Developing a microscopic city model in SUMO simulation system

E R Maiorov¹, I R Ludan¹, J D Motta¹ and O N Saprykin¹

¹Samara National Research University, Moskovskoye shosse 34, Samara, Russia, 443086
e-mail: benjamin1437@mail.ru, ludanilya@gmail.com

Abstract. Modern high rates of urbanization affect the economic and social processes in cities around the world. One of the main problems in cities is the weakness of transport systems. To solve this problem, it is necessary to use specialized software for modeling, in consequence of the fact that there is no possibility to influence the situation in real time. The main problem in creating a transport model of cities is the lack of data on the mobility of residents. The aim of this work is to develop a method and software for automated creation of microscopic models of cities using open data. We propose the processing pipeline that consumes the open data about city population and transport infrastructure and produces routes for agents of microscopic simulation. The developed method was implemented using the Python programming language in the Zeppelin interactive environment. The software was tested on data from the city of Samara (Russia). The obtained results show acceptability of proposed method and possible ways to improve the accuracy of the model.

1. Introduction
Due to the active development and growth of the urban population, there is an acute problem of compliance with the transport infrastructure with the mobility needs of citizens. The number of city inhabitants is growing in the same way as the size of cities. Transport systems still remain one of the main problems of modern megacities. According to statistics, citizens of large cities every 365 days spend 3 days waiting in traffic jams, which is an average of about 16% of the total driving time due to the poor condition of urban road networks. These statistics indicate the existing problem of urban road networks and the necessity to solve the problem. In this regard, there are several modern approaches. The first method is to improve the road network by its reconstructing [1]. Another approach to solving the problem is to create comfortable conditions in public transport in order to attract citizens. This makes sense since the desire to change people from their personal transport to public transport will certainly lead to discharging of the city’s road networks.

Transport systems and their investigation become more difficult in the course of time and require accurate knowledge of this subject. Nowadays there is a lot of research dedicated to this area. Discussing the transport system, it is worth noting that it consists of many interrelated parts. That is why in the world there is a big amount of research groups engaged in research in a particular field.

The most attention should be paid to the construction of Origin-Destination (OD) matrices which are developed to contain information about population movements in the entire city and to be able to
apply this data in modeling. The data driven-approach and model-driven approach are the most popular approaches to build OD matrices.

Applying GPS data to scientific research in transportation systems is very popular nowadays. Therefore, the data-driven approach is frequently used because it is based on the fact that the model is created from dynamic data of residents’ movement in the entire city using GPS, social networks, etc. This approach is very modern due to the widespread use of GPS navigation in our life. In addition, the installation of the GPS-navigator in cars is now very common, thus OD matrices can be formed from received data from the GPS device. It is worth noting that the application of this method substantially decreases the possibility to obtain an inadequate model. This area of research is relevant in our time [2-6].

Information from social networks is another source of population movement data. Nowadays Foursquare and Facebook are one of the best for this purpose. This method is based on geolocation data processing included in posts and photos. Having processed a rather wide amount of data, it is possible to make statistics showing the intensity of population movement in certain areas of the city at certain intervals. In this area, there are a lot of researches as new social networks are creating and being used [7-10].

The most accurate method for obtaining data of residents’ mobility was proposed by Klaas Friso [11]. High accuracy of the method is ensured by using data obtained from cellular communication of city residents. This technique allows for obtaining the most current data, however, the processing of large datasets from different operators is very tough. The difficulties of applying this method include the fact that it may be limited to the level of the Legislation of some countries, as well as the disagreement of subscribers to the processing of their personal data.

The model-driven approach is also popular among the researchers and used when mobile data or GPS are unobtainable. Filippo Simini [12], who uses the universal model for mobility and migration uses the gravitational method to obtain data on population movements across the country. This method allows tracing the trade flows up-country, as well as between countries, migration flows and intercity telephone calls. The gravitational method was elaborated by W. Reilly [13]. An idea for constructing a gravitational-type model is based on the universal law of gravitation. Applying this law to the transport system, it turns out that the objects are the points generating and absorbing traffic flows [14], and the object mass is a total volume of the incoming and outgoing flow. It is worth noting that there is a possibility to substitute the physical distance with any costs [15] associated with the movement.

The entropy model [16] is specific for the second law of thermodynamics and was suggested by physics. The law states that any closed physical system tends to achieve a stable equilibrium state, which is characterized by the maximum entropy of this system. The first application of the method was carried out to the transport processes of A. Wilson [17]. The principle of the entropy method is that the system of residents’ movement on the city road network has a fairly large number of uncontrolled elements. Due to the fact that our research is devoted to city transportation, it is necessary to consider the transport system as a closed one.

