Building a Large-Scale Micro-Simulation Transport Scenario Using Big Data

Joerg Schweizer [1,*] , **Cristian Poliziani** [1] , **Federico Rupi** [1] , **Davide Morgano** [1] and **Mattia Magi** [2]

1   Department of Civil, Chemical, Environmental and Materials Engineering, University of Bologna, 40126 Bologna, Italy; cristian.poliziani2@unibo.it (C.P.); federico.rupi@unibo.it (F.R.); davide.morgano@studio.unibo.it (D.M.)

2   Righetti & Monte—Ingegneri e Architetti Associati, 40126 Bologna, Italy; mattia.magi2@studio.unibo.it

*   Correspondence: joerg.schweizer@unibo.it

Abstract: A large-scale agent-based microsimulation scenario including the transport modes car, bus, bicycle, scooter, and pedestrian, is built and validated for the city of Bologna (Italy) during the morning peak hour. Large-scale microsimulations enable the evaluation of city-wide effects of novel and complex transport technologies and services, such as intelligent traffic lights or shared autonomous vehicles. Large-scale microsimulations can be seen as an interdisciplinary project where transport planners and technology developers can work together on the same scenario; big data from OpenStreetMap, traffic surveys, GPS traces, traffic counts and transit details are merged into a unique transport scenario. The employed activity-based demand model is able to simulate and evaluate door-to-door trip times while testing different mobility strategies. Indeed, a utility-based mode choice model is calibrated that matches the official modal split. The scenario is implemented and analyzed with the software SUMOPy/SUMO which is an open source software, available on GitHub. The simulated traffic flows are compared with flows from traffic counters using different indicators. The determination coefficient has been 0.7 for larger roads (width greater than seven meters). The present work shows that it is possible to build realistic microsimulation scenarios for larger urban areas. A higher precision of the results could be achieved by using more coherent data and by merging different data sources.

Keywords: large scale; agent-based; micro-simulation; mode choice model; big data; GPS traces; OpenStreetMap; GTFS; SUMO

1. Introduction

Tracing the exact movements of individuals and vehicles from door to door is feasible with today's computers—even with a population of larger urban areas. The microsimulation of a virtual copy of the real population, buildings, and infrastructure, called synthetic population or digital twin, is becoming a reality thanks to the availability of big data, larger random-access memories, and faster CPUs. Modeling the interactions between neighboring vehicles or between vehicles and pedestrians results in precise trip times and speed profiles, enabling accurate performance evaluations and transport impact analysis. Microsimulations are sensitive to the individual's trip experience as details of infrastructure and transport-services can significantly change travel times. Furthermore, emerging transport technologies such as intelligent traffic light systems, platooning, and driver assistance, as well as alternative means of transportation such as bike sharing or shared autonomous vehicles (SAVs), can be integrated in and evaluated by micro-simulations in a realistic environment.

The "transport technology development" and "transport planning" can be seen as the main drivers behind the creation of ever more realistic transport models, even though they approach the problem from completely different angles: while transport technology development is primarily interested in realistically evaluating the performance of the deployed

technology, the transport planner is primarily interested in predicting the behavior of the transport system as a whole in order to identify the best possible future transport scenario. This approach includes all alternative transport modes and the transport choices made by the users, meaning the planner pays more attention to the demand models, such as trip generation, activity location choice and mode choice, while the technology developer is interested in accurate transport supply models, such as vehicle controls or communications between vehicles (V2V) and between vehicle and infrastructure (V2I).

The "problem" is that there are apparent difficulties to create large-scale, city-wide simulation models that include all the details of the technologies and devices, as mentioned before. In theory, a microsimulation can provide both accurate demand and precise supply models. However, in practice, it is challenging to create such a complex microsimulation scenario as this would require a myriad of real-world details such as a refined road network with speed limits, lane access rights, pedestrian crossings, restricted turns at intersections, traffic light-phases, and parking facilities; the traffic generation would require public transport lines with timetables, vehicle types and frequencies, population data, mobility plans of individuals, and much more.

The increasing availability of "big data" does certainly facilitate the creation of microsimulation scenarios. In the present context, big data stands for large, disaggregate, area-covering databases, which are often (but not always) publicly available. Examples are the OpenStreetMap (OSM) database [1], GPS traces recorded by citizens, geo referenced cell phone data, social network activities, or vehicle flow measurements from road-side detectors. Big data, together with more aggregate data such as origin-to-destination matrices (OD matrices), may play different roles in the scenario building process: OSM does often contain most of the required information on the transport networks. It appears more difficult to use big data for modelling the transport demand, as GPS traces or social networks are usually not linked to reliable user profiles; cell phone operators are in possession of attribute-rich georeferenced data but cannot release user information for privacy reasons. This means that user behavior cannot be directly calibrated from this "poor" data as it is the case for properly designed traditional surveys. However, big data from different sources may be merged prior to a model calibration in order to increase the information content; or big data can be used to enrich high-quality travel surveys.

The brief literature review below captures the historic development towards microsimulation while tracing two approaches: the planner, starting traditionally with macroscopic models and moving on to activity-based, mesoscopic, and finally to microscopic models; and the technology developer, starting with microscopic models from the beginning.

"Macroscopic models" are still in use by transport planners and have also relevance for microscopic models, as explained below. A main characteristic of macroscopic models is the aggregated traffic flow between a zone of origin and a zone of destination. These zone-to-zone flows, which are typically represented by an OD-matrix, are used for the traffic assignment. Different traffic assignment methods have been developed, see [2] for a comprehensive overview. The simplest assumption is that all users follow the shortest route. A more realistic traffic assignment, formulated by Wardrop [3], is called the user equilibrium (UE), where the link flows are determined in such a way that no user can reduce its travel time by changing his/her route. Thanks to efficient algorithms (for example Dijkstra's shortest path assignment or the Frank and Wolfe algorithm for the UE assignment [4]), traffic assignment problems can be solved almost instantly with today's computers, even for large urban areas. Moreover, the traffic assignments are used in a loop to iteratively calibrate or relax trip generation, trip distribution and mode choice models—these are the models which allow the transport planner to predict user behavior and traffic flows due to changes in transport infrastructure or transport services. For a comprehensive collection of conventional demand and supply models see [5]. However, the above mentioned emerging "intelligent" transport technologies are generally difficult to cast in conventional framework of macroscopic models. Nevertheless, there are valid attempts to integrate microscopic effects of new services in aggregate, macroscopic model

using certain idealizing or extreme assumptions: for example, in [6] a multi-modal traffic assignment is modeled; in [7] the link flows of autonomous vehicles (AVs) are modeled by increasing the link capacities; in [8] the empty and occupied vehicle flows of SAVs are determined under system optimum flow constraints by solving a linear programming problem and in [9] the stability of the UE with AVs is examined by means of Lijapunov functions.

The introduction of "activity-based models" has been a major step towards modeling the decision-making of individuals: each individual pursues a specific sequence of activities throughout the day and makes mobility plans to travel from one activity to the next in the best possible way [10]. The mobility plans of an entire population can be executed by simulating each individual on a transport network. The "mesoscopic simulation" is the preferred simulation method for activity-based demand models. Mesoscopic simulation means that the traffic flow is implemented as a dynamic queue simulation, where each road-link is represented as a FIFO (first-in first-out) queue with three restrictions [11,12]: (1) each agent (vehicle or person) has to remain for a certain time on the link, corresponding to the free flow speed travel time; (2) the outflow rate of a link is constrained by its flow capacity; and (3) a link storage capacity is defined, which limits the number of agents on the link; if it is filled up, no more agents can enter the link and spillback may occur.

Such a simulation-model produces time varying link flows and permits to track a person from one activity location to the next. The mesoscopic method allows modelling more details with respect to the macroscopic model, by enabling the determination of individual trip times and waiting times. Mesoscopic simulations are slower compared with macroscopic assignments but still fast enough to simulate large urban areas [13–16]. Mesoscopic models are also used to determine a dynamic user equilibrium (DUE) by running simulations iteratively while updating link travel times [10].

In activity-based demand modeling frameworks, mesoscopic simulations are employed to iteratively optimize the activity sequencing, plan generation, and to determine the DUE [12]. Flötteröd et al. (2011) applies such algorithms to the city of Zurich, Switzerland [17], and Meister (2010) performs a mesoscopic simulation on whole Switzerland [18], where some link-flows are validated with real counts. Numerous publications use mesoscopic simulations for assessing the impact of AVs on a city scale. For example, Zhao (2012) [13] simulated Buffalo and Niagara Region, while Hsueh et al. (2021) simulated the whole San Francisco Bay Area, California (about 18.000 km^2), Childress (2015) examined AVs in the Seattle region using the SoundCast software [19]; the user preferences with respect to AVs have been studied by simulating the entire Paris region [20], for a recent review see [21].

A "microsimulation" reproduces the acceleration, speed and position of each vehicle and person at a fixed sampling rate by solving the difference equations of underlying physical processes. Dynamic vehicle models do typically include human driver behavior. It is also possible to implement vehicle control algorithms of any kind, for example, to correctly model the headways of AVs [22]. Moreover, communication channels can be integrated as well in order to simulate V2V and connected autonomous vehicles (CAVs) [23–26]. It is worth noting that link capacity limits are not explicitly imposed but are a consequence of the vehicle-headways resulting from the difference equations. In addition, infrastructure characteristics like the number of lanes or traffic light cycles are details that directly impact achievable vehicle flows. This closeness to the physical world has made microsimulations the natural choice for technology developers. Line capacities of AVs and CAVs are estimated in [27], while safety aspects of CAVs are investigated in [28], see [29] for an overview of different microsimulation approaches. Such analyses are typically made with small networks and an artificially generated demand.

