

Article

Combining Telecom Data with Heterogeneous Data Sources for Traffic and Emission Assessments—An Agent-Based Approach

Nastasija Grujić ^{1,*} , Sanja Brdar ¹ , Sjoukje Osinga ² , Gert Jan Hofstede ^{2,3} , Ioannis N. Athanasiadis ⁴ , Miloš Pljakić ⁵ , Nikola Obrenović ¹ , Miro Govedarica ⁶  and Vladimir Crnojević ¹ 

- ¹ BioSense Institute, University of Novi Sad, 21000 Novi Sad, Serbia; sanja.brdar@biosense.rs (S.B.); nikola.obrenovic@biosense.rs (N.O.); crnojevic@biosense.rs (V.C.)
- ² Information Technology, Wageningen University & Research, 6706 KN Wageningen, The Netherlands; sjoukje.osinga@wur.nl (S.O.); gertjan.hofstede@wur.nl (G.J.H.)
- ³ Optentia, North-West University, Potchefstroom 2351, South Africa
- ⁴ Geo-Information Science and Remote Sensing Laboratory, Wageningen Data Competence Center, Wageningen University & Research, 6708 PB Wageningen, The Netherlands; ioannis.athanasiadis@wur.nl
- ⁵ Faculty of Technical Sciences, University of Priština in Kosovska Mitrovica, 38220 Kosovska Mitrovica, Serbia; milos.pljagic@pr.ac.rs
- ⁶ Department for Computing and Control Engineering, Faculty of Technical Sciences, University of Novi Sad, 21000 Novi Sad, Serbia; miro@uns.ac.rs
- * Correspondence: n.grujic@biosense.rs



Citation: Grujić, N.; Brdar, S.; Osinga, S.; Hofstede, G.J.; Athanasiadis, I.N.; Pljakić, M.; Obrenović, N.; Govedarica, M.; Crnojević, V. Combining Telecom Data with Heterogeneous Data Sources for Traffic and Emission Assessments—An Agent-Based Approach. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 366. <https://doi.org/10.3390/ijgi11070366>

Academic Editors: Wolfgang Kainz and Jamal Jokar Arsanjani

Received: 16 April 2022

Accepted: 20 June 2022

Published: 28 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: To create quality decision-making tools that would contribute to transport sustainability, we need to build models relying on accurate, timely, and sufficiently disaggregated data. In spite of today's ubiquity of big data, practical applications are still limited and have not reached technology readiness. Among them, passively generated telecom data are promising for studying travel-pattern generation. The objective of this study is twofold. First, to demonstrate how telecom data can be fused with other data sources and used to feed up a traffic model. Second, to simulate traffic using an agent-based approach and assess the emission produced by the model's scenario. Taking Novi Sad as a case study, we simulated the traffic composition at 1-s resolution using the GAMA platform and calculated its emission at 1-h resolution. We used telecom data together with population and GIS data to calculate spatial-temporal movement and imported it to the ABM. Traffic flow was calibrated and validated with data from automatic vehicle counters, while air quality data was used to validate emissions. The results demonstrate the value of using diverse data sets for the creation of decision-making tools. We believe that this study is a positive endeavor toward combining big data and ABM in urban studies.

Keywords: ABM; air quality; big data; big data GIS applications; CDR; telecom data; data-driven decision-making; emission; traffic; urban studies

1. Introduction

Traffic congestion is one of the major problems in most cities worldwide, especially in developing regions producing increased fuel wastage, time, and monetary losses [1]. According to the World Health Organization, air pollution is responsible for approximately 4.2 million premature deaths every year [2]. Augmented duration and severity of traffic congestion has a more negative impact on pollution than free traffic flow. A higher number of speedups, slowdowns, stops, and starts increases the emission, as well as lower vehicles speeds [3]. To face the challenges, many diverse applications are developed and policies tested, such as the effects of electric cars, telecommuting [4], and car-pooling [5] on CO₂ emission, the outcome of banning old diesel cars on NO_x emission [6], changing the speed limit on several emissions types [7], congestion pricing [8], etc. However, all the studies mentioned rely on synthetic or static data that are usually biased, expensive, and time-

consuming to collect, and suffer from a lack of human dynamics [9]. In addition, static data are rarely available in near real-time.

With an increasing number of devices that passively collect data about people's spatial-temporal behavior, new opportunities arise. The widespread use of mobile phones, even in developing countries, makes passively collected data on user movements a promising method for tracking people's mobility. Unlike other location data such as Facebook, GPS traces, etc., telecom data cover the majority of the population of all age groups since most people carry their mobile devices all the time, which makes them suitable for travel patterns generation [9].

While numerous studies have demonstrated the value of leveraging telecom data in a variety of applications, reaching operation level is not near [10]. Those studies show the potential of telecom data in various domains, such as urban planning [11], pandemic monitoring of infectious diseases [12–17], human dynamics studies [18–25], socio-economic analysis [26–28], and disaster risk management [29–31]. In the context of traffic, studies have been focused on the assessment of traffic composition [32], travel speed and time duration [33], the number of passengers [34], origin-destination matrices [35–37], route choice [38,39], and mode of transport [40,41]. In only a few papers, authors use telecom data to extract mobility flows and populate a traffic model [42–44]. However, to our knowledge, we are the first ones to link telecom data to traffic emission. Moreover, we simulate traffic and calculate its emission on a precise spatial-temporal resolution (traffic: 1-s resolution, street level; emission: 1-h resolution, street level) in the whole city area, together with its subareas. Contrary to models that simulate only a part of a city, we believe that some traffic policies and regulations affect the whole city dynamics and because of that, the whole system should be taken into account while testing it. As a consequence, we use more diverse data sources to support our model, such as data from automatic vehicle's counters, air quality data, GIS data, and official population data.

Since traffic behaves like a complex system in which dynamics depend on all of its actors, agent-based modelling (ABM) imposes itself as a well-suited tool for its simulation. As such, it has been applied to urban transport [4,6,42,43], bus routing [45], car-pooling [5,46], planing of public transport [47], etc. In comparison to aggregated travel models, ABM can explain the causes of global phenomena that emerged by acting at a local level and then investigate the system-level consequences of appropriate local changes [48]. Insights and understanding of the phenomena in this way are necessary in order to create quality regulations that would improve the functioning of urban areas [49]. Empirical agent-based models can help not just to explain, but to predict the outcome of concrete scenarios. Nevertheless, for realistic simulation of individual dynamics, the data are needed on that level as well, which is not always available. Moreover, these models are usually data hungry and suffer from increased complexities [50]. In addition, the model's dynamics created by adding details from the data enriches the model and its complexity, so the output is more difficult to map to real systems [51]. We overcome the challenges mentioned through the use of diverse data sets and by applying calibration and validation processes.

The aim of this research is to propose an agent-based methodology that (1) simulates traffic and assesses its emissions on a precise spatial-temporal resolution during the whole day, (2) shows the potential of telecom data for large-scale traffic modeling, (3) addresses the challenges of combining telecom data with heterogeneous data sources, and (4) replicates the real scenario as closely as possible while tackling the challenges of data availability and data resolution. We use the GAMA platform [52] to build a traffic model, and utilize its output for emission assessment. As the proposed methodology is agent-based, it reveals global-local dependency, which is important for identifying and understanding the drivers of complex systems' behavior. Additionally, it enables the prediction of traffic and emission dynamics at the global level, which would be a consequence of changes at the local level. We believe that the proposed model could be used as a decision-making tool. Taking the second biggest city in Serbia, Novi Sad, we successfully simulate and verify the proposed methodology.

2. Methodology (Model Architecture)

In this section, we will first provide an overview of the complete methodology used, and then we will describe each part in detail.

The scheme of the proposed methodology is shown in Figure 1. It is based on three steps. The first step implies spatial-temporal travel pattern generation, which further serves as a basis for traffic and emission estimation. The second and third steps include the traffic ABM simulation, and its emission calculation at a street level. The traffic ABM model, built in the GAMA platform [52], is based on the car-following model. Every agent represents a vehicle that has an origin and destination, together with its scheduled time for commuting. It realistically simulates traffic conditions, by including crossings, lane changes, other vehicles, drivers' behavior in the time step of 1-s. The output of the model is the assessed number of vehicles and their speed at every traffic link at 1-h resolution, which is further used to assess their emission with coefficients from the HBEFA handbook [53]. We calculated the CO, NO_x, and PM pollutants.

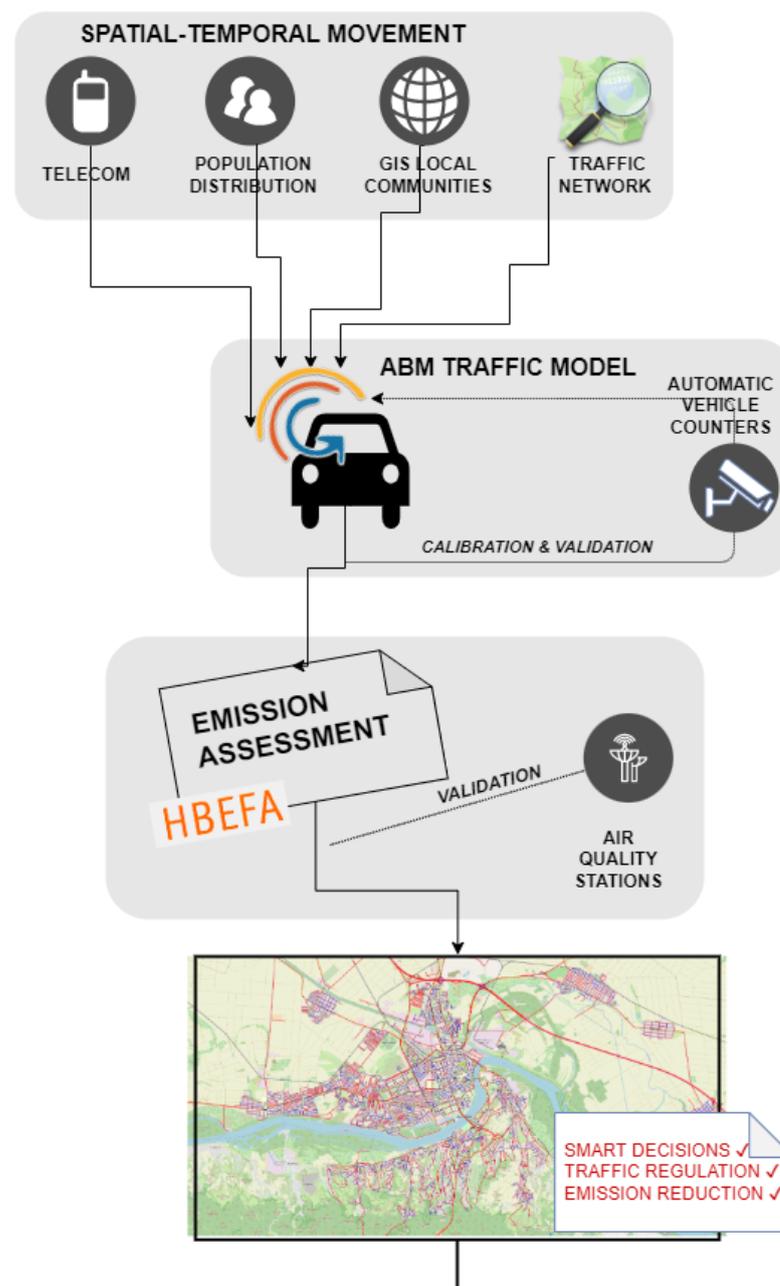


Figure 1. The scheme of the proposed methodology.

To feed up, calibrate, and validate the model, we used diverse data sets with heterogeneous spatial and temporal resolutions listed in Table 1. Using a telecom data set, we calculated the probabilities of spatial-temporal movement in the case study area. The probabilities of the spatial movement were calculated among local communities, while the temporal movement was computed at 1-h resolution and presented an agent's probability of starting commuting at a certain hour. Furthermore, we used that output together with official population data and GIS data at the local community level as an input to our traffic ABM model. We calibrated the traffic model by finding a combination of model parameters that best corresponded to a day selected from the automatic vehicle counters data set. We validated our model by comparing the model's output with selected parameters to a different day chosen from the automatic vehicle counters data set. Calibration and validation were accomplished at temporal resolution of 1-h. After that, the model output was used together with estimated coefficients from the HBEFA handbook to calculate the CO, NO_x, and PM pollutants produced by traffic. The output is again compared with ground truth data, or to be more precise, with emission data from two available stations in the case study area. With that, we confirmed the proposed methodology.

Table 1. Data sources.

