Alma Mater Studiorum · Università di Bologna

Dottorato di Ricerca in
Ingegneria Civile, Chimica, Ambientale e dei Materiali

Ciclo XXXIV

Settore Concorsuale (SC): 08/A3 - INFRASTRUTTURE E SISTEMI DI TRASPORTO, ESTIMO E VALUTAZIONE
Settore Scientifico Disciplinare (SSD): ICAR/05 - TRASPORTI

Modeling and Implementation of Digital Twins for the Analysis of Transportation Systems

—

Supervisore:
FEDERICO RUPI

Presentata da:
CRISTIAN POLIZIANI

Co-supervisore:
JOERG SCHWEIZER

Coordinatore Dottorato:
ALESSANDRO TUGNOLI

Esame Finale Anno 2022

About Cristian Poliziani

Cristian is at his final year of the PhD Course at University of Bologna - in the field of Transportation Engineering. His research activity mainly focuses on facing the current challenges related to the use of advanced approaches for the analysis of a transport system in large-scale urban contexts through a digital twin, in order to more in detail estimate its performances and impacts and those related to different scenarios that may include the introduction of new technologies, services and mean of transport of the future, as well as more specific studies: evacuation strategies for catastrophic events, analysis of the resilience of the road infrastructure following unforeseen events and so on.

After concluding the PhD studies, Cristian has been accepted for a post doctoral position at Lawrence Berkeley National Lab, California, where he can continue his research activity on the construction, analysis and application of transportation digital twins.

Contents

Summary Abstract

Transport engineers, authorities, companies, stakeholders, and all experts involved on transportation planning work every day to improve the trip experience of people, and to reduce the impacts on the collectivity. Traffic congestion, pollution and energy consumption are core problems for the transportation system. On the one hand a congested road cause an exponential increase in the energy consumption and waiting times, while a not-congested road can be more attractive and slowly become congested over time; on the other hand, the environmental capacity of road links is not perceived, thus generating high emission and distribution of pollutants. In fact, the ability of transportation planners is to avoid traffic congestion but offer at the same time pleasant, accessible and sustainable trips. This may include several planning techniques which may involve traffic calming and limitations, lane reservation, changing in the road network and geometry, as well as apply new technologies, services and means of transportation. It is worth noting that trying these features directly on a city context can be very expensive and produce irreparable damages to the transportation system. Moreover, it is important to highlight the importance - and difficulty - of studying large-scale scenarios that can consider the spatial propagation of certain effects, which is not negligible; this clearly requires an enormous modeling effort, especially for estimating the transportation demand, but also to manage the transportation network data; for these reasons most studies focus on small scale scenarios. However, it has been recently introduced the digital twin, a digital reproduction of a city to be used as a test platform for 'what if' scenarios. In fact, a transport digital twin is not just a digital reproduction of the transportation system, but usually contains advanced models that consider reaction of humans to the changes applied to the system, in order to make a comparison between different sce-

narios. Indeed, to answer complex transportation planning questions, it is necessary to use advanced approaches to calibrate the digital twin: either with mesoscopic or microscopic level of detail. This because traditional - or macroscopic - approach considers a stationary scenario, aggregated trips and empirical equation to simulate interaction between vehicles and infrastructure. Testing the new features of a transportation system means unlocking the variable time, but also consider agent-based or vehicle-based trips, and more precisely simulate the interaction between transportation system entities, as well as allowing live decisions. The literature contains some isolated large-scale case studies, some models and some software to perform either a mesoscopic or microscopic analysis of a transportation system. However, there are only few already calibrated and simple microscopic large-scale digital twins, only few open software - which are usually not user friendly and in a beta version - and lack of models that exploit different data structures which can be useful to estimate a disaggregated demand; in general, there is a lot of room for improvement from all sides. The present work aims at enriching the literature on transportation system simulation. In particular, the research activity focuses on calibrating new models and software implementations for time-dependent large-scale digital twins modeling and construction, particularly exploiting different structure of big data. On the one hand, the new models allowed the modeling and construction of a state-of-the-art and licensed microscopic large-scale digital twin of the city of Bologna on the software Simulation of Urban MObility - python interface (*SUMOPy*) to be used as a test platform for future research. On the other hand, they allowed facing a practical application on a large-scale mesoscopic digital twin of the Bay Area, California, with the novel software Behavior, Energy, Autonomy, and Mobility (*BEAM*). In particular, the present research found that both mesoscopic and microscopic approaches have they own pros and cons, and they are necessary to simulate the mobility of the future; therefore, future research should focus on their combination to take advantages of the benefit of their respective advantages, while eclipsing their biggest challenges. From the mesoscopic side, one of the first applications of the Bay Area's digital twin developed at Lawrence Berkeley National Lab (*LBNL*) has been conducted while implementing and testing the efficacy of on-demand fleets for first and last mile service. This study allowed to demonstrate the *BEAM* software capabilities and highlight possible software developements, as well as provide information to directionate future studies and investments on first and last mile services. From the microscopic side, the state-of-the-art large-scale digital twin with an activity based demand composed by a virtual population has been created while developing new models that exploit big data for the microscopic digital twin calibration: Global Positioning System (*GPS*), General Transit Feed Specification (*GTFS*) files, traffic counts, traffic light schemes, Open Street Map (*OSM*) data, Origin/Destination (*OD*) matrices. Even if new models have been calibrated and digital twins have been built and tested, further validation and some more applications of the digital twins are necessary to respectively confirm the model validity and to exploit its potentialities for practical case studies. Also, future research can aim at expanding the digital twin in terms of both size and interval time. However, big data are expensive to use in terms of time and resources, and are not always available; for these reasons future research should also focus on the generalization of the digital twin construction process, with less efforts and using only data available for all the study area - for example using only the population density and the land use information.

Timeline

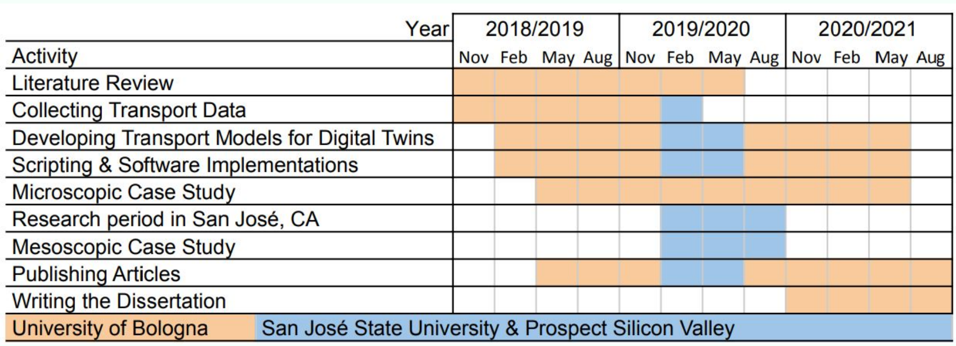


Figure 1: Timeline of the research activities

Chapter 1

Introduction

Travel is a human's primary need. Every single day, almost all people need to make trips for different purposes and with different means of transportation. Rodrigue (2020), shows that in 2017, each households in the united States performed about 3,000 trip. At the same time, transports contributes to most of the pollutant emissions in the planet; in Europe, transport is guilty of 22% of CO (Carbon Monoxide) emission and 63% of NOx (Nitrogen oxide) emissions[1]. For these reasons, the goal of the transportation system theory is being able to predict and simulate the trips of people, as well as estimate trips performances and the generated environmental impacts, in a way to at the same time support the transport planning decision-making problem to satisfy people desires - less travel and waiting time, more connectivity, accessibility, comfort, simplicity, safety and security and so on - and promotes a sustainable development from the economic, environmental and social point of view for the collectivity. A transportation system is historically defined as the conjunction of the transportation demand and supply and their interactions with the system of activities in the territory (see Cascetta (2009)). In fact, a transportation system analysis may include the determination of the current transportation impacts of a study area, testing impacts of either small or radical changes in the transportation supply, estimating consequences due to transportation demand variation - for example the traffic reduction due to the COVID-19 pandemic, adapt the

[1]European Environment Agency (*EEA*): https://www.eea.europa.eu/ (accessed: February 16, 2022)

transportation supply to different demand configurations, experiment different planning options and so on. Some recurring planning optics to improve the travel experience in a sustainable way can be: the so called Mobility as a service (*Maas*) - which tries to simplify the trip planning and ticket purchase, especially for multimodal trips; the ambition to a Smart City, where the city makes part of the user's trip decision, in order to optimize the total system's impact, thus avoiding the egotistical behavior of people that brings the total system to a lower performance; the European Sustainable Urban Mobility Plan (*SUMP*)[2], which tries to help urban contexts in promoting actions to encourage the use of sustainable means of transportation; follows the 17 European development goals[3]; designing a resilient city; include new means of transportation and technologies; planning the people evacuation due to catastrophic events; reduce barriers to people with disabilities as well as for soft mobility; increase transport accessibility; use a digital hub to manage and provide all traffic information to users based on their trip. The approaches used for the analysis of a transportation system can be classified in *macroscopic*, *mesoscopic* and *microscopic* approach. These models are respectively characterized by an increasingly complexity of data, tools and models required for the analysis, but at the same time allowing to estimate more in detail transportation impacts and being able to analyze advanced technologies, means of transportation and strategies, which can not be simulated with traditional macroscopic approach. The main difference between these approaches is that the macroscopic approach analyses the transportation system under a stationary regime, while mesoscopic and microscopic approaches consider the evolution of the system with time. It is worth noting that predicting users decision is a complex task; travel preferences and choices can vary during the same day or even during the trip. In fact, the main challenge for a transportation system analysis consists on modeling the transportation demand; in general, more data are known regarding users' trips, behavior and preferences and more efficiently can be estimated. The transport sector can currently count on different types of big data that are easily detectable - tracing of telephone cards, *GPS* tracks, navigators, floating car data, and so on, but it is expensive either to obtain them from public or private authorities that usually own

[2] Osservatorio PUMS: https://www.osservatoriopums.it/bologna (accessed: February 16, 2022)

[3] Sustainable Development Goals (*SDGs*): https://ec.europa.eu/international-partnerships/sustainable-development-goals$_e n(accessed : February 16, 2022)$

these type of data or build other databases through in-site inspections. For this reason, the transportation system theory is composed by many models which step by step try to estimate users' trips in terms of starting time, trip purpose, origin position, end position, means of transportation and route. From the other hand, knowing certain type of data which can help to better estimate the transportation demand can requires specific adaptions to the context. In addition, the transportation system is currently evolving into a more dynamic and variegate environment with many technology features, new means of transportation and different travel habits; the transportation demand is more difficult to estimate and certain peculiarities of the system can not be captured by traditional models. Another crucial aspect is related to the study area: certain effects on the transportation system can be captured only considering a large-scale and real scenario, despite increasing at the same time the analysis complexity. For these reason, current research is focusing more on mesoscopic and microscopic models to analyse the transportation system on large contexts, thanks also to the recent development of computers that allows to manage all data and process these approaches need.

The literature review (see Section 1.1) mainly focuses on exploring software, models and applications for the analysis of a transportation system with different approaches, while the literature needs and research goals are illustrated on Section 1.2. Chapters 2 and 3 summarize my deductions derived from both the literature review and practical application involving software testing, development and implementations, model calibrations and digital twin construction. Chapter 4 shows the current available open or commercial software for analysing a transportation system with an advanced approach. The core applications are illustrated on Chapter 5 and 6, related respectively to a microscopic digital twin of the city of Bologna and a mesoscopic digital twin of the whole Bay Area, California. Final thoughts and conclusions, as well as intellectual contributes, policy implications and future research are illustrated on chapters 7. Finally, my dissemination and presentations at conferences are illustrated on Chapter 7.3.