Execution times of microsimulations are considerably longer compared with mesoscopic or macroscopic models, in particular when using sub-second time steps. Another criticality of microsimulation models is that they require a huge amount of data, while small modeling errors can lead to significant errors of the simulated traffic. This is why microsimulation networks need to be checked carefully, which is a time-consuming task.

These are probably the main reasons why microsimulations are less used as traffic assignment method for activity-based models. Indeed, there are very few validated large-scale microsimulations reported in literature (see Table 1).

The transport demand for microsimulations is usually defined by routes and departure times of all agents participating in a scenario. Most large-scale studies either determine the dynamic user equilibrium iteratively or enable real time routing/re-routing option for a certain share of vehicles. The bulk of large scale microsimulation scenarios is not validated in any way: a simple random trip generation has been used in a simulation of a 1.5 km^2 area of Budapest [30] using the open source simulator SUMO [31]; random trips have also been generated for simulating a 9 km^2 area of Manhattan, Paris, Berlin, Rome, and London [32] with SUMO, but results show unrealistically low average speeds; a more realistic demand generation method is the disaggregation of OD matrices from official surveys; examples are the simulations of North Leeds [33] using the DRACULA software [34] and the simulation with AVs of Halifax [35] using the commercial software VISSIM [36]; a synthetic population with mobility plans have been generated by SUMO's activity generator, based on demographics and land use data, for the city of Monaco [37]. The latter simulation is the only large-scale simulation including "soft modes" such as bicycles and pedestrians, while all other studies are focused on cars and AVs only. An alternative approach attempts to reconstruct the traffic flows of Modena, Italy, by calibrating a flow model based on traffic counter data at specific links [38]; even though this approach does not provide realistic vehicle routes, it is well suited to estimate pollutant emissions.

There are also numerous studies on "wide scale" scenarios, analyzing specific sub-networks of an entire city, for example the main roads of Riga city [39] or the New Jersey Turnpike scenario with tolled highways [40].

Two publications on validated large-scale micro-simulation could be found by the authors, see Table 1. Surprisingly, only few realistic large scale micro-simulations exist to date, despite the importance of emerging technologies such as AVs. Note that the validation methods of those scenarios are not standard methods applied in transport planning.

The present work tries to enrich the literature with a properly validated microsimulation scenario. In order to identify the "scientific contribution", the characteristics of the present scenario is compared with those found in literature, see Table 1: the demand is generated by a suitable fusion of reliable data such as OD matrices and GPS traces, it includes all major transport modes of the city and the traffic flows are validated against traffic counts on a link-by-link basis. A simple mode choice model is also provided. To the knowledge of the authors, no validated large-scale simulation with active modes has ever been published.

Table 1. Comparison of published validated large-scale micro-simulation and present work.

Pub/Year	Simulator/Demand Model	Network	Demand Generation	Modes	Validation Method
[41] 2011	SUMO/DUE	Cologne, Germany from OSM 400 km^2	Activity generator based on 7000 surveys, 700,000 trips in 24 h	Car	Qualitative comparison of flows with observed data
[37] 2017	SUMO + activity-gen/stochastic assignment	Luxembourg, OSM, 156 km^2, 931 km roads	Activity generator based on public data demographics, POIs, etc., 24 h	Car, bus	Comparison of average link speeds from floating car data
This article 2021	SUMO + SUMOPy/DUE, Mode choice	Bologna, Italy, OSM, 12 × 7 km	Activity base, disaggregation of OD matrix, GPS traces, GTFS, peak hours	Car, Bus, motorcycle, bike, pedestrian	Car/motorcycle link flows compared to link traffic counts

Given the difficulties to create large-scale microsimulation models, why would it not be reasonable if planners built realistic demand models using simplified, macro/mesoscopic networks, while technology developers estimated critical parameters, such as the lane capacity, using smaller, microscopic models? Such critical parameters would then be used as constants or cost functions in macro/mesoscopic models, as it is practice to date [7].

For some important cases, there is a strong inter-dependency between microscopic events and macroscopic quantities (flows or densities), suggesting that a separation between local microscopic simulations and large-scale macroscopic models would give unrealistic results.

One example, is the lane capacity increase of AVs with respect to manually driven cars. It turns out that capacity increases are significant only if there is a high share of CAVs circulating [42]. In this case, vehicle platoons can be organized, average headways decrease and capacity increases. Shladover, (2012) [27] who has micro-simulated CAVs on a one-lane, intersection-free highway at steady-state traffic flows, has shown an 80% increase in capacity, assuming all vehicles are CAVs. However, micro-simulating CAVs in an urban environment with random trips results in much lower capacity gains of approximately 16%, due to the network-level effect [30]. Clearly, the dynamics in intersections and the durations of platoons (the time vehicles stay together while traveling on a common route) have a dramatic effect on the capacity [9]. This means route-choice, capacity gains and travel times are interdependent.

Another example concerns the interaction between vehicles and pedestrians on mixed access roads or at pedestrian crossings, where the average travel speed reduces for both pedestrians and vehicles, dependent on the vehicle flows and pedestrian flows. Changes in travel time will in turn alternate demand and consequently flows of vehicles and pedestrians. See [43–45] for pedestrians-bicycles interactions and [46] for gap acceptance of pedestrians crossing a road with platooned CAVs.

These examples suggest that, in general, small, microscopic and large-scale macroscopic models cannot be simulated separately, which means only a large-scale microscopic model will ensure that microscopic dynamics will correctly alternate traffic flows and vice versa, thus network-level effects are taken into account.

However, as realistic large-scale microsimulations are rare (see Table 1), there appears to be a real research gap and a need for such scenarios—the only publicly available scenario of this kind is the LuST scenario [37] on Github [47] which has already been used in many research projects (61 citations in 3 years).

The main challenge for creating microsimulations is the demand modeling. There are recent articles suggesting a new, data driven approach to transport modeling [48–50] or the use of "big data" to improve traditional surveys [51]. It is also worth noting that for evaluating the impact of many future scenarios, there is no need to calibrate complex demand models; there are use-cases where the transport services remain almost unaltered, for example, when electric vehicles substitute gasoline vehicles or when AVs replace manually driven cars or when floating bike sharing schemes replace private bikes.

Given this research gap, the "research question" is whether it is possible to build a traffic scenario that covers an entire urban area while modelling at the same time details on the device level. Therefore, this article has the aim to, at least partially, fill the above-mentioned research gap by providing a validated microsimulation model for the medium size city of Bologna, Italy, including all modes except trains. A further research question is how to calibrate a useful mobility plan choice model, as part of a microsimulation model, while using a limited amount of computational resources or computing time. For this reason, a computationally efficient mobility plan choice is calibrated with the aim (1) to predict user behavior beyond the route choice and (2) to match official modal split data, while improve the consistency between the individual's transport environment and the individual's mode choice.

The purpose of the elaborated traffic scenario is the development of a test platform where town planners and transport system developers can meet to evaluate and optimize

new technologies and services—the scenario is freely available on-line [52]. Even though the scenario building process is specific to the data available for Bologna, it should also serve as a blueprint for creating scenarios for other cities.

In Section 2 the scenario and the modeling processes of transport supply and demand are explained and in Section 3 a simple plan choice model is calibrated. In Section 4 the calibration and simulation results are presented, validated, and discussed and in Section 5 final conclusions are drawn and future research directions are suggested.

2. The Scenario Building Process

Various big data sources led to the construction of a large-scale microsimulation scenario for the metropolitan area of Bologna, Italy, with a population of approximately 1.02 million inhabitants, whereas Bologna city itself counts 308 thousand inhabitants [53]. This section explains how the data has been processed to represent the supply and demand of the transport systems using the SUMOPy/SUMO simulation suite [52,54]. While it is good practice to describe agent based models with the ODD protocol (Overview, Design concept, and Description) defined by the Grimm et al. protocol [55], this protocol is hardly applicable to the present case as the number of parameters and the dimensions of the state space is relatively high. Nevertheless, transparency is guaranteed as the scenario and software are published online.

In particular, Sections 2.1 and 2.2 describe the transport supply of road and public transport, Sections 2.3 and 2.4 explain the data preparation of ODMs and GPS traces, while Sections 2.5 and 2.6 explain the external and internal demand created from ODM and GPS data.

2.1. The Road Network Model

The road network of Bologna city has been converted from OSM in a SUMO XML format by SUMO's "netconvert" [56] program and edited manually with SUMO's "netedit" [57] software, using both satellite images in the background and street-level graphical information from Google maps, as well as some on-site inspections. In addition, connectivity problems have been identified by matching GPS traces to the network: matching errors occurred often at locations where network links are not properly connected, see [58] for details. The road network data contains the directed road network graph made of links and nodes; each link consists of one or several lanes. The most important lane attributes are maximum speed, width, and access rights; all the values are determined by analyzing the OSM attributes of the respective way. Moreover, SUMO assigns a priority level to each link which depends on the link attributes and range from 1 (footpath) up to 13 (national motorway). The connectivity of lanes at intersections is also derived from OSM or guessed from heuristics; all connections have been manually checked, together with road attributes and geometry. Traffic lights are an OSM node attribute, but the signals have been generated by heuristics. Large traffic light systems in and around the center have been edited manually based on traffic light plans provided by the city of Bologna.

