

CDSSES 2020**IV International Scientific Conference "Competitiveness and the development of socio-economic systems" dedicated to the memory of Alexander Tatarkin****GIS-MONITORING OF REGIONAL TRANSPORT NETWORK**

Yulia Shitova (a)*, Yury Shitov (b), Dmitry Vlasov (c),
Anton Mitroshin (d), Pavel Mitroshin (e)
*Corresponding author

- (a) Russian State University for the Humanities, 6, Miusskaya Sq., Moscow, Russia, yu.shitova@rggu.ru
(b) Joint Institute for Nuclear Research, 6, Joliot Curie St., Dubna, Russia, shitov@jinr.ru
(c) Dubna State University, 19, Universitetskaya St., Dubna, Russia, tier6969@gmail.com
(d) Dubna State University, 19, Universitetskaya St., Dubna, Russia, mitaa@yandex.ru
(e) Dubna State University, 19, Universitetskaya St., Dubna, Russia, mitpa@yandex.ru

Abstract

The work is devoted to GIS-monitoring of the transport network of the Moscow metropolitan area (MMA). This direction is an actual trend in the modern regional economy, as it can be seen from a brief overview of the scientific publications presented in work. Special software has been developed for the automatic collection, accumulation, and processing of data on MMA commuters (a sample of 20 thousand virtual people living at a distance of up to 80 km from Moscow) from the region to the center. Data is collected hourly through the Yandex platforms (Maps and Traffic), which allows us to monitor the regional transport network's load status in time and space. Indicators of commuter's losses are discussed, and an analysis of time and financial losses of MMA commuters is carried out. The methods of collecting, storing, and analyzing the data are described, as well as the first results obtained are presented. Prospects for further work are discussed.

2357-1330 © 2021 Published by European Publisher.

Keywords: Big Data, commuting, GIS, regional economy, transport



1. Introduction

The use of GIS technologies for the scientific analysis of transport systems in general, and commuting in particular, continues to be a popular trend. In particular, for an extensive overview of studies using GIS data (Blind et al., 2018). Dauth and Haller (2016) investigated changes in wages depending on home-work distance (due to job change). The source is data from the German federal agency, which included the exact coordinates of housing and work coordinates, 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) investigated employer compensation for center-periphery commuting travel.

Frimmel et al. (2017); Paetzold and Winner (2016) 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 employee's zip codes and the employer (the geographic center of the region was taken for each index). The honesty of employees' declaration of tax benefits for MTM was investigated. As a result, the travel length was overstated in 30% of declarations!

A standard GIS system offers commuting routing with a single vehicle. Djurhuus et al. (2016) propose the implementation of multimodal commuting trips (more than one mode of transport) by integrating information from the planner (timetables) into GIS using SQL. This approach has been successfully applied in Denmark's metropolitan area with very complex multimodal commuting trips: bus, train, light metro, metro, ferry, bicycle, and walking for 30 thousand people.

In (Obaco et al., 2016), GIS systems (QGIS and Stata) were used to define cluster functional urban areas in Ecuador. GIS was used to determine the length and time of travel between given geo-coordinates. In (Rehak & Kacer, 2018), the GIS system is used to assess the proximity of Bratislava districts (according to simulated routes (time and length) from the district to the center using the Google Map API) to the center for the subsequent comparison of this indicator with housing prices determined by sales proposals on the Internet.

The variety of applications demonstrates the fact that GIS technology has evolved into a powerful data analysis tool in economic research. Its use for analyzing the transport network will be discussed below.

Monitoring the region's transport network using modern digital technologies is a strategical direction of the development for a modern metropolis within the framework of the concept of a "smart city". Increasing the mobility of the population, the quality of urban infrastructure, and the quality of management of the urban environment are among the hot topics and important applied problems of the modern regional economy.

2. Problem Statement

Monitoring the region's transport network using modern digital technologies is a strategical direction of the development for a modern metropolis within the framework of the concept of a "smart city". Increasing the mobility of the population, the quality of urban infrastructure, and the quality of

management of the urban environment are among the hot topics and important applied problems of the modern regional economy.