1.1 Literature Review

The literature review below captures the historic development towards microsimulations for the analysis of a transportation system 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. The 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 Patriksson (1994) for a comprehensive overview. The simplest assumption is that all users follow the shortest route. A more realistic traffic assignment, formulated by Wardrop (see Wardrop (1952)), 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 - see Frank and Wolfe. (1956)), 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 Cascetta (2001). 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 Ömer Verbas et al. (2016) a multi-modal traffic assignment is modeled; in Kloostra and Roorda (2019) the link flows of autonomous vehicles (AVs) are modeled by increasing the link capacities; in Schweizer et al. (2016b) the empty and occupied vehicle flows of SAVs are determined under system optimum flow constraints by solving a linear programming problem and in Lee et al. (2017) 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 (see Bowman and Ben-Akiva (2001)). 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 (see Charypar et al. (2007) and Michael Balmer (2006)): (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 modeling 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 (see Zhao and Sadek (2012), Hsueh et al. (2021), Mtoi et al. (2014) and Pi et al. (2019)). Mesoscopic models are also used to determine a dynamic user equilibrium (DUE) by running simulations iteratively while updating link travel times (see Bowman and Ben-Akiva (2001)). In activity-based demand modeling frameworks, mesoscopic simulations are employed to iteratively optimize the activity sequencing, plan generation, and to determine the DUE (see Michael Balmer (2006)). Flötteröd et al. (2011) applies such algorithms to the city of Zurich, Switzerland , and see W. et al. (2010) performs a mesoscopic simulation on whole Switzerland, 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 and Sadek (2012) simulated Buffalo and Niagara Region, while Hsueh et al. (2021) simulated the whole San Francisco Bay Area, California (about 18,000 km2), Childress et al. (2015) examined AVs in the Seattle region using the SoundCast software; the user preferences with respect to AVs have been studied by simulating the entire Paris region (see Kamel et al. (2019)), for a recent review see Do et al. (2019). 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 (see Ramezani et al. (2017)). Moreover, communication channels can be integrated as well in order to simulate V2V and connected autonomous vehicles (CAVs) (see Milanés et al. (2013), Calvert et al. (2017), Haas and Friedrich (2017) and Fernandes and Nunes (2010)). 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 Shladover et al. (2012), while safety aspects of CAVs are investigated in Papadoulis et al. (2019), see juan Liu et al. (2020) 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 Figure 1.1): Schweizer et al. (2021) has been developed during my research activity (see Chapter 6). 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 km2 area of Budapest (see Lu et al. (2019)) using the open source simulator (see Behrisch et al. (2011)); random trips have also been generated for simulating a 9 km2 area of Manhattan, Paris, Berlin, Rome, and London (see Mavromatis et al.) with Simulation of Urban MObility (*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 (see Liu et al. (2006)) using the DRACULA software (see Liu (2006)) and the simulation with AVs of Halifax (see Alam and Habib (2018)) using the commercial software Verkehr In Städten - SIMulationsmodell (*Vissim*) (see Fellendorf and Vortisch (2010)); 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 (see Codeca et al. (2017)). 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 (see Po et al. (2019)); 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 (see Savrasovs et al. (2018)) or the New Jersey Turnpike scenario with tolled highways (see Bartin et al. (2018)). A part for the case study developed during with my research activity (see Schweizer et al. (2021) and Chapter 6), two publications on validated large-scale micro-simulation could be found by the authors, see Figure 1.1. 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. In fact, the developed case study (see Schweizer et al. (2021)) 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 Figure 1.1: 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. 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 (see Kloostra and Roorda (2019)). For some important cases, there is a strong inter-dependency between microscopic events and macroscopic quantities (flows or densities), suggesting that a separation

Publication	Simulator/Demand Model	Network	Modes
Uppoor and Fiore (2011)	*SUMO/DUE*	Cologne, Germany from *OSM* 400 km2	Car
Codeca et al. (2017)	*SUMO* + activity-gen/stochastic assignment	Luxembourg, *OSM*, 156 km^2, 931 km roads	Car, bus
Schweizer et al. (2021)	SUMO + *SUMOPy/DUE*, Mode choice	Bologna, Italy, *OSM*, 12 * 7 km	Car, bus motorcycle, pedestrian, bike

Publication	Demand Generation	Validation Method
Uppoor and Fiore (2011)	Activity generator based on 7,000 surveys, 700,000 trips in 24 h	Qualitative comparison of flows with observed data
Codeca et al. (2017)	Activity generator based on public data demographics, *POIs*, etc., 24 h	Comparison of average link speeds from floating car data
Schweizer et al. (2021)	Activity base, disaggregation of OD matrix, *GPS* traces, *GTFS*, peak hours	Car/motorcycle, link flows compared to link traffic counts

Table 1.1: Comparison of published validated large-scale micro-simulation and present work.

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 (see Martínez-Díaz and Soriguera (2018)). In this case, vehicle platoons can be organized, average headways decrease and capacity increases. Shladover et al. (2012) 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 (see Lu et al. (2019)). Clearly, the dynamics in intersections and the duration of platoons (the time vehicles stay together while traveling on a common route) have a dramatic effect on the capacity (see Lee et al. (2017)). 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 Wierbos et al. (2021), Luo et al. (2014) and Bernardi et al. (2015), for pedestrians-bicycles interactions and Woodman et al. (2019) 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 Figure 1.1), there appears to be a real research gap and a need for such scenarios — a part for the case study developed during my research activity (see Schweizer et al. (2021) and Chapter 6), the only publicly available scenario of this kind is the LuST scenario on Github[4], 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 (see Roulland

[4]Github. LUST Scenario: https://github.com/lcodeca/LuSTScenario (accessed: February 16, 2022)

et al. (2015), Wilson (2018) and Cuauhtemoc et al. (2017)) or the use of "big data" to improve traditional surveys (see Croce et al. (2019)). 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.

1.2 Literature Needs and Research Goals

The literature needs a better understanding of the capabilities of using either a microscopic or mesoscopic approach to analyze a transportation system, as well as more more reliable, detailed, user friendly and flexible *open* software and flexible and transferable models, as well as having open built-in detailed, large-scale and validated scenarios to be used as a test platform for 'what if' scenarios.

Said that, this elaborate aims at expanding the knowledge of mesoscopic and microscopic approaches for the analysis of a Transportation system; in particular, thanks to a the literature review (see Section 1.1) and practical applications during my PhD studies (see Chapter 5 and Chapter 6), I explored both traditional and advanced approaches for the analysis of a transportation system, to identify their differences, their relatives pro and cons, their potentialities (see Chapter 2 and Chapter 3), as well as available software (see Chapter 4). The main goal of my PhD dissertation was represented by developing new models and software implementations to model both transportation supply and demand for an advanced transportation system analysis. In particular, the main activities focused on exploiting different type of big data while approaching two practical applications of digital twin have been developed using a microscopic and a mesoscopic approach for the transportation system's analysis (see Chapter 5 and Chapter 6). The first application have been developed at University of Bologna since my Master's Internship and thesis. The second application has been developed as a Visiting Researcher in the field of transportation Engineering at the MBA Department of Marketing and Business Analytics - Lucas College and Graduate School of Business of the San José State University, and to the not-for-profit company Prospect Silicon Valley. During this period I was involved in the Mineta transportation Institute (MTI) Research Project

'Estimating Impacts of Automated Shuttles' and the Silicon Valley Community Foundation (SVCF) Research Project 'Automated Shuttles as First and Last Mile Service in San Mateo County'.

Chapter 2

Traditional Approach for a Transportation System Analysis

Traditional models use a macroscopic approach and opt for deriving the transportation system performances and emissions as well as travelers experience quantities mainly from traffic flows - as an average number of vehicles crossing the considered road sections in a fixed time interval (stationary hypothesis), which can be either collected from traffic counters, or estimated. To estimate traffic flows, it is initially requested to estimate transportation demand, which is usually schematized as an *OD* matrix which indicates the number of people who want to make a trip between each couple of points in the study area - aggregated by zones called Traffic Assignment Zones (*TAZs*), and associated to dummy nodes (centroids) - in a fixed time interval, for a travel purpose, with a specific means of transportation. Based on the available data, zones can be even very small or even coincide with single road links or buildings.

At the same time, the transportation supply is modeled through the transport graph, which can either contain only the main roads (see Figure 2.1), or all the transportation network.

Clearly, more the transportation demand and supply are detailed and more the analysis is complex, time consuming and then expensive, especially until a few years ago when processes were compiled manually.

The recent and fast development of software and computer engineering allowed on the one hand to automatize these processes, and on the other hand to obtain the maximum outputs to balance the high efforts in modeling up to a complete road network and agent-based demand.

For this reason, a macroscopic approach is usually convenient for faster and aggregated analysis with just an extracted simple road network and with the transportation demand modeled as an *OD* matrix as small as possible, while for more detailed analysis it is more convenient to use advanced models (see Section 3.1 and Section 3.2).

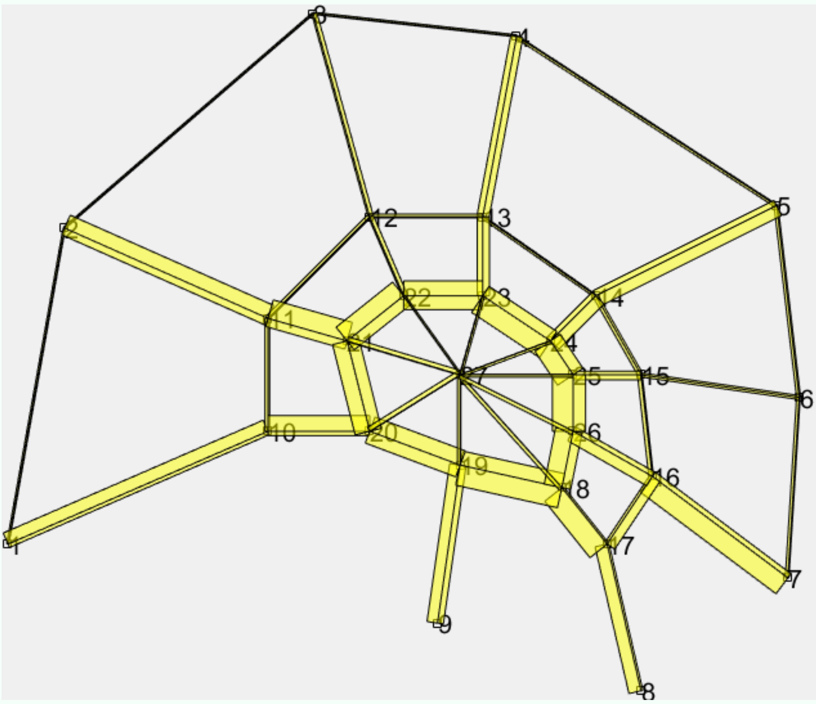


Figure 2.1: Example of a simplified transportation graph of the city of Bologna, Italy, with link flows, using a macroscopic approach for the analysis of a transportation system

Chapter 3

Advanced Models

Recently, transportation analysis are moving from the traditional macroscopic approach to the new mesoscopic and microscopic approach thanks to the recent availability of big computational power and willingness to estimate more in detail the transportation system and their performances and impacts as well as the need to study the effect of new technologies, strategies and means of transportation which are difficult to model with the traditional methods (see Chapter 2).

Thanks to the literature review (see Section 1.1), and to the practical application with simulation software (see Chapter 4) and specific cases studies (see chapters 5 and 6), I could experiment with my hands and deeply understand the advanced approaches for the analysis of a transportation system (see Section 3.1 and 3.2), comprehend the effective needs and potentialities of using these tools (see Section 3.5), as well as identify the pros and cons and their challenges (see Section 3.6), and make a comparison between different approaches (see Section 3.7), as illustrated on next sections.

3.1 Microscopic Approach

Using a microscopic approach to analyze a transportation system essentially means considering the evolution of the system with time, model a vehicle/agent/based demand and simulate it considering interactions between vehicles (e.g. car-following models,), and between these latter with the infrastructure (e.g. Intelligent Transport System (*ITS*), platooning). A simulation is essential for considering all possible variables which can conditionate the system (e.g. agents' behavior, propagation of congestion).

