GIS-monitoring of Regional Transport Network Traffic as a Method to Study Commuting: Moscow Region Case

P A Mitroshin1*, Yu Yu Shitova2, Yu A Shitov3, A A Mitroshin4 and D N Vlasov5

1 Dmitrov Institute of Continuing Education branch of Dubna State University, 23, mikrorayon DZFS, Dmitrov, Moscow Region, Russia
2 Russian State University for the Humanities, Moscow, Chyanova str,15, Russia
3 Joint Institute for Nuclear Research, Dubna, Jolliot Curie, 6, Moscow Region, Russia
4 Dmitrov Institute of Continuing Education branch of Dubna State University, 23, mikrorayon DZFS, Dmitrov, Moscow Region, Russia
5 Dubna State University, Dubna, Universitetskaya,19, Moscow Region, Russia

*Email: mitpa@yandex.ru

Abstract. The paper is aimed to analyse the labour mobility in the region using GIS services. Within the framework of the project, a methodology for continuous automatic collection and accumulation of information on the state of the transport network and weather conditions was developed, tested and implemented. Based on this information, it is possible to analyse the important factors of the regional economy: temporary and financial losses of residents of the Moscow Region on the home-to-work paths, as well as identify patterns of traffic congestion from various factors. The first results of the analysis of the accumulated data are presented, which demonstrate the commuting effects in the regional transport network.

1. Introduction

The study of the mobility of commuters in the region is an important task of the theory and practice for regional management, while the lack of primary data is the main problem. Software platforms of geographic information systems (GIS) are complex work tool which can be used in scientific research for a variety of tasks.

Research of the mobility of commuters is one of the urgent task in present days. In work [1], using Swedish data (GeoSweden), the length of a car trip via a Google router (Google Map API) was compared with direct distance (Euclidean length) and the length of public transport trips (API "ResRobot", an online travel planning service - collaboration of Samtrafiken, Stockholms Lokaltrafik and Viktoria ICT, http:\\ www.trafiklab.se, where you can request the shortest path between two geo-points at a specific time of the day and day). Reaserch which used a sample of 704 thousand people living in the metropolitan area of Stockholm gave the result that in Stockholm 2014 the share of commuting aged 25-65 years is 58%.

In Swedish statistics, the locations of work and residence of people are available with an accuracy of 100 m2. Elder et al. (2012) [2] aggregated this data in cells of 500 m2, combined it with the database (DB) of the Swedish road system (timetables and routes of public transport) and using GIS software calculated and compared routes for private and public transport of different groups (high-paying men and low-paid women) 4500 hospital workers in Gothenberg (Sweden).
Dauth and Haller (2016) [3] studied changes in wages depending on changes in home-work distance (due to job change). The source is data from the German federal agency which for some time included the exact coordinates of places of residence and work and the distance was considered in the GIS platform OpenStreetMap Routing Machine as the shortest distance to travel between points. Based on the same data, Heuermann et al. (2017) [4] studied employer compensation for center-periphery travel.

Paetzold and Winner (2016) [5] and Frimmel et al. (2017) [6] combined tax data from the Austrian Ministry of Finance and employee-employer data from the Austrian Social Security Database. The GIS system was used to calculate the travel distance to work based on the zip codes of the employee and the employer (the geographic center of the region was taken for each index). The honesty of employees’ declaration of tax benefits was investigated. As a result, 30% of declarations were overstated.

The study [7] investigated the study trips of students of the Ritsumeikan University (Ritsumeikan, Kyoto, Japan). Residence address data was collected from an online university study and then student routes were reconstructed using ArcGIS 10 Network Analyst taking into account the public transport network (from HokkaidoChizu Co., Ltd., from National Land Numerical Information of Japan), followed by geovisualization. The goal is to determine the spatial patterns of trips and the location of traffic jams.

Study [8] uses a GIS approach to study the choice of travel of schoolchildren - between the bus and the car of the parents in Dublin (Ireland). The main factors of choice are distance, car availability and classmates living nearby.

Study [9] GIS models the home-work path for a sample of 15% (220,000) Irish people (known as the POWAR DB).

Thus, the use of GIS-approaches for solving economic problems is a relevant modern trend.

One of the main tasks solved within of this project (the previous stages of which are described in [10, 11]) was the development, debugging and launch of an automatic data collection system for continuous monitoring of the transport network of Moscow and the Moscow region based on information provided by the Yandex.Maps GIS platform. Obtaining of primary data makes it possible for further analysis and processing of the information collected to identify patterns and correlations for various factors.