The road-network of the city of Bologna with surrounding towns is the core simulation area, covering approximately 50 km^2. The core area has a detailed street network, including bikeways and footpath, see Figure 1a. The metropolitan area of Bologna covers a wider area of 3703 km^2, see Figure 1b. Figure 1 also shows the traffic assignment zones (TAZs) of the core area and the metropolitan area. The TAZs are derived from the 2001 national population census [59]. There is a substantial traffic between the core simulation area and the extra-urban TAZs. For this reason, the city's road network has been manually expanded in order to capture the external demand: using again SUMO's network editor and satellite images, a simplified road network has been created linking all major towns and villages with the core network of Bologna; this network consists predominantly of motorways, major federal roads, and provincial roads.

(**a**)　　　　　　　　　　　　　　　　　　　　(**b**)
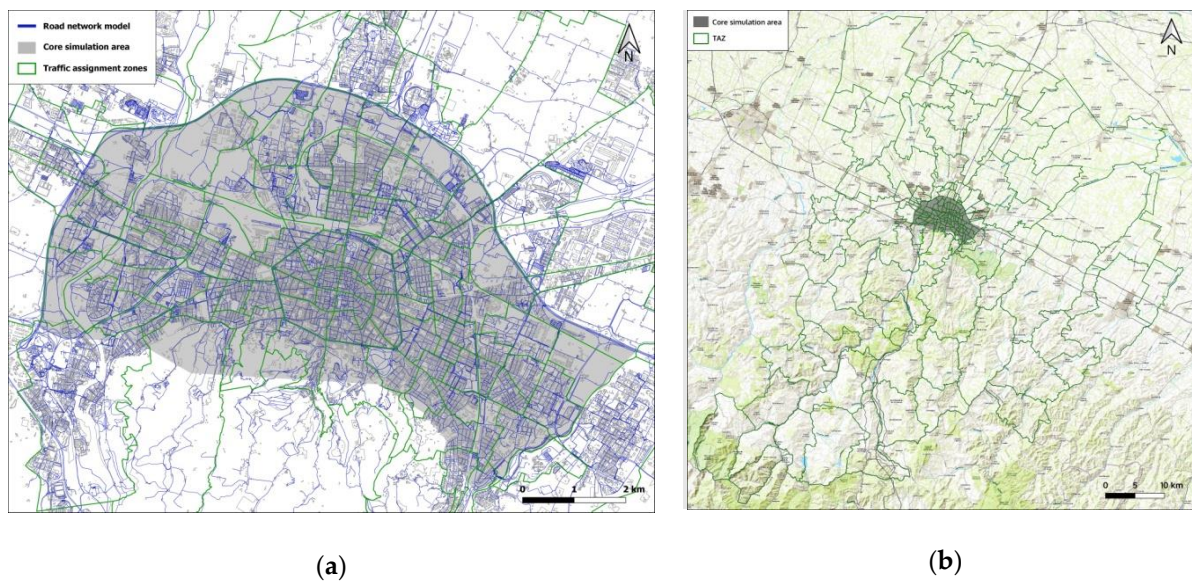
Figure 1. Maps of simulated area and traffic assignment zones (TAZs) in green: (**a**) Core simulation area which is the city of Bologna with some bordering towns ~50 km^2; (**b**) Entire simulation area including the extra-urban areas of the Bologna ~3700 km^2 metropolitan area.

The total number of road links is 32,409 with a total length of 3316.20 km. The share of major road (with priority level greater than 7) is 20.11% of the total length or 667.05 km. Moreover, there are 59,218 link connections within 14,724 intersections, 530 of which are controlled by a traffic light. The geometric shapes, heights, and type of 58,421 buildings in the core simulation area have also been imported from OSM. Buildings will be associated with activity locations of persons in the synthetic population model, see Section 2.6. In addition, on-street parking lots have been created with some heuristics along roads with at least two lanes and road priority below eight.

2.2. Public Transport Services

The entire public transport (PT) provided by the local operator (Tper) has been realistically modelled within the core simulation area by generating bus lines based on data from GTFS (General Transit Feed Specification). The used GTFS represents the timetable valid for spring 2018 and contains geographic information of bus stops and bus routes as well as precise times for bus runs. Bus stops with ID and name have been positioned on the network links. Bus routes have been identified as a sequence of network links using the mapmatching procedure from SUMOPy, as described in [58]. Bus stops play an important role in the microsimulation as they represent the point where people of the synthetic population access public transport services. Successively, bus runs of all urban bus lines have been imported from the GTFS for a workday in May 2018 during the time from 6:00 to 9:00 a.m. for the purpose of realizing a steady state bus service for the analyzed simulation time (from 7:00 to 8:00 a.m.). For all PT lines, a constant service frequency has been determined by averaging the time delays between all runs in the considered time interval. One-off or infrequent bus lines with service times below 30 min have been excluded. The constant service time is needed to generate the service in the microsimulation but also to estimate the waiting time during the plan generation, see Section 2.6. After this import procedure the ID, name, stop sequence, route, and service frequency of 234 bus lines are present in the scenario.

2.3. Transport Demand from OD Matrices

The disaggregation of OD matrices presents a major method to generate trips and routes for different modes of transport, see Sections 2.5 and 2.6. The raw OD matrix has

been available for the time interval 7:00–8:00 a.m. and for the following transport modes: car drivers, car passengers, public transport, and scooters. The corresponding TAZs are more refined in the core simulation area (116 TAZs) and larger in the extra-urban areas (61 TAZs), see Figure 1.

The raw OD matrices for the different modes have been obtained from the 14th population census, conducted by the Italian institute for statistics (ISTAT) during the year 2001 [59]. The OD matrices have been updated to the year 2018 by considering the population increase in the various zones: the OD flows within the core simulation area have been increased by 5.5%, while the flows from or to extra-urban areas have been increased by 8.5%.

Applying the above procedure, the following five matrices have been created for the scenario: one OD matrix for each of the modes car, scooter, bus, and walking, with demand flows only between TAZs inside the core simulation area: these OD matrices have been successively disaggregated to create the synthetic population, see Section 2.6; one OD matrix for cars with origins or destinations in the extra urban TAZs were used to create the external traffic of the scenario, see Section 2.5.

2.4. Transport Demand from GPS Traces

Bicycle demand has been estimated from GPS traces recorded by citizens on a volunteer bases using Smartphone. Each GPS trace describes the movements of each participating cyclist through a sequence of time-stamped and georeferenced Lat/Lon locations. For the present study, the GPS traces recorded during the European Cycling Challenge campaign in Bologna in May 2016 have been used. Only traces during morning rush hours have been relevant, more precisely between 8:30 and 10:30 a.m. The GPS traces underwent a filtering process where inconsistent traces have been eliminated, such as traces with over speed, too long waiting times or too big spatial gaps. Further, the typical point clouds at the beginning and at the end of cyclist traces have been cut off. Successively, a mapmatching process has been applied to identify for each GPS trace the sequence of road network links, resulting in one or several routes per participant.

The estimation of transport demand from GPS traces recorded by volunteers has the obvious problem that the share of the recording population is generally unknown. For this reason, the number of GPS trips need to be scaled to the effective number of trips. In a previous publication [60] the scaling has been performed by means of bicycle flow counts at dedicated links of the road network. In particular, the scale factor has been estimated as the ratio between the observed bicycle flows and the bicycle flows generated by the mapmatched GPS traces. In order to match the scaled number of trips, the mapmatched routes needed to be replicated by a certain number. For replicating a matched GPS trip, the first and last link of the replicated trip has been located randomly around the mapmatched trip extremities, while the mapmatched route has been entirely kept. The departure times of the trips are defined by the first timestamp of the GPS traces.

The above procedure has led to a model of all cyclist trips during morning rush hour, including routes and departure times.

2.5. Construction of External Demand

The external demand comprises all car trips between the core simulation area and the extra-urban areas as well as car trips between extra-urban areas which probably pass through the core simulation area. All other modes were neglected, as car has been the dominant mode for these typically long-distance trips. Further, low-frequency extra-urban bus services have been judged to have only a minor impact on the overall traffic flows.

The external trips for cars have been generated by disaggregating the relative OD matrix with origins or destinations in the extra urban TAZs: the demand flow f_{od} from a zone of origin o to a zone of destination d has been used to generate f_{od} trips, between those zones; the first and last link of the f_{od} trips have been distributed proportionally to their link length, in zone o and d, respectively. This procedure assumes that the number of

residences or workplaces along a link is proportional to the road length. Inaccessible links for cars or links with maximum speeds above 50 km/h have been excluded. Road links in traffic limited zones (TLZ), mainly located in the historic center, are not accessible for ordinary passenger cars, but are allowed for taxis, buses, scooters, and bicycles. In order to allow cars with origin or destination on a TLZ link, the passenger type "car" has been converted into a "taxi" for specific vehicles. In this way ordinary cars without origin or destination in the TLZ cannot drive through the historic center, while it remains accessible for workers and residents with origin or destination in the TLZ, just as in reality.

The disaggregation of the car ODM has produced a total of 71,680 external trips. For each trip, an initial route is generated by connecting the first and last link of each trip with the shortest time route, where the estimated link travel times assume free flow conditions. The departure times of the vehicles have been uniformly distributed within the interval 7:00 to 8:00 a.m.

Furthermore, mapmatched and scaled bicycle GPS trips (see Section 2.4) which goes through the near suburb have been kept, even if partially out of the core area. A total of 616 bike trips have been identified, where either the first or the last link lays within an external zone. Note that vehicles performing external trips do not carry people of the synthetic population. They are merely used to generate a background traffic in the core simulation area which adds up with the traffic from the synthetic population.