Attempts to build smart city subsystems are being carried out in the world's leading cities (Barcelona, New York, London, Beijing, Moscow, etc.). Permanent monitoring and analysis of the quality of the transport system in the region is in current scientific research. An increasing number of objects of the system (public and private transport, roads, bridges, tunnels, etc.) are equipped with inexpensive GPS devices for recording information (photo and video recorders), which, being used massively (using Big Data technology), will allow conducting a global analysis of the traffic situation. If earlier such studies were quite expensive (Scharenbroich et al., 2009; Zhu & Levinson, 2015), then in the future, their cost will decrease as the cost of the final recording devices becomes significantly cheaper. GIS technologies have been used to construct commuters' traffic flows in New York and Amsterdam (Arribas-Bel & Michiel, 2012). GIS analysis was used to assess the correlation of business geography with the levels of development of territories and infrastructure in India (Vaz, 2013). An interesting and well-developed GIS study was the analysis of the impact on the external environment depending on commuting trips, transport alternatives of models and traffic scenarios based on data from a special transport study by Vodafone, which covered 1829 respondents (Wood, 2003). An interesting analysis of daytime pictures of workers' trips in Flanders (a polycentric agglomeration in the north of Belgium, (Dalton et al., 2015), where the difference in travel times of people during and outside rush hour was investigated, while the routes were simulated in a GIS program.

Overall, it is clear that the collection of big data on the transport network will only grow. Finally, the smartphone became the main generator of data (GPS-sensor, which is always with you!). Mobile telecom operators have complete information about users' mobility, and this data is already used for scientific data analysis (Makhrova et al., 2016).

Thus, the proposed methodology is in the modern mainstream of using big data obtained online. It will be discussed in detail below.

3. Research Questions

Data from the monitoring of the transport network by residents will allow us to study several topical scientific issues.

- Investigation of the cyclicity of the transport network load at different time scales - day, week, month, year. In this case, the qualitative nature will be determined, but quantitative estimates of the process parameters will be made.
- Estimation of time and financial losses incurred by the region as a whole, and people in particular, from traffic jams.
- Analyse the spatial structure and temporal dynamics of the process.
- Learn to predict the transport network state based on the available data using innovative machine learning technologies (a task in the long term).

4. Purpose of the Study

The Yandex.Maps and Yandex.Traffic web platforms (Yandex Map Platform, 2020; Yandex Traffic Platform, 2020) calculate user-specified traffic routes with and without traffic jams via API (application program interface) in real time. Starting the Yandex.Traffic service, the user sees the state of the transport system only for his route (Figure 1, a).

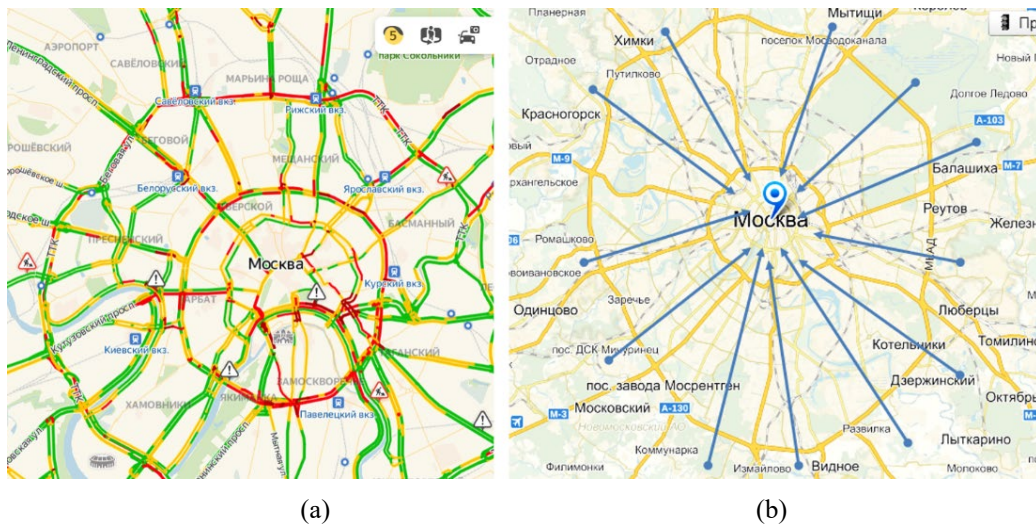


Figure 1. (a) Interactive interface of Yandex.Traffic web application. (b) An example of a commuter sample for monitoring the Moscow Ring Road transport network. Commuters live just outside the Moscow Ring Road and work in the centre of Moscow (the beginning and end of the arrow, respectively, for each commuter)

The purpose of our project is to capture information simultaneously for thousands of virtual users dispersed throughout the region in order to assess the situation with the transport network as a whole. And, accordingly, assess the macroeconomic effect of traffic jams for each resident in particular and the region's economy as a whole.