Michael Balmer compares two types of modeling in his work [18]: micro-simulation model and a traditional assignment model. The traditional assignment model is based on the use of OD matrices to reflect the movement of large population flows [19]. The main advantage of the micro-simulation model is that it is possible to simulate the movement of each agent individually [20].

Nowadays modern transport simulation systems allow creating the visual representation of traffic flows [21]. Currently, there is a wide range of transport simulation software, among which may be noted more popular software such as T7F/TSIS, SUMO [22], TRANSYT, VISUM, MATsim, Aimsun. They allow the verification of transport infrastructure changes, without special economic and time costs. Additionally, there is the research of OD matrices building from microscopic simulation outcome [23].

Despite a large number of researches in the field of transport modeling, some issues still remain. The most important problem is related to the automation of model creation. This paper is devoted to developing methods and software of automated creation of transport microscopic models for any urban area based on open data. The developed system provides the ability to obtain up-to-date information from open Internet resources, process it, and apply it for transportation model creating or
updating. As an example of open resource, we consider OpenStreetMap that provides information about city infrastructure collected by volunteers. Using information from OpenStreetMap, we have defined the transportation supply and demand for building a microscopic transportation model. The proposed approach allows getting models with acceptable quality using only data from public resources.

2. Transportation model of the city
In this paper, it was decided to use modeling that allows us to study possible changes in the structure of the road network in order to increase its capacity. Modeling is used to study real objects by building models to reveal their latent characteristics, and also influence them. There are several types of modeling that allow creating a model of a specific object. Applied to this field of science the following types of modeling are used: mathematical, structural-functional, conceptual, physical, and simulation. The choice of these types of modeling depends on the specific purpose of the research.

In our study, we use simulation, which allows us to build models describing processes, in our case, transportation processes. The reason for choosing this method is the dynamic properties of the transport system, as well as the impossibility of creating a static model.

At the moment, there are a large number of software applications for creating a transport model of cities of different levels of detail - from macroscopic to microscopic [24]. In this work, we use microscopic simulation, to create a microscopic model. The main advantage of microscopic modeling is the ability to model individual elements, in this case, vehicles, as well as objects of transport infrastructure [25].

The main goal is to create a universal method (figure 1), which would allow to modeling the transport processes of any city using open data.

![Figure 1. The method of activity chains constructing.](image)

To create a model that most accurately describes a real object, it is necessary to solve several problems [26]: partitioning a city into traffic analysis zones (TAZ), constructing OD matrices, and building activity chains. To create the most accurate model, it is necessary to split the city into
transport areas of such size that within the given area, residents do not use transport, but use public or personal transport to move to another transport area.

2.1. Traffic Analysis Zones

The division by TAZ is the foundation of correct elaboration of transport flow models [27]. Starting from these divisions, we can elaborate an image that allows the identification of several economic, political and social aspects of the city in question.

The concept of TAZ can be understood as the elementary area used for the analysis of transport models. With them, we can determine the flow of vehicles and pedestrians moving from region A to region B. At the time of the splitting of zones, the main criterion to be taken into account the points of the largest density flow, around which regions are formed, playing down such criteria as the physical size of the regions as suggested in fundamentals of transport modeling [28].

In the establishment of the Traffic Analysis Zone, it is important to choose areas where a relationship between the origin and destination exists. For the determination of the scale of the transport regions, it is necessary to keep in mind that the more the size of said area increases, the less effective the model will be.

On the criteria described above, the regions were extracted according to the main road arteries and natural boundaries (figure 2). The proposed approach allows the model's performance to be more accurate and provides better possibilities for improvement. In addition, each region can be subdivided, to correspondingly make a model on this specific area. Therefore, we obtain a scalable model which can be adapted to solve transportation problems at different levels with less handling of data and information.

2.2. Origin-Destination Matrices

We used the gravitation model to construct the OD matrix because it has sufficient quality for urban areas and is not very sensitive to computational resources. The transport gravity model relates to the intensity of the flow. Between the total number of departures from the i-th zone and arrivals to the j-th zone and the cost of travel between zones i and j (1).

\[
T_{ij} = \frac{Q_i D_j}{c_{ij}} i = 1, ..., N, j = 1, ..., M, \tag{1}
\]

where N is the total number of departure areas, and M is the total number of arrival areas. This model assumes that the cost of moving between areas is considered as the distance between the centers of these areas.