2.6. Construction of the Activity Based Synthetic Population

A synthetic population has been built for people living in the core simulation area, based on the previously described demand elements. A basic assumption is that the external demand is independent from the travel behavior of the synthetic population, except for the route choice.

Essentially the synthetic population consists of a database of people, each person with its own attributes (e.g., home/work location, activity pattern, vehicle ownerships, preferred mode, and socioeconomic attributes) and a set of feasible mobility plans. A plan describes a door-to-door trip between successive activities and consists of a series of stages, where each stage represents a movement with a single mode of transport [61]. The estimated or effective execution time of plans allows people to choose their optimal mobility solution for their specific activities, including travel modes and routes.

This section describes the generation of the synthetic population with a primary plan, which is the plan that uses their preferred mode. The preferred mode of each person depends on the data source. The generation of alternative plans for each person together with a plan choice model are treated in Section 3. Due to the available data, the presented construction focuses on the activity pair home-work during the morning peak hour.

The share of the population who uses the modes car, scooter, bus and walking is generated by disaggregating the respective ODMs in the following way: the number of people living in a certain zone corresponds to the sum of trips leaving the zone with all the aforementioned transport modes. The home activity location of individual persons has been associated with buildings, such that the probability to depart from a building in the zone of origin is proportional to its surface. The same reasoning has been applied to identify the building associated with work location inside the destination zone. The building surfaces have been determined from the imported shapes. The generation of pedestrians has received a special treatment: their generation between a particular OD pair took only place if the distance between the center of the respective pair of TAZ was less than 1.5 km. This somehow arbitrary threshold is insensitive as it simply avoids unrealistically long walks. The departure times of all persons created with ODMs have been uniformly distributed within the interval 7:00 to 8:00 a.m. The preferred mode of each person is set by the mode of the ODM that has been used to generate the person. Each person received the vehicle required to travel with his/her preferred mode, e.g., all car drivers received a car, and all scooter drivers received a scooter.

The cyclist population has been generated from the processed GPS traces (see Section 2.4) where the first and last links are within the core simulation area. For each of these trips the home activity building and the work activity building have been picked randomly within a radius of 50 m around the first and last trip links, respectively. Obviously, all cyclists do own a bicycle.

At this point, the entire population has been created for the core simulation area, which performs trips during rush hour. The synthetic population statistics with absolute numbers and shares of the preferred mode are shown in Table 2. Note that despite the different data sources, the mode share of the population is similar to the official statistics obtained from the Sustainable Mobility Plan (PUMS) of Bologna [62].

Table 2. Statistics of the synthetic population with share of preferred modes (M_s) and observed mode share O_s provided by the Sustainable Mobility Plan (PUMS) of Bologna [62]. The last two columns refer to the added number of plans and the total number of plans generated per mode for the mode choice model, see Section 3.

Transport Strategy	N° of People per Strategy Assigned with Preferred Mode	Share of People Assigned with Preferred Mode M_s	Observed Mode Share O_s	Additional Feasible Plans	Total Feasible Plans
Car	17,337	30.50%	30.70%	13,923	31,260
Bicycle	2424	4.26%	7.20%	20,310	22,734
Bus	17,557	30.89%	27.70%	39,280	56,837
Scooter	6199	10.91%	11.40%	5168	11,367
Walking	13,320	23.44%	23.00%	12,467	25,787
Total	56,837	100.00%	100.00%	91,148	147,985

Successively, a primary plan for the home-work activity pair has been created for each person, based on the previously acquired person attributes and the preferred mode. A plan with the mode "car" consists of the following stages: home activity-walk to car parking-drive to car parking-walk to work location-work activity. A general network location is defined in terms of link and position on link. The two parking lots have been chosen to minimize the distance to the home and work location, respectively. A plan with the modes "scooter" or "bicycle" does not require a parking, hence the stages have the shape: home activity-drive to work location-work activity. The initial vehicle routing between two network links equals the shortest time route. There is one exception: the routes of bicycles are already determined by the mapmatched GPS traces.

Similarly, the plan for walking includes a simple walk stage between activity locations. The plan for "bus" mode includes a walk to and from the bus stop, a bus ride, and intermediate walks, depending on the number of transfers. In general, SUMOPy allows creating plans for any mobility strategy, which can also include several modes, such as "bike + bus".

As the initial shortest time routing is not realistic in a congested city, the deterministic dynamic user equilibrium (DUE) has been determined for all modes except bikes and buses, which have their fixed routes. The determination of the DUE involves the simulation of the entire scenario, including all persons and vehicles from the synthetic population, all trips from the external demand as well as the urban bus lines. It has been found that the latter have a significant influence on the traffic flows of other modes. The DUE has been calculated using SUMO's "duaiterate" assignment tool [63] with default parameters and choosing the c-logit stochastic traffic assignment as assignment method during each iteration. After 20 simulation iterations, link travel times have converged and traffic congestion, which occurred with the initial shortest time routing, have been significantly reduced. After the DUE assignment, link travel times and plan execution times have become more realistic. Finally, the entire synthetic population has been created, including plans for the preferred mode with realistic plan execution times.

3. Calibration of a Simple Plan Choice Model

The proposed plan choice model attempts to predict the used transport mode of individuals, such that the modal split of the simulation corresponds to the observed modal split. The developed calibration method is specifically suited for microsimulations, as it avoids simulation runs in every iteration step. For this purpose, for each person of the population, all feasible plans (or likewise all feasible modes) are generated. In the present context, a mode is feasible if the person possesses the required vehicle—walking and bus is feasible for all. For this reason, it is of fundamental importance that vehicle ownerships correctly reflect statistical data reported in [64], as stated by Grimm et al. [65]: in Bologna 53% are car owners, 20% are scooter owners, and 40% are bicycle owners. In order to fit this statistic, the appropriate vehicles have been randomly assigned to people, in addition to the vehicle corresponding to their preferred mode.

The model consists of utility functions, where each function is associated to a mobility plan. The utility function is composed of a travel time proportional component, the value of time (VoT), and a mode specific parameter. Indeed, the travel time is the most important factor when choosing an urban transport mode. The model calibration phase uses an evolutionary minimization algorithm and requires the generation of all feasible plans for each person and the computation of the respective plan execution times.

The application of the model calibration succeeds in two steps: in a first step, the travel times for all feasible mobility strategies for all persons are determined. An iterative algorithm has been developed that selects one of the feasible plans of each person during each iteration and runs the simulation with the selected plans, as depicted in the flow chart of Figure 2a; the iterations of plan (re)-selection and simulating are continued until all plans of all people have been simulated at least once.

(a)

(b)
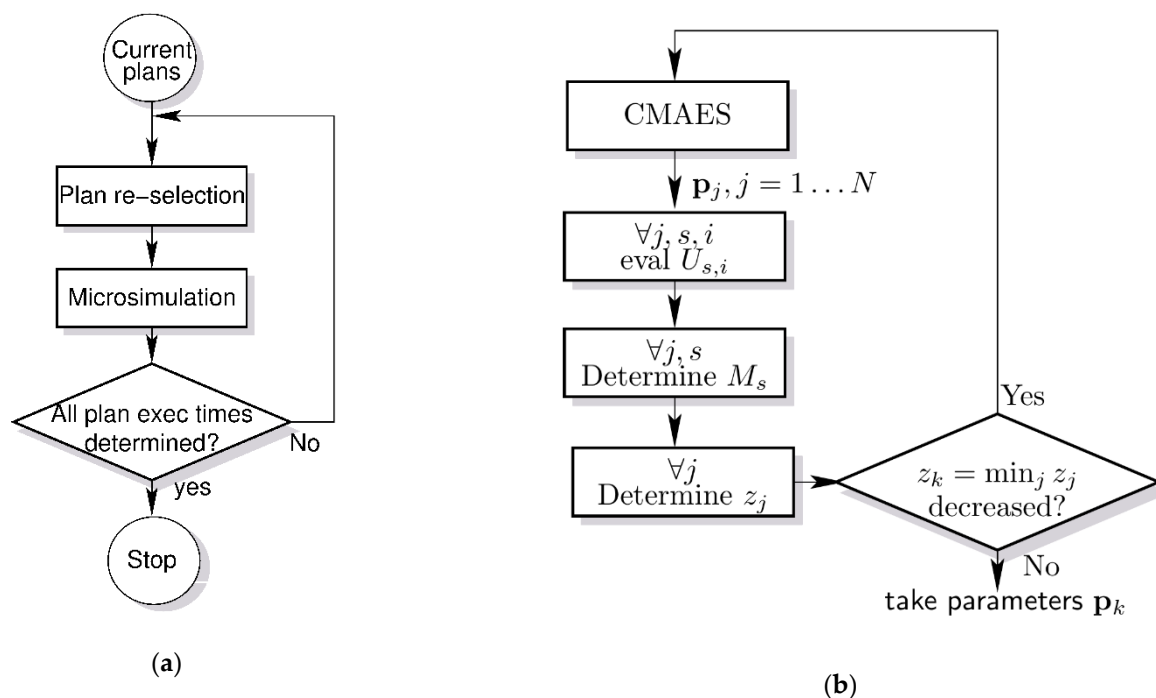
Figure 2. Application of the plan choice calibration model: (**a**) Scheme of algorithm that calculates the plan execution time of all plans while maintaining the official modal split. (**b**) Iteration of CMASE minimization algorithm; determination of objective function z_j and choice of final parameter vector p_k.