Next, we move on to a detailed presentation of the research methodology.

5. Research Methods

The methodology is based on generating a basic fixed set of commuters whose home-job paths cover the region of interest (the MMA in this work). And further, regular in time (in this work, hourly) collection of information about the home-work-home paths for all commuters from the basic set. The last step is to analyse the time series accumulated from the base set of commuters.

5.1. Base set and controlled region

The base set includes a set of N virtual commuters with places of residence and work defined by geographic coordinates (latitude and longitude) on the map (GIS coordinates). It makes no sense to set the coordinates of real people (hence virtualization), since the Yandex.Maps router builds a route along real

roads between any two points specified on the map, which is what we need. The main thing is that the places of residence and work of virtual people in the base sample should be chosen in such a way that their home-work routes would evenly cover the transport network of the study area, for example, the Moscow Ring Road (Figure 1, b). The reconstructed commuter routes of the basic sample will cover the entire transport network within the Moscow Ring Road and, thus, take into account the situation on the roads in a comprehensive manner.

5.2. Size of base sample

Base sample size is a trade-off between precision and the complexity of large data. The main problem is that the Yandex platform bans the user's IP due to a large number of requests treating it as a DDOS hacker (Hulme, 2020). As a result, the sample size should not exceed 20 thousand people, while it is necessary to take measures to parallelize the data collection process.

5.3. Basic set fixation

The basic sample is specially fixed for a set of correct time series by analogy with a fixed sample of panel research respondents. The collection of a fixed set allows you to analyse the temporal dynamics of data obtained by statistical methods at any time scale, from minutes to years.

5.4. Generated base sample

The basic sample for the Moscow agglomeration was generated from 20 thousand virtual MMA commuters living at a distance of 20, 50 and 80 km from the centre of Moscow city, specified by the GIS coordinate $C_0 = (55.7520, 37.6166)$, and their places of work were selected within a radius of 5 km from the centre Moscow (Figure 2).

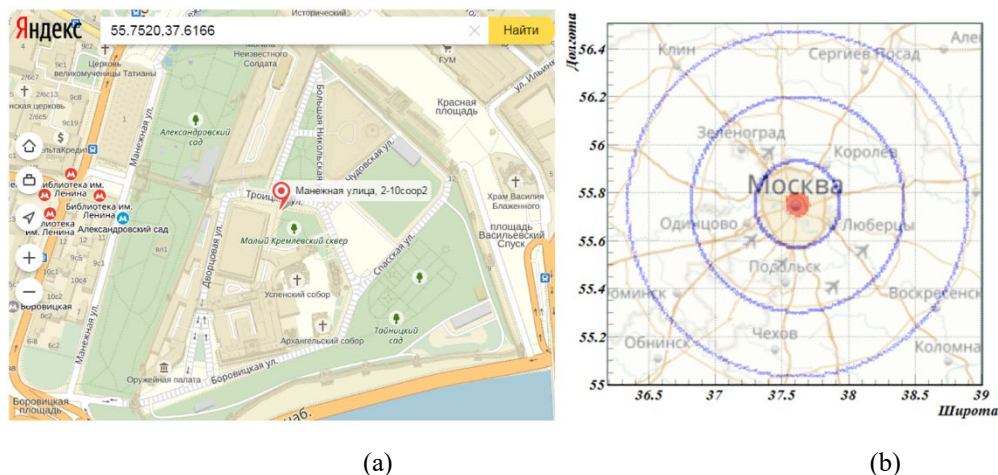


Figure 2. (a) The Moscow centre, the coordinate $C_0(\text{longitude, latitude}) = (55.7520, 37.6166)$, taken in the calculations. (b) The location of houses (blue concentric circles) and places of work (red spot in the centre) of commuters of basic set used in this study

Thus, the routes of the basic sample evenly cover the region-centre radial roads within a radius of 80 km from Moscow. In doing so, we cover three main types of citizen who make short, medium and long commuting trips, in order to compare the performance of these three groups of commuters.