A large amount of primary data and the results of their processing, based on the workload of distributed traffic routes, weather conditions, emergency situations and time intervals, makes it possible to accurately predict based on the use of machine learning methods.

2. Methods of Collecting Information on the Workload of the Transport Infrastructure
The study of commuting in the region is an important task of the theory and practice of regional management while the lack of primary data is the main problem. Primary data collection of the traffic infrastructure using is a laborious process. Today, GIS systems are the best way for obtaining traffic information. Based on the services Yandex.Maps and Google.Maps it can be getting up-to-date information about the load of the transport infrastructure at any time. Web-based mapping services provide an opportunity to obtain information by using specialized sets of services, which allows to apply cartographic data and technologies in different solutions. Such sets of specialized services are called API Yandex.Maps and API Google.Maps. As part of the study, the transport network of Moscow and the Moscow region was selected. The study the Yandex.Maps service is used to collect information on road congestion. An important factor in favor of choosing Yandex GIS services is a large audience of users in the source region, which gives more accurate results.

2.1. Formation of Initial Data of Commuting.
For the subsequent collection and accumulation of initial data, the structure of stored parameters was determined, which characterizes each conditional employee moving within the study region:

- `mtm_id` - the unique identifier of the employee. Key field to link to basic traffic data;
- `h_lat` - longitude of residence;
- `h_lon` - latitude of residence;
• w_lat - longitude of the place of work;
• w_lon - latitude of place of work;
• dist_id - group id by distance 1/2/3 – 20/50/80 км;
• rad_id - Radial (azimuth) direction identifier from Moscow. The number corresponds to the clock on the round dial. For example, 3 - means the direction for three hours (+/- 15 degrees, the general cone is 30 degrees), that is, east of Moscow. 6 hours south, etc. Needed for grouping by direction (Figure 1);
• r_0 - Direct distance between home-to-work (measured in kilometer);
• path_min - minimum reconstructed distance between home and work (measured in meter);
• ut_min - Minimum specific travel time of one kilometer (measured in minute/meter).

An example of the received data looks like this:
1^55.9318^37.6156^55.7453^37.6114^1^6^20.7487^23752.1^0.905298
2^55.8937^37.4198^55.7622^37.6131^1^8^18.9626^20781.2^0.939132
3^55.9068^37.4539^55.7813^37.6453^1^8^18.3719^23767.8^0.897452

, where the character "^" is a separator of the previously defined collected parameters.

![Figure 1. Distribution of commuting by region and zonal division of the sample.](image)

The initial sample of conditional commuting for further monitoring of the transport network of the Moscow Region consists of 20 thousand people living at radii of 20, 50 and 80 km from Moscow (Figure 1) and working in the center (within a radius of 5 km from point $C_0 = [55.7520; 37.6166]$). The distances are chosen to monitor the near, middle and long distance mobility of the inhabitants of the agglomeration. There is no need to take the coordinates of real people, since the GIS platform based on Yandex.Traffic services builds a path along real roads between any two points specified on the map, which means that it takes into account the state (congestion) of real roads in its calculations.

2.2. Collection and Formation of Data on Road Congestion
An important component of the study is the collection, formation and analysis of data on the congestion of highways along the home-to-work and work-home routes, for each conditional participant during commuting. For each of the residents of the sample, according to the given GIS coordinates of theirs living place and work, the home-to-work path is reconstructed taking into account traffic jams and without taking into account traffic jams at the current time and date when the reconstruction is carried out (figure 2).
Figure 2. Reconstruction of the route using GIS services.

This technique was developed and successfully applied in previous studies (RFBR grants No. 11-06-00323 and No. 14-06-00249). In particular, was created a software package that allows this operation to be carried out in a semi-automatic mode for large amounts of data. Information characterizing the workload of the transport system when forming a work-to-home route is formed in the form of the following parameters:

- **mtm_id** - employee unique identifier;
- **Datetime** - date of data collection in MySQL format YYYY-MM-DD HH:MM:SS;
- **path** - length of the route between home and work, reconstructed by the Yandex.Maps service router taking into account the transport network (measured in meters);
- **time_jam** - travel time by car, taking into account traffic jams (measured in minutes).

A variable generated for each route in a single copy as a constant:

- **time_wojam** - travel time by car without traffic jams (measured in minutes).