3.1.1 Microscopic Supply

The transportation supply should be as detailed as possible, in order to make a more realistic simulation, and to consider the effect of more variables: see 3.1 and 3.2 for respectively an example of a complex intersection opened on *SUMO*'s NETwork EDITor (*NETEDIT*)[1] and a large-scale *SUMO*'s network schematization opened on *SUMOPy*, which have been used in this research (see Chapter 6). For example, a *SUMO*'s network infrastructure is modeled as represented on Figure 3.3, and mainly contain information obtainable from *OSM* database through SUMO's NETwork CONVERTer (*NETCONVERT*)[2] regarding:

- *modes* - Vehicle modes used for describing the vehicle typologies that can access each road lane
- *nodes* - Intersections where two or more roads converge
- *edges* - Roads between couples of intersections
- *lanes* - Road's lanes where only certain vehicles are allowed
- *roundabouts* - Roundabouts automatically revealed from *SUMOPy*
- *connections* - Maneuvers connecting couples of lanes inside an intersection
- *crossing* - Pedestrian crossings
- *Traffic Light System (*TLS*)*
- *ptstops* - Public transport stops

[1]NETEDIT: https://sumo.dlr.de/docs/Netedit/index.html (accessed: February 16, 2022)

[2]SUMO's Netconvert: https://sumo.dlr.de/docs/netconvert.html (accessed: February 16, 2022)

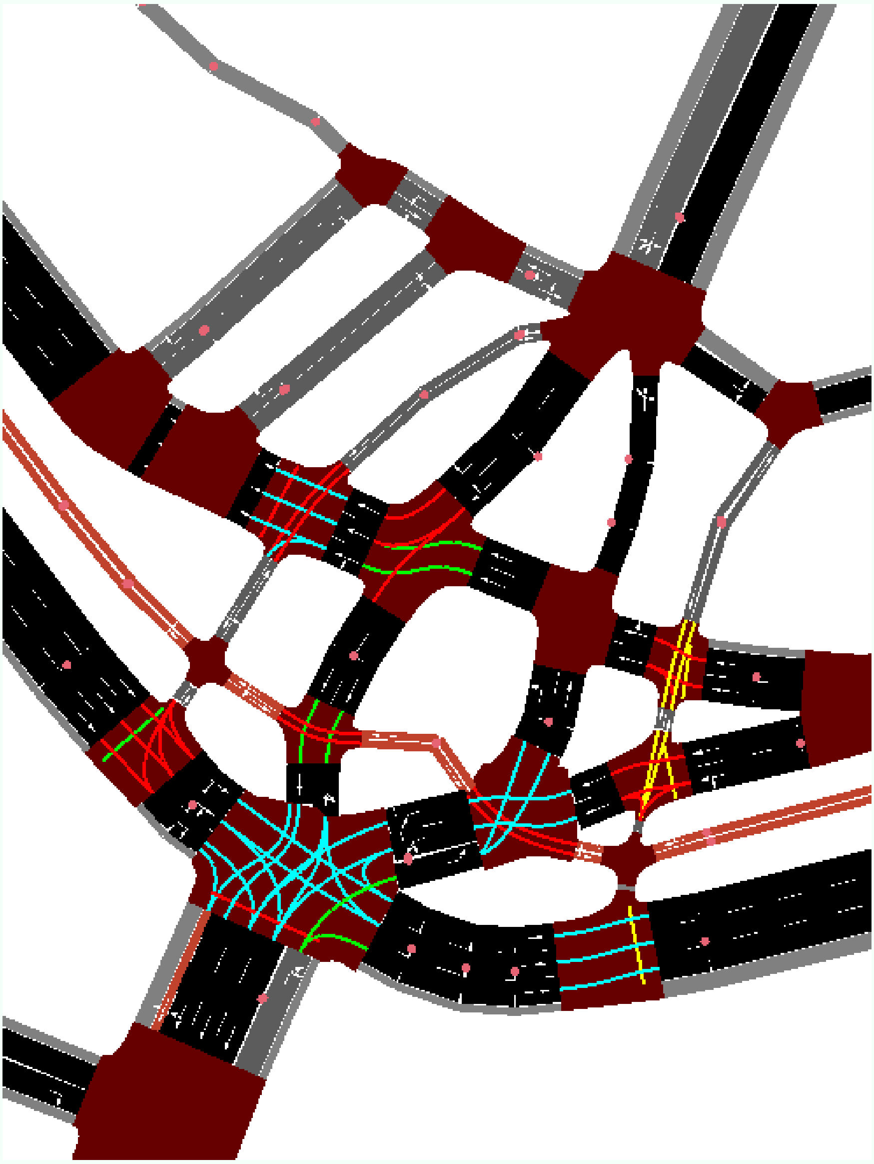


Figure 3.1: SUMO's network opened on *SUMO*'s NETEDIT - 'Porta San Felice' intersection scenario in Bologna, IT

Figure 3.2: SUMO's network opened on *SUMOPy* together with landuses - City of Bologna, IT

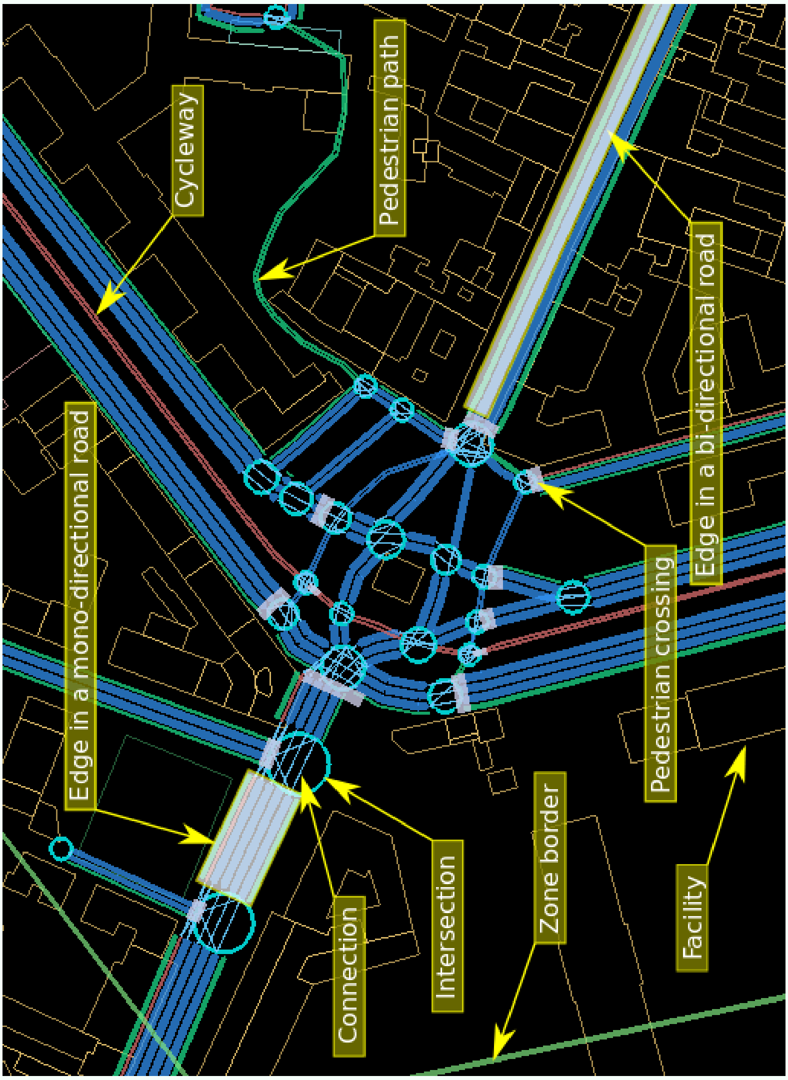


Figure 3.3: Illustration of the objects modeled on a SUMO's network - 'Porta San Felice' intersection scenario in Bologna, IT

Generally, simulation software bases the transportation network on *OSM*[3] open information (see Figure 3.13 and sections 5.1.1 and 6.1.1) and try to convert it on the desired format, but even other open and commercial information can be used, as well as local available data (see Section 3.3). The strength behind *OSM* is that it is free to consult, it is quite reliable, it covers all the globe and it contains many detailed information, even related to the territory and land-use; in any case, it is build often from passionate (not-experts) volunteers, and it may contains some errors.

Other supply's features may include the public transportation system implementation (e.g. stops, services, schedules and so on; see Figure 3.4 and 3.5), other public services (e.g. shared vehicles), charging points (electric grid and gas stations), traffic light, transportation technologies (e.g. Vehicle-to-Everything (*V2X*) communication), parking and so on.

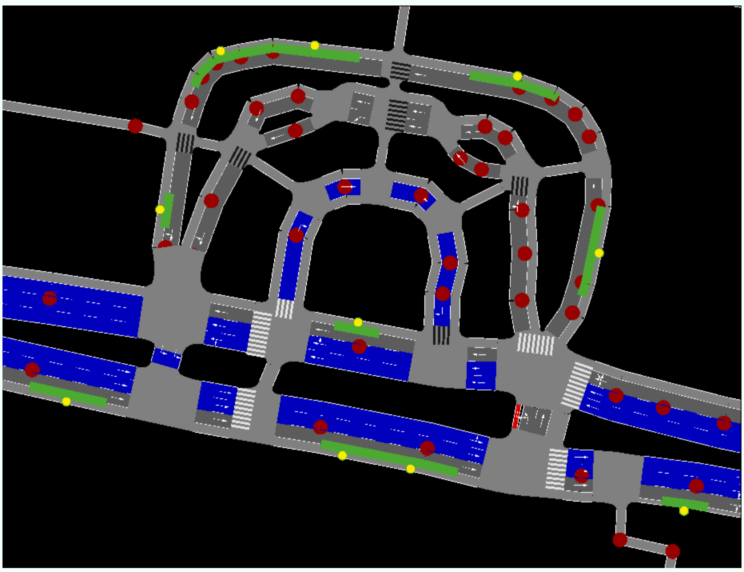


Figure 3.4: Modeling the public transportation stops at the main train station in the city of Bologna - 'Stazione Centrale'

[3] *OSM*: https://www.openstreetmap.org/map=9/37.0541/-121.3852 (accessed: February 16, 2022)

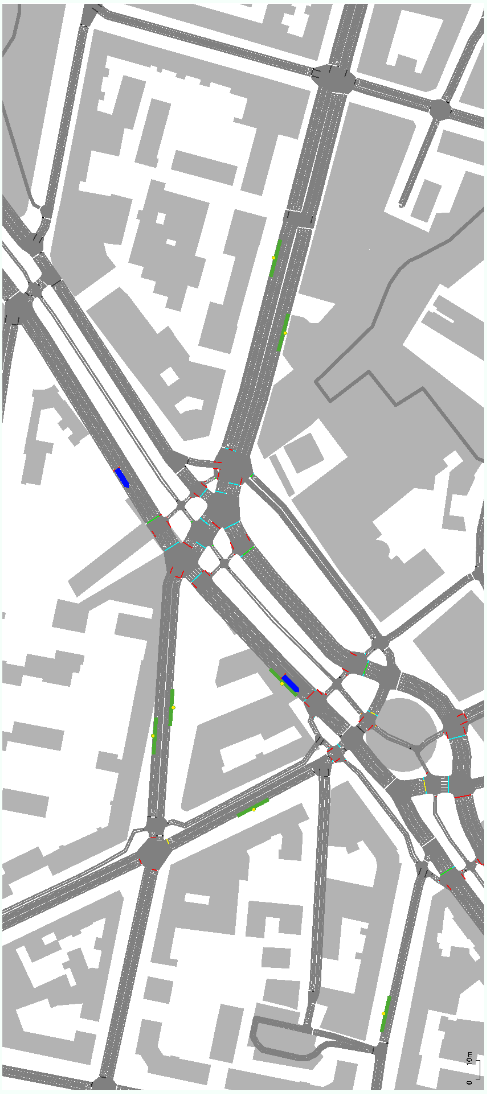


Figure 3.5: Example of a public transportation line simulation with *SUMO*. It is possible to see some buses (blue color) and some public transportation stops (green color)

3.1.2 Microscopic Demand

The transportation demand should be agent-based or vehicle-based, and the so called 'trips' should include the exact starting and ending position and time of trips in the transportation network, as well as used mode of transportation. More sophisticated demand models could include activity-based trips, multi-modal trips, live mode choice, trip plans and concatenated trips, as the case studies showed on chapters 5 and 6. After the routing process, the transportation demand is ready to be simulated in the transportation supply.

The transportation demand can be estimated from different data sources. Usually, more data are available, for example big data 8see Section 3.3), and more accurate is the analysis. In the case no data are available for a certain study area, it would be necessary to count on accurate models that build agents, plans and infrastructures from advanced models. An example could be to estimate agent-based demand either from the disaggregation of *OD* matrices (see section6.1.3) or from simple population statistics (see Section 5.1.1).

3.2 Mesoscopic Approach

A Mesoscopic approach can be considered as a simplified microscopic approach: the interactions between vehicles and between these latter with the infrastructure during the traffic simulation are more approximated and based on empirical equations that mainly estimate travel time on network links and waiting time at intersection based of traffic flows and road infrastructure: e.g. queue theory and circulation theory. In fact, it is not possible to model the physical vehicle in the network. but we can just monitor its position over time, and a graphical illustration can show only the flow distribution over time and, if desired, the exact position of entities and travel choices (see Figure 3.6).