5.5. Data collection

Specially developed software is used to collect information (the method was developed by us, RFBR grants No. 11-06-00323 and No. 14-06-00249). For each person from the basic sample the home-work-home paths are reconstructed through Yandex.Maps API (Figure 3).

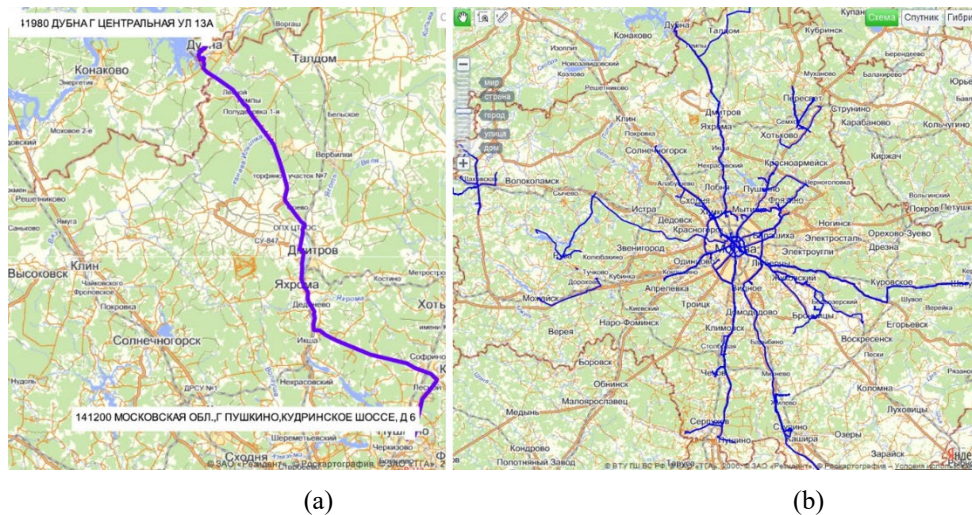


Figure 3. (a) Build of one route between home and work addresses of commuter. (b) A home-work route map for a small sample of MMA commuters

The database stores routing date and time, route length and travel times with and without traffic jams. The program is called hourly in automatic mode using standard operating system tools. More frequent calls can again be the reason for blocking. Collecting data on an hourly basis is a reasonable compromise, as it allows you to investigate the state of the roads with an accuracy of about an hour (24 points per day for assessing dynamics).

5.6. Indicators

The collected data allows the calculation of a variety of indicators. Here we use the simplest two - the specific U_I and relative W_I time losses by the I-th commuter of the base sample of N people, due to traffic jams per route unit:

$$U_I = \frac{T_I^R - T_I^0}{L_I}, \quad I = 1, \dots, N, \quad 1a)$$

$$W_I = \frac{T_I^R}{T_I^0}, \quad I = 1, \dots, N, \quad 1b)$$

where T_I^R and T_I^0 are times with and without traffic jams, respectively, and L_I is the total length of the route. Dimensional indicator U_I (typical unit is min/km) is used for quantitative estimates, while dimensionless indicator W_I is convenient for visualization of time series.

Further, in a natural way, you can use aggregation, in particular, the average time loss over the sample at the j -th moment of time:

$$\bar{U}(t_j) = \sum_{I=1}^N U_I(t_j) \quad 2)$$

Since the base sample is fixed, the time series of indicator (2):

$$\{\bar{U}(t_0), \bar{U}(t_1), \dots, \bar{U}(t_j), \dots\} \quad 3)$$

shows the temporal dynamics of the status of the transport network in the region covered by the routes of the base sample.

Finally, the basic sample of people can be studied not only as a whole, but also in the context of a number of variables:

$$U = f(L, DIR, REG, \dots) \quad 4)$$

where L , and DIR are the length and angular (azimuthal) direction of the route (to the cardinal points in relation to the centre of the region), and REG is the region of residence, etc. The study of subgroups of the base sample gives a more detailed monitoring of the transport network behaviour.

The first results obtained during the development and implementation of the method described above are discussed below.

6. Findings

The accumulation of data according to the procedure described above was launched in September 2015, this work presents the results of processing the data accumulated up to February 2020 - more than 500 million records.

6.1. Commuting times and losses

Commuting times and losses for the three study groups of the base sample are presented in Figure 4.