To collect and store data obtained as a result of queries to the Yandex.Maps service, a database management system based on the freeware MySQL software is used. The principle of data collection is to contact the Yandex.Maps service on an hourly basis using a specialized API and obtain information about the route load. For all conditional participants in the sample of 20,000 people, a route is built and the time of its passage is estimated, after which the time of the API services query and the traffic load of the route with its length are entered into the database. The number of records and the amount of data received since 2016 allows us to assert that BigData approaches will be used in processing and analysis. The received data is entered and stored in the following form:

1^2016-12-17 11:50:04^28212.8^3459.79
2^2016-12-17 11:50:04^25842^2774.03
3^2016-12-17 11:50:04^24399.7^3139.2
4^2016-12-17 11:50:04^22408^3310.98
5^2016-12-17 11:50:04^28073.7^3412.5

where the character "^" is a separator of the previously defined collected parameters.

3. Analysis and Systematization of Data

In the early stages of collecting information, it was assumed that the route is always formed the same. That is, it was initially believed that for a fixed pair of coordinates, the Yandex.Maps router always returns the same route (either with or without traffic jams), and only the estimate of its travel time differs. Then, the assessment of losses would be quite simple for each participant in labor migration:

\[
Time_{losses}[i] = time_{jam}[i] - time_{nojam} \tag{1}
\]
where \( i \) is the number of records in the database on the congestion of the transport network on a particular route.

At the stage of information analysis, it turned out that, depending on the workload of transport nodes, the service of Yandex.Maps returns various routes with the optimal travel time. Thus, formula 1 has lost its relevance for research. The time to complete the route without traffic jams \( \text{time}_{\text{no jam}} \) is the minimum specific travel time for this route, taking into account the traffic jams in all statistics.

\[
\text{time}_{\text{no jam}} = \text{ut}_{\text{min}} [\text{mtm}_{\text{id}1}] = \text{MIN}(\frac{\text{time}_{\text{jam}}[1]}{\text{path}[1]}, \frac{\text{time}_{\text{jam}}[2]}{\text{path}[2]}, \ldots, \frac{\text{time}_{\text{jam}}[N]}{\text{path}[N]})
\]  

(2)

In addition, for the convenience of information perception, it can be used the reciprocal for \( \text{ut}_{\text{min}} \), the maximum speed of the route:

\[
\text{ut}_{\text{max}} = \frac{1}{\text{ut}_{\text{min}}}
\]  

(3)

Considering the fact that the data collection is produced every hour and the data was analyzed for a year so we can conclude that \( \text{ut}_{\text{min}} \) – is the estimate that was obtained at the best moment in time:

* current specific travel time:

\[
\text{ut} = \frac{\text{time}_{\text{jam}}}{\text{path}}
\]  

(4)

* specific losses:

\[
\text{ut}_{\text{losses}} = \text{ut} - \text{ut}_{\text{min}}
\]  

(5)

* to visualize the dynamics of losses, it is convenient to use relative losses equal to 1 in the absence of traffic jams and taking values > 1 in proportion to the road load.

\[
\text{t}_r_{\text{el losses}} = \frac{\text{ut}}{\text{ut}_{\text{min}}}
\]  

(6)

* absolute time loss:

\[
\text{time}_{\text{losses}} = \text{ut}_{\text{losses}} \times \text{path}
\]  

(7)

* monetary losses:

\[
\text{money}_{\text{losses}} = \text{time}_{\text{losses}} \times \text{UdMO}_{\text{salary}}
\]  

(8)

where \( \text{UdMO}_{\text{salary}} \) – is the unit price of a working hour in the Moscow region, which is calculated based on a 40-hour working week (9600 minutes):

\[
\text{UdMO}_{\text{salary}} = \frac{\text{Avgsalary}}{9600}
\]  

(9)

, where \( \text{Avgsalary} \) – is the average salary in the Moscow region.

### 4. Visualization Tools for Collected Information

Visualization of the received and processed information is an important task when working with big data for interpreting the results.

To display the data that collected as a result of the research are using a web-based system released on technologies: JS, CSS, Bootstrap, PHP, MySQL, Yandex.Maps.API, Highcharts. The processed data, at the stage of collecting information on the traffic congestion, are placed in the MySQL database. The software interface is developed by JavaScript using Bootstrap visualisation technology. Maps of the selected routes are visualized using Yandex.Maps services. The results of data processing are displayed both as a separate table and as Hightcharts charts based on PHP queries to the database.
Figure 3. Interface for selecting information output elements.

