design robustness as well as justify its updates to overall resilience.

3. On proposed prediction model

Our model is based on migration data obtained from telecom operators and information about smart cards validation in public transport. In our case, it was possible to restore origin-destination movement for mobile devices in a square with a side of 250 meters.

![Figure 3. Public transport and railroad](image)

After obtaining (restoring) the information about the movements (migration data), we must allocate (share, distribute) these trips to the different modes of transport.

As the first step (it's the simplest one), we can separate pedestrians. This, as a rule, is based on a simple assumption about the length of the route. If this length (and routes begin in a block and end in some other block – see Fig. 1) is less than some boundary value (for example, 1 km), it can be assumed that this is a walking route.

The remaining "migrants" are either passengers for public transport or take advantage of cars. The question about the proportion of drivers (motorists) in this flow is, of course, debatable. The above-mentioned papers (e.g., [2]) again used data from polls (Boston, San Francisco), where residents were directly asked for the ways they are moving around the city. For Moscow, such data is not available.
but we can use the advantage of smart cards. The most accurate way to determine the number of people using cars is to take all the transport migrants and remove (deduct) the public transport migrants. And for public transport passengers, we can use data from smart cards (travel card Troika, in Moscow, for example) validation. Transport cards in Moscow are using for landing (entry) only. It is an analogue of the check-in (the presence) in social networks [11].

Naturally, some of the users of cars will switch to the railway. This is noted in all works on urban transport (see [1], for example). In reality, the city railway is practically the only transport that is not subject to traffic jams. It is the reason, why drivers could switch to urban railways. The percentage of switched drivers is debatable (see review [1] again). The historical data is 5–20 %.

After receiving a real number of public transport passengers (we took into account both pedestrians and drivers), we need to distribute them according to the types of public transport. We assume that each passenger will choose the shortest route. Accordingly, for the distribution of passengers, we need a geo-information system with all stops of public transport (see Fig. 3).

On this geo-information system, we should place a new railway line too (bold line in Fig. 3). On this geo-information system, we should place our squares for which the movement routes were determined. In the process of determining the shortest paths for passengers, those who will use the railway will be identified (their shortest path implies moving by rail). Naturally, in this case, the used railway stations will be accurately determined too.

We can also calculate the movement and without a railway. After that, we can calculate the difference in time that will be spent by passengers on trips without taking into account the railway and with it. This difference presents a time that passengers will save on a daily basis in case of putting the road into operation. Multiplying this time by the average hourly wages, we can get a direct economic effect, expressed in money terms.

4. Conclusion
In this article, we have described a complete diagram of the representation of the passenger flow for a new urban rail line. The model is based on migration data, which was estimated (collected) with the help of mobile operators. Data on movements from mobile operators are compared with data on the use of travel documents in public transport. In the work, we managed to build a mashup for telecom data and data on the validation of travel documents. This allows, in the end, to identify both users of public transport, and those who use the car to travel around the city. To identify those passengers of public transport who will use the new railway line, a geo-information system with stops and public transport lines is used. This system allows us to calculate the shortest travel paths for passengers. We define the shortest path as the path which incurs the least amount of travel time between two stations. By the criterion for choosing the shortest path, the possible passengers of the new railway line are identified. The predicted passenger traffic allows us to calculate such metrics of the new railway line, as the distribution of the duration of the routes and centrality for the stations. The proposed model has been verified on the existing lines of the city railways, when the results of the simulation were compared with the actually recorded passenger traffic.

References
[1] Namiot D et al 2018 On the assessment of socio-economic effects of the city railway International Journal of Open Information Technologies 6 (1) 92–103
[2] Wang J et al 2013Vulnerability analysis and passenger source prediction in urban rail transit networks PloS one 8 (11) e80178
[3] 2010 US Census Bureau
[4] 2010 Bay Area Census
[5] Steenbruggen J et al 2013 Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities GeoJournal 78 (2) 223–43
[6] Ratti C et al 2006 Mobile landscapes: using location data from cell phones for urban analysis
Environment and Planning B: Planning and Design 33 (5) 727–48
[7] Calabrese F et al 2011 Real-time urban monitoring using cell phones: A case study in Rome IEEE Transactions on Intelligent Transportation Systems 12.1 141–51
[8] Mearian L 2016 Ford, MIT use Bostonians’ cellphone location data for traffic planning Computerworld
[9] Calabrese F, Ferrari L and Blondel V 2015 Urban sensing using mobile phone network data: a survey of research Acm computing surveys (csur) 47.2 25
[10] Namiot D and Sneps-Sneppe M 2014 On the analysis of statistics of mobile visitors Automatic Control and Computer Sciences 48.3 150–8
[11] Namiot D and Sneps-Sneppe M 2011 Customized check-in procedures Smart Spaces and Next Generation Wired/Wireless Networking (Springer, Berlin, Heidelberg) 160–4
[12] Ramli MA et al 2014 A method to ascertain rapid transit systems’ throughput distribution using network analysis Procedia Computer Science 29 1621–30

Published under licence in Journal Title by IOP Publishing Ltd