There is an opportunity to receive data from any GIS required for finding the intensity of the flow based on information about residential buildings and organizations of the city. To involve organizations into a model we need to receive geographic and attribute data about them. We also need information about residential buildings: number of stores and coordinates. With all of this information,
we can compare the average number of working or living people in each organization and residential building. Coordinates will allow the location of residential buildings and organizations to be displayed on the city map.

2.3. Method of Activity Chain Building
After dividing the city into traffic analysis zones and constructing the OD matrix, it is necessary to distribute the agents in the model according to the obtained data and to build a route for each of them. This is the process of activity chains building. To obtain them, it is necessary to determine which TAZ region each road section belongs to. For this purpose, the operation intersection is performed, which allows overlaying the two tables and intersecting the data sets. After that, agents are distributed along the road network. Start, finish and intermediate points are assigned according to the OD matrix. The last step is to combine all obtained data to visualize the simulation process, and then to compare the resulting model with the actual situation on the city's road network. If necessary, the model is corrected using additional data from the actual transport network.

3. Implementation
The automated method of microscopic transport models building was implemented using Python programming language. We have used the data for the city of Samara (Russia) for debugging the model.

Innovations in GIS make possible the modeling of transport processes using open data, (e.g. OpenStreetMap. The interface can be used and adapted to modeling and editing Traffic Analysis Zones with OSM Slippy Map Generator (figure 3). For the division of the city into traffic analysis zones in this study, the main highways of the city and natural borders were used. Extracted zones are used as source and target areas in the gravity model of the origin-destination matrix.

For visualization, the resulting matrix was colored in characteristic colors (figure 4). The less the intensity between the two transport areas, the shade of the matrix cell is closer to red, and vice versa, the higher the intensity, the closer to green.

To build the activity points it is necessary to separate road network sections by TAZ districts. Using the OSMnx library [29] is a way to automatically get geodata of the city. Geodata can be obtained only by just entering the name of the city. There we can receive data about coordinates and type of organizations and residential buildings of the city of Samara in the format of GeoDataFrame (GDF). The next step of the method is to divide the city into transport areas. To implement this method, the GeoPandas library [30] was used to perform various manipulations with a geographic database, which contains various tables with initial data: names of streets, districts, roads and so on.

After obtaining a database with the coordinates of roads and areas, it becomes necessary to determine the affiliation of roads to a certain area. In the library, there are special operations to perform various overlays of two spatial tables, as well as visualize them. In the study, by means of the sjoin operation, the traffic analysis zone and their roads are united.
The next step in the method of activity chain constructing is to create a file containing general data on the affiliation of all the city roads to a particular transport area. The final step of the method is the entry of all the data obtained into the SUMO program for further modeling.

The next steps are processing the received data in programs OD2TRIPS and DUAROUTER. OD2TRIPS allows defining the start and end points of a path, but it does not have the ability to compile route on real roads. In other words, program OD2TRIPS creates the starting point of the agent's movement and the end point of its path. In turn, the DUAROUTER program serves to compile a route along those roads, the data of which were obtained in the TAZ file. These programs are complementary to each other and create a well-constructed direction of traffic flows.

After using the created method of constructing activity chains and putting all the data in the SUMO program, it is possible to start modeling urban transport processes (figure 5). SUMO uses Krauss model of agents’ dynamic [31]. Model is based on the computation of the safe distance between master and slave vehicles. One of the main features of the Krauss model is the assumption that the driver does not necessarily have to reach the desired velocity, which adds important features to the driver's behavior. For example, the random difference in vehicle velocity from the desired ones leads to the spontaneous formation of congestion and the slow start of vehicles at traffic lights, which is characteristic of real conditions. The application of this method is possible for any city, subject to obtaining data on the transport areas of the city and their residents.
4. Conclusions
The developed method allows easily building a microscopic model of an interesting urban area. However, at this moment, it has certain shortcomings and limitations. The main issue is the inconsistency of the model sections with real traffic in some road network. For example, in the case of the city of Samara, most of the agents are moving along the Moskovskoye and Zavodskoye highways, ignoring roads that are no less used in real life, such as Novo-Sadovaya or Kirovskaya. There is also the problem of the emergence of new agents in the middle of the road, while they should appear near residential buildings. Generally, this issue is related to the inaccuracy of open data or missing some information about transport infrastructure objects. This limitation requires manual verification and fixing the model, so it can only work automatically in the simplest cases.

Despite all the shortcomings listed above, the method of constructing activity chains completely works. Even today it can be used to create a model of transport processes in other cities. The advantage of the developed method is that only initial data is needed in order to create activity chains and OD matrices for the city (i.e. information on transport areas and the number of residents of a particular transport area).