Concerning the plan (re)-selection, a fundamental constraint is that the initial modal split of the simulation is preserved, meaning that the number of plans for each mode does not change with the iterations. This is necessary since the different plan execution times must be determined under the same traffic conditions, otherwise, some plan alternatives

would have advantages/penalties due to different traffic situations in successive simulations. For this reason, at each iteration, the algorithm swaps the selection of feasible mobility plans between all those people having the same pairs of strategies, giving priority to those plans not yet simulated—thus, allowing the modal split to remain unchanged in each iteration.

In a second step, the model is actually calibrated, e.g., model parameters are determined as to maximize an objective function, see Figure 2b. Typical utility based mode choice models consider, in addition to travel time, numerous other attributes such as trip related costs (e.g., fuel, and ticket), fixed costs and also non-quantifiable attributes (e.g., convenience, privacy, etc.) [5]. However, in contrast with conventional mode choice models based on surveys, it is the mode share produced by the model that is calibrated to match the observed mode share. This means that there are only five values (corresponding to the five strategies) available to compare with, which is limiting also the number of coefficients that can be calibrated. For this reason, the utility function of plan *s* represents the monetary value of a plan choice and has the form:

$$U_{s,i} = \alpha_s - \beta T_{s,i} \tag{1}$$

where $U_{s,i}$ is the utility function of strategy *s* for person *i*, $T_{s,i}$ is the plan execution time of strategy *s* of person *i* and β represents a universal value of time (*VoT*), valid for all people and strategies. The coefficient α_s is a mode specific parameter that accounts for all unobserved attributes. In the present model α_s is expressed in monetary terms and can be understood as a price to be paid (if negative) or a reward given (if positive) when choosing the respective strategy *s* and assuming the travel time is the only decision criteria otherwise. The car is the reference strategy (*s* = 1), where α_1 is set to zero. Once all utilities of all plans are known, each person *i* chooses the plan of strategy *s* if $U_{s,i}$ is the maximum utility of all feasible strategies for this person. Let M_s be the mode share of people choosing strategy *s* and let O_s be the observed mode share of strategy *s* from official statistics (see third column of Table 2), then the calibration algorithm needs to adjust all the parameters, $\alpha_2 \ldots \alpha_5$ such that the geometric differences between the model mode shares and the observed mode shares are minimized.

$$z = \sum_{s=1}^{5} |M_s - O_s| \tag{2}$$

This is not a simple minimization problem as the resulting objective function is not smooth and gradient decent algorithms could fail. Instead, a stochastic minimization algorithm (CMAES) has been applied. In brief, the iterative algorithm works as follows, for details see [66]: in each iteration a set of $j = 1, \ldots, N$ parameter vectors are drawn from a finite parameter space by the CMAES algorithm. For each parameter vector $\boldsymbol{p}_j = [\alpha_2 \ldots \alpha_5]$ the objective function z_j is determined by evaluating the utilities $U_{s,i}$ for each person *i* and plan *s*, the plan choice for each person, and the mode choice M_s; the CMAES algorithm selects a set of new parameter vectors for the successive iteration, dependent on which parameter vectors $\boldsymbol{p}_j$ have produced the lowest objective function z_j. The algorithm stops if the lowest of all objective functions z_j during an iteration can no longer be decreased significantly with respect to the previous iteration, see Figure 2b.

4. Results and Discussions

4.1. Mode Share Model Calibration Results

Once the execution times of all feasible plans of all people have been determined, the actual calibration process can start, as described in the previous section. For the present study, the value of time is assumed to be $\beta = 0.07 \, €/min$ [5] and the parameter space for all four parameters has been limited to the interval $(-5, 5)$. Figure 3 shows that a good convergence has been achieved after 4000 iterations, highlighting that the objective function tends to zero, which means the observed modal split has been reproduced by the simulation. As the plan execution times $T_{s,i}$ have been predetermined, no microsimulation

run is required during the calibration phase, which means that results can be obtained in a reasonable time (approximately 120 min on a i7 processor computer).

(**a**)

(**b**)
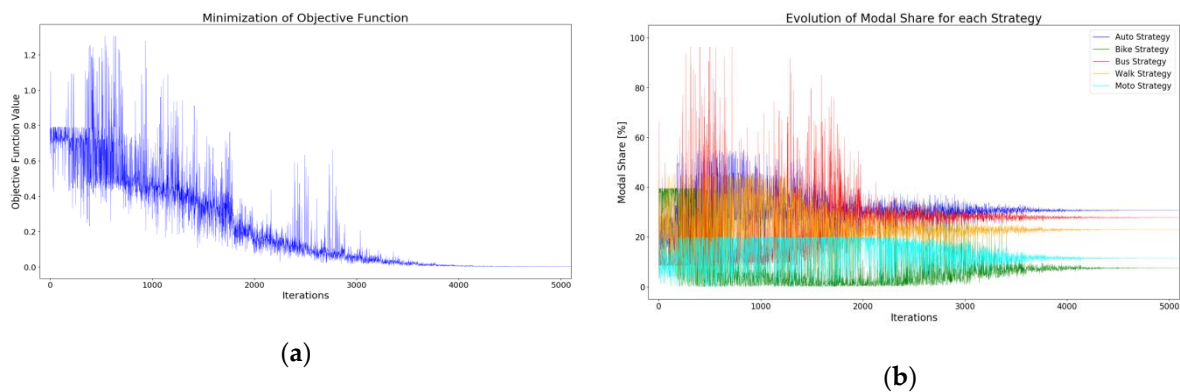
Figure 3. Convergence of calibration process: (**a**) Convergence of objective function *z* over iterations. (**b**) Calibrated mode share M_s over iterations, converging to the official mode share, as shown in Table 3.

Table 3. Calibrated parameters of utility function (Equation (1)) of mode share model by minimizing objective function z

α_1 Car (ref.)	α_2 Bike in €	α_3 Busin €	α_4 Walking	α_5 in € Scooter	β in €/min
0.0000	−0.5604	0.3727	−0.0556	−0.0161	0.0700

There is an observation regarding data consistency: after the calibration, there are people for whom the plan utility corresponding to the preferred mode is no longer the highest, which means that a plan different from the originally assigned mode is selected for the final simulation. Of course, in order to preserve the predefined modal split, other persons may choose these preferred modes, just because the respective plan shows the highest utility of their feasible plans. This means that the mode choice of individuals becomes more adapted to the person's particular environment. For example, if the person's home and work activities were located near a bus line and a "car" has been assigned as preferred mode, then the bus strategy may receive the highest utility and the person would change mode from "car" to "bus"; for another person, the contrary may happen when the "bus" is the preferred mode, and the home or work activity are far away from any bus stop. Therefore, we can state that the calibration process does increase the consistency of the mode choice behavior of the synthetic population, even if the modal split was already similar to the reality.

Looking at the numerical values of the alphas, it appears that bicycle riders are penalized, and bus users are incentivized, meaning that if only the door-to-door trip time was important, more people would choose the bike and less people would use the bus.

It is interesting to note how, by replacing the car as reference strategy by another strategy, the proportionality between the parameters related to the different modes of transport remains completely unchanged, thus highlighting that the crucial aspect for the calibrated mode choice model does not concern the absolute value of each parameter, but primarily the relative difference between the various pairs of constants.

It is further worth mentioning that by changing the value of time β, the calibrated α parameters will change too, but the changes are always proportional to the change of β. This means a variation of β does only scale the utility function, which does not affect the strategy choices made upon it.

4.2. Microsimulation Results and Model Validation

After each person has chosen the plan with the highest utility (with utility from Equation (1) and the calibrated parameters from Table 3), a final microsimulation run has

been launched for the morning rush hour between 7:00 and 8:00 a.m. and travel times and link flows have been recorded on the entire network. For evaluation purposes, only the traffic data of the second half hour has been recorded, in order to avoid unrealistic flows due to transition effects while the network fills with vehicles and people. The link-flows in the core simulation area are visualized on Figure 4.
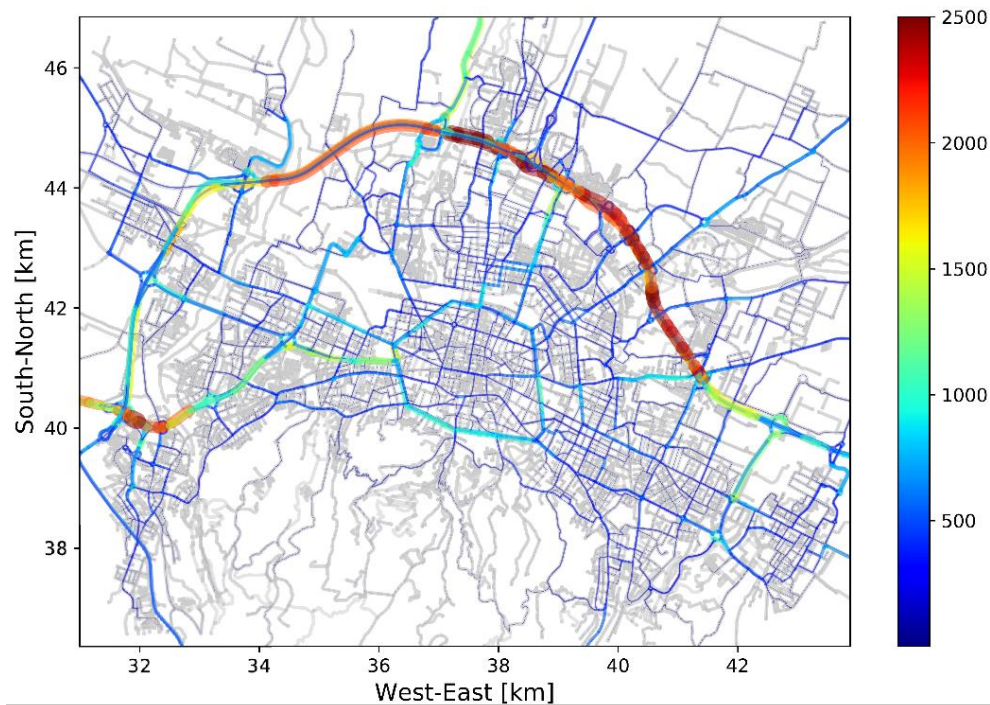
Figure 4. Measured simulated flows in the core simulation area as number of vehicles entered into a link in 30 min.