Data Set	Time Period	Temporal Resolution	Spatial Resolution	Purpose
Telecom data	3–11/7/2017	1 s	antenna level	ABM—input
Population data	2019	-	local communities	ABM—input
GIS local communities data	-	-	—	ABM—input
Traffic network from Open-Street maps	-	-	—	ABM—input
Data on the number of cars and average speed from automatic vehicle counters	1/11–17/11/2019 and 3/12–9/12/2019	1 h	at main crossings in the city	ABM—validation and calibration
Emission coefficients from HBFA handbook	calculated for 2020 year	-	country level	emission—calculation
Air quality data	available for everyday	1 h	2 stations	emission—validation

The remainder of the section is organized as follows. First, we present the mobile phone data processing and generation of the origin-destination probability matrix, together with the probabilities of temporal movement. After that, the ABM model design is presented, together with its emission calculation.

2.1. Mobile Phone Data Processing

Mobile phone data are rich in user behavior data and are passively collected by telecom providers for billing purposes. Whenever a user uses a service from a provider, one Call Detail Record (CDR) is created and stored in a telecom database. Apart from their primary function, telecom data are extremely valuable, as they enable near-real-time monitoring of the spatial-temporal dynamics of a large number of people. Due to privacy issues, telecom operators perform rigorous procedures of data anonymization before they give data to third parties.

In this research, we utilized CDR to assess the probabilities of spatial-temporal movement in the case study area. The data set consisted of a set of telecommunication records (SMS In/Out, Call In/Out, and Internet activity) performed by randomly selected anonymized users. Besides the type of telecommunication records, the data set included the approximate coordinates of a Radio Base Station (RBS) that registered traffic, time, and duration of an activity, as well as the country code and number of digits in a

telecommunication number, which were used for record creation. From those records, we reconstructed the user mobility paths.

To generate the origin-destination (OD) probability matrix, which further served us to generate the traffic, we relied on several methodologies presented in the literature [24,28,35,43] with appropriate modifications adjusted to our case. The steps are presented here in three sub sections.

2.1.1. Data Preprocessing

The first step implied data cleaning:

- We eliminated RBSs that officially did not belong to the municipality region.
- We excluded landline numbers as they are not representative of user mobility.
- Records made by numbers with four digits or less were eliminated, as they probably belong to public services (e.g., parking services).
- Foreign numbers were excluded since they were probably from tourists who attended a festival held during a few days of the time period for which we had the data. We treated those records as anomalies, as these tourists did not contribute to the city's everyday traffic and emission.
- Finally, we selected users that had records during the day and night period, since we wanted to estimate users' regular trips that occurred during a whole day. Therefore, we wiped out users with only a few records.

2.1.2. Stay Extraction & Activity Inference

The second step was to extract the users' trips from the CDRs by selecting each user individually and ordering their records over time. To estimate the time a user spent on each location, we calculated the time difference between two successively visited antennas, which we then split and assigned each half to an RBS [24]. According to [35], we classified locations as "stay" and "pass-by" and finally kept only the stay locations since we were only interested in the origin and destination user point. Locations were classified as stay only if a user spent more than 10 min there. To estimate origin (home) and destination locations, we split the data set into two parts:

- A data set for estimating origin locations—it contained records obtained on weekdays between 7 p.m. and 8 a.m., and weekends.
- A data set for estimating destination locations—it contained records made during the weekdays between 8 a.m. and 7 p.m.

Unlike the authors in [35], we summed up the duration each user spent on each antenna in both data sets and utilized it to estimate the origins and destinations. As a result, the origin location was identified as the location where a user spent the majority of their time in the first data set, while the destination location was identified as the location with the highest $d1*d2$, where $d1$ is the distance from the estimated origin location and $d2$ is the duration a user spent in the range of an antenna. This assumption is adopted from the paper [35] and is based on the testimony from the literature that, in accordance with time spent at a location, destination locations, such as work, are more likely to be further away from an origin (home) location than closer locations [54,55].

2.1.3. Rescaling

As a final step, we rescaled the data to the local community level. The reasons for that were twofold. Previous studies showed a higher correlation between trips extended from telecom data and traffic surveys when aggregating trip origins and destinations to areas larger than one square mile [35,56]. Moreover, we also had the population distribution at the local community level. For the abovementioned reasons, we further grouped the RBS points and assessed movement among communities. To achieve that, we first approximated the domains of RBSs with Voronoi polygons [57], as it was performed more often in previous research [18,21,24,28,36,39,41,43]. We assumed that a user has a uniform

probability of being located at any point of a Voronoi polygon. Therefore, each user for their origin and destination locations got a point assigned inside corresponding polygons according to a uniform distribution. Next, we overlapped the areas of Voronoi and local community polygons, and in accordance with the users' points locations and intersections between the polygons, each user was assigned a local community for their origin and destination locations [28]. We further transformed the extracted number of trips among local communities into probabilities of movement among them. There are two reasons. The telecom provider that gave us the data set did not have a full market share. Therefore, we assumed users of the telecom provider were uniformly allocated in the case study area. Since we did not distinguish the users' transport modes (e.g., passengers, cyclists, vehicles, etc.), we made an assumption that the probabilities calculated depict the likelihood of vehicle movements within the municipality. The probability of telecom activity is also inferred from telecom data by calculating the proportion of active users per hour during working days.

2.2. Agent-Based Traffic Model

Urban systems in general are complex systems and to achieve higher sustainability, we need to understand their complexity [49]. The systems are strongly defined by decentralized, local interactions among sets of independent entities. They exhibit collective intelligence without the existence of a central authority, i.e., they tend to co-organize and adapt to changes in the environment, thus optimizing their behavior over time. In order to comprehend these systems, we need to study them together at two levels of abstraction, the local component level and the system level. ABM is a modeling paradigm that is suitable for modeling complex systems using a bottom-up approach by programming entities, their behavior rules, and the environment [58]. As the benefits of a bottom-up approach have been recognized, a lot of ABM frameworks for transportation modelling have been developed [59]. Among them, we chose the GAMA platform [52] to model the traffic as it has the ability to model large spatial complex systems and built-in actions for traffic modeling.

Model implementation The model implementation is based on a built-in plugin for traffic modeling [60]. The plugin consists of three built-in action components for three different types of agents that declare them. Those plugins provide agents with a set of defined attributes and actions. The types of agents defined by the plug-in are:

Road agents—Each road is a polyline composed of a set of road sections (segments). Road agents have a target and a source node and have information on all the input and output roads. They are directed. Therefore, if a road segment is bidirectional, two roads will be created for each direction. Road agents could have several lanes, which would allow vehicles to change them at any time. They take the *road skill* action that provides them with a set of variables, such as attributes that define the maximum allowed speed on the road, linked roads, and connected nodes.

Road's node agents—They define the beginning and/or end of traffic links. They adopt the *road node skill* that supplies the agents with attributes related to the linked road agents as well as attributes that define crossings, such as stop signs or a list of driver agents that block a node.

Driver agents—Each driver agent has a planned trajectory that consists of a succession of road links. A driver picks a lane according to the traffic density, favoring the rightmost lane. Driver agents use the *advanced driving skill* that gives them many attributes that characterize drivers' commuting (e.g., target, vehicle length, maximum acceleration of a vehicle, the distance they keep from another driver, etc.) and drivers' personal characteristics (probabilities of respecting stop signs, traffic rules, changing lanes, etc.). It also provides them with actions that define vehicle commuting. They are based on the car-following model. In the paper, we used all three agents with their skills and adjusted them for a specific case. The specifications of the proposed model will be described below. However, for further implementation details on the plug-in, please refer to [60].

Model rules The model follows simple rules. Each agent represents a vehicle and each has an origin and destination point and a time of departure that defines the agents' commuting. The model setup completely relies on data. The simulation starts at 00:00, and lasts until the next day with a simulation step of 1-s. Even though the output is needed at 1-h resolution, we chose to use a built-in traffic plugin that works at 1-s resolution, as it has the ability to realistically and dynamically simulate vehicle speeds, which are needed for precise emission assessment. Before the simulation starts running, each agent gets an origin and destination point, together with a departure time. When the simulation starts, agents are at their origin or destination locations, which are located somewhere in the case study area, depending on the time window an agent spends at its destination (e.g., working hours) allocated to each agent during the initialization. When the agent's departure time comes, it drives on the traffic network toward its destination. An agent enters the traffic network at a point that represents the crossing that is closest to its origin/destination location, and in the same manner, it exits the network. If an agent is on a destination point, its stay is defined by a model input parameter—*time duration*. After the duration of time passes, it returns to its origin point. Agents' commuting is completely defined by the built-in plugin for traffic modeling in the GAMA platform, which is based on the car-following model [60]. For commuting, they use the shortest path, which is calculated by the Dijkstra algorithm. While commuting, they respect the maximum allowed speed on the road and adjust their speed according to other agents around them, their attribute values, and the road network. Agents have certain attributes which define the probabilities of changing a lane, accelerating, decelerating, respecting the traffic rules (priorities, stop signs, maximum allowed speed), keeping a higher or lower distance from other cars, stopping for no reason. Furthermore, when an agent slows down due to a car in front of it, the likelihood of changing lanes or using a linked road increases. Traffic lights are not included in the model, but nevertheless, agents are programmed to stop for a second before they enter the intersection and to respect the traffic rules on intersections (with a certain probability). As we need the output to be at 1-h resolution to match the output to the observed pattern in the automatic vehicle counters data set, the output of the ABM model is an estimated number of cars and average speed at 1-h resolution per road link. However, since the model works at 1-s resolution, it can be easily exported and used for studying traffic composition at a more precise level.

Model setup For the model setup, we used diverse data sets, such as the traffic network, population data, and GIS data of local communities, as well as probabilities of movement among local communities and over time, assessed from telecom data. These data defined our model assumptions and had a significant impact on model dynamics. Model input parameters were *time duration* that agents spend at their destination locations, and *percent of simulated population*. These were tuned to the real traffic scenario during the calibration processes.

Traffic network The traffic network was downloaded from the Open-Street Maps (OSM). Aside from the spatial representation of roads, the data from the OSM contains other relevant information, such as the number of lanes, maximum allowed speed, a road's width, whether it is a one-way or two-way street, and other traffic network features. We created a buffer of 1km around the city in the municipality case study area and downloaded drivable road types (e.g., we excluded pedestrian roads). For the surrounding villages that also belong to the municipality case study area, we downloaded only the main roads. This was undertaken in this way because we wanted to model traffic pressure in the city through traffic inflow and outflow, and commuters from the surrounding villages significantly contribute to it, making the roads at the city-entrance suffer the most pressure. We preprocessed the data and adjusted it to our traffic model, creating a road link for each direction and lane, as well as road's nodes for each intersection on the network. Moreover, the missing road features were fulfilled (e.g., maximum allowed speed).

Population distribution and trip generation Since we got the population number per local community level, the probabilities of movement between communities (including a

self-loop) were assessed from telecom data. In the model, the population of agents was generated according to the population distribution per local community and the *percent of simulated population* input parameter. During the model setup, the number of agents was created for each community by multiplying the community population with the *percent of simulated population* and each agent got a randomly assigned point inside a community polygon with the condition that it was located between 100 and 500 m from a traffic link. The assigned point was marked as an agent's origin location. According to the extracted movement probabilities from each local community, each agent was assigned a second local community and a random point inside it with the mentioned condition that represented its destination location. Exported probabilities of temporal telecom activity served to allot the hour of departure, while the exact minute and second were uniformly chosen for each agent.

Drivers' behavior The variables listed in the first column of Table 2 were used to characterize the personal characteristics of the drivers. We characterized them with probability distributions extracted from [61] and listed them in the second column of Table 2. Every driver has a personal probability of changing a lane, respecting priorities and stopping signs, blocking a crossing node for no reason. With the security distance coefficient, the minimal distance a driver keeps from another driver is determined. The speed coefficient represents the speed they opt to reach according to the maximum allowed speed on the road. When a driver's speed falls below 25 km/h, its probability of using a linked road increases with every simulation step. When a driver reaches a speed of more than 25 km/h, the probability is set to be zero.

Table 2. Drivers' behavioral attributes.

Driver's Attributes	Value
Changing upper lane	$X \sim \mathcal{U}(0.1, 1)$
Changing lower lane	$X \sim \mathcal{U}(0.5, 1)$
Security distance coefficient	$X \sim \mathcal{U}(1, 3)$
Respecting priorities	$X \sim \mathcal{U}(0.8, 1)$
Respecting stop signs	1
Blocking a node for no reason	0
Maximum acceleration of its vehicle	1.39
Driver's speed coefficient	$X \sim \mathcal{U}(0.8, 1.2)$
Using a linked road	$p = \begin{cases} 0, & \text{if speed} \geq 25 \text{ km/h} \\ \min([1.0, p + \text{rnd}(0.10, 0.33)]), & \text{speed} < 25 \text{ km/h} \end{cases}$

2.2.1. Model Calibration & Validation

Since we did not have information on the number of active drivers or the duration of time that agents should spend at their destination locations, we assessed those parameters by comparing the model output with data from automatic vehicle counters, which included the number of vehicles passed and their average speed at 1-h resolution, together with coordinates of the counters' location. For every different combination of the model parameter values (every scenario), we produced a model output that was further compared with the real situation captured by the data from automatic vehicle counters in the case study area.