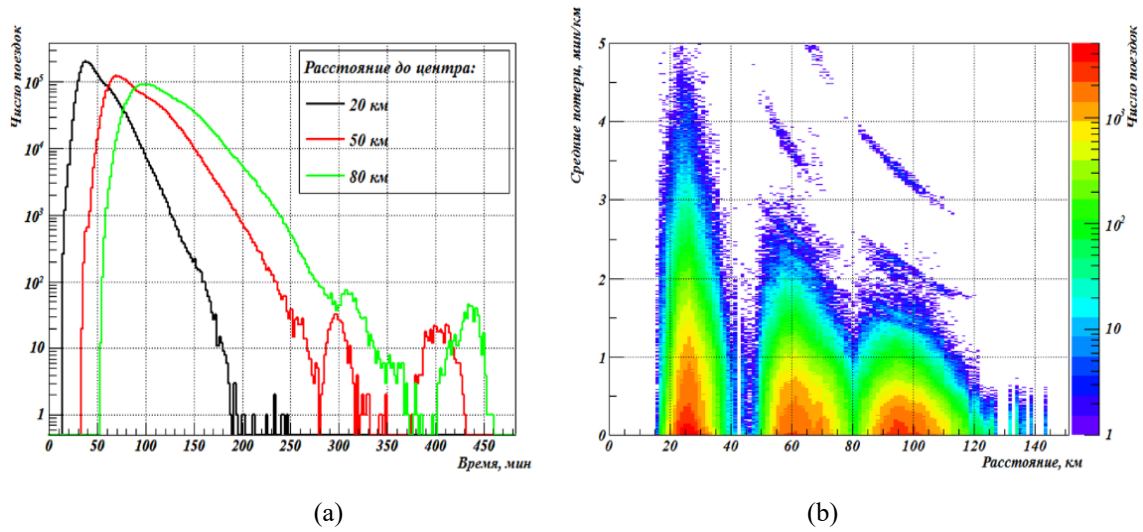


Figure 4. (a) Travel time of commuters from the basic sample (in logarithm scale). (b) Average losses on their commuting trips due to traffic jams according to formula (1a)

As can be seen from Figure 4, the specific losses of commuters driving long, medium and short distances, respectively, are increasing, which is consistent with expectations. Residents living closest to Moscow suffer the most from traffic jams in terms of wasted time.

Traffic network loading patterns have a pronounced cyclical nature, and in two different time frames at once - within a day and a week. Further we show the results of the first group of commuters living 20 km from the centre. We have the longest time series for this group, and the analysis of the remaining groups is planned in the future.

6.2. Intraday cycle

The intraday transport network load cycle is shown in Figure 5a. The graph clearly shows peak loads at 6-10 hours in the direction from the region to Moscow. And return rush hours in the evening from 16-19 hours are in the opposite direction.

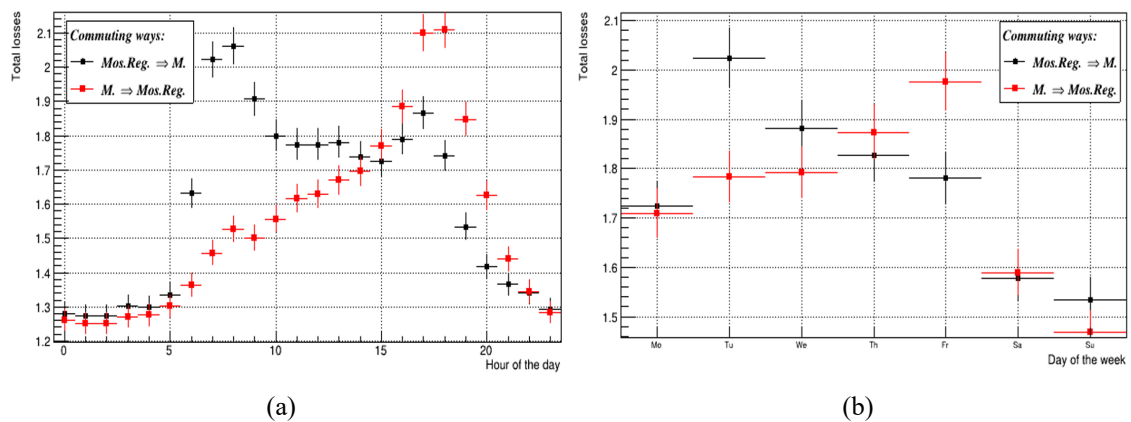


Figure 5. Time losses by commuters depending on the hour of travel in January 2020 (a) and the day of the week in September 4-10, 2017 (b). Black and red dots - centre-area and centre-area roots respectively. Along the vertical axis, the ratio of losses compared to a trip without traffic jams, calculated using formula (1b)

6.3. Intra-week cycle

Intra-week cycle is shown in Figure 5b, which demonstrates the classic picture of less traffic on the weekend. The increased load at the exit from Moscow on Friday and at the entrance to Moscow on Tuesday is associated with "dacha traffic".