The transportation network should be entirely modeled, thus including all transportation link and nodes, but can contain less information than microscopic approach - e.g. real road geometry - and some features can not be modeled - e.g. platooning.

The transportation demand can be modeled as for the microscopic approach.

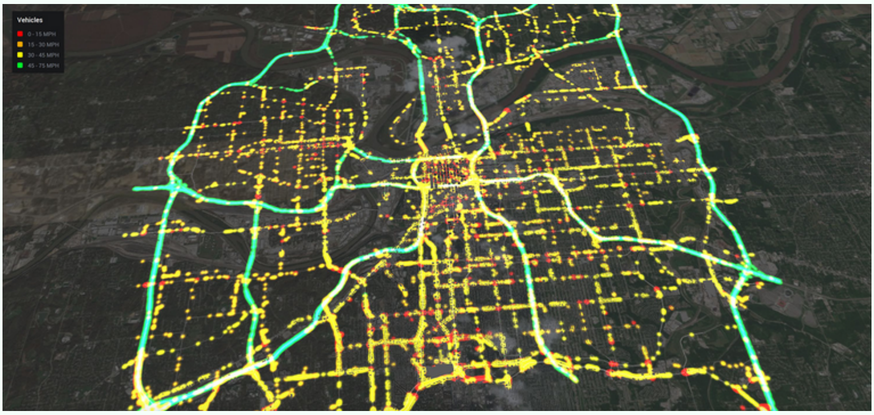


Figure 3.6: Graphical illustration of a mesoscopic simulation with *BEAM* software - Bay Area, California. In this case, the colors represent the current average speed on road links, from red (low speeds), to green (high speeds)

Therefore, the difference with microscopic approach lies on the fact that trips are simulated with a higher approximation, and results may be less realistic: see Section 3.7 for punctual comparison between a microscopic and a mesoscopic approach.

3.3 Big Data

Advanced approaches for the analysis of a transportation system undoubtedly prefer exploit big data and different source of information. The infrastructure modeling process can count on *OSM* data or other open network, as explained on 3.1.1 and applied on Section 6.1.1 and 5.1.1, but demand data are location-based and are either expensive or difficult to obtain. The difficulty on having transportation demand data for performing a microscopic traffic analysis is mainly related to the fact that they are usually owned by companies (e.g. insurance companies, transportation companies, telephone companies and so on); these data can assume different forms, for example *OD* matrices (see Section 6.1.3), *GPS* traces (see Section 6.1.3 and figure 3.7), statistics (see Section 5.1.1), survey reports, censuses, traffic counts (see Section 6.2.2) and so on. It is important to highlight that these data can be useful for both the transport demand estimation and for the model validation.

Figure 3.7: *GPS* big data overlapped with the transportation network of Bologna. The *GPS* zomes from the 'Bella Mossa' initiative and are related to cyclists: each *GPS* trace is composed by *GPS* points recorder on average every f seconds

Other infrastructure data which can increase the accuracy of a standard analysis can be the traffic light schemes, parking, charge points and public transportation schedules (see Section 3.1.1). It is worth noting that having information about the whole scheme of traffic lights in a big area can be a big issue if not collected by either the Municipality or other entities. Regarding the public transportation service, recently most cities provide the *GTFS* files that includes information about all the public transportation services in the area (see sections 6.1.2 and 5.1.1). Parking, if not contained in a open database, are difficult to count even manually: many are hidden from the road, while charge points can be easily identified.

3.4 Simulation

Advanced approaches are clearly characterized by a high level of complexity that can be solved by simulating the system. In fact, a simulation is generally used to analyze a complex system, where the simplest thing to do is to replicate it by modeling all his rules and managing all the entities that enter, stay and leave the system (see Figure 3.8).

A simulation model applied to a transportation problem, usually called 'traffic micro-simulation', should first of all focus on the 'transportation time' of each user, considered as the periods of the days of each users dedicated to a movement from a place to another; these places usually coincide with an activity - work, school, supermarket and so on - and depending on the scale of the analysis certain movements can clearly be ignored: e.g. people walking inside a big shopping center from one shop to another in a city level simulation. The simulation model can be transferred to the traffic simulation case as modeled in Figure 6.2, and described as follow:

- System: the system is delimited to the study area
- System schematization : road network and activities; there is the possibility to move from some places to others with different means of transport
- Entities: population that carry out different type of activities and sometime want to move from one activity to another by either walking or driving a vehicle.
- Rules: users carrying out their activities and when they have to move from one point to another they use a means of transportation. Users and vehicles have to respect the road rules by traveling the road network and interfering with other users, vehicles and the transport infrastructure.

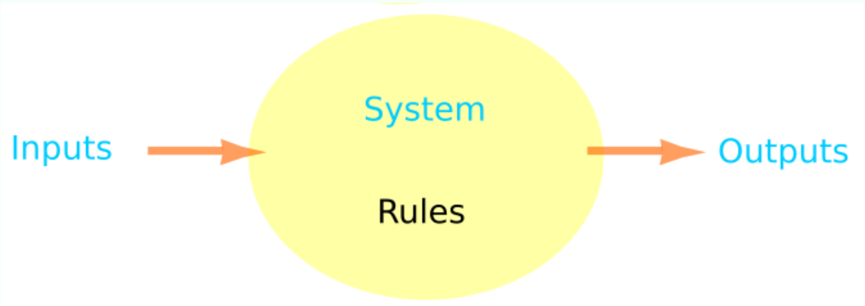


Figure 3.8: Simualtion scheme

Figure 3.9: Traffic simulation scheme

3.5 Towards Advanced Approaches

Current research in the transportation field is particularly focused on advanced - either mesoscopic or microscopic - traffic analysis, as described in Section 1.1, this is a consequences of several factors and advantages.

The main need of applying an advanced approach derives from obtaining a more detailed analysis, under several point of view:

- Graphics: two or three dimensional animated simulation rendering (see Figure 3.12)
- Simulate the traffic evolution during an interval time
- Better represent transport supply: road geometry, lane-detail (access,

connections, public transport service, Traffic lights)

- Reproduce better the interactions between agents, vehicles and the transport network
- Better modeling user experience: walking and driving skills, performances and behavior as well as real-time decisions
- Step by step position monitoring to determine the hourly low of motion
- Possibility to estimate pollutant emissions based on the vehicle and his hourly low of motion
- Track user experience: multi agent and agent based simulation
- Monitors all simulation parameters step by step: user speeds, edge flows, total waiting time, and so on
- Simulate chained trips - Activity based demand
- Simulate multi-modal trips

Also, advanced approaches are needed to model new transportation technologies like:

- Flow-based and coordinated traffic light systems (see Figure 3.11)
- $V2X$
- Automated Vehicles (AV), Electric Vehicles (EV),Connected Automated Vehicles (CAV), Personal Rapid Transit (PRT), Group Rapid Transit (GRT)
- Platooning
- Driver assistance (see Figure 3.10)

and new transportation services like:

- On-demand vehicles
- First and last mile services

- Shared vehicles: car, bike (see Poliziani et al. (2022c)), E-bike, E-scooters
- Real-time route-planning
- Electric grid

It is worth noting that this considerably increase the complexity of the analysis, so it is not suggested to use advanced approaches to analyze a transportation system in the case it is not necessary.

The full process is generally supported by simulation suites and software (see Chapter 4): during last years, computer engineering field has experienced a considerable development in terms of software, computational speed, data storage, hardware components and so on, thus allowing to make even bigger processes affordable to most people; this is essential for performing traffic simulations.

3.6 Challenges

The advanced approaches have many advantages but the complexity of the analysis brings up many challenges that are not still fully argued in literature:

- *Transportation demand estimation*: the transportation demand estimation remains the biggest problem, as for the macroscopic approach. An advanced approach requires even more detailed information (see Section 3.1.2), that can for example derive from big data or from advanced models that estimate the transportation demand from available statistics. There is a lack of demand models in literature due to the data variety and also to the different level of detail that can be required by the advanced models' demand, that are generally high.
- *Transportation supply schematization*: an advanced approach require also a more detailed modeling of the transportation supply (see Section 3.1.1), depending on the case study, but this is more accessible than transportation demand.
- *Data availability*: see Section 3.3

Figure 3.10: Testing driver assistance technologies with a real BMW. The Startup fka Silicon Valley developed this driving simulator which allows to track drivers approach to different travel conditions and driving assistance technologies. The simulator is composed by a real car, a driving simulator software that project an artificial scenario front and back the car and tracking controls of the driver behavior. The startup is hosted by Prospect Silicon Valley and the CEO is Christian Roth - Photo taken during my Visiting Scholar period at San José State University in collaboration with the MBA Department and the not-for-profit Prospect Silicon Valley.

- *Data quality*: big data does not means good data; usually, even if it is possible to have a big data, this can not be representative to the population. From the demand side, a simple example can be having *GPS* traces of all the trips carried out for a survey campaign that rewards users with discounts in concerts tickets for recording the bicycle trips for a sport purpose, to encourage cycling as a physical activity: it is

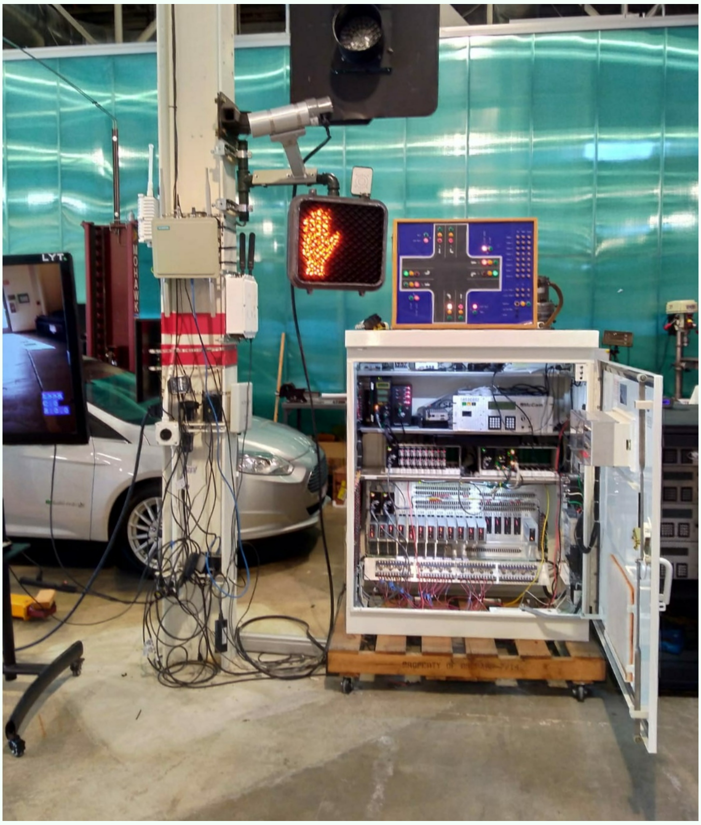


Figure 3.11: Example of *ITS*: external box to connect directly to the traffic light in order to manage the duration of its phases. The Startup LYT developed this system in order to prioritize either transit or emergency vehicles at traffic light. LYT is hosted by Prospect Silicon Valley and the CEO is Timothy Menard - Photo taken during my Visiting Scholar period at San José State University in collaboration with the MBA Department and the not-for-profit Prospect Silicon Valley.

clear that this campaign can affect only a certain part of the population and this type of data can be used only for a specific study and

Figure 3.12: Three dimensional simulation rendering with an example scenario in *Vissim*

not to simulate the bicycle traffic during the morning peak hour of week days. Demand data useful for a advanced approach can assume a big variety of formats and should be carefully analyzed, also from a critical point of view, before using them, focusing on their representativeness, accuracy and form. From the supply side lot of data are required and they should be handled diligently. Regarding the transportation network, even if it is possible to adopt open source data, they are mainly built from volunteer and they can contain errors as well as they can not schematize well the real infrastructure. Regarding information related to other supply components like traffic lights, public transportation services and so on, they are usually drawn up by competent people, but they can also contain errors of any type due to their complexity.

- *Data management*: performing an advanced analysis require a good organization handling a big quantity of various data typologies. The data management process is usually handled by microsimulation suite software (see Chapter 4), that can for instance allow either to store data in a scenario file or to write configuration files where all the data file paths are identified.

- *Time consuming*: whatever is the scale of the analysis, an advanced approach require a time-consuming analysis.

- *Find the simulation software*: many simulation software are already developed and are currently developing; the challenge consists on identifying a software suitable for the available demand and supply data as well as for their provided outputs.