Calibration After we preprocessed the automatic vehicle counters data set, we selected one working day and compared it with every scenario produced by the model. For every scenario, the number of vehicles and their speed were assessed at 1-h resolution and compared with values in the automatic vehicle counters data set in a corresponding hour. Moreover, we calculated correlation measures between observed traffic circumstances and every scenario produced by the model. We used the Pearson and Spearman correlation coefficients. The model parameters from the scenario that were the best fit for the day chosen from the automatic vehicle counters data set were selected. Through the calibration processes, we assessed the global values that define the fleet composition during a day (the

number of vehicles on the road network across one day).

Validation To assess the reliability of our model, we compared the traffic volume of the selected model scenario in the calibration step with the other selected day from the automatic vehicle data. Validation is performed in the same way as calibration, that is, by calculating correlation measures between the observed and predicted number of vehicles and their speed.

2.3. Emission Evaluation

To estimate vehicle emissions produced by traffic, we used publicly-available coefficients from the HBEFA handbook [53]. The handbook contains diverse emission factors calculated for different types of vehicles (such as passenger cars, heavy-duty vehicles, light-duty vehicles, motorcycles, coaches, and urban buses) with different types of engines (such as diesel, petrol, electricity, and CNG) in Switzerland, Austria, Germany, Norway, Sweden, and France. The database contains factors calculated for the following pollutants: CO, HC, NO_x, PM, several components of HC (CH₄, NMHC, benzene, toluene, xylene), fuel consumption (gasoline, diesel), CO₂, NH₃, N₂O, PN, and PM in g/km. The pollutants are calculated for a wide range of traffic situations, such as cold start and warm emission events. Besides that, it includes the aggregated values of pollutants per type of vehicle and country.

We simplified the emission assessment since we did not have precise data on vehicle types or engines or factors assessed for Serbia. We chose to use factors estimated for Austria, as it is geographically the closest country. We used aggregated values provided for passenger cars and calculated emissions of pollutants CO, NO_x, and PM at 1-h resolution. We applied Formula (1) to calculate the emissions per traffic link. As reported in the literature mentioning that coefficients are approximately twice as large when the traffic is in a stop&go regime [62], we added the *k* coefficient to illustrate that. For every road link on which the assessed speed was $< 1/5 * max_allowed_speed$, *k* value was set to be 2, otherwise 1.

$$l_{h,p} = n_{h,l} * d_l * agg_p * k_{h,l} \quad (1)$$

where:

<i>h</i>	hour
<i>l</i>	traffic link
<i>p</i>	pollutant, can be CO, NO _x , and PM
<i>l_{h,p}</i>	calculated emission for pollutant <i>p</i> on traffic link <i>l</i> in hour <i>h</i>
<i>n_{h,l}</i>	number of vehicles on traffic link <i>l</i> in hour <i>h</i>
<i>d_l</i>	length of traffic link <i>l</i> in km
<i>agg_p</i>	aggregated coefficient from the HBEFA handbook for pollutant <i>p</i> in unit g/Vehkm
<i>k_{h,l}</i>	congestion coefficient for traffic link <i>l</i> in hour <i>h</i> . It takes value 2 for congested links, and 1 otherwise

Emission Validation

To assess the reliability of the estimated emissions, we compared the model output with the available air quality data measured by stations located in the case study area. We calculated the correlation measures between the assessed and obtained values of PM, CO, NO_x from air quality stations on the closest road link to the corresponding station. Again, we selected a day that is the same as the day selected for the traffic model calibration, as we believe that is the most realistic simulated day.

3. Case Study

Novi Sad is the second biggest city in Serbia with a positive growth rate (Figure 2) and more than 300,000 inhabitants [63]. The city is located on the border of the Backa and Srem geographical regions, which is defined by the Danube River. Furthermore, it faces the northern slope of Fruska Gora Mountain. Due to its geographical position, its road

network, and its relatively small available area, Novi Sad has become very crowded with considerable daily traffic congestion, which has led to increasing pollution. Congestion is mainly present at the extension of bridges. It is estimated that the city needs two more car bridges and one more pedestrian bridge. Unfortunately, due to financial reasons, it will get only one car bridge by the end of the year 2030 [64], which is quite a long period for a city that is constantly growing. For the abovementioned reasons, policy-makers need to find another way to optimize traffic to minimize congestion. With the proposed model, policy-makers have the possibility to explore the effects of various traffic regulatives to find the most optimal solution in the whole case study area.

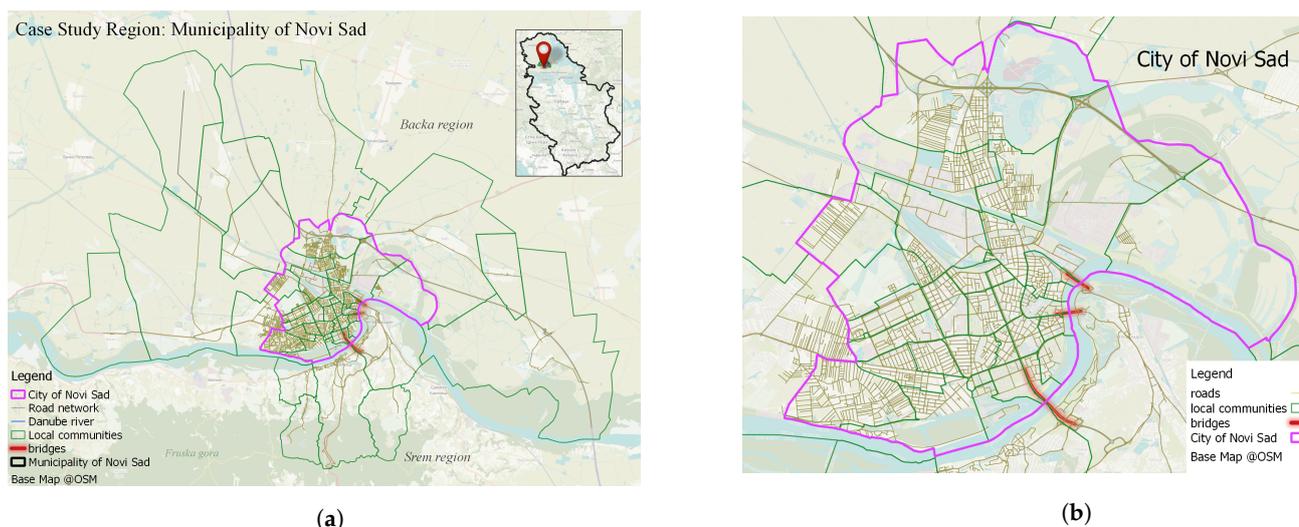


Figure 2. Case study area. (a) The municipality of Novi Sad is shown with its local communities. In the Figure, the city of Novi Sad (marked with pink polygon) can be distinguished from suburban areas which were included with its main roads to model traffic pressure in the city. (b) The city of Novi Sad, shown with its local communities and existing bridges.

3.1. Data Acquisition & Processing

3.1.1. Telecom Data

The data set was provided by the operator Telecom Serbia for the time period 3–11 July 2017. It contained approximately a million records per day generated by 197,950 individual users. Moreover, spatial resolution was defined by 80 RBS. After data preprocessing, we finally got 80,542 individual users, 77 antennas, and less than a million records per day, which were further used for the estimation of the OD probability matrix and probability of telecom activity per hour. The number of estimated origin (home) locations together with official population data is depicted in Figure 3a, while the calculated probability of temporal telecom activity is depicted in Figure 3b. The origin-destination probability matrix is shown in Figure A1.

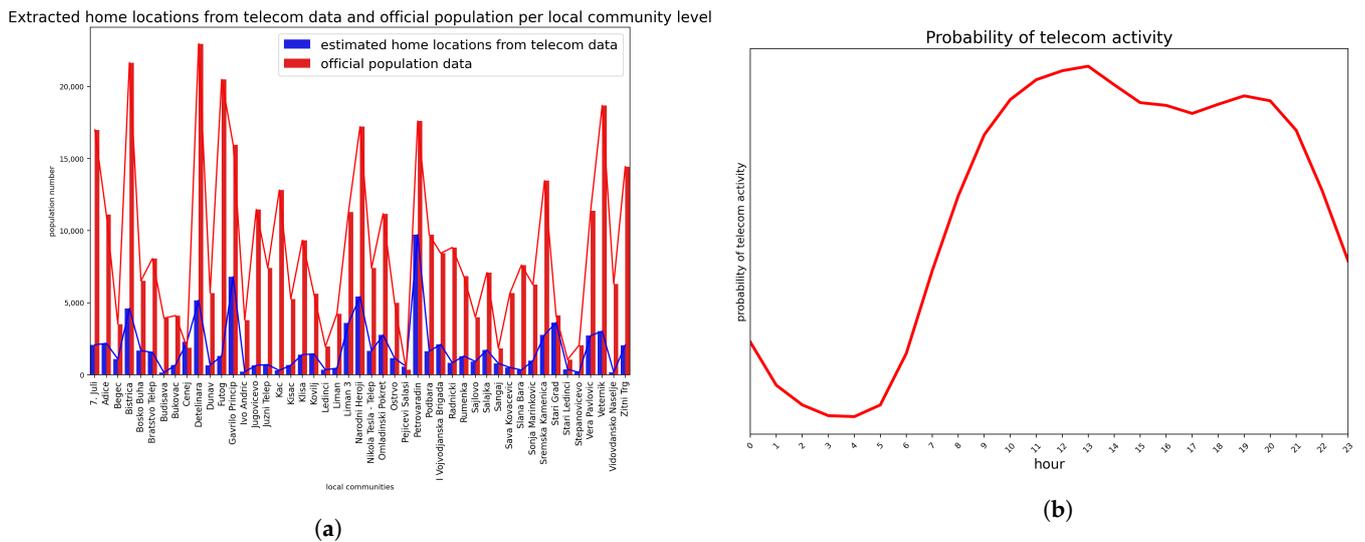


Figure 3. Telecom output. (a) The number of origin (home) locations extracted from telecom data compared to census data at local community level. (b) Extracted probability of temporal telecom activity across hours in a working day for the whole population.

3.1.2. Automatic Vehicle Counters Data Set

The data set from automatic vehicle counters was available for the time periods 1–17 November 2019 and 3–9 December 2019. It contained the number of cars at 1-h resolution from 26 counters, and the average speed from 16 counters located at main crossings in the city (Figure 4). However, as implementing those was a pilot project in the case study, some interruptions in data collection were present, and thus we first cleaned the data. Moreover, interruptions and inconsistent data were more present when measuring speed, compared to counting the number of vehicles.

We eliminated counters with invalid records. We set a condition that a counter must have at least four records when counting the cars or measuring the speed in one day (4 h) to be kept for validation or calibration. The number four was chosen for the possibility of calculating the correlations. If during a day, most of the counters had an interruption in the data collection, we wiped it out. In addition, we eliminated days that we identified as extreme outliers, as we assumed that there was an error in data collection. Finally, we ended up with 7–17 counters per day (depending on the day and measure (number of cars or average speed); measures from different counters were available) and 16 usable days. We calculated the mean number of cars and their average speed on a daily basis for the usable days and available stations and showed their variability in Figure 5.

Nonetheless, as previous literature reported that there are four different types of days in urban areas: working days (Monday–Thursday), Fridays, weekends, and holidays [25], and we wanted to simulate a regular working day, we only considered those that fall into the mentioned range. For the model calibration, we selected Wednesday—6 November 2019, since it had the largest number of counters available. The model output was then validated by comparing it with Thursday—3 December 2019, the next day with the largest number of available counters.