6.4. Dynamics of commuting time losses

Average relative intraday commuting time losses due to traffic jams from December 2019 to August 2020 are shown in Figure 6. The chart clearly shows the cyclical weekly loss structure with noticeable fluctuations on weekends (discussed above). Another noticeable effect is that there are more difficulties with leaving Moscow than with entering (black dots are systematically higher than red ones) till March 2020.

At the end of March, strict restrictions were announced in Moscow and the region due to coronavirus infection. And we see that during the month from the end of March to the end of April 2020, the loss of traffic was minimal - the loss of time was kept to a minimum.

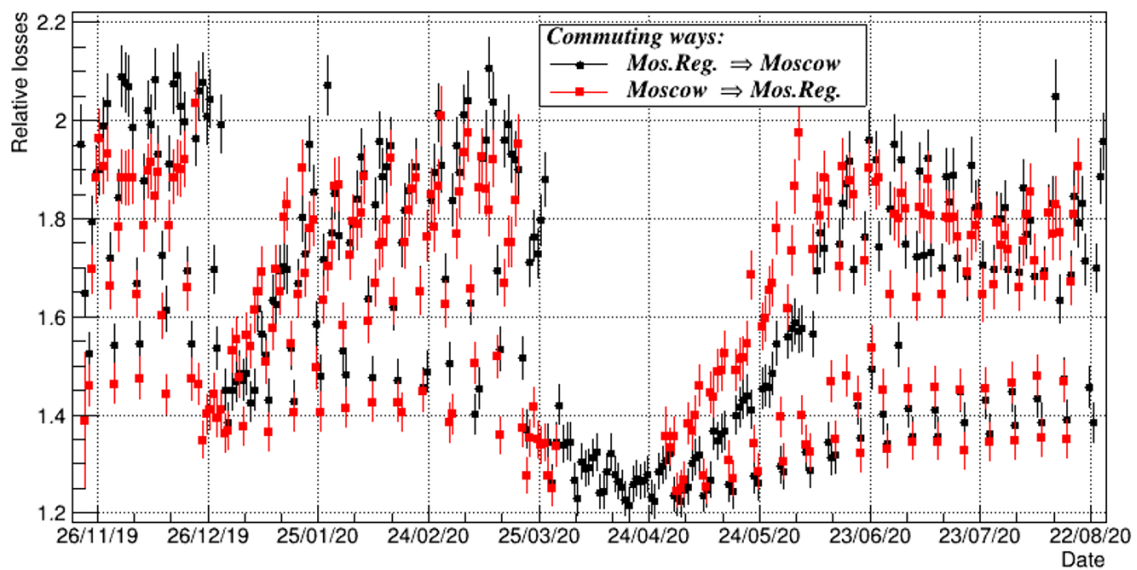


Figure 6. Average daily travel time losses due to traffic jams (black dots - center-area, red dots - center-area). On the vertical axis, the ratio of losses in comparison with a trip without traffic jams, calculated by formula (1b)

The recovery of traffic on the roads of the Moscow region took place throughout the May-June 2020, while during this period the difficulties in entering Moscow exceeded the load on leaving the centre. After the recovery of usual activity in July 2020, the situation has been returned to its previous picture: traffic jams at the exit from the centre overcome the difficulties in entering Moscow. Finally, there is a growing dynamic in the gap between travel on weekdays and weekends.

7. Conclusion

This paper presents the results of the original methodology of continuous GIS monitoring of the transport network of the Moscow agglomeration. The methodology for collecting, accumulating, storing and processing GIS information received from the online platforms Yandex.Maps and Yandex.Traffic is described, as well as the quantitative results of the analysis. A software package has been developed, with the help of which the transport network is automatically monitored in real time.