Further work is intended to reduce the scale of TAZ areas, which should lead to improvement of the model because it will reflect the urban area in more detail. Another pending task is developing a method for obtaining information about city organizations and clarifying information about residents.

5. References
[1] Bundesministerium für Verkehr 2004 Bau und Wohnungswesen (BMVBW): Neubau von Bundesautobahnen
[2] Gasnikov A and Klenov S 2010 Introduction to mathematical modeling of traffic flows (MFTI)
[3] Moreira-Matias L 2016 Time-evolving O-D matrix estimation using high-speed GPS data streams Expert Systems With Applications 44 275-288
[4] Russell J and Ye M 2017 Cooperative Localisation of a GPS-Denied UAV in 3-Dimensional Space Using Direction of Arrival Measurements IFAC 50 8019-8024
[5] Singh S and Sujit P 2016 Landmarks based path planning for UAVs in GPS-denied areas IFAC 48 396-400
[6] Kim B and Kim D 2016 Automated Complex Urban Driving based on Enhanced Environment Representation with GPS/map, Radar, Lidar and Vision IFAC 49 190-195
[7] Kheiri A 2015 Intra-urban movement flow estimation using location based social networking data The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, International Conference on Sensors & Models in Remote Sensing & Photogrammetry XL-1/W5
[8] Zhang Z and Wang Z 2017 The data-driven null models for information dissemination tree in social networks Physica A: Statistical Mechanics and its Applications 484 394-411
[9] Yu C and Xiao B 2017 Using check-in features to partition locations for individual users in location based social network Information Fusion 37 86-97
[10] Jiao Z and Ran L 2017 Data-Driven Approach to Operation and Location Considering Range Anxiety of One-Way Electric Vehicles Sharing System Energy Procedia 2287-2294
[11] Friso K 2015 Enriching the transport model of the Rotterdam region by cell phone data Models and Technologies for Intelligent Transportation Systems (MT-ITS)
[12] Simini F 2012 A universal model for mobility and migration patterns (Macmillan Publishers Limited)
[13] Reilly W 1931 The law of retail gravitation (New York)
[14] Wiedemann R 1974 Simulation des Straßenverkehrsflusses (University of Karlsruhe, Germany)
[15] Beckmann M and McGuire C 1955 Studies in the economics of transportation (Santa Monica: RAND Corporation)
[16] Fang S and Rajasekera J 1997 Entropy optimization and mathematical programming (Kluwer Academic Publisher)
[17] Wilson A 1970 Entropy in urban and regional modeling (London: Pion)
[18] Balmer M 2005 Generating Day Plans Based on Origin-Destination
[19] Roughgarden T and Tardos 2002 E How bad is selfish routing? Journal of the ACM
[20] Krajzewicz D and Erdmann J 2012 Recent Development and Applications of SUMO – Simulation of Urban Mobility International Journal On Advances in Systems and Measurements 5(3) 128-138
[21] Agafonov A and Yumaganov A 2018 Short-term traffic flow forecasting using a distributed spatial-temporal model CEUR Workshop Proceedings 2210 402-409
[22] Hofer C and Jäger G 2018 Generating Realistic Road Usage Information and Origin-Destination Data for Traffic Simulations: Augmenting Agent-Based Models with Network Techniques Complex Networks & Their Applications VI 1223-1233
[23] Saprykin O and Saprykina O 2015 Multilevel Modelling of Urban Transport Infrastructure, Proceedings of the 1st International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS) 78-82
[24] Chen M, Bernstein D H, Chien S I J and Mouskos K C 1999 Simplified formulation of toll design problem Transportation Research Record 1667(1) 88-95
[25] Agafonov A and Myasnikov V 2018 Numerical route reservation method in the geoinformatic task of autonomous vehicle routing Computer Optics 42(5) 912-920 DOI: 10.18287/2412-6179-2018-42-5-912-920
[26] Martinez L and Viegas J 2009 A traffic analysis zone definition: a new methodology and algorithm (Springer Science+Business Media, LLC)
[27] Gorev A and Buttger K 2015 Transport modeling fundamentals (St. Petersburg, Kosta) p 168
[28] Saprykin O and Saprykina O 2017 Validation of Transport Infrastructure Changes via Microscopic Simulation Proceedings of the 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems 788-793
[29] OSMnx URL: https://github.com/gboeing/osmnx
[30] Introduction to data analysis with pandas URL: https://habr.com/ru/post/196980/
[31] Song J, Wu Y, Xu Z and Lin X 2014 Research on car-following model based on SUMO The 7th IEEE/International Conference on Advanced Infocomm Technology 47-55