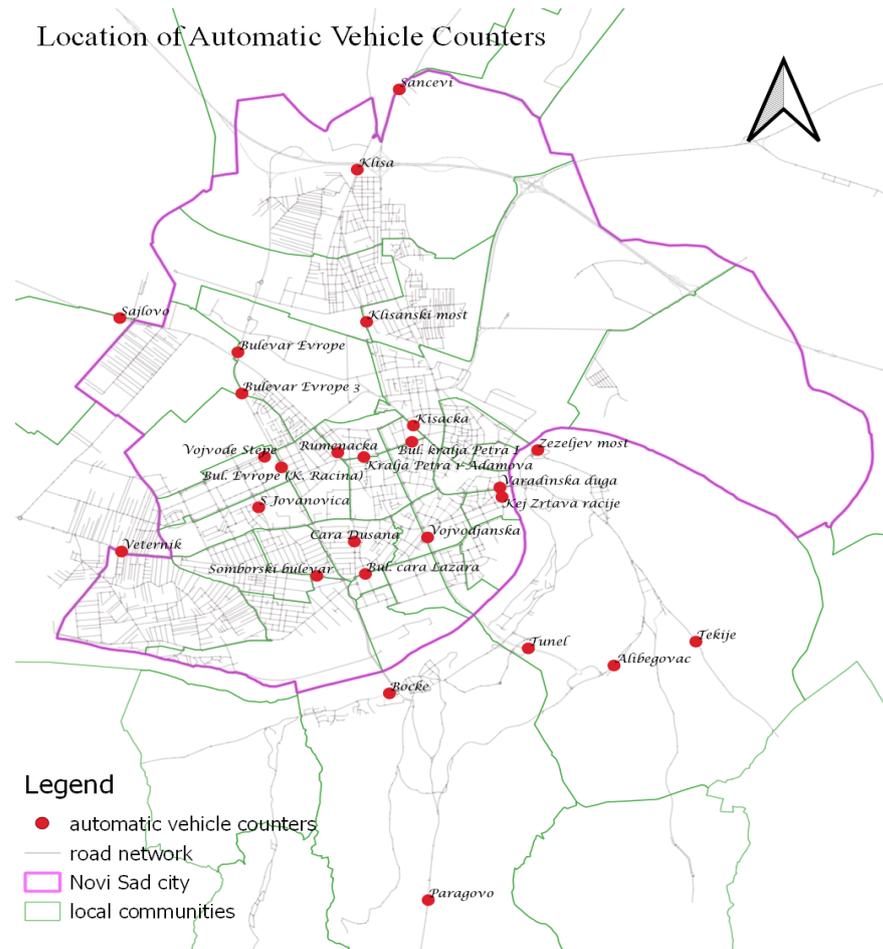


Figure 4. Location of automatic vehicle counters in the case study area.

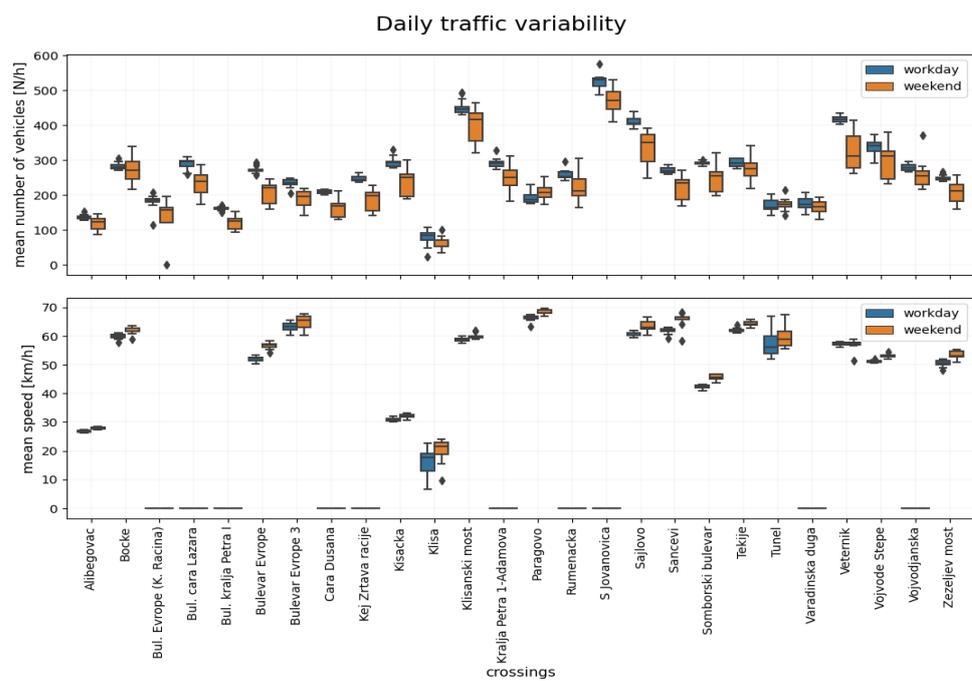


Figure 5. On the first subplot, the variability of the daily number of vehicles is shown, while on the second plot, the variability of average speed on a daily level is presented. The dashes in the second subplot present counters with no speed data.

3.2. Traffic Simulation

The traffic ABM model is based on data and the car-following model. The traffic network was downloaded from the OSM and imported into the model. For the city and its 1 km buffer, all drivable roads were imported, while for the peripheral parts of the city that belong to the municipality, only the main roads were included. In this way, the model is simplified, and the traffic entering the city is included. Agents are generated according to population distribution per local community. Their destination location and movements were inferred according to the probabilities calculated from telecom data. For further implementation details, please refer to Section 2.2.

The values of two input parameters that define the number of drivers and traffic dynamics were estimated during the calibration processes. Those parameters were *time duration*, which defines an agent's stay at a destination location, and *percent of simulated population*, which determines the number of agents in the model. In the calibration process we ran a manifold of experiments with the following values for parameters *time duration* = [4, 5, 6, 7, 8, 9] and *percent of simulated population* = [0.1, 0.2], chosen in accordance with expert opinion in the traffic domain. Since the stochasticity in the model is not considerable at the 1-h output, we ran the model five times for every combination of parameter settings and averaged the results. Then, as a part of the calibration process, we compared the outputs generated for each combination of parameters values to the chosen day in the vehicle automatic counters data set. We obtained the highest values of correlation between model output and data from automatic vehicle counters with parameter values [9, 0.1]. In the Figures 6 and A2, the comparison between the observed and predicted output could be observed for the mentioned parameter values. In Figure 7, the congestion maps are shown at 6 a.m., 10 a.m., 12 p.m., and 3 p.m. respectively. Congestion is calculated as a difference between the maximum allowed speed and the speed predicted by the model at a given hour for every traffic link. To validate the proposed model with the selected model parameters, we compared the results with the second day selected from automatic vehicle counters. Validation results are enclosed in the Appendix A and shown in Figures A3 and A4.

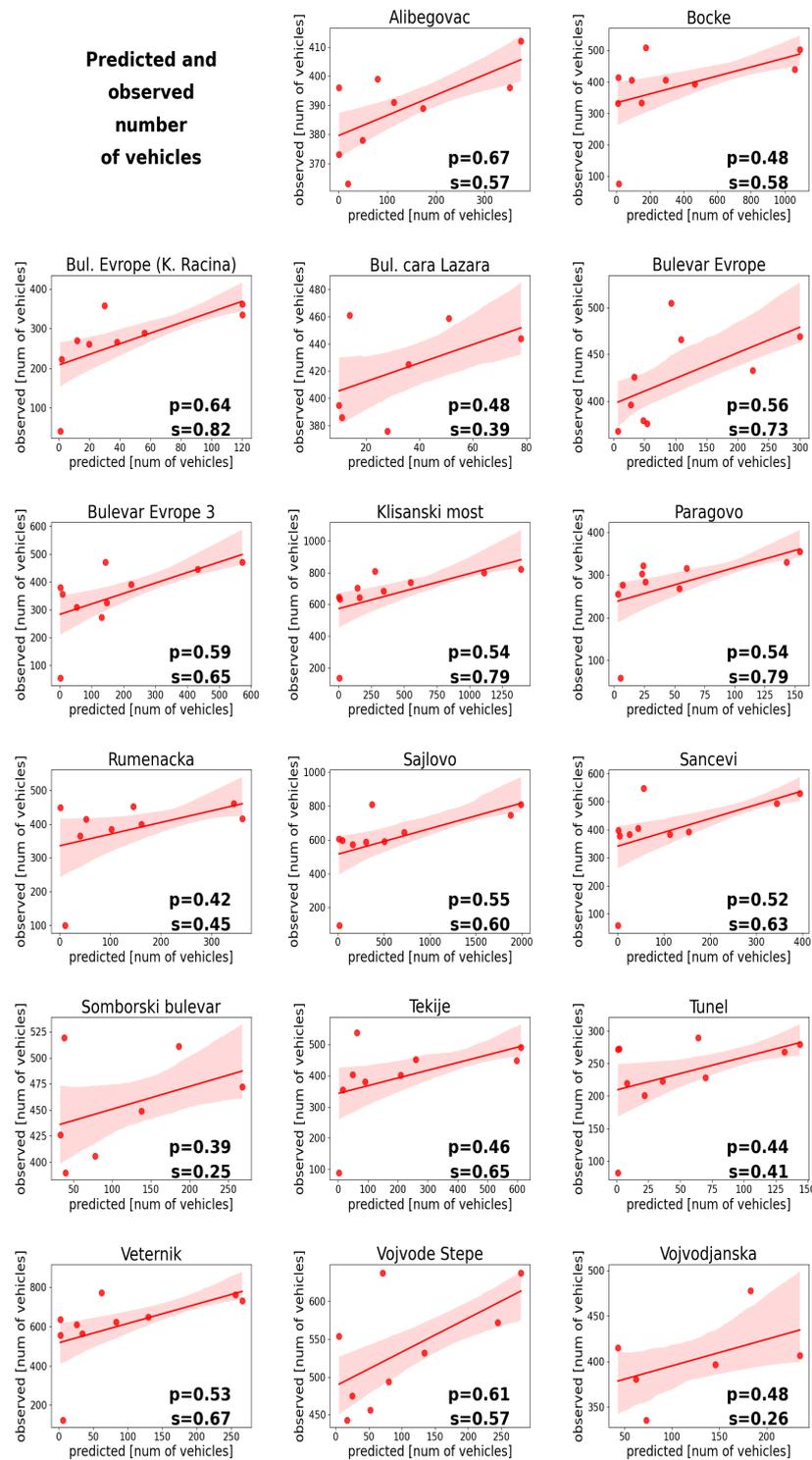


Figure 6. Calibration—Pearson (p) and Spearman (s) correlation coefficients between predicted and observed numbers of cars in automatic vehicle counters data set. The selected day from automatic vehicle counters data set was 6 November 2019. Please note: the axes are not in the same scale to emphasize the captured trend.

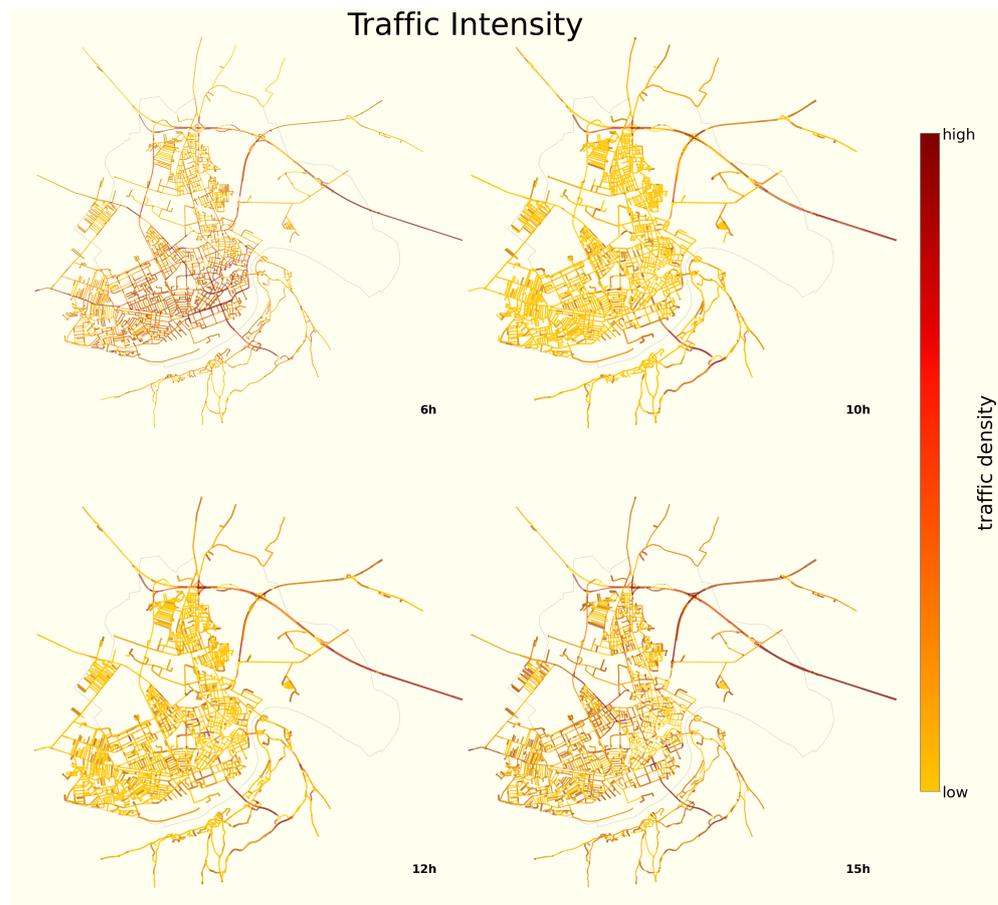


Figure 7. Traffic density at different hours.

3.3. Emission Assessment

Model output is further used to calculate the traffic emission by using the HBEFA aggregated coefficients and Formula (1). Emission is calculated for every traffic link and every hour. For the validation of calculated emission, we used measurements from two available *Serbian Environmental Protection Agency (SEPA)* [65] stations that measure different air quality parameters from traffic at 1-h resolution: station Novi Sad Rumenacka (45.262626, 19.819016)—CO and station Novi Sad SPENS (45.24506, 19.84119)—CO, NO_x, PM. The validation of emission output is shown in Figure 8, while the heat emission maps for NO_x, CO, and PM are depicted in Figures 9, A5 and A6, respectively.