The maximum flows of 2500 vehicles/h per lane on Figure 4 can be observed on the outer ring road, the "Tangenziale", where also real traffic flows are indeed close to capacity limits in the morning rush hour. Flows on the inner ring of 1000 to 1500 vehicles/h are also realistic. In order to validate the simulation, the simulated flows shown in Figure 4 are compared to the flows measured by induction loop based detectors, scattered around the city, as shown in Figure 5. The 459 detectors counted average hourly flows on a work day in February 2014 during 7:00 to 8:00 a.m.

Note that the flows measured by the detectors include cars, buses, and trucks, while two-wheeled vehicles (bicycles and motorcycles) are not detected. On the other hand, the link-flows determined during the simulation consider all vehicle categories, consistent with the generated demand, including cars, scooters, bicycles, and public transport buses. Another systematic error source is the fact that the counts have been recorded in year 2014 while the demand has been calibrated for year 2018. Other difficulties are related to the association of detectors with road links and the malfunctioning of some of the detectors: these detectors have been eliminated from the evaluation.
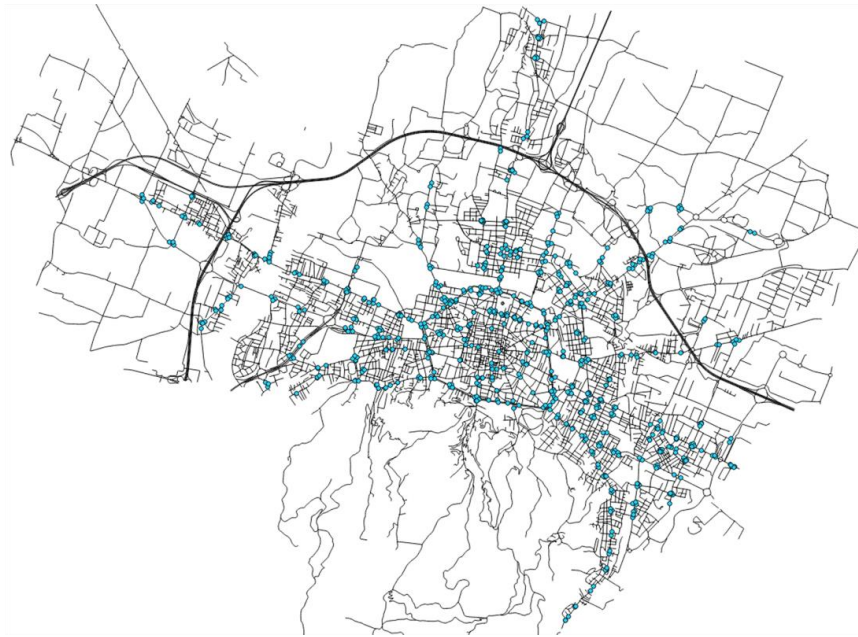
Figure 5. Location of flow detectors (cyan circles).

The plot of link-flows from the microsimulation run over the link-flows from the detectors is shown in Figure 6, where the simulated flows are doubled so that both, simulated and detected flows, are expressed in vehicles per hour. Ideally the flows should be validated not only at different points on the network, but also at different time instances [67]. However, the observed flows at disposal were only available as an hourly average, which did not permit a more refined analysis during the simulated rush hour. Three indicators, typically used in transport science, have been considered to validate the simulated link flows:
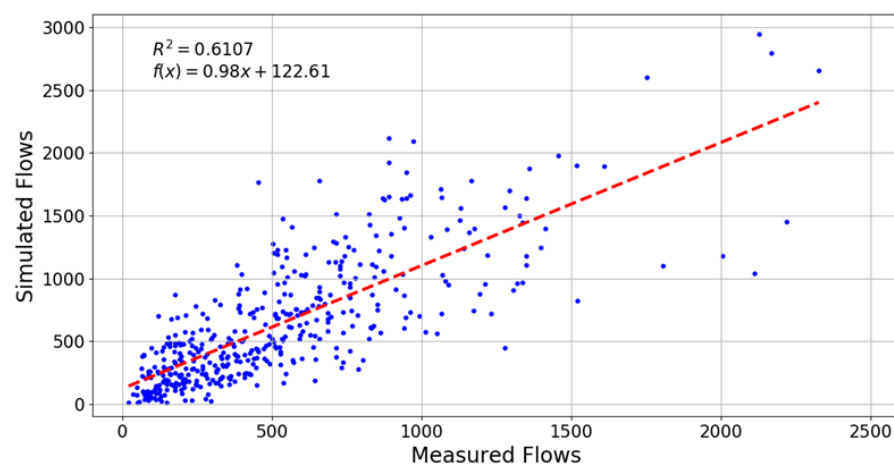
Figure 6. Simulated flows over measured (observed) flows in vehicles per hour.

A first indicator verifies whether there is in average a good correspondence between simulated and observed flows. For this purpose, a linear regression line is calibrated, as shown in Figure 6. The regression line has a slope of $m = 0.98$ and an intercept of 123 vehicles per hour. Both parameters are significant as p-values are 1.47×10^{-91} for the slope and 8.53×10^{-7} for the intercept. According to literature [5], this slope m is acceptable because it is within the range of $0.9 < m < 1.1$.

The second indicator is the determination coefficient R^2 which verifies the correlation between simulated and observed flows. The resulting coefficient of $R^2 = 0.6107$ is below the

suggested acceptance level [5] of $R^2 > 0.8$. The relatively low R^2 can be partially explained with the above mentioned error sources due to the available data.

A third measure is the GEH statistic, which is a modified chi squared statistic that incorporates both, relative and absolute differences, in comparison of simulated and observed flows. It is a well-established indicator that has been used to validate other microsimulation scenarios [68]. In brief, the GEH of link i is determined by

$$GEH_i = \sqrt{\frac{(f_i - f_i^*)^2}{0.5(f_i - f_i^*)}}. \tag{3}$$

Links with GEH values below 5 represent a good fit, links with values in the range $5 < GEH < 10$ are considered questionable and links with $GEH > 10$ are not a good fit. The evaluation of the present study reveals that 31% of all observed links are in the range GEH < 5; 28% are in the range $5 < GEH < 10$ and 41% are in the range GEH > 10. The relatively high percentage of links which do not show a good fit does again reflect the low correlation coefficient.

For further evaluations, the above indicators have been calculated for different road types. As criteria to differentiate road types, the road width and number of lanes have been used. The indicators shown in Table 4 clearly suggest that the flows on larger roads are modelled with a higher precision with respect to smaller roads. Roads larger than seven meters achieved the maximum R^2 of 0.7. The share of well-fitting links with GEH < 5 does also increase with road size, except for roads with more than three lanes. Apparently, larger roads are often fast and straight connections, without efficient alternatives; whereas smaller roads are likely to have many alternatives in an urban network. For this reason, the traffic assignment has a higher chance of picking the correct large road than choosing the correct small road. Other classification with more subjective criteria (for example road priority) did not show coherent evaluation results.

Table 4. Flow comparison data and indicators for different road widths and number of lanes.

Road Link Type	# Links	Slope m	Intercept (veh/h)	R^2	GEH < 5	5 < GEH < 10	10 < GEH
0 m < width < 5 m	128	0.87	157.22	0.34	29%	30%	41%
5 m < width < 7 m	203	0.83	150.94	0.51	33%	27%	40%
width > 7 m	116	1.06	146.27	0.7	32%	26%	42%
1 lane	8	0.53	97.36	0.33	25%	25%	50%
2 lanes	191	0.87	125.32	0.46	32%	31%	37%
3 lanes	154	0.77	228.4	0.43	34%	23%	43%
>3 lanes	86	1.02	203.97	0.62	26%	27%	48%
All links	**439**	**0.98**	**122.61**	**0.61**	**31%**	**28%**	**41%**

5. Conclusions

A large-scale, agent-based microsimulation scenario including the transport modes car, bus, bicycle, scooter, and pedestrian, has been built and validated for the city of Bologna during the morning peak hour. The activity-based model allows simulating and evaluating door-to-door trip times with different mobility strategies.

Transport network, bus services and the transport demand have been extracted from different "big data" sources. Many data processing steps were necessary to homogenize the data and to make it coherent. Microsimulations are sensitive to small modeling errors, particularly with congested networks; for this reason, much attention has been paid to modeling details such as external traffic, parking spaces, traffic light programs, access of roads for different vehicle types, and in particular vehicle access to traffic limited zones in the city center.

A simple mode choice model has been calibrated which successfully reproduces the modal split from official statistics and increases the consistency between modal choice and the transport environment of the individual. The scenario has been validated by comparing simulated traffic flows with observed flows from road-side detectors. The quality of the simulated flows is satisfactory even though different systematic error sources have impeded a higher correlation coefficient: a main source of error is that the different data sources (e.g., network, ODMs, GPS traces, and traffic counts) stem from different years and the updating to the year 2018 contains many assumptions. Further improvements are expected when more recent data become available. In addition, more sophisticated data fusion methods [48–51] have also the potential to reconstruct the synthetic population more precisely. It would be further interesting to make comparative studies with other available microsimulators as there are differences in link capacities [69].