- *Software availability*: not all the software are open source (see Chapter 4). Open source software are valid as well as commercial software but the main difference is that they are usually in a beta version and can require programming competences to run microsimulations. Contrarily, the strength of commercial software is mainly due to the user-friendly and easy approach to the process usually managed by a Graphical User Interface (*GUI*).

- *Simulation software usage*: each simulation software requires a preliminary study of its documentation as well as simulation tests in order to understand all the software capabilities and tools.

- *Inputs preparation*: all the input files should be adapted based on software requirements; transportation demand and supply information should coincide with the information required by the simulation software based on its level of detail and format, that should be previously investigated.

- *Outputs elaboration*: not all the software provide user-friendly result formats; the output provided by the software should be previously verified.

- *Computational power*: the simulation process usually require a big computational power, particularly for large-scale scenarios.

- *Simulation accuracy*: each traffic simulator has its own algorithms describing the users and vehicles behavior that are conditioned by several parameters that should be calibrated, at least in the case of a sophisticated analysis, instead of performing a 'black box' analysis. The calibration of these parameters will be particularly useful when simulation a traffic congestion that, as unstable situation, can be sensitively conditioned by even small changes of this parameters, a part of even small errors in the transportation infrastructure model. Parameters can affect for example driver impatience, lane change models, headway perception, emergency deceleration, car-following model and so on.

- *Results validation*: as for the traditional approach, it is always important not to use all the available demand data for the transportation demand estimation, even if this latter generally require a big quantity of data, especially with an advanced approach. This why it is important to successively validate the model.

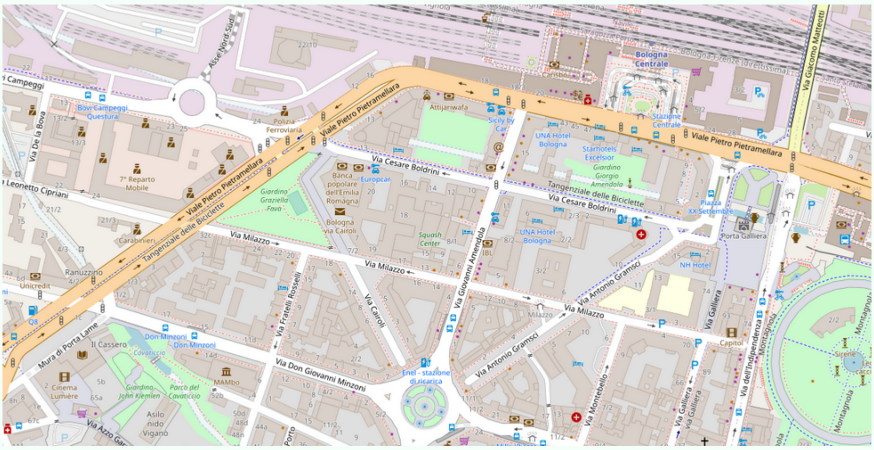


Figure 3.13: extract of the exportable *OSM* network from its web page visualization tool; the case of the train station area in Bologna, Italy

3.7 Microscopic vs Mesoscopic Approaches

In this section, the mesoscopic and microsocpic approaches for the analysis of a transportation system are compared. The pluses of using a microscopic approach respect to a mesoscopic approach can be:

- Reproduction of interactions between vehicles, network and agents.

- Tracking of step-precision speed and acceleration profile.
- Can potentially estimate simulation outputs with more accuracy - energy consumption, pollutant emissions, noise, travel/waiting time (see Poliziani et al. (2022a) and Poliziani et al. (2022b)) and so on: vehicles and agents behaviour is simulated, rather than modeled with empirical formulas.
- Consider traffic dynamics at intersections.
- Simulate the congestion effect and its propagation in the network.
- Model certain technologies (Traffic light sync., platooning, V2X and so on).
- Higher sensibility with system changes and more degree of freedom.
- Better highlights any eventual error.
- Ability to simulate border line scenarios.
- Eye catching graphics.

While the minus of using a microscopic approach respect to a mesoscopic approach can be:

- Model creation is more time-consuming due to the more detail of the analysis.
- More susceptible to errors.
- Requires a bigger computational power.

Chapter 4

Simulation Suites and Software

From a literature emerges some software available to perform macro, meso and microscopic analysis of a transportation system and demand simulation; they differ mainly for 1) their policy - open or commercial software; 2) level of development - most are a beta-version in continuous development, and others are already quite developed and documented; 3) required inputs and provided outputs; 4) level of detail - number of considered variables; 5) user friendliness - most does not even have a graphical interface; 6) reliability - some software have been already applied to many case studies and other are not; 7) support - assistance from the development team.

4.1 Employed Software

The used or/and developed software during my PhD research have been: *SUMOPy*[1] , *SUMO*[2] and BEAM[3].

The beta software *BEAM* has been developing for some years at *LBNL*[4]

[1]Github. SUMOPy: https://github.com/schwoz/sumopy/ (accessed: February 16, 2022)

[2]SUMO: https://www.eclipse.org/sumo/ (accessed: February 16, 2022)

[3]BEAM: http://beam.lbl.gov (accessed: February 16, 2022)

[4]LBNL: https://www.lbl.gov/ (accessed: February 16, 2022)

and allows to realize a large-scale traffic mesoscopic and agent-based simulation as an extension of the open source software Multi-Agent Transport Simulation (*MATSim*)[5], as well as provide the simulation outputs in a table form. The *BEAM* developers recently calibrated and validated a traffic scenario related to the San Francisco Bay Area, California (see Figure 4.1). The *BEAM* software together with the San Francisco Bay Area scenario are available from the relative open source GitHub repository[6]. The study area is 18,000 square kilometers wide and hosts almost 8 billion of inhabitants, with an average density of more than 400 people per square kilometer. The scenario uses an *OSM* network and contains activity-based demand created for the 10% of the population, based on average activities profile and population demography, that can choose via a probabilistic assignment to use: private car, either private or pooled ride-hail vehicles, public transportation service, bike, go by walk, or use a combination of public transportation service with a walk, car or ride-hail trip. The means of transportation is chosen based on a within-day evaluation of plans and a multinomial logit choice function with a person-specific value of time, and it is refined through an iterative process that consist of running several times the simulation, allowing the possibility to try different plan.

SUMO is an open source, highly portable, microscopic and continuous multi-modal traffic simulation package designed to handle large networks[7]. *SUMO* rapidly developed into a flexible and powerful open-source microsimulator for multi-modal urban traffic networks. The features and the number of tools provided are constantly increasing, making simulations ever more realistic. However, the different functionalities consist at the present state of a large number of binaries and scripts that act upon a large number of files, containing information on the network, the vehicles, districts, trips routes, configurations, and many other parameters. Scripts and data files exist in a dispersed manner. In practice, a master script is necessary to hold all processes and data together in order run a simulation of a specific scenario in a controlled way. This approach is extremely flexible, but it can become very time consuming and error prone to find the various tools, combine their input and output and generate the various configura-

[5]MATSim: https://www.matsim.org/https://sumo.dlr.de/docs/Contributed/SUMOPy.html (accessed: February 16, 2022)

[6]BEAM's documentation: https://beam.readthedocs.io/en/latest/users.html (accessed: February 16, 2022)

[7]*SUMO*: https://www.eclipse.org/sumo/ (accessed: February 16, 2022)

Figure 4.1: Bay Area scenario simulation with *BEAM* software

tion files. Furthermore, it reduces the user-base of *SUMO* to those familiar with scripting and command line interfaces. Instead, *SUMO* has the potential to become a multi-disciplinary simulation platform if it becomes more accessible to disciplines and competences. This problem has been recognized and different graphical user interfaces have been developed.

SUMOPy[8] is written entirely in the object-oriented script language Python, it uses wxWindows with Python Open Graphics Library (*PyOpenGL*) as *GUI* interface and = Numerical Python library (*NumPy*)[9] for fast numerical array-type calculations: Figure 4.2 summarize the main features of both *SUMO* and *SUMOPy*, as well as their communication strategy. It is similar to the traffic generator in that it simplifies the use of *SUMO* through a *GUI* (see Figure 4.5). But *SUMOPy* is more than just a *GUI*, it is a suite that allows to access *SUMO* tools and binaries in a simple unified fashion. The distinguishing features are:

- *SUMOPy* has Python instances that can make direct use of tools already available as Python code.

[8]Github. SUMOPy: https://github.com/schwoz/sumopy/ (accessed: February 16, 2022)
[9]*NumPy*: https://numpy.org/ (accessed: February 16, 2022)

- *SUMOPy* has a Python command line interface that allows direct and interactive manipulation of *SUMOPy* instances.
- *SUMOPy* provides a library that greatly simplifies the scripting.

The more sophisticated process in *SUMOPy* currently consists on micro-simulating a multi-modal agent-based and activity-based demand, where a virtual population is created and each user can decide what transport strategy - either mono- or multi-modal - to use for travel between the planned daily activities. All the processes are based on a detailed transportation network imported from *OSM*.

To better understand the potentialities and the user-friendliness of *SUMOPy*, below is illustrated a tutorial to rapidly create a test-scenario (see figure4.3) with just ten steps - where 10,000 users travel from home to work by either walk or driving his own bike, motorbike or car in a 8X8 blocks artificial city.

1. Open sumopy_gui.py
2. Insert Name (e.g. Hello Sumopyans!!) and Description (e.g. Getting started with *SUMOPy*) and *Scenario/Safe as...* the document with standard characters and without spaces (e.g. hello_sumopyans.obj)
3. Create a grid network with *Network/Generate/Grid network.../Run* and click the magnifying glass *Zoom to fit* below the network editor (See figure4.5), for better visualize it
4. Create random facilities with *Landuse/Facilities/Generate facilities.../Run*
5. Create random parkings with *Landuse/Parking/Generate parking.../Run*
6. Create a random population in the city and their activities with *Demand/Virtual population/Configure population/Generate.../Run*
7. Provide vehicles to the population with *Demand/Virtual population/Configure vehicles/Provide vehicles.../Run*
8. Create a plan for each user that want to move from one activity to another, for each strategy that he can use with *Demand/Virtual population/Plans/Generate.../Run.../Done*

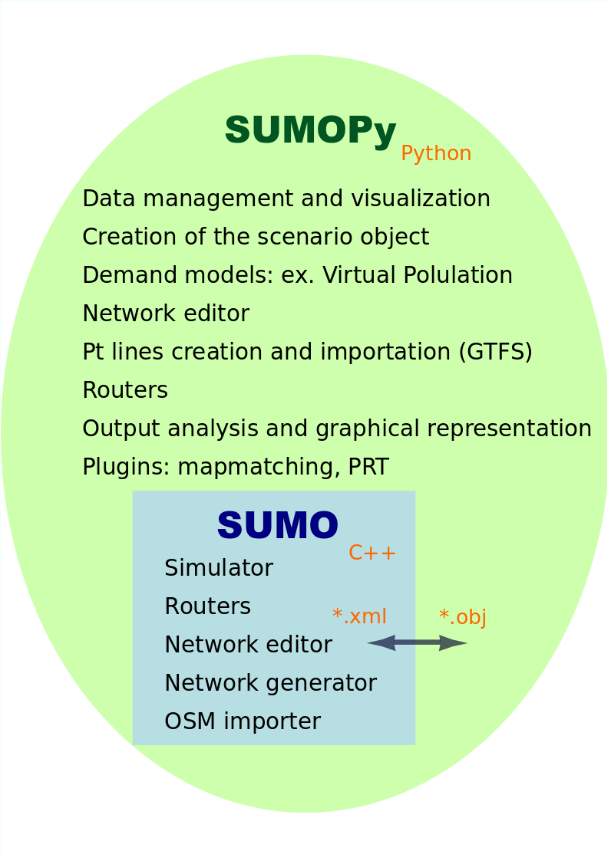


Figure 4.2: *SUMOPy* and *SUMO* software

9. Select a random plan for each user with
   Demand/Virtual population/Plans/Select current plans.../Run
10. Micro-simulate the scenario with
    Simulation/Microscopic simulation/SUMO.../Run

Start the simulation with the *play* button (see figure4.4); zooming the network it is possible to see every user that leave the home, goes by walk

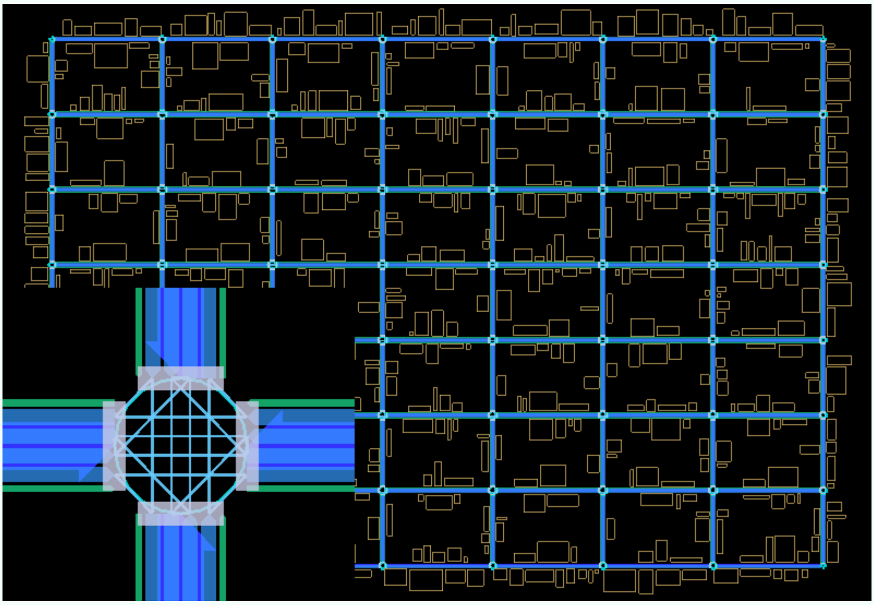


Figure 4.3: SUMOPy*'s* GUI with the *Hello Sumopyans!!* scenario

up to either the work-building or the bike/moto/car parking, then drive the vehicle up to the work-building, park the vehicle and reach the building by walk.