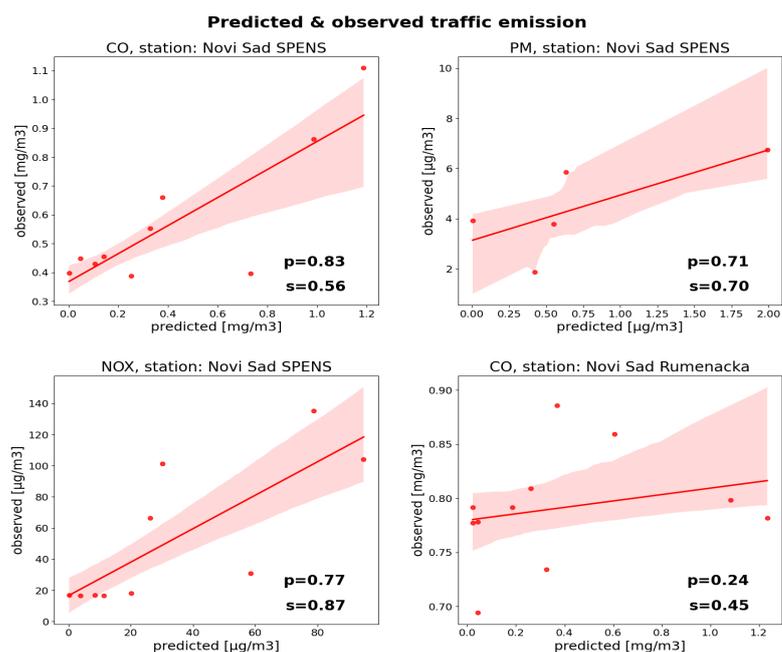


Figure 8. Validation—Pearson (p) and Spearman (s) correlation coefficients between predicted and observed emission in air quality stations. The selected day—6 November 2019 *Please note: there were interruptions in measurements of PM at the Novi Sad Spens station. Moreover, the axes are not in the same scale to emphasize the captured trend.*

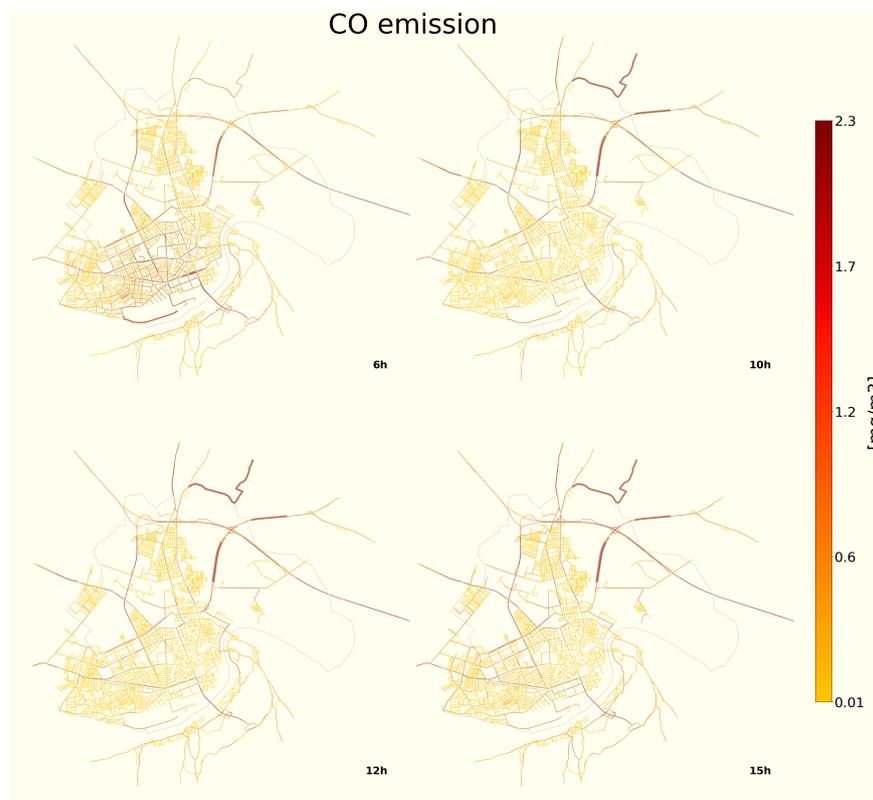


Figure 9. CO emission intensity at different hours.

4. Discussion

The model setup and achieved accuracy were dictated by the available data. We achieved good correlation measures for most of the crossings, meaning that we managed

to assess traffic composition and its emission. However, for some locations, the number of cars or emission was not predicted in the same range as observed by automatic vehicle counters and emission stations. Moreover, while the Pearson correlation for most of the plots was around 0.5–0.6, the Spearman caught higher variability, and its values went from 0.1 to 0.9. The Pearson correlation uses actual values, compared to the Spearman correlation, which calculates the correlation based on ranks. In terms of the predicted emission, we obtained good correlation measures for both coefficients, but we only had one station to compare NO_x and PM pollutants, and two stations for CO pollutants. Although the number of stations was not big enough to present the overall pollution in the city, we did reproduce it to some extent and showed the possible joint use of different data sets, ways of data processing, and overcoming their versatile data forms for the purpose of creating a decision-making tool, which was the main goal all along.

There are several possible reasons for not being able to predict absolute numbers at some crossing points. First, we had the data set used for the OD probability matrix generation from July 2017, while the rest of the data were available in November and December 2019. In addition, the chosen days affected the model validation. It is reported in the literature that traffic patterns are distinct across days and cannot be generalized [66]. Therefore, to get a reliable model, a policy-maker should use data sets for assumptions, calibration, and validation from the same time period, preferably the same day. Moreover, the reported accuracy depended on the nature of data used in calibration and validation processes, and its spatial and temporal density. Contrary to the papers that compared the overall pattern produced by a model with the congestion or emission maps or a global number [4,6,42,67], we calculated the measures on a more precise spatial-temporal level and compared the real numbers, and not only the visual patterns. Even though this is a more precise way to go, there was a possibility that available data did not capture correctly the complete interactions of individuals. Presumably, the best option would be to combine both ways. With respect to the emission calculation, we chose to use a simplified aggregated approach due to the lack of the traffic fleet data and measured emission coefficients in Serbia. However, there is an open question to what extent the aggregated coefficients for Austria from the HBEFA book agree with the traffic conditions in Serbia. As we only had two stations for result verification, this question remains open for further research. From the perspective of modeling, the best way would be to dynamically simulate the vehicle emissions for every simulation step in the model, simulating the cold start and warm emission activities for every type of vehicle and its engine. Nonetheless, this increases the model's complexity and lengthens the simulation time, which potentially leads to other issues. Finally, we simulated traffic and assessed its emission at the municipality level, and generally, it is more difficult to tune the model to the real-world scenarios when is larger and more complex, compared to the models that simulated the traffic on a smaller area as it is one district.

The reported model accuracy, in addition to data, depends on the way it is validated. The literature contains numerous examples of various traffic modeling practices [68] and their validation methods differ. There are methods in which validation is only based on a visual comparison of a traffic heat map with the traffic map from services, such as Google or Baidu maps [4,6,42,67]. Moreover, there are papers that use only one global number to validate the overall pattern produced by a model, such as a comparison of the overall CO or NO_x emission produced by a model with the overall emission reported in cities [4,67]. The question is how accurate the models are and to what extent the reported accuracy is influenced by the data and validation methodology used. On the other hand, some papers do validate the produced pattern with actual numbers on a higher spatial-temporal scale, but their approaches also differ. Some examples are comparisons of hourly traffic to annual average daily traffic [32], total number of vehicles with one number in several locations [43], traffic volumes with 600 sensors [44], a model output with an output from another model [69], etc. All these models are distinguished by the level of detail,

assumptions, and research questions they include. From this point of view, comparing the accuracy of different models is difficult to undertake.

Despite not being able to predict the absolute numbers and scarce data sets used for validation, we believe the presented methodology could serve for the creation of a decision-making tool and be used for gaining insights when comparing different policies in order to find the optimal outcome. Relying on an evidence-based conclusion promises smoother and more intelligent decision-making. However, the inclusion of more data also contributes to the complexity of the model, leading to other problems such as long execution time and increased complexity. Therefore, it is necessary to find a compromise between the level of reality in the model and its goal. Although there are no perfect models, they could support to some extent decision making processes and provide insights into phenomena of interest. Policy-makers just need to be aware of the limitations of the model. In addition, computer resources and capacities together with algorithms are advancing, so we are moving toward more realistic models.

Additionally, this study gives a retrospective of the challenges of combining disparate data sources with different spatial-temporal resolutions. Using heterogeneous data sources brings many open questions and challenges [9,10,70]. On the one hand, we can argue that passively collected data are mostly free of bias and do not suffer from the selection of sampling methods. As they cover a larger sample size, it is usually possible to obtain a broad picture of behavioral patterns. Nevertheless, using data not primarily made to support some subject, like we did with telecom data in traffic studies, introduces gaps and needs to be fused with other data sources in order to unlock their value. Combining different data sets brings a lot of uncertainties and challenges caused by mismatches in data resolution and the multimodal and dynamic nature of data and introduces an additional effort to overcome them [10]. To get from telecom data to traffic emission, we went through various steps of combining the data sources. To match the spatial resolution of telecom data with official population data, we upscaled the domain at the local community level. We populated the ABM model at the local community level and connected their trips to the traffic network. Furthermore, to compare the traffic or its emissions with the stations and counters in the case study area, we sum up traffic or its emissions on the closest traffic link and calculate the correlations. In addition, different data sources have different error distributions, and modelers should keep this in mind when using them [9]. In order to overcome these issues, modelers should support their solutions with as much data as possible and expert opinions. In terms of telecom data, we recognized that the main limitation was data availability. As a consequence, in order to replicate the methodology, other case studies would require similar data sources. Telecom operators are usually reluctant to give data to third parties, as they need to protect user privacy. If they decide to share, they need to make an extra effort to anonymize and aggregate the data. Another approach was suggested by the Open Algorithms (OPAL) initiative, which advocates moving the algorithms to data [71]. Therefore, raw data is never revealed to outside parties, only vetted algorithms run on telecom companies' servers. Nevertheless, with this study, we believe that we promote the value of using telecom data, in addition to its main use. Besides this, telecom data introduce other gaps and issues when it comes to mobility studies. They are defined by the resolution of RBSs domains, which are not homogeneous across the case study. Regarding the trip generation patterns, spatial resolution is not big enough to capture spatially detailed movements. Moreover, it is difficult to distinguish drivers from other traffic participants. It is likely that using some advanced techniques that include analysis of the speed of changing locations could distinguish pedestrians from others and overcome the problem. However, there is an issue in urban zones, where traffic intensity is high and speed is very low. It is also a challenge to differentiate between passengers and drivers. To circumvent the issues mentioned above, we did not extract precise spatial-temporal patterns but probabilities of movement for each local community. With the given results, we argue that this is sufficient when it comes to the spatially larger case study area, such as the municipality level we chose to model.

In the context of transport studies, the results represent the estimated traffic flow and its emissions during different time periods of the day. Traffic flow assessment is very useful in traffic planning as a measure of exposure. In many traffic safety studies, exposure measures are the most important factor for modeling traffic accidents and determining the influencing factors that contribute to traffic accidents [72,73]. In addition, exposure measures play an important role in traffic planning and traffic infrastructure development, where they help transport professionals to better understand the mechanisms and factors for smooth traffic flow and alleviation of its emission. To create quality decisions with respect to traffic regulation and its emission, data are needed at the precise spatial-temporal level in the near real-time, which is not always available. This study emphasizes the value of data and makes a step forward in using big data in transport studies.

5. Conclusions

In the study, an agent-based methodology for traffic simulation and emission estimation was presented. The methodology relies on the use of heterogeneous data sources. Telecom data was used for temporal and spatial movement probability assessments between different local communities in the case study area. Using the probabilities together with an official census, GIS data, and data about road networks from the OSM, we built the Traffic Agent-Based Model in the GAMA platform. Furthermore, we used the model output and estimated traffic emissions. Using data from automatic vehicle counters, air quality data measured from several stations in the case study area, we proved that the model results are consistent with actual traffic and emission conditions. However, the achieved accuracy was dictated by the available data. Nevertheless, this study represents a positive step towards using and combining heterogeneous big data in urban studies.