The user can choose to display the results both for a separate element of the list of studied individual elements of the set of 20,000 conditional respondents, and for directions that summarize the information of the respondents that included into them. The selection of individual respondents from the database is carried out by entering a value in a special field (figure 3). To display the generalized data, it is necessary to select a special marker that is responsible for the direction and distance (figure 3).

The output of the selected information can be seen on the graph and map (figure 4).

Figure 4. Building a route and a schedule for visualizing information by dates.

5. Time Series Examples
The dynamics of losses for interval 2016-2020 describe travel during peak hours which averaged over months for the first group of commuting labor migrants living 30 km from Moscow, is shown in figure 5. This is the roughest estimate of losses that can give us many interesting conclusions.

First, the characteristic annual cycles of road load are clearly traced - decreased in January of each year, followed by an increase in load and reaching a maximum in late autumn and December. Secondly, there is a decreasing of traffic intensity in April-May 2020, associated with the introduction of quarantine measures due to the situation with the pandemic.
A more detailed analysis of the collected data is currently underway, the detailed results of which will be published shortly.

**Figure 5.** Average monthly losses (calculated by formula 6) when traveling during peak hours for residents 30 km from the center, along the home-work routes (black dots) and work-home (red dots). The omissions and increased errors are related to a lack of data that arose due to the blocking of the software by service Yandex. Maps.

6. **Summary**

Thus, the paper has described the problem of systematically collecting a large amount of data about congestion of the transport network of the Moscow region. In addition, a mathematical method for processing the collected data and their interpretation is presented. This project is being carried out in accordance with the global trends in the analysis of big data for solving economic problems of various types, including monitoring population mobility. It should be noted that this topic is being actively developed by Russian and international experts for using in predictive models.

**References**

[1] Blind I, Dahlberg M, Engström G, and Östh J 2018 Construction of register-based commuting measures CESifo Economic Studies 64 (2) 292-326 https://doi.org/10.1093/cesifo/ify014

[2] Ellde’r E, Sola’ A G, Larsson A 2012 Featured Graphic. Spatial Inequality and Workplace Accessibility: The Case of a Major Hospital in Go’teborg, Sweden Environment and Planning A 44 2295–7.

[3] Dauth W, Haller P 2016 The Valuation of Changes in Commuting Distances: An Analysis Using Georeferenced Data IAB Discussion Paper no 43/2016 (The Research institute of the German Federal Employment Agency, Nuremberg, Germany)

[4] Heuermann D, Assmann F F, Vom Berge P, Freund F 2017 The Distributional Effect of Commuting Subsidies - Evidence from Geo-Referenced Data and a Large-Scale Policy Reform Regional Science and Urban Economics 67 11–24

[5] Paetzold J, Winner H 2016 Taking the High Road? Compliance with Commuter Tax Allowances and the Role of Evasion Spillovers Journal of Public Economics 143 1–14

[6] Frimmel W, Halla M, Paetzold J 2017 The Intergenerational Causal Effect of Tax Evasion: Evidence from the Commuter Tax Allowance in Austria Working Papers 2017-01, Faculty of Economics and Statistics (University of Innsbruck, Innsbruck, Austria)

[7] Kazumasa H and Tomoki N and Keiji Y and Shigeru I 2014 Network-based spatial interpolation of commuting trajectories: application of a university commuting management project in Kyoto Japan Journal of Transport Geography 34 274–281

[8] Kelly J A and Fu M 2014 Sustainable school commuting – understanding choices and identifying opportunities Journal of Transport Geography 34 221-230

[9] Meredith D, Charlton M, Foley R, Walsh J 2007 Modelling commuting Catchments in Ireland: A Hierarchical Approach using GIS RERC Working Paper Series 07-WP-RE-12
[10] Shitova Yu Yu, Shitov Yu A, Vlasov D N Digital 2019 monitoring of transport in the Moscow agglomeration using geoinformation technologies *Moscow Government University Bulletin*. 3(45) 54-59

[11] Shitova Yu Yu, Shitov Yu A, Mitroshin P A, Mitroshin A A, Vlasov D N 2019 GIS-monitoring of labor mobility in the region using the example of the Moscow region *Component of Scientific and Technological Progress* 3(45) 45-49