Finally, the built microsimulation scenario represents a test-platform for transport technology developers as the used microsimulator SUMO has already been employed to evaluate a wide range of transport technologies, such as battery electric vehicles, ride-sharing schemes, V2X communication, platooning of automated vehicles, or intelligent traffic light systems. Thanks to a high-level programming interface called TraCi, it is possible to interact with a running simulation using custom-made code. However, even transport planners can make use of the scenario to test how different technologies and new means of transportation interact with transport demand, while taking advantage of the growing availability of big data. The concept of mobility strategies allows adding any kind of new technology or service. SUMO with SUMOPy enable an easily access to microsimulations, edit scenarios and track all simulation events, step by step, through a user-friendly interface, and a rich spectrum of analysis tools.

Even though the present scenario-building is a special case and leaves ample room for improvements, it starts narrowing the gap between different research areas and allows planners, data scientists and technology developers to work together more effectively on the same transport scenario with the common goal to realistically evaluate and improve future sustainable transport systems.

Author Contributions: Conceptualization, Joerg Schweizer, Cristian Poliziani and Federico Rupi; Methodology, Joerg Schweizer, Cristian Poliziani, and Davide Morgano; Software, Joerg Schweizer, Cristian Poliziani, Davide Morgano, and Mattia Magi; Validation, Davide Morgano and Mattia Magi; Data Curation, Davide Morgano, Mattia Magi, and Cristian Poliziani; Writing—Original Draft Preparation, Joerg Schweizer; Writing—Review and Editing, Joerg Schweizer, Cristian Poliziani, Federico Rupi, Davide Morgano, and Mattia Magi; Supervision, Joerg Schweizer; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The simulation software as well as the entire transport scenario described in this article, hence all information needed to replicate the simulation, are published or linked online at https://github.com/schwoz/sumopy. The source data of the GTFS is available online at the TPER web-site https://solweb.tper.it/web/tools/open-data/open-data-detail.aspx?source=&filename=gommagtfsbo. The publication of the source data of the OD matrices, the traffic light plans, the GPS traces of cyclists and the traffic counts is prohibited by contract.

Acknowledgments: We are grateful to SRM (Società Reti e Mobilità, Bologna) for providing the GPS traces related to the European Cycling Challenge campaign and to the Traffic Department of the city of Bologna for providing the detector flows. Special thanks go to the following graduated students for their dedication on the construction of the scenario from both the demand and supply side: Sara Castaldini, Francesco Filippi, Marco Sermasi, Caterina Ciabatti, Ginevra Antignano, Roberto Todisco, Enrico Pio Troiano, Alessandro Nalin.

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

References

1. Open Street Map (OSM). Available online: https://www.openstreetmap.org (accessed on 6 July 2020).
2. Patriksson, M. The Traffic Assignment Problem: Models and Methods. In *Topics in Transportation*; Grafton, J., Ed.; VSP, Linköping Institute of Technology: Gothenburg, Sweden, 1994; ISBN 978-0486787909.
3. Wardrop, J.G. Road paper. Some theoretical aspects of road traffic research. *Proc. Inst. Civ. Eng.* **1952**, *1*, 325–362. [CrossRef]
4. Frank, M.; Wolfe, P. An algorithm for quadratic programming. *Nav. Res. Logist. Q.* **1956**, *3*, 95–110. [CrossRef]
5. Cascetta, E. *Transportation Systems Engineering: Theory and Methods*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 2001; Volume 49.
6. Verbas, I.O.; Mahmassani, H.S.; Hyland, M.F.; Halat, H. Integrated Mode Choice and Dynamic Traveler Assignment in Multimodal Transit Networks: Mathematical Formulation, Solution Procedure, and Large-Scale Application. *Transp. Res. Rec. J. Transp. Res. Board* **2016**, *2564*, 78–88. [CrossRef]
7. Kloostra, B.; Roorda, M.J. Fully autonomous vehicles: Analyzing transportation network performance and operating scenarios in the Greater Toronto Area, Canada. *Transp. Plan. Technol.* **2019**, *42*, 99–112. [CrossRef]
8. Schweizer, J.; Parriani, T.; Traversi, E.; Rupi, F. Optimum Vehicle Flows in a Fully Automated Vehicle Network. In Proceedings of the International Conference on Vehicle Technology and Intelligent Transport Systems; SCITEPRESS—Science and Technology Publication: Rome, Italy, 2016; pp. 195–202. Available online: https://www.scitepress.org/PublicationsDetail.aspx?ID=W0rOwi3OENM=&t=1 (accessed on 5 March 2021).
9. Lee, S.; Heydecker, B.G.; Kim, J.; Park, S. Stability analysis on a dynamical model of route choice in a connected vehicle environment. *Transp. Res. Procedia* **2017**, *23*, 720–737. [CrossRef]
10. Bowman, J.; Ben-Akiva, M. Activity-based disaggregate travel demand model system with activity schedules. *Transp. Res. Part A Policy Pract.* **2001**, *35*, 1–28. [CrossRef]
11. Charypar, D.; Axhausen, K.W.; Nagel, K. Event-Driven Queue-Based Traffic Flow Microsimulation. *Transp. Res. Rec. J. Transp. Res. Board* **2007**, *2003*, 35–40. [CrossRef]
12. Balmer, M.; Axhausen, K.; Nagel, K. Agent-Based Demand-Modeling Framework for Large-Scale Microsimulations. *Transp. Res. Rec. J. Transp. Res. Board* **2006**, *1985*, 125–134. [CrossRef]
13. Zhao, Y.; Sadek, A.W. Large-scale Agent-based Traffic Micro-simulation: Experiences with Model Refinement, Calibration, Validation and Application. *Procedia Comput. Sci.* **2012**, *10*, 815–820. [CrossRef]
14. Hsueh, G.; Czerwinski, D.; Poliziani, C.; Becker, T.; Hughes, A.; Chen, P.; Benn, M. *Using BEAM Software to Simulate the Introduction of On-Demand, Automated, and Electric Shuttles for Last Mile Connectivity in Santa Clara County*; MTI Pub.: San Jose, CA, USA, 2021; p. 343. [CrossRef]
15. Mtoi, E.T.; Moses, R.; Ozguven, E.E. An Alternative Approach to Network Demand Estimation: Implementation and Application in Multi-Agent Transport Simulation (MATSim). *Procedia Comput. Sci.* **2014**, *37*, 382–389. [CrossRef]
16. Pi, X.; Ma, W.; Qian, Z.S. A general formulation for multi-modal dynamic traffic assignment considering multi-class vehicles, public transit and parking. *Transp. Res. Procedia* **2019**, *38*, 914–934. [CrossRef]
17. Flötteröd, G.; Chen, Y.; Nagel, K. Behavioral Calibration and Analysis of a Large-Scale Travel Microsimulation. *Networks Spat. Econ.* **2011**, *12*, 481–502. [CrossRef]
18. Meister, K.; Balmer, M.; Ciari, F.; Horni, A.; Rieser, M.; Waraich, R.A.; Axhausen, K. Large-scale Agent-based Travel Demand Optimization Applied to Switzerland, Including Mode Choice. In Proceedings of the 12th World Conference on Transportation Research, Lisboa, Portugal, 11–15 July 2010.
19. Childress, S.; Nichols, B.; Charlton, B.; Coe, S. Using an Activity-Based Model to Explore the Potential Impacts of Automated Vehicles. *Transp. Res. Rec. J. Transp. Res. Board* **2015**, *2493*, 99–106. [CrossRef]
20. Kamel, J.; Vosooghi, R.; Puchinger, J.; Ksontini, F.; Sirin, G. Exploring the Impact of User Preferences on Shared Autonomous Vehicle Modal Split: A Multi-Agent Simulation Approach. *Transp. Res. Procedia* **2019**, *37*, 115–122. [CrossRef]
21. Do, W.; Rouhani, O.M.; Miranda-Moreno, L. Simulation-Based Connected and Automated Vehicle Models on Highway Sections: A Literature Review. *J. Adv. Transp.* **2019**, *2019*, 1–14. [CrossRef]
22. Ramezani, M.; Machado, J.A.; Skabardonis, A.; Geroliminis, N. Capacity and Delay Analysis of Arterials with Mixed Autonomous and Human-driven Vehicles. In Proceedings of the 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Naples, Italy, 26–28 June 2017; IEEE: New York, NY, USA, 2017; pp. 280–284.
23. Milanes, V.; Shladover, S.E.; Spring, J.; Nowakowski, C.; Kawazoe, H.; Nakamura, M. Cooperative Adaptive Cruise Control in Real Traffic Situations. *IEEE Trans. Intell. Transp. Syst.* **2013**, *15*, 296–305. [CrossRef]
24. Calvert, S.C.; Schakel, W.J.; Van Lint, J.W.C. Will Automated Vehicles Negatively Impact Traffic Flow? *J. Adv. Transp.* **2017**, *2017*, 1–17. [CrossRef]
25. Haas, I.; Friedrich, B. Developing a micro-simulation tool for autonomous connected vehicle platoons used in city logistics. *Transp. Res. Procedia* **2017**, *27*, 1203–1210. [CrossRef]
26. Fernandes, P.; Nunes, U. Platooning of Autonomous Vehicles with Intervehicle Communications in SUMO Traffic Simulator. In Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems, Funchal, Portugal, 19–22 September 2010; IEEE: New York, NY, USA, 2010.
27. Shladover, S.E.; Su, D.; Lu, X.-Y. Impacts of Cooperative Adaptive Cruise Control on Freeway Traffic Flow. *Transp. Res. Rec. J. Transp. Res. Board* **2012**, *2324*, 63–70. [CrossRef]