The proposed methodology has several constraints that need to be addressed. Travel demands (such as pedestrians, trucks, etc.) from telecom data were not extracted. Instead, we calculated the movement probabilities and incorporated them into the model. In order to simulate the traffic, we use *Advanced Traffic Plugin*, offered by the GAMA framework, which uses precise traffic conditions by including lanes, directionality of streets, drivers' behavior, etc. The model needs to be run at a 1-s time resolution and without the possibility to parallelize the processes within the model. If we add a larger number of agents and traffic networks that should be simulated on that, we could end up with a really complex and time-consuming model. For the proposed model with 10% of the simulated population, simulation lasted up to 4 days, which is a relatively long period. Within the model, different types of vehicles cannot be distinguished. Traffic lights were excluded. To simplify the model, we only included the main roads in the peripheral parts of the municipality. We did not incorporate emission calculus within the model, but instead, we calculated it as a posterior, which is inconsistent with the resolution of the built traffic model. We did this because of the lack of data and to avoid additional model complexity. We did not include any external factors, such as weather conditions (wind, temperature, etc.) and chemical reactions.

In future work, we intend to incorporate more data into the model and thus enrich it. In addition, testing different traffic policies and observing their impact on traffic flow and emissions during one day is planned.

Author Contributions: Conceptualization, Nastasija Grujić, Ioannis N. Athanasiadis, Gert Jan Hofstede, Nikola Obrenović and Sjoukje Osinga; methodology, Nastasija Grujić; software, Nastasija Grujić; validation, Nastasija Grujić, and Miloš Pljakić; formal analysis, Nastasija Grujić; investigation, Nastasija Grujić resources, Nastasija Grujić; data curation, Nastasija Grujić, Sanja Brdar and Miloš Pljakić; writing—original draft preparation, Nastasija Grujić; writing—review and editing, all authors; visualization, Nastasija Grujić; supervision, Sanja Brdar, Miro Govedarica, and Vladimir Crnojević; project administration, Sanja Brdar; funding acquisition, Vladimir Crnojević. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by DRAGON (Data Driven Precision Agriculture Services and Skill Acquisition) project funded from European Union’s Horizon 2020 research and by the financial support of the Ministry of Education, Science and Technological Development of the Republic of Serbia (Grant No. 451-03-68/2022-14/200358).

Acknowledgments: The authors acknowledge all companies that participated in providing the data, especially Telecom Serbia, for sharing anonymized telecom data.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Results

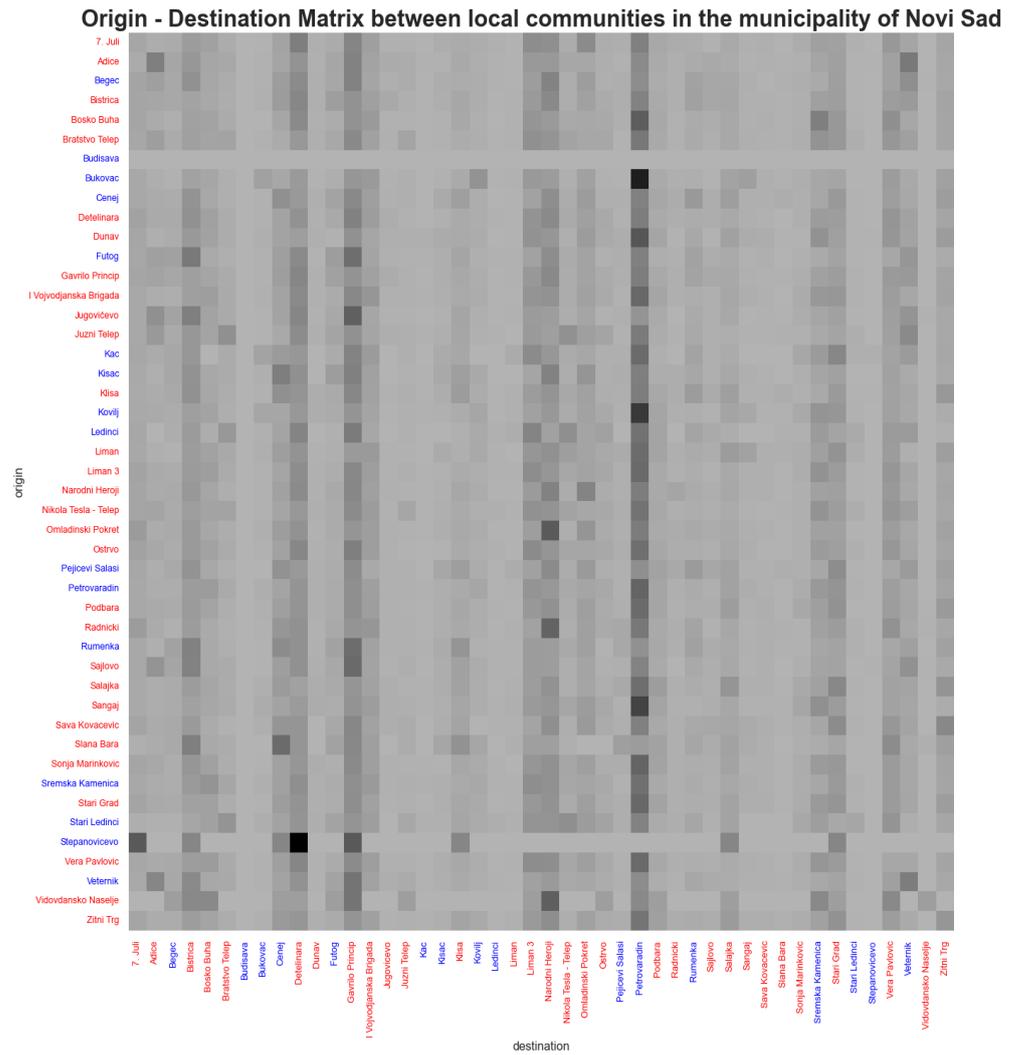


Figure A1. Origin-Destination Probability matrix between local communities in the case study area; red ticks indicate local communities that are in the city area, while the blue ticks represent local communities located outside the city.

Predicted and observed speed

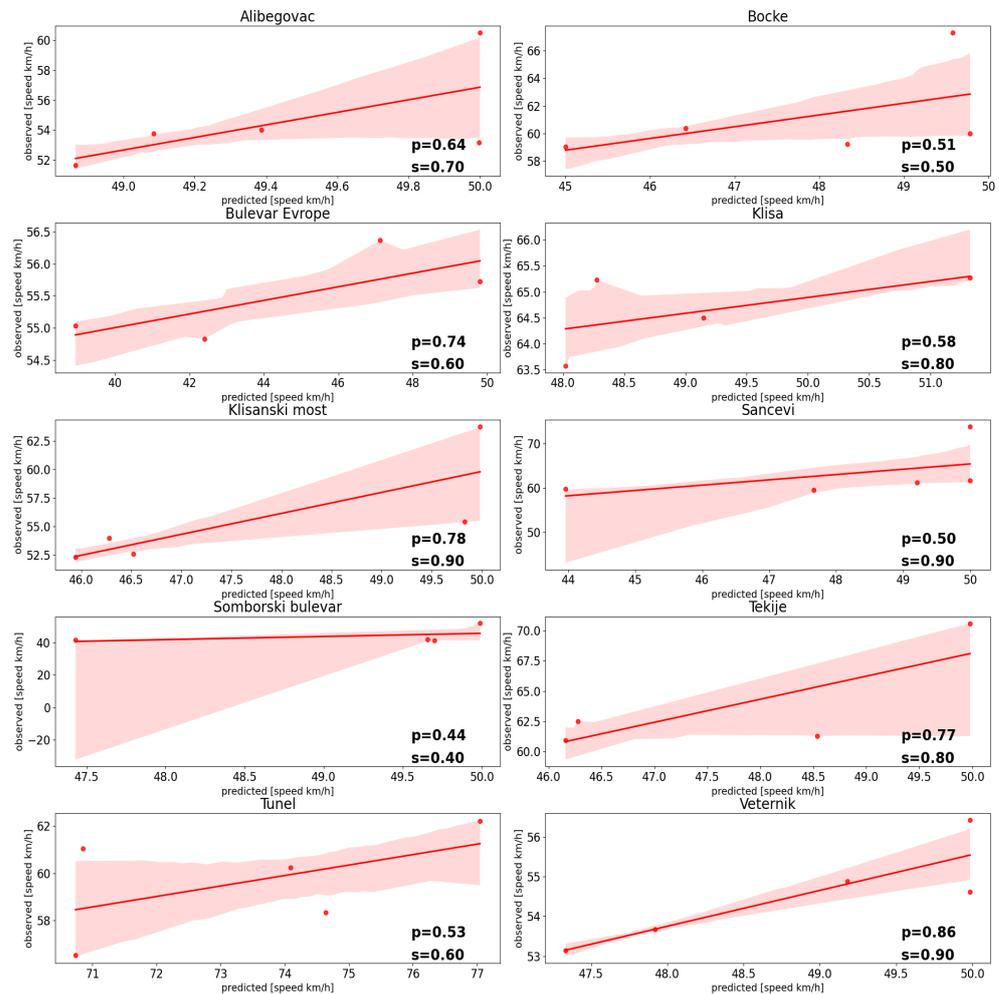


Figure A2. Calibration—Pearson (p) and Spearman (s) correlation coefficients between predicted and observed speed in automatic vehicle counters data set. The selected day from automatic vehicle counters data set was 6 November 2019. Please note: the axes are not in the same scale to emphasize the captured trend.

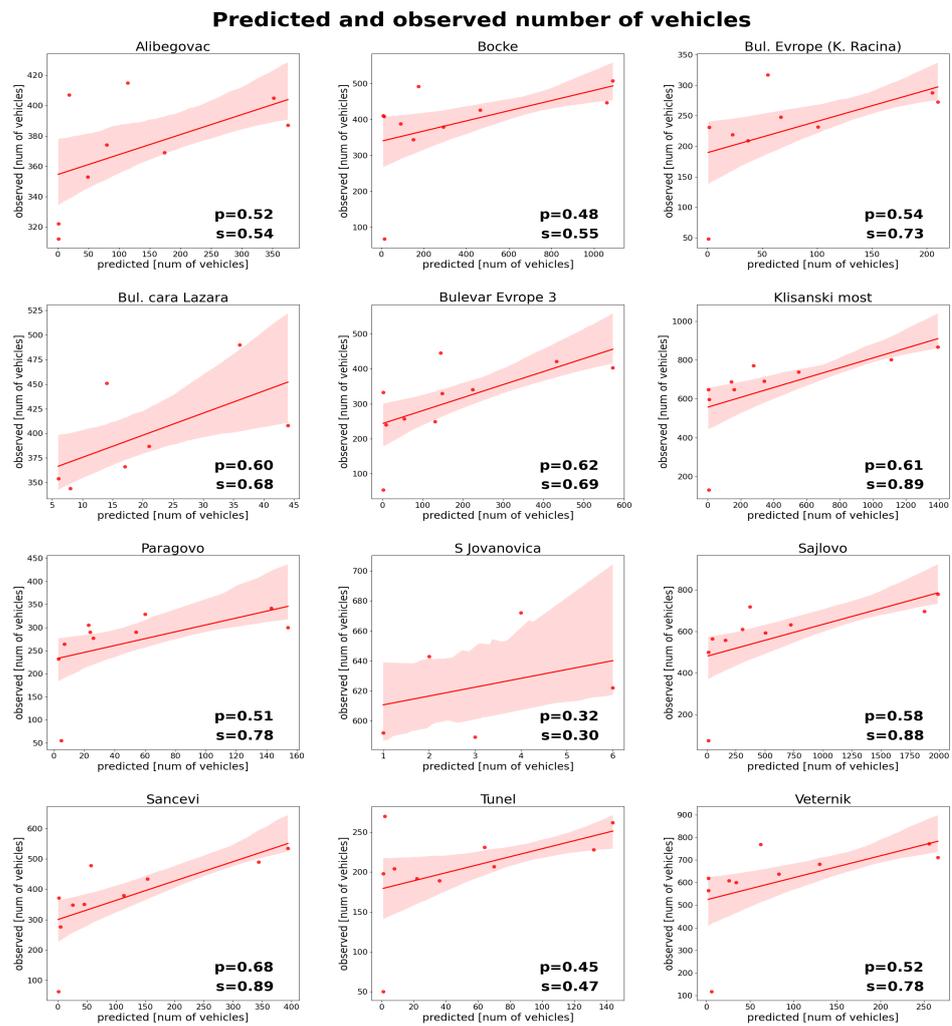


Figure A3. Validation—Pearson’s (p) and Spearman’s (s) correlation coefficients between predicted and observed number of cars in automatic vehicle counters data set. The selected day from automatic vehicle counters data set was 3 December 2019. *Please note: the axes are not in the same scale to emphasize the captured trend.*