The technique is scalable, the value of the data grows as it accumulates. In the future, it is planned to use Big Data techniques for forecasting purposes. Besides, visualization and online monitoring and automatic publication of the results are planned.

In conclusion, we emphasize that GIS analysis of transport data is an actual modern trend in scientific research and commercial projects. With its unique data and methodology, this work is in trend, at the forefront of research in this scientific direction.

Acknowledgments

This work was supported by the Russian Foundation for Basic Research under Grant No. RFBR-19-010-00794.

References

- Arribas-Bel, D., & Michiel, G. (2012). Featured graphic. Monocentricity? Commuting flows visually. *Environment and Planning, A44*, 2041-2042. <https://doi.org/10.1068/a44561>
- Blind, I., Dahlberg, M., Engström, G., & Östhet, J. (2018). Construction of Register-based Commuting Measures. *CESifo Economic Studies, 64*, 292-326. <https://doi.org/10.1093/cesifo/ify014>
- Dalton, A. M., Jones, A. P., Panter, J., & Ogilvie, D. (2015). Are GIS-modelled routes a useful proxy for the actual routes followed by commuters? *Journal of Transport & Health, 2*, 219-229. <http://dx.doi.org/10.1016/j.jth.2014.10.001>
- Dauth, W., & Haller, P. (2016). The Valuation of Changes in Commuting Distances: An Analysis Using Georeferenced Data. *IAB Discussion Paper, 43*. <http://doku.iab.de/discussionpapers/2016/dp4316.pdf>
- Djurhuus, S., Hansen, H. S., Aadahl, M., & Glümer, C. (2016). Building a multimodal network and determining individual accessibility by public transportation. *Environment and Planning B: Planning and Design, 43*, 210-227. <https://doi.org/10.1177%2F0265813515602594>
- 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*. <https://www2.uibk.ac.at/downloads/c4041030/wpaper/2017-01.pdf>
- Heuermann, D. F., Assmann, F., Berge, Ph., & 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. <https://doi.org/10.1016/j.regsciurbeco.2017.08.001>
- Hulme, G. V. (2020, July 2). DDoS explained: How distributed denial of service attacks are evolving. <https://www.csoonline.com/article/3222095/ddos-explained-how-denial-of-service-attacks-are-evolving.html>
- Makhrova, A. G., Kirillov, P. L., & Bochkarev, A. N. (2016). Mayatnikovyye trudovyye migratsii naseleniya v Moskovskoy aglomeratsii: opyt otsenok potokov s ispol'zovaniyem dannykh sotovykh operatorov [Commuting of the Population in the Moscow Agglomeration: Experience in Estimating Flows Using Data from Cellular Operators]. *Regional Studies, 3(53)*, 71-82.

- Obaco, M., Royuela, V., & Vítores, X. (2016). Computing functional urban areas using a hierarchical travel time approach. *56th Congress of the European Regional Science Association: "Cities & Regions: Smart, Sustainable, Inclusive?"*, 23-26. http://www-sre.wu.ac.at/ersa/ersaconfs/ersa16/Paper238_MoisesObaco.pdf
- 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. <https://doi.org/10.1016/j.jpubeco.2016.08.001>
- Rehak, S., & Kacer, M. (2018). Estimating distance gradient in Bratislava with different types of distance measurements. European Real Estate Society (ERES). <https://eres.architexturez.net/doc/oai-eres-id-eres2018-259>.
- Scharenbroich, M., Iacono, M., & Levinson, D. M. (2009). How local is travel? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1736090>
- Vaz, E. (2013). Spatial Business Landscape of India. *Journal of Spatial and Organizational Dynamics*, 1(4), 241-253. http://www.cieo.pt/journal/J_4/article2.pdf
- Wood, G. (2003). Modelling the ecological footprint of green travel plans using GIS and network analysis: from metaphor to management tool? *Environment and Planning B: Planning and Design*, 30, 523-540. <https://doi.org/10.1068%2Fb12995>
- Yandex Map Platform (2020, September 15). <https://yandex.ru/maps>
- Yandex Traffic Platform (2020, September 15). <http://maps.yandex.ru/traffic>
- Zhu, S., & Levinson, D. M. (2015). Do people use the shortest path? An empirical test of Wardrop's first principle. *PLoS ONE*, 10(8), e0134322. <https://doi.org/10.1371/journal.pone.0134322>