All the SUMOPy's parameters tools and objects have their own definition explorable by moving the mouse-cursor over them. As illustrated in Figure 4.5, it is possible to decompose the *GUI* in nine main chapters:

1. *Scenario's name and description* - Indicates the name of the scenario and his description

2. *Main menu* - Allows to access all SUMOPy's tools

3. *Object broswer* - Organizes and makes all the objects imported in SUMOPy explorable: mainly all the network, landuse and demand objects as well as the simulation results. While navigating in the object browser, both the attributes referred to a particular object and

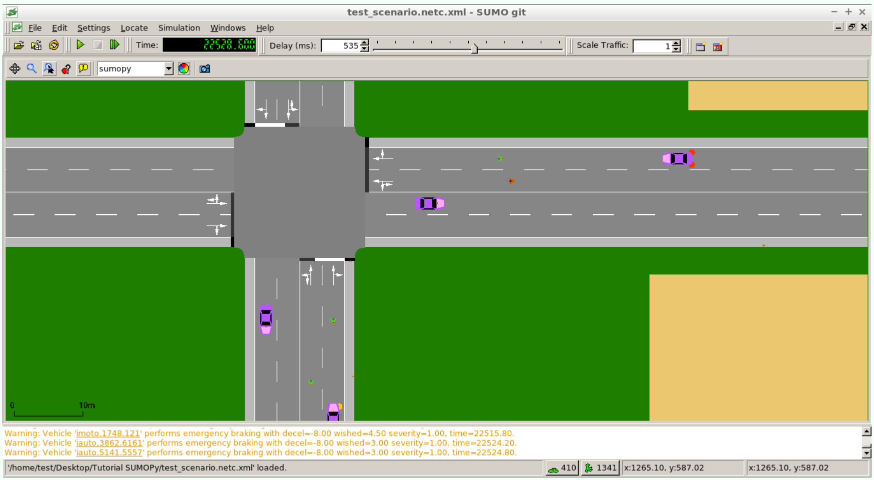


Figure 4.4: *SUMO's GUI* during the micro-simulation of the *Hello Sumopyans!!* scenario

the description of these attributes appear by double-left clicking on a object's ID.

4. *Network manipulation tools* - Allow to directly edit the transportation network in the network editor:

 *Info* - Allows to select every object in the network editor for retrieving their information in the object browser.

 *Stretch* - Allows to edit the object shapes in the network editor by left-clicking on their corners

 *Move* - Allows to move objects in the network editor by left-clicking on them

 *Configure* - Allows to configure the visualization of objects in the network editor, in terms of color, width, shape and so on

 *Delete* - Allows to delete objects from the network editor

Add zone - Allows to draw different type of zones by left-clicking on their corners and double-left-clicking the last one

Add facility - Allows to draw different types of facilities by left-clicking on their corners and double-left-clicking the last one

Add OD - Tool for adding *OD* flows by interacting directly with the zones in the network editor. The flows will allow to create trips to be simulated.

Add turn-flows - Tool for adding turn-flows by interacting directly with the edges in the network editor. The flows will allow to create trips to be simulated.

Add PT line - Allows to insert a bus-line by interacting directly with the public transport stops in the network editor.

5. *Result viewer* - Allows to visualize the simulation results in the network editor

6. *Network editor* - This is the interactive 3D network visualization. This allows to better visualize and interact with the scenario - thanks to the network manipulation tools. The following mouse-key combination allow to navigate the network:

   - *Zoom in/out* - *Hold down Ctrl + Wheel*
   - *Panning* - *Hold down Ctrl + Shift + Button-Left*
   - *Rotate* - *Hold down Ctrl + Shift + Wheel*

7. *Zoom buttons and selection of visible objects* - Facilitate the navigation in the network editor, allowing the following functionalities:

Zoom to fit - zooms the network to fit approximately the boundaries of the window

Zoom in - enlarges the window

Zoom out - contract the window

Select drawing components - allows to select the objects drawn in the window to better inspect the network

8. *Coordinates* - Indicates the coordinates of the mouse pointer in the SUMOPy's local coordinate system in meters

9. *General information* - Provide information about SUMOPy's tools under the mouse pointer

4.2 Other Software

Other software which is worth mentioning are:

- *Vissim*[10]: is an advanced and flexible traffic simulation software belonging to the Planung Transport Verkehr AG - Planning Transportation Traffic (PTV)[11] traffic suite. It simulates complex vehicle interactions realistically on a microscopic level, it models demand, supply, and behaviour in detail, it simulates new forms of mobility such as *CAV* and *Maas* and it can be integrated with the traffic planning tool PTV Visum.

- Visum[12]: is a traffic planning software belonging to the PTV traffic suite designed for transport planners. It conducts macroscopic traffic analyses, forecasts and GIS-based data management, it models all road users and their interactions, it plans public transport services, it develops advanced and future-proofed transport strategies and solutions.

- Paramics Discovery - originally S-Paramics (*Paramics*)[13]: is a microsimulation 3D software developed to help transport professionals design, evaluate and present new solutions. *Paramics* Microsimulation is state of the art traffic modeling software to enable transport

[10] *Vissim*: https://www.ptvgroup.com/it/soluzioni/prodotti/ptv-vissim/?utm$_campaign = AdstrVissimITdsagclid = Cj0KCQjwjer4BRCZARIsABK4QeXjkYyoaJwmLiaQe1b5KbD0W8DvIIbjOcWR_gUNiJyfvXmeWy6r-ugaAimCEALw_wcB(accessed: February16, 2022)$

[11] PTV: https://www.openstreetmap.org/map=9/37.0541/-121.3852 (accessed: February 16, 2022)

[12] Visum: https://www.ptvgroup.com/it/soluzioni/prodotti/ptv-visum/?utm$_campaign = AdstrVisumITdsagclid = Cj0KCQjwjer4BRCZARIsABK4QeW9r6FmyIDEGCBQwdV$ $February16, 2022)$

[13] *Paramics*: https://www.paramics.co.uk/en/ (accessed: February 16, 2022)

professionals to design, evaluate and present solutions. *Paramics* Microsimulation offers: 1) fast network construction, editing, visualisation and simulation; 2) fewer parameters required for calibration than other products; 3) streamlined workflow to match users project workflow; 4) easy to understand outputs for clients. Originally developed by SIAS Ltd, *Paramics* Microsimulation has been at the forefront of microsimulation since the 1990s. S-*Paramics* (the original) and *Paramics* Discovery (the new) have been used on thousands of transport planning projects for over 20 years. They have been used to: 1) design and test major extensions to the M74 and M8 motorways in Scotland; 2) test innovative ATM schemes on the M25 in London; 3) plan major events including the 2014 Ryder Cup.

- Advanced Interactive Microscopic Simulator for Urban and Non-Urban Network (*Aimsun*)[14]: allows to model transportation networks of any dimension: from a single intersection to an entire region. *Aimsun* offer four different packages: *Aimsun* Next, *Aimsun* Live, *Aimsun*, Auto *Aimsun* Ride. *Aimsun* Next mobility modeling software to analyze anything from a single intersection to an entire region. *Aimsun* Live simulates mobility in real time, allowing traffic managers to anticipate congestion on our roads, and stop it before it happens. *Aimsun* Auto is a new software platform for large-scale design and validation of path planning algorithms for self-driving vehicles. *Aimsun* Ride simulates *Maas*, Demand Responsive Transportation (*DRT*), and City Logistics applications. Velasquez-Martínez et al. (2022) made an interesting comparison between the commercial software *Aimsun* and the ope software *SUMO*.

- *MATSim*[15] *MATSim* provides a framework to implement large-scale agent-based transport simulations. The framework consists of several modules which can be combined or used stand-alone. Modules can be replaced by custom implementations to test single aspects of your own work. Currently, *MATSim* offers a framework for demand-modeling, agent-based mobility-simulation (traffic flow simulation), re-planning, a controler to iteratively run simulations as well as methods to analyze the output generated by the modules.

[14] *Aimsun*: https://www.aimsun.com/ (accessed: February 16, 2022)

[15] MATSim: https://www.matsim.org/ (accessed: February 16, 2022): is an open-source framework for implementing large-scale agent-based transport simulations

- SUMO Activity GenerAtion (*SAGA*)[16]: is an activity-based multi-modal mobility scenario generator for the Simulation of Urban MObiltiy (SUMO). Starting from an *OSM* file, *SAGA* extracts the data required to build a multi-modal scenario, and in a step-by-step fashion, generates the configurations needed to execute it, including the intermediate steps required to refine the scenario with additional data, allowing the iterative improvement of realism and representativeness. The workflow implemented, extended, and automated by *SAGA* was developed while hand-crafting the Monaco *SUMO* Traffic (MoST) Scenario. Based on the fast prototyping capabilities added by SAGA, the creation of a multi-modal mobility scenario is readily achievable, and the incremental process to fine-tune it is supported by a workflow instead of being solely based on expert knowledge and experience (see Codeca et al. (2017)).

- TRansportation ANalysis SIMulation System (*Transims*)[17]: is an integrated set of tools originally developed by Los Alamos National Lab (*LANL*) to conduct regional transportation system analyses. With the goal of establishing *Transims* as an ongoing public resource available to the transportation community, it is made available under the United States National Aeronautics and Space Administration (*NASA*) Open Source Agreement Version 1.3.