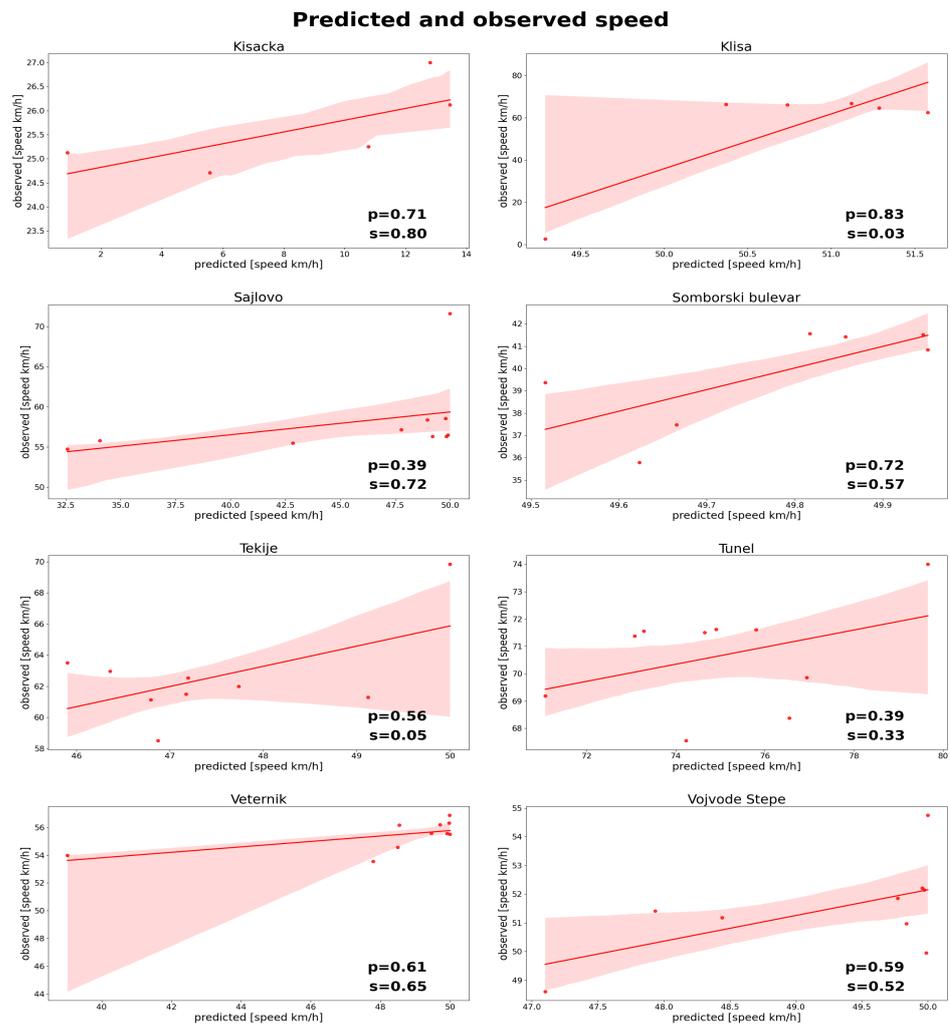


Figure A4. Validation—Pearson’s (p) and Spearman’s (s) correlation coefficients between predicted and observed speed in automatic vehicle counters data set. The selected day from automatic vehicle counters data set was 3 December 2019. *Please note: the axes are not in the same scale to emphasize the captured trend.*

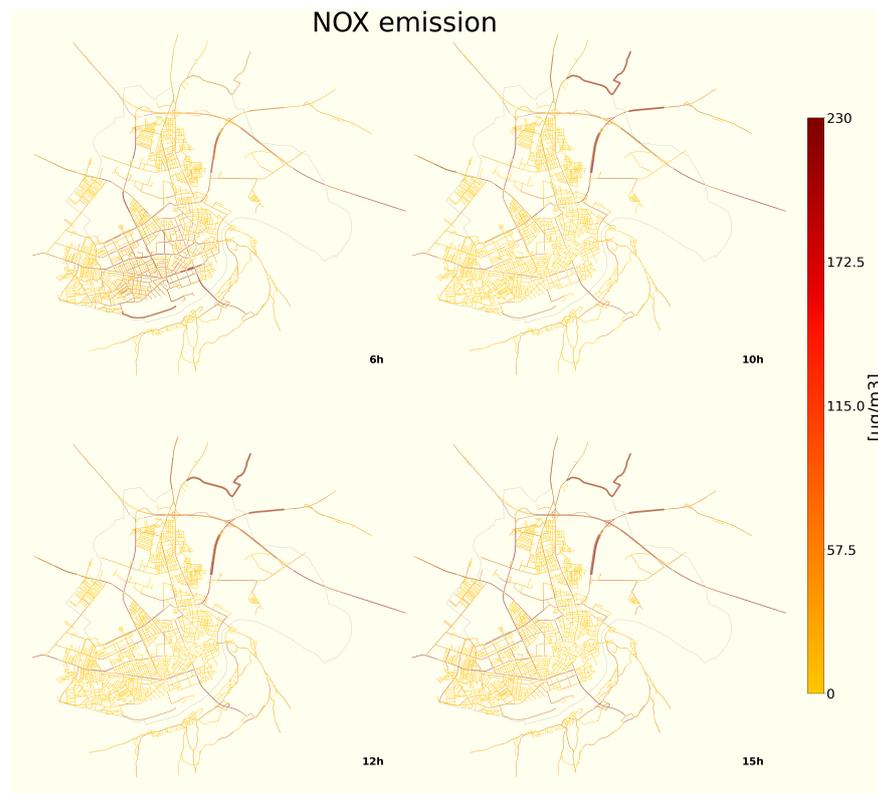


Figure A5. NO_x emission intensity at different hours.

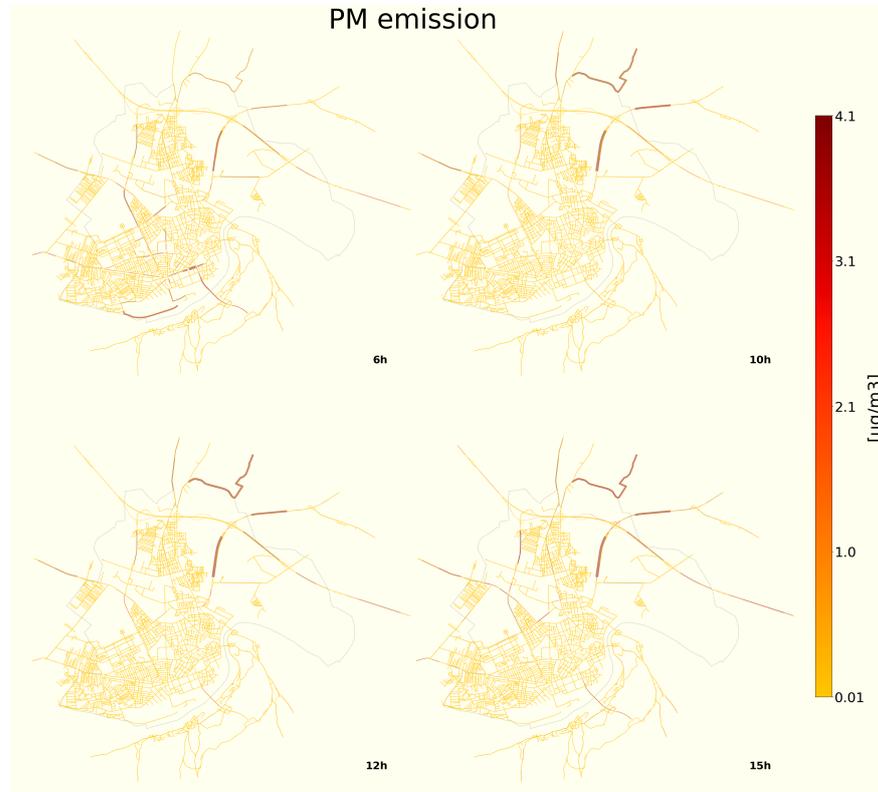


Figure A6. PM emission intensity at different hours.

References

- Jain, V.; Sharma, A.; Subramanian, L. Road traffic congestion in the developing world. In Proceedings of the 2nd ACM Symposium on Computing for Development, Atlanta, GA, USA, 11–12 March 2012; pp. 1–10.
- WHO. World Health Organization—Urban Population Growth, Global Health Observatory. 2014. Available online: <https://www.who.int/> (accessed on 18 November 2021).
- Khalfan, A.; Andrews, G.; Li, H. Real World Driving: Emissions in Highly Congested Traffic. In Proceedings of the SAE Powertrain Fuels and Lubricants Meeting 2017, Beijing, China, 16–18 October 2017; SAE International: Warrendale, PA, USA, 2017.
- Hofer, C.; Jäger, G.; Füllsack, M. Large scale simulation of CO₂ emissions caused by urban car traffic: An agent-based network approach. *J. Clean. Prod.* **2018**, *183*, 1–10. [[CrossRef](#)]
- Martinez, L.M.; Viegas, J.M. Assessing the impacts of deploying a shared self-driving urban mobility system: An agent-based model applied to the city of Lisbon, Portugal. *Int. J. Transp. Sci. Technol.* **2017**, *6*, 13–27. [[CrossRef](#)]
- Hofer, C.; Jäger, G.; Füllsack, M. Including traffic jam avoidance in an agent-based network model. *Comput. Soc. Netw.* **2018**, *5*, 1–12. [[CrossRef](#)] [[PubMed](#)]
- Kickhöfer, B.; Nagel, K. Towards high-resolution first-best air pollution tolls. *Netw. Spat. Econ.* **2016**, *16*, 175–198. [[CrossRef](#)]
- Zheng, N.; Waraich, R.A.; Axhausen, K.W.; Geroliminis, N. A dynamic cordon pricing scheme combining the macroscopic fundamental diagram and an agent-based traffic model. *Transp. Res. Part A Policy Pract.* **2012**, *46*, 1291–1303. [[CrossRef](#)]
- Willumsen, L. *Use of Big Data in Transport Modelling*; OECD/ITF: Paris, France, 2021.
- Brdar, S.; Novović, O.; Grujić, N.; González-Vélez, H.; Truić, C.O.; Benkner, S.; Bajrovic, E.; Papadopoulos, A. Big Data Processing, Analysis and Applications in Mobile Cellular Networks. In *High-Performance Modelling and Simulation for Big Data Applications*; Springer: Cham, Switzerland, 2019; pp. 163–185.
- Becker, R.A.; Caceres, R.; Hanson, K.; Loh, J.M.; Urbanek, S.; Varshavsky, A.; Volinsky, C. A tale of one city: Using cellular network data for urban planning. *IEEE Pervasive Comput.* **2011**, *10*, 18–26. [[CrossRef](#)]
- Arai, A.; Witayangkurn, A.; Kanasugi, H.; Fan, Z.; Ohira, W.; Pedro, S. *Building a Data Ecosystem for Using Telecom Data to Inform the COVID-19 Response Efforts*; Zenodo: Geneva, Switzerland, 2020.
- Bosetti, P.; Poletti, P.; Stella, M.; Lepri, B.; Merler, S.; De Domenico, M. Reducing measles risk in Turkey through social integration of Syrian refugees. *arXiv* **2019**, arXiv:1901.04214.
- Brdar, S.; Gavrić, K.; Čulibrk, D.; Crnojević, V. Unveiling spatial epidemiology of HIV with mobile phone data. *Sci. Rep.* **2016**, *6*, 19342. [[CrossRef](#)]
- Lima, A.; De Domenico, M.; Pejovic, V.; Musolesi, M. Disease containment strategies based on mobility and information dissemination. *Sci. Rep.* **2015**, *5*, 10650. [[CrossRef](#)]
- Lu, X.; Tan, J.; Cao, Z.; Xiong, Y.; Qin, S.; Wang, T.; Liu, C.; Huang, S.; Zhang, W.; Marczak, L.B.; et al. Mobile Phone-Based Population Flow Data for the COVID-19 Outbreak in Mainland China. *Health Data Sci.* **2021**, *2021*, 9796431. [[CrossRef](#)]
- Wesolowski, A.; Qureshi, T.; Boni, M.F.; Sundsøy, P.R.; Johansson, M.A.; Rasheed, S.B.; Engø-Monsen, K.; Buckee, C.O. Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 11887–11892. [[CrossRef](#)] [[PubMed](#)]
- Novović, O.; Brdar, S.; Mesaroš, M.; Crnojević, V.; Papadopoulos, A.N. Uncovering the relationship between human connectivity dynamics and land use. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 140. [[CrossRef](#)]
- De Nadai, M.; Staiano, J.; Larcher, R.; Sebe, N.; Quercia, D.; Lepri, B. The death and life of great Italian cities: A mobile phone data perspective. In Proceedings of the 25th International Conference on World Wide Web, Montreal, QC, Canada, 11–15 May 2016; pp. 413–423.
- Grujić, N.; Novović, O.; Brdar, S.; Crnojević, V.; Govedarica, M. Mobile Phone Data visualization using Python QGIS API. In Proceedings of the 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Republic of Srpska, 20–21 March 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
- Doyle, J.; Hung, P.; Farrell, R.; McLoone, S. Population mobility dynamics estimated from mobile telephony data. *J. Urban Technol.* **2014**, *21*, 109–132. [[CrossRef](#)]
- Louail, T.; Lenormand, M.; Ros, O.G.C.; Picornell, M.; Herranz, R.; Frias-Martinez, E.; Ramasco, J.J.; Barthelemy, M. From mobile phone data to the spatial structure of cities. *Sci. Rep.* **2014**, *4*, 5276. [[CrossRef](#)]
- Brdar, S.; Grujić, N.; Obrenović, N.; Novović, O.; Lugonja, P.; Minić, V.; Bajić, Ž.; Milovanović, M.; Rokvić, N. *Project-Depopulation Sensing by Integrative Knowledge Discovery from Big Data*; Biosense: Novi Sad, Serbia, 2021.
- Grujić, N.; Brdar, S.; Novović, O.; Govedarica, M.; Crnojević, V. Evidence of urban segregation from mobile phone data: A case study of Novi Sad. In Proceedings of the 2019 27th Telecommunications Forum (TELFOR), Belgrade, Serbia, 26–27 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–4.
- Grujić, N.; Brdar, S.; Novović, O.; Obrenović, N.; Govedarica, M.; Crnojević, V. Biclustering for uncovering spatial-temporal patterns in telecom data. In Proceedings of the EGU General Assembly Conference Abstracts, Online, 19–30 April 2021; p. EGU21-14423.
- Pappalardo, L.; Pedreschi, D.; Smoreda, Z.; Giannotti, F. Using big data to study the link between human mobility and socio-economic development. In Proceedings of the 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 29 October–1 November 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 871–878.