28. Papadoulis, A.; Quddus, M.; Imprialou, M. Evaluating the safety impact of connected and autonomous vehicles on motorways. *Accid. Anal. Prev.* **2019**, *124*, 12–22. [CrossRef] [PubMed]

29. Liu, C.-J.; Liu, Z.; Chai, Y.-J.; Liu, T.-T. Review of Virtual Traffic Simulation and Its Applications. *J. Adv. Transp.* **2020**, *2020*, 1–9. [CrossRef]

30. Lu, Q.; Tettamanti, T.; Hörcher, D.; Varga, I. The impact of autonomous vehicles on urban traffic network capacity: An experimental analysis by microscopic traffic simulation. *Transp. Lett.* **2019**, *12*, 540–549. [CrossRef]

31. Behrisch, M.; Bieker-Walz, L.; Erdmann, J.; Krajzewicz, D. SUMO—Simulation of Urban MObility: An Overview. In Proceedings of the SIMUL 2011, Barcelona, Spain, 23–28 October 2011.

32. Mavromatis, I.; Tassi, A.; Piechocki, R.J.; Sooriyabandara, M. On Urban Traffic Flow Benefits of Connected and Automated Vehicles. In Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 25–28 May 2020; IEEE: New York, NY, USA, 2020; pp. 1–7.

33. Liu, R.; Van Vliet, D.; Watling, D. Microsimulation models incorporating both demand and supply dynamics. *Transp. Res. Part A Policy Pract.* **2006**, *40*, 125–150. [CrossRef]

34. Liu, R. The DRACULA Dynamic Network Microsimulation Model. In *Human Resource Development and Information Technology*; Springer International Publishing: Berlin/Heidelberg, Germany, 2006; pp. 23–56.

35. Alam, J.; Habib, M.A. Investigation of the Impacts of Shared Autonomous Vehicle Operation in Halifax, Canada Using a Dynamic Traffic Microsimulation Model. *Procedia Comput. Sci.* **2018**, *130*, 496–503. [CrossRef]

36. Fellendorf, M.; Vortisch, P. Microscopic Traffic Flow Simulator VISSIM. *Harvey J. Greenberg* **2010**, *145*, 63–93.

37. Codeca, L.; Frank, R.; Faye, S.; Engel, T. Luxembourg SUMO Traffic (LuST) Scenario: Traffic Demand Evaluation. *IEEE Intell. Transp. Syst. Mag.* **2017**, *9*, 52–63. [CrossRef]

38. Po, L.; Rollo, F.; Bachechi, C.; Corni, A. From Sensors Data to Urban Traffic Flow Analysis. In Proceedings of the 2019 IEEE International Smart Cities Conference (ISC2), Casablanca, Morocco, 14–17 October 2019; IEEE: New York, NY, USA, 2019; pp. 478–485.

39. Savrasovs, M.; Pticina, I.; Zemlyanikin, V. Wide-Scale Transport Network Microscopic Simulation Using Dynamic Assignment Approach. *Comput. Devices Commun.* **2018**, *36*, 241–251.

40. Bartin, B.; Özbay, K.; Gao, J.; Kurkcu, A. Calibration and validation of large-scale traffic simulation networks: A case study. *Procedia Comput. Sci.* **2018**, *130*, 844–849. [CrossRef]

41. Uppoor, S.; Fiore, M. Large-scale Urban Vehicular Mobility for Networking Research. In Proceedings of the 2011 IEEE Vehicular Networking Conference (VNC), Amsterdam, The Netherlands, 14–16 November 2011; IEEE: New York, NY, USA, 2011; pp. 62–69.

42. Martínez-Díaz, M.; Soriguera, F. Autonomous vehicles: Theoretical and practical challenges. *Transp. Res. Procedia* **2018**, *33*, 275–282. [CrossRef]

43. Wierbos, M.J.; Knoop, V.L.; Hänseler, F.S.; Hoogendoorn, S.P. A macroscopic flow model for mixed bicycle–car traffic. *Transp. A Transp. Sci.* **2021**, *17*, 340–355. [CrossRef]

44. Luo, Z.; Liu, Y.; Guo, C. Operational characteristics of mixed traffic flow under bi-directional environment using cellular automaton. *J. Traffic Transp. Eng.* **2014**, *1*, 383–392. [CrossRef]

45. Bernardi, S.; Krizek, K.J.; Rupi, F. Quantifying the role of disturbances and speeds on separated bicycle facilities. *J. Transp. Land Use* **2015**. [CrossRef]

46. Woodman, R.; Lu, K.; Higgins, M.D.; Brewerton, S.; Jennings, P.A.; Birrell, S. Gap acceptance study of pedestrians crossing between platooning autonomous vehicles in a virtual environment. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *67*, 1–14. [CrossRef]

47. Github. *LUST Scenario*. 2019. Available online: https://github.com/lcodeca/LuSTScenario (accessed on 15 November 2020).

48. Roulland, F.; De Souza, C.; Ulloa, L.; Mondragón, A.; Niemaz, M.; Ciriza, V. Towards Data-Driven Simulations in Urban Mobility Analytics. In Proceedings of the 14th ITS Asia Pacific Forum, Nanjing, China, 27–29 April 2015.

49. Wilson, A. The Future of Urban Modelling. *Appl. Spat. Anal. Policy* **2018**, *11*, 647–655. [CrossRef]

50. Anda, C.; Erath, A.; Fourie, P.J. Transport modelling in the age of big data. *Int. J. Urban Sci.* **2017**, *21*, 19–42. [CrossRef]

51. Croce, A.I.; Musolino, G.; Rindone, C.; Vitetta, A. Transport System Models and Big Data: Zoning and Graph Building with Traditional Surveys, FCD and GIS. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 187. [CrossRef]

52. Github. SUMOPy. 2020. Available online: https://github.com/schwoz/sumopy/ (accessed on 15 November 2020).

53. Dati ISTAT. Available online: http://dati.istat.it/ (accessed on 15 November 2020).

54. Contributed/SUMOPy. Available online: https://sumo.dlr.de/docs/Contributed/SUMOPy.html (accessed on 15 November 2020).

55. Grimm, V.; Berger, U.; DeAngelis, D.L.; Polhill, J.G.; Giske, J.; Railsback, S.F. The ODD protocol: A review and first update. *Ecol. Model.* **2010**, *221*, 2760–2768. [CrossRef]

56. SUMO Netconvert. Available online: https://sumo.dlr.de/docs/netconvert.html (accessed on 15 November 2020).

57. SUMO Netedit. Available online: https://sumo.dlr.de/docs/netedit.html (accessed on 15 November 2020).

58. Schweizer, J.; Bernardi, S.; Rupi, F. Map-matching algorithm applied to bicycle global positioning system traces in Bologna. *IET Intell. Transp. Syst.* **2016**, *10*, 244–250. [CrossRef]

59. Censimento Popolazione e Abitazioni. 2001. Available online: https://www.istat.it/it/archivio/3847 (accessed on 29 December 2020).

60. Rupi, F.; Poliziani, C.; Schweizer, J. Data-driven Bicycle Network Analysis Based on Traditional Counting Methods and GPS Traces from Smartphone. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 322. [CrossRef]

61. Schweizer, J.; Rupi, F.; Poliziani, C. Generating Activity Based, Multi-modal Travel Demand for SUMO. In Proceedings of the SUMO User Conference 2018, Berlin, Germany, 14–16 May 2018.

62. Osservatotio PUMS. Available online: https://www.osservatoriopums.it/bologna (accessed on 15 November 2020).

63. SUMO. Demand/Dynamic User Assigment. Available online: https://sumo.dlr.de/docs/Demand/Dynamic_User_Assignment.html (accessed on 15 November 2020).

64. Grimm, V.; Revilla, E.; Berger, U.; Jeltsch, F.; Mooij, W.M.; Railsback, S.F.; DeAngelis, D.L. Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science* **2005**, *310*, 987–991. [CrossRef]

65. I Numeri di Bologna Metropolitana. *Il Parco Veicolare di Bologna al 31.12.2017.* Available online: http://inumeridibolognametropolitana.it/studi-e-ricerche/il-parco-veicolare-di-bologna-al-31122017 (accessed on 29 December 2020).

66. Hansen, N. The CMA Evolution Strategy: A Tutorial. *Inria* **2011**, 1–34. Available online: https://hal.inria.fr/hal-01297037/document (accessed on 29 December 2020).

67. Kang, J.-Y.; Aldstadt, J. Using multiple scale spatio-temporal patterns for validating spatially explicit agent-based models. *Int. J. Geogr. Inf. Sci.* **2018**, *33*, 193–213. [CrossRef] [PubMed]

68. Tawfeek, M.H.; Mohamed, E.; Khaled, E.-A.; Hatem, A.-L. Calibration and Validation of Micro-Simulation Models Using Measurable Variables. In Proceedings of the Canadian Society for Civil Engineering 2018 Annual Conference, Fredericton, NB, Canada, 13–16 June 2018.

69. Maciejewski, M. A comparison of microscopic traffic flow simulation systems for an urban area. *Transp. Probl.* **2010**, *5*, 27–37.