[16] *SAGA*: https://www.eurecom.fr/en/publication/6224 (accessed: February 16, 2022)
[17] *Transims*: https://en.wikipedia.org/wiki/Transims (accessed: February 16, 2022)

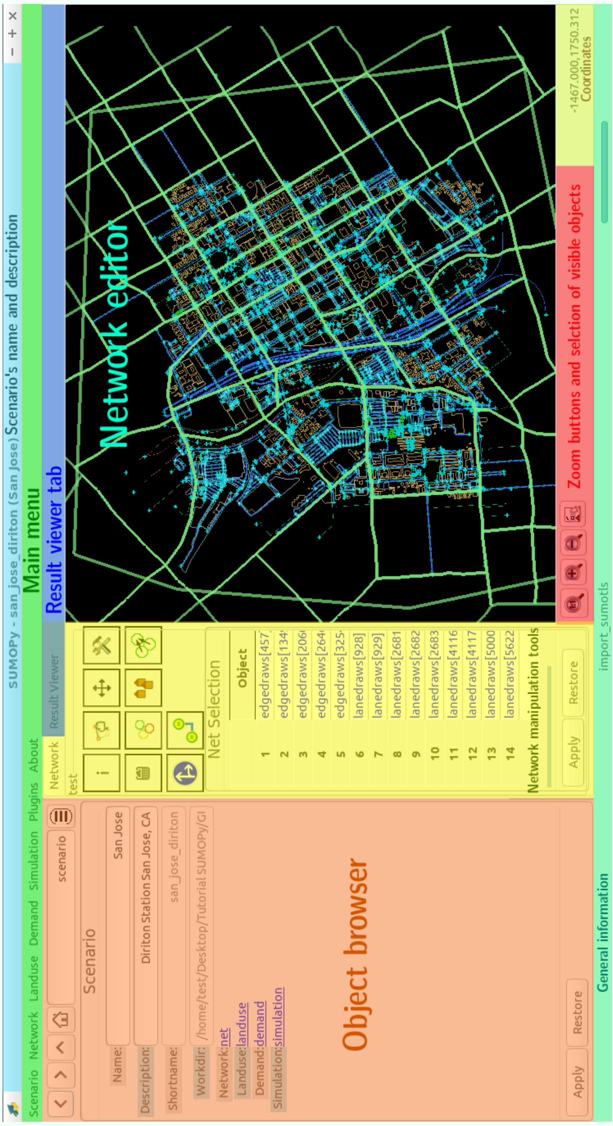


Figure 4.5: SUMOPy*'s* *GUI* with Diriton Station scenario in San Jose, CA

Chapter 5

Mesoscopic Transport Digital Twin of the Bay Area, CA

The growing interest in automated shuttles as a method for providing first and last mile connections to public transport services, together with the scarcity of studies and pilot deployments based on such a transportation mode, led to the idea of using a traffic simulation to model a range of scenarios with low-speed automated vehicles in Santa Clara County, California. The present study aims at testing one of the first practical application of the *BEAM* software while implementing and investigating the potential of automated shuttles that are meant to provide first and last mile connections to higher-order public transit services. The service is assumed to be automated, on-demand, and low-speed—up to 40 km/h (25 mph)— and each vehicle is confined to operate within a round catchment area centered on a high-quality bus stop or transit station, characterized by a high frequency of buses or trains. The first intent of the study is to simulate and estimate the impacts of such a transport system in Santa Clara County. The second goal is to report the experience of conducting research using the beta software *BEAM*, which allows one to perform a traffic simulation as an extension for MATSim, which is mesoscopic in scale characterized by an intermediate level of detail between the classical macro-approach and the recent micro-approach and agent-based, in which an artificial population is reproduced, and each agent can autonomously make their own travel decisions and interact with other agents. The software was developed at Lawrence Berkeley National Laboratory and a baseline scenario was calibrated by the *BEAM* development team to simulate the nine-county San Francisco Bay Area region, which includes the study area of Santa Clara County. This application has been published in the Mineta Transportation Institute publications (see Hsueh et al. (2021)). The main weakness of a public transport network is clearly represented by its accessibility, mainly for low-density areas where fixed-route local bus services are inefficient and generate unsustainable ridership. Hess (2009) observed that the accessibility of a transport network greatly affected the choice of transport mode, especially for older people. As described by Farber and Grandez (2017) for the city of Toronto, the accessibility of transit considerably affects its efficacy. Zuo et al. (2020a) state that transit service can be greatly extended by improving first and last mile access to transit, and disadvantaged residents receive better and more equitable transit accessibility to jobs than others. Kanuri et al. (2019) state that first and last mile connectivity is an important factor in enabling greater integration and accessibility of mass transit networks to the largest number of urban residents. For these reasons, authorities are addressing the problem of the first and last mile connection for a public transport network as a method to improve the accessibility of the public transport service and hence its level of usage[1]. Micro transit, or transit service using smaller on-demand vehicles, represents a valuable alternative for the first and last mile connection, and in recent years public transit agencies including the Santa Clara Valley Transportation Authority and the Alameda-Contra Costa Transit Authority, both in the San Francisco Bay Area, tested this model[2]. For example, the AC Transit Flex pilot program, that serves the catchment area around two BART stations in Alameda County, provides an operating example of the flexible-route, on-demand last mile shuttle service contemplated in this study (except it uses manned, not automated vehicles) that improves service in low den-

[1]Santa Clara Valley Transportation Authority - Transit Service Guidelines: http://www.vta.org/proj-ects-and-programs/programs/tran (accessed: February 16, 2022)

[2]Eno Center for Transportation - UpRouted: Exploring Microtransit in the United States: https://www.enotrans.org/eno-resources/uprouted-exploring-microtransit-united-states/ (accessed: February 16, 2022)

sity and low demand areas while demonstrating cost neutrality and higher efficiency compared to the previous fixed-route service. A survey in 2013 performed in Santa Clara County reported that 10% of riders drive or carpool to the local public transport service (see Corey (2014)). This share of users be a reasonable market segment to capture using new micro transit systems. In fact, through a model of existing public transit options and a hypothetical level 4 shared autonomous vehicle, Moorthy et al. (2017) explored the possibility of automated vehicles being used to solve the last mile problem, concluding that they significantly improved transit sustainability by promoting mode shifts to public transit. Ride hail systems are an important competitor to many public transport services, potentially replacing transit rides[3]. However, Wang and Mu. (2018) analyzed data of Uber ride hail activity in Atlanta, highlighting that the whole ride hail fleet cannot always guarantee sufficient accessibility in terms of wait time in opposition to a well-supported public transport service. Chee et al. (2020) state that automated bus service competes with existing last-mile services. Zuo et al. (2020b) state that transit accessibility to jobs can be improved with bicycles as the first-and-last mile mode, however, Rupi et al. (2019), Rupi et al. (2020) and Schweizer et al. (2020) found that cyclists try to avoid riding in the presence of buses because doing so impacts cyclist safety, speed, and waiting times. Therefore automated shuttles may offer a competitive first and last mile service. In order to provide support for the current public transport service and not to replace it, automated shuttles should be bound within a predefined, geographically-bounded "catchment area," but it is not clear how to define these; a general methodology to determine catchment area is not yet present in literature. Biba et al. (2014) used a parcel-network method for estimating the population with walking access to bus stop locations using spatial and a spatial data (i.e. location and demographics) and the network distances from parcels to bus stop locations in order to deploy the catchment area in a 100-square-mile portion of the Dallas Area Rapid Transit (DART) system covering two Texas communities. Guerra et al. (2012) defined half a mile as the maximum distance from a stop or station that makes a public transport service desirable to users. El-Geneidy et al. (2013) declared that the 85^th percentile walking distance to bus transit service is around 524 m for home-based trip origins. Eom et al. (2019)

[3]International Transport Forum - Shared Mobility Simulations for Helsinki: https://www.itf-oecd.org/sites/default/files/docs/shared-mobility-simulations-helsinki.pd (accessed: February 16, 2022)

found that about 90% of passengers traveled within 3.6 km from railway stations, suggesting a catchment area with a radius of less than 4 km. AC Transit recommended service zones of approximately 5–7 square miles[4]. In any case, as stated by Lin et al. (2019) the catchment area can increase commensurate with changes to first and last mile service: they proved that for a dock-less bike sharing system. However, it is important to take into account, as stated by Roy and Basu (2020), that poor first and last mile performance is generally observed in more suburban locations, where poor sidewalk and bus stop infrastructure combined with the extra cost and time needed to make longer first and last mile trips using feeder buses contribute to low first and last mile quality. Moreover, Chee et al. (2020). declare that frequency of the primary transit routes is critical to the last-mile automated bus service usage. For this reason, the catchment areas are centered only on high frequency bus stops in this study.

On Section 5.1 are explained the methodology used for achieving the main goals . Attention is particularly focused on the pros and cons of using the *BEAM* software (see Section 5.1.1, describing what the team learned by using it and providing suggestions to future users. On Section 5.1.2, the authors describe the methodology to model the automated shuttles in the *BEAM* software, while Section 5.1.3 shows the calibration process of some model's parameters: radius of the catchment areas, pricing of the automated shuttle service, time of simulation, initial position of the shuttles inside the catchment areas, vehicle size, and maximum speed. These parameters were calibrated by running many simulations of scaled-down scenarios and with the support of the *BEAM* developers' group. The last part of the Chapter describes the methodology used for analyzing the outputs of the model. In that Section, the authors explain the processes used in the software R (see Section 5.1.4). The results and final remark are respectively reported on Sections 5.2 and 5.3. The results of this study support the wide-scale adoption of on-demand, flexible-route shuttles to address the first/last mile problem and improve accessibility to high-quality public transit throughout Santa Clara County. Communities, particularly those lacking many fixed-route transit services, can look to on-demand shuttles as an alternative approach. Regional and county transportation agencies can use these high-level conclusions as a basis to support more robust scenario testing and

[4] AC Transit Flex Pilot Program: http://www.actransit.org/flex/ (accessed: February 16, 2022)

planning with their official models. Given the beta status of the software and user interface limitations, it is not recommended for practitioners to use “off the shelf,” but owing to its Bay Area calibration, can be applied to other Bay Area counties for high-level planning of similar scenarios.

5.1 The Method

5.1.1 The BEAM model

The major element of the methodology involves applying the travel demand model called *BEAM*, which has been under continuous development for the past two years by the*LBNL*, sponsored by the U.S. Department of Energy (*DOE*)'s Systems and Modeling for Accelerated Research in Transportation (*SMART*) Mobility Consortium. The *BEAM* model[5] adds an important discrete choice modeling extension to *MATSim*, a widely used open-source software for large-scale agent-based and mesoscopic traffic simulations. *BEAM* also incorporates new transportation modes, most notably ride hailing, into human choice modeling, and it includes calculators for energy impacts. *BEAM* offers the ability to simulate a combination of several features critical for this study: defining a new vehicle type representing the automated electric shuttle; assigning the ride hail choice function to the shuttles; establishing geofenced areas throughout a broad area such as a county; and assigning fleets of shuttles to operate only within their specified geofenced areas. *BEAM*, due to its local development origins, is pre-populated with data and calibrated for the San Francisco Bay Area. The data and model parameters are scaled to represent 10% of the population. In order for the 10% sample to generalize to the entire population, the capacity of the road network and public transit vehicle capacities are also scaled down. Therefore, even though only 10% of the population is being simulated, the agents encounter realistic levels of congestion and travel times on the roadways. Agents also encounter realistic levels of crowding on public transit vehicles. The scaling parameters were tuned and calibrated by the *BEAM* development team. The *BEAM* model was perfectly suitable for the case study. The behavior of users in the simulation and the overall results are considered robust because the algorithms in the *BEAM* software, on which the simulation depends, have been validated by the *BEAM*

[5]BEAM: http://beam.lbl.gov (accessed: February 16, 2022)

working group. But, as with all beta-level software, there are a few lessons learned about the use of the software that could be useful for future users:

- *BEAM* requires a powerful computer. For this reason, the research team used a dedicated Linux server from Amazon Web Service (*AWS*) with 256 GB of random-access memory. The computation cost was more than 5 dollars per hour in addition to storage costs.

- The process of installing *BEAM* from the Git repository was quite long and required precise versions of different packages.

- *BEAM* does not have a graphical interface. As a result, it was not always easy to understand what was wrong when the software did not start, how to prepare the input files, and where the output was located.

- This research project represents the first application of the *BEAM* model for the automated shuttle use case and in fact the *BEAM* development team created or updated some functions specifically to aid this study. Thus, the model is still under development: the online documentation is not always updated[6], and there are still open issues to be fixed. In fact, many proceedings required the support of the *BEAM* developers.

- Many of *BEAM*'s internal processes—for example, the cost associated with each alternative during the mode choice process, waiting times, noise pollution and emissions, and so on—are not currently traceable, and the output files are limited to standard attributes.

- The output file is not intuitive, and the analysis required a significant post-processing phase.

- There are still some limitations and minor inconsistencies to be fixed. For example, simulations can currently start only at midnight. Also, if a user uses their private car for the first part of a round trip, they are unconstrained to choose another means of transport for the return trip, which is not typical and could skew the results in favor of non-automobile modes.

[6] BEAM's documentation: https://beam.readthedocs.io/en/latest/users.html (accessed: February 16, 2022)

- The calibration of certain parameters for the Bay Area scenario changed from version to version. The parameters, and changes made to them, are often obscure and not documented. Moreover, updating the version of *BEAM* from the Git repository may yield a version that does not work.

- The software could clearly be used for different scenarios, but there are not enough guidelines for easily using it outside of the Bay Area, since a lot of different input files must be prepared in the right format.

Despite these challenges and drawbacks, the *BEAM* software produced crucial results for the case study. The main benefits encountered by using the software have been as follows.