27. Steele, J.E.; Sundsøy, P.R.; Pezzulo, C.; Alegana, V.A.; Bird, T.J.; Blumenstock, J.; Bjelland, J.; Engø-Monsen, K.; De Montjoye, Y.A.; Iqbal, A.M.; et al. Mapping poverty using mobile phone and satellite data. *J. R. Soc. Interface* **2017**, *14*, 20160690. [[CrossRef](#)]
28. Galiana, L.; Sakarovitch, B.; Smoreda, Z. Understanding socio-spatial segregation in French cities with mobile phone data. *DGINS18 2018, unpublished manuscript*.
29. Lu, X.; Wrathall, D.J.; Sundsøy, P.R.; Nadiruzzaman, M.; Wetter, E.; Iqbal, A.; Qureshi, T.; Tatem, A.J.; Canright, G.S.; Engø-Monsen, K.; et al. Detecting climate adaptation with mobile network data in Bangladesh: anomalies in communication, mobility and consumption patterns during cyclone Mahasen. *Clim. Chang.* **2016**, *138*, 505–519. [[CrossRef](#)]
30. Pastor-Escuredo, D.; Morales-Guzmán, A.; Torres-Fernández, Y.; Bauer, J.M.; Wadhwa, A.; Castro-Correa, C.; Romanoff, L.; Lee, J.G.; Rutherford, A.; Frias-Martinez, V.; et al. Flooding through the lens of mobile phone activity. In Proceedings of the IEEE Global Humanitarian Technology Conference (GHTC 2014), San Jose, CA, USA, 10–13 October 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 279–286.
31. Wilson, R.; zu Erbach-Schoenberg, E.; Albert, M.; Power, D.; Tudge, S.; Gonzalez, M.; Guthrie, S.; Chamberlain, H.; Brooks, C.; Hughes, C.; et al. Rapid and near real-time assessments of population displacement using mobile phone data following disasters: The 2015 Nepal earthquake. *PLoS Curr.* **2016**, *8*, ecurrents.dis.d073fbee328e4c39087bc086d694b5c. [[CrossRef](#)]
32. Järvi, O.; Ahas, R.; Saluveer, E.; Derudder, B.; Witlox, F. Mobile phones in a traffic flow: A geographical perspective to evening rush hour traffic analysis using call detail records. *PLoS ONE* **2012**, *7*, e49171.
33. Bar-Gera, H. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Transp. Res. Part C Emerg. Technol.* **2007**, *15*, 380–391. [[CrossRef](#)]
34. Sørensen, A.Ø.; Bjelland, J.; Bull-Berg, H.; Landmark, A.D.; Akhtar, M.M.; Olsson, N.O. Use of mobile phone data for analysis of number of train travellers. *J. Rail Transp. Plan. Manag.* **2018**, *8*, 123–144.
35. Alexander, L.; Jiang, S.; Murga, M.; González, M.C. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transp. Res. Part C Emerg. Technol.* **2015**, *58*, 240–250. [[CrossRef](#)]
36. Zagatti, G.A.; Gonzalez, M.; Avner, P.; Lozano-Gracia, N.; Brooks, C.J.; Albert, M.; Gray, J.; Antos, S.E.; Burci, P.; Zu Erbach-Schoenberg, E.; et al. A trip to work: Estimation of origin and destination of commuting patterns in the main metropolitan regions of Haiti using CDR. *Dev. Eng.* **2018**, *3*, 133–165. [[CrossRef](#)]
37. Iqbal, M.S.; Choudhury, C.F.; Wang, P.; González, M.C. Development of origin–destination matrices using mobile phone call data. *Transp. Res. Part C Emerg. Technol.* **2014**, *40*, 63–74. [[CrossRef](#)]
38. Tettamanti, T.; Demeter, H.; Varga, I. Route choice estimation based on cellular signaling data. *Acta Polytech. Hung.* **2012**, *9*, 207–220.
39. Sakamane, P.; Phithakkitnukoon, S.; Smoreda, Z.; Ratti, C. Methods for inferring route choice of commuting trip from mobile phone network data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 306. [[CrossRef](#)]
40. Wang, H.; Calabrese, F.; Di Lorenzo, G.; Ratti, C. Transportation mode inference from anonymized and aggregated mobile phone call detail records. In Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems, Beijing, China, 12–15 October 2008; IEEE: Piscataway, NJ, USA, 2010; pp. 318–323.
41. Doyle, J.; Hung, P.; Kelly, D.; McLoone, S.F.; Farrell, R. *Utilising Mobile Phone Billing Records for Travel Mode Discovery*; Maynooth University: Kildare, Ireland, 2011.
42. Wu, H.; Liu, L.; Yu, Y.; Peng, Z.; Jiao, H.; Niu, Q. An agent-based model simulation of human mobility based on mobile phone data: How commuting relates to congestion. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 313. [[CrossRef](#)]
43. Bassolas, A.; Ramasco, J.J.; Herranz, R.; Cantú-Ros, O.G. Mobile phone records to feed activity-based travel demand models: MATSim for studying a cordon toll policy in Barcelona. *Transp. Res. Part A Policy Pract.* **2019**, *121*, 56–74. [[CrossRef](#)]
44. A Generative Model of Urban Activities from Cellular Data. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 1682–1696. [[CrossRef](#)]
45. Zhang, L.; Huang, J.; Liu, Z.; Vu, H.L. An agent-based model for real-time bus stop-skipping and holding schemes. *Transp. A Transp. Sci.* **2021**, *17*, 615–647. [[CrossRef](#)]
46. Hussain, I.; Knapen, L.; Galland, S.; Bellemans, T.; Janssens, D.; Wets, G. Organizational-based model and agent-based simulation for long-term carpooling. *Future Gener. Comput. Syst.* **2016**, *64*, 125–139. [[CrossRef](#)]
47. Motieyan, H.; Mesgari, M.S. An agent-based modeling approach for sustainable urban planning from land use and public transit perspectives. *Cities* **2018**, *81*, 91–100. [[CrossRef](#)]
48. Bonabeau, E. Agent-based modeling: Methods and techniques for simulating human systems. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7280–7287. [[CrossRef](#)]
49. Batty, M. Agents, cells, and cities: New representational models for simulating multiscale urban dynamics. *Environ. Plan. A* **2005**, *37*, 1373–1394. [[CrossRef](#)]
50. Crooks, A.; Malleon, N.; Manley, E.; Heppenstall, A. *Agent-Based Modelling and Geographical Information Systems: A Practical Primer*; Sage: Thousand Oaks, CA, USA, 2018.
51. Heppenstall, A. *Agent-Based Models for Geographical Systems: A Review*; University College London: London, UK, 2019.
52. Taillandier, P.; Gaudou, B.; Grignard, A.; Huynh, Q.N.; Marilleau, N.; Caillou, P.; Philippon, D.; Drogoul, A. Building, composing and experimenting complex spatial models with the GAMA platform. *GeoInformatica* **2019**, *23*, 299–322. [[CrossRef](#)]
53. Handbook on Emission Factors for Road Transport. Available online: <https://www.hbefa.net/e/index.html> (accessed on 3 February 2022).

54. Levinson, D.M.; Kumar, A. The rational locator: Why travel times have remained stable. *J. Am. Plan. Assoc.* **1994**, *60*, 319–332. [[CrossRef](#)]
55. Schafer, A. Regularities in travel demand: An international perspective. *J. Transp. Stat.* **2000**, *3*, 1–31. [[CrossRef](#)]
56. Landmark, A.D.; Arnesen, P.; Södersten, C.J.H.; Hjelkrem, O.A. Mobile phone data in transportation research: Methods for benchmarking against other data sources. *Transportation* **2021**, *48*, 2883–2905. [[CrossRef](#)]
57. Okabe, A.; Boots, B.; Sugihara, K.; Chiu, S.N. *Concepts and Applications of Voronoi Diagrams*; John Wiley & Sons Ltd.: Chichester, UK, 2000.
58. Salgado, M.; Gilbert, N. Agent based modelling. In *Handbook of Quantitative Methods for Educational Research*; Brill Sense: Paderborn, Germany, 2013; pp. 247–265.
59. Bazzan, A.L.C.; Klügl, F. A review on agent-based technology for traffic and transportation. *Knowl. Eng. Rev.* **2014**, *29*, 375–403. [[CrossRef](#)]
60. Taillandier, P. Traffic simulation with the GAMA platform. In Proceedings of the Eighth International Workshop on Agents in Traffic and Transportation, Paris, France, 5–6 May 2014; p. 8.
61. Shaharuddin, R.A.; Misro, M.Y. Traffic simulation using agent based modelling. In Proceedings of the AIP Conference Proceedings, Chiangmai, Thailand, 20–22 February 2021; AIP Publishing LLC: Melville, NY, USA, 2021; Volume 2423, p. 020035.
62. Kickhöfer, B.; Hülsmann, F.; Gerike, R.; Nagel, K. Rising car user costs: Comparing aggregated and geo-spatial impacts on travel demand and air pollutant emissions. In *Smart Transport NEtworks*; Edward Elgar Publishing: Cheltenham, UK, 2013.
63. Novi Sad. Available online: https://en.wikipedia.org/wiki/Novi_Sad (accessed on 21 September 2021).
64. New Bridge. Available online: <https://www.021.rs/story/Novi-Sad/Vesti/268583/Planiranje-novog-mosta-u-Novom-Sadu-Dokument-na-javnom-uidu-prigovori-do-11-aprila.html> (accessed on 21 September 2021).
65. SEPA—Serbia Environmental Protection Agency. Available online: <http://www.sepa.gov.rs/> (accessed on 1 February 2022).
66. Ni, D. *Traffic Flow Theory: Characteristics, Experimental Methods, and Numerical Techniques*; Butterworth-Heinemann: Oxford, UK, 2015.
67. Plakolb, S.; Jäger, G.; Hofer, C.; Füllsack, M. Mesoscopic urban-traffic simulation based on mobility behavior to calculate NOx emissions caused by private motorized transport. *Atmosphere* **2019**, *10*, 293. [[CrossRef](#)]
68. Azlan, N.N.N.; Rohani, M.M. Overview of application of traffic simulation model. In Proceedings of the MATEC Web of Conferences, Penang, Malaysia, 23 February 2018; EDP Sciences: Les Ulis, France, 2018; Volume 150, p. 03006.
69. Zhao, B.; Kumar, K.; Casey, G.; Soga, K. Agent-based model (ABM) for city-scale traffic simulation: A case study on San Francisco. In Proceedings of the International Conference on Smart Infrastructure and Construction 2019 (ICSIC) Driving Data-Informed Decision-Making, Cambridge, UK, 8–10 July 2019; ICE Publishing: London, UK, 2019; pp. 203–212.
70. Wang, H.; Xu, Z.; Fujita, H.; Liu, S. Towards felicitous decision making: An overview on challenges and trends of Big Data. *Inf. Sci.* **2016**, *367*, 747–765. [[CrossRef](#)]
71. Lepri, B.; Oliver, N.; Letouzé, E.; Pentland, A.; Vinck, P. Fair, transparent, and accountable algorithmic decision-making processes. *Philos. Technol.* **2018**, *31*, 611–627. [[CrossRef](#)]
72. Aguero-Valverde, J.; Jovanis, P.P. Spatial analysis of fatal and injury crashes in Pennsylvania. *Accid. Anal. Prev.* **2006**, *38*, 618–625. [[CrossRef](#)]
73. Quddus, M.A. Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accid. Anal. Prev.* **2008**, *40*, 1486–1497. [[CrossRef](#)]