- The *BEAM* development team has been very available and generously provided support for this research. They fixed some problems that the researchers uncovered, added new functions to the model, helped with the interpretation of parameters, and suggested the best values to use. All the communications occurred in the *BEAM* user group[7], and the *BEAM* developers were very responsive. There is clearly a mutual interest concerning this study, because it represents one of the first practical applications of the *BEAM* model to a case study. It is important for the software developers to receive outputs resulting from the use of their software, as this aids in identifying where to improve the software and what issues should be fixed.

- *BEAM* is one of the few existing software programs that allows for the traffic simulation for a large-scale area and then provides simulation outputs. Moreover, a model defined for the whole Bay Area has been already calibrated and it is ready for use, even if it is still under development: calibration often presents a significant challenge.

- Despite its being a mesoscopic model, *BEAM* takes into consideration a detailed transport network through *OSM*, and the full public transport schedules of all the transportation agencies are pulled from their *GTFS* files. Moreover, the vehicle typologies are very detailed and well calibrated for the area of study, and the model considers

[7]Beam User Group: https://groups.google.com/forum/!forum/beam-model-users (accessed: February 16, 2022)

many transportation modes (bus, ride hail, bike, car, walk) as well as multimodal combinations (walk to transit, ride hail to transit, drive to transit). This means that the model is very flexible for different typologies of macroscopic studies, even if it currently requires the support of the development team.

The total computational and storage cost using *AWS* was on the order of several thousands of dollars, which in addition to the final scenarios included many previous scenario runs for learning about the model behavior and sensitivity testing.

5.1.2 Modeling the Automated Shuttles

For the case study, the authors employed a careful process to select and size the catchment areas based on the literature review. To provide support to the bus network with automated shuttles, the research team inserted into the transportation network fleets of 20 shuttles. Each fleet was bound to provide service within a circular area with a fixed radius. The areas were placed in proximity to high-quality transit stops as a first and last mile supply for the current public transport service. Automated shuttle operational assumptions were principally derived from the Federal Transit Administration's Strategic Transit Automation Research (STAR) Plan, which describes existing EasyMile, Local Motors, and Navya models. Based on these existing models, the theoretical automated shuttles that the team created in *BEAM* have SAE level 4 autonomy (can operate fully autonomously within specified operational domains), travel at low speeds (up to 25 mph), and have a capacity of 12 passengers. The literature review provided additional operational assumptions including average speed (13 km/h), capacity (5–15 passengers/hour), typical hours of operation (6:30am–6:30pm), frequency (15–30 minutes), flexible routing, and on-demand boarding. Both the Caltrain and the Vally Transportation Authority (*VTA*) *GTFS* datasets were used to identify high-quality transit stops. These dataset consist of several text files that describe the entire public transport service, including details such as hours of activities, bus routes, frequency, and stop locations. The high-quality stops were defined as bus stops or transit stations which are visited on average by at least four buses or trains per hour, except for Caltrain, which has lesser frequency but represents major transit stations. Once the high-quality stops were identified, the research team used Aeronautical Reconnaissance Coverage Geographic Information System (*ArcGIS*) Network Analyst to define the size of the catchment areas. The catchment areas were based on 15-minute isochrones from the high-quality stops, with a speed of 25 mph. The catchment areas were then approximated by circles, as required by *BEAM*. Based on this approach, the radius of each catchment area in Santa Clara County was fixed at 3,900 meters as the maximum reachable distance with these conditions. The centers of the catchment areas were placed at the center of the individual isochrone polygons. Initially, the *BEAM* software did not allow ride hail vehicles to be confined to operate only within fixed areas. For this reason, the main intervention of the *BEAM* developers for this research was allowing certain ride hail vehicles to travel only inside a fixed round area, described by the coordinates of the center of the catchment area and its radius. The 218 locations that were defined as the catchment areas for the automated shuttles are represented in Figure 5.1. With 20 vehicles in each catchment area, in all, 4,360 automated vehicles were deployed to serve these areas in the simulation

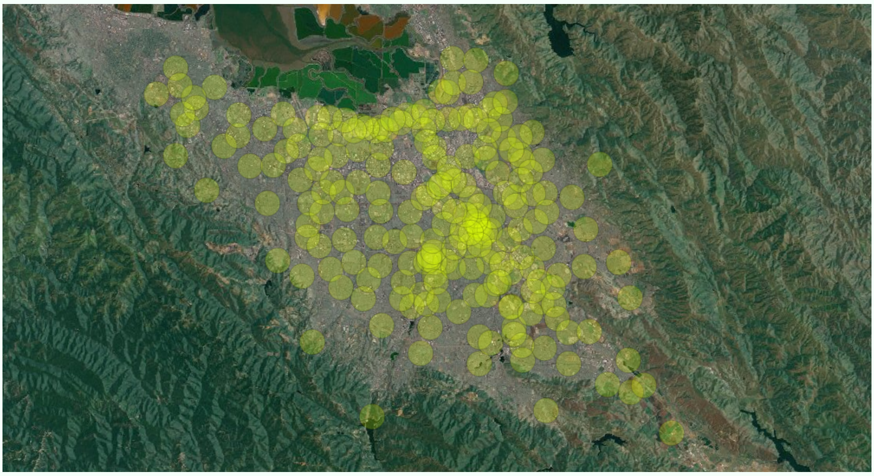


Figure 5.1: Catchment areas for automated shuttles in Santa Clara County

5.1.3 Parameters Calibration

Before running the various scenarios, the authors performed a sensitivity analysis of different parameters of the *BEAM* model to understand the magnitude of impact of various input parameters and to minimize computational time: the initial position of the automated shuttles inside the catchment areas, size of the catchment areas for the automated shuttles, pricing, and number of hours simulated. In a scenario in which the radius of the catchment areas was fixed to 3,000 meters and all of the automated vehicles started at the centers of the catchment areas, 176 trips on the automated shuttle were registered from midnight to 11am — the reader should keep in mind that this was obtained using a 10% population sample. A second scenario showed that when the initial positions of the automated vehicles are changed, trips can increase. In particular, when placing the vehicles randomly inside the relative catchment area rather than the centroid, trips increased by 50% (from 176 to 266). Increasing the size of the catchment areas led to an exponential increase in the number of trips on the automated shuttles. This is because the automated shuttles start replacing other modes of transport for point-to-point trips within the catchment areas: only 40 trips were made with a catchment area radius of 2,000 meters, 989 trips were made when the radius was increased to 6,000 meters, and 5,879 trips were made when the radius was unbounded. The pricing of all vehicles including ride hail vehicles is considered during the mode choice process in *BEAM* using utility values per mode. An increase in the utility value of the ride hail vehicles corresponds to a decrease in price. When the utility was progressively increased, this resulted in a gradual increase in the use of automated shuttles up to the capacity of the system, as expected. However, this application of the *BEAM* model required the automated shuttles to share the same utility value as the other ride hail vehicles. The authors did not have the opportunity to calibrate values for the automated shuttles separate from the values previously calibrated for the other ride hail vehicles. Therefore, the authors devised a method within *BEAM* to differentiate the fares for the two modes by assigning travel time cost only to the automated shuttles and travel length cost only to the other ride hail vehicles, while using the same utility function. Finally, some small-scale scenarios were simulated to test the computational time required for the simulations. The authors found that the computational times related linearly with the number of iterations, number of hours of the simulation, and percentage of simulated population. After the sensitivity analysis, the following parameters were chosen for the main scenarios.

- Simulation from midnight to 8pm, in order to take into consideration the daily activity of users, including morning and afternoon peak hours.
- Use 10% of the real population, since the whole Bay Area model has been calibrated with this percentage of population, and using higher values could compromise the results.
- Automated shuttles with a capacity of 12 people and a maximum speed of 25 mph.
- A fixed radius of 3,900 meters for the catchment areas.
- A supply of 20 automated shuttles per catchment area, intended to represent an unconstrained supply, as the computational cost limited our ability to optimize the number of shuttles to match the estimated demand.
- A random initial position of automated vehicles inside the catchment areas in order to increase the initial accessibility of the shuttles.
- A lower average cost for automated shuttle trips compared to ride hail trips, made possible by the method described above. This reflects an assumed policy of subsidized automated shuttle fares similar to regular public transit fares and market-rate, distance-based fares for the classic ride hail services.

5.1.4 Outputs Elaboration

Two different main scenarios were simulated: the baseline scenario of the entire San Francisco Bay Area without automated shuttles and the same scenario with the automated shuttles deployed in Santa Clara County. Each scenario required 20 iterations and 12 hours of computational time in order to reach an equilibrium of the outputs. The main outputs provided by *BEAM* at the end of the simulation are stored in a table with than 11 million event records concerning the simulation. The main challenge in processing the output data arises due to the fact that the most useful information is categorized into four types of events that must be correlated back to each other to discern each person's whole trip across vehicles, necessary for reporting the transportation statistics shared in the Findings. The event types are: ModeChoice events, PathTraversal events, and PersonEntersVehicle and PersonExitsVehicle events. The ModeChoice event records contain information about the overall trip of each user (with person ID, time of departure, chosen mode of transport, and total length being the most important details for this analysis), excluding the paths the user takes throughout their trip. Instead, the PathTraversal event records contain information about the paths, or portions thereof, of each vehicle (with type of vehicle and vehicle ID, trip length and duration, geographic coordinates of departure and arrival points, type of fuel, and energy used being the most important). For example, regarding the buses, there is a PathTraversal event between every pair of bus stops traveled between by each person on the bus. The PersonEntersVehicle and PersonExitsVehicle events relate the individual users to specific vehicles by recording the time when each person enters and leaves a vehicle. Therefore, each mode choice event contains one or more vehicle-path events and has to be merged with the other event types in order to provide complete trip information for each user and for each vehicle. The variables that are available to match the four sets of events are the person ID, vehicle ID, departure time, and arrival time. First, we merged information about the PathTraversal events with the PersonEntersVehicle and PersonExitsVehicle events in order to associate the 'ID person' attribute to the PathTraversal events. The result of this merger was a VehicleTrips table. This table was then aggregated in order to have one unique row for each vehicle used by each person. The aggregated table was moved to the QGIS platform in order to associate each vehicle trip to a starting and ending county. For this purpose, the team used the shape file from the California Open Data Portal[8], which describes the cartographic boundaries of the nine Bay Area counties. With a tool within QGIS that joins attributes by location, the authors associated a county with the starting and ending points of each vehicle trip. Finally, with the person ID and departure time attributes, the VehicleTrips table was matched with the ModeChoice events in order to obtain information about the total duration and energy of each trip, the time at which the automated shuttles were used for ride hail trips, and the starting and ending county of each trip. At the end of the aggregation process, two main tables useful for further analysis

[8]California Open Data Portal: https://gis.data.ca.gov/ (accessed: February 16, 2022)

were built.

- Trip table: each row represents a full person trip, indicating the total length, time and energy used, starting and ending county, departure time, and the transportation mode used. Multimodal trips are included in the model, and the possible modes or modal combinations available are 'Bike,' 'Car,' 'Drive and Transit (meaning a multimodal trip with driving plus transit),' 'Ride Hail,' 'Ride Hail AV ,' 'Ride Hail Pooled,' 'Ride Hail Pooled AV,' 'Ride Hail and Transit (a multimodal trip with ride hail plus transit),' 'Ride Hail AV and Transit,' 'Walk,' and 'Walk and Transit (a multimodal trip with walking and transit).' The modes with 'AV' refer to trips that include a ride hail on the automated shuttle vehicle fleet. These AV trips are distinguished from the other ride hail systems (e.g., Lyft and Uber).

- Vehicle paths table: each row represents a full vehicle trip completed by a person, indicating the vehicle type used, fuel type, total length, time and energy used, the starting and ending county, and the departure time. The vehicle types are 'AVshuttle,' 'Bike,' 'Bus,' 'Car,' 'Rail,' 'Ride Hail,' 'Tram,' and 'Walk.' The merging process was fully realized in the R platform and required about 10 hours per scenario, while the QGIS tool required one hour per scenario.

5.2 Results and Findings

In the baseline scenario, as well as the scenario with the automated shuttles, about 627,000 trips were performed by the 10% population sample of the Bay Area (see Table 5.1). Of these trips, 30% either started or arrived in Santa Clara County. Figure 5.2 shows the symmetry of the trips with respect to noon and the heat map of both the home and work activities of Santa Clara inhabitants. Moreover, 37% of the trips involving Santa Clara County started or ended in another county (see Table 5.2). Alameda County accounted for 48.7% of the cases and San Mateo County for 34.5%. Figure 5.3 represents a heat map of the departure point of trips involving Santa Clara County: that is, trips inside, from, or to Santa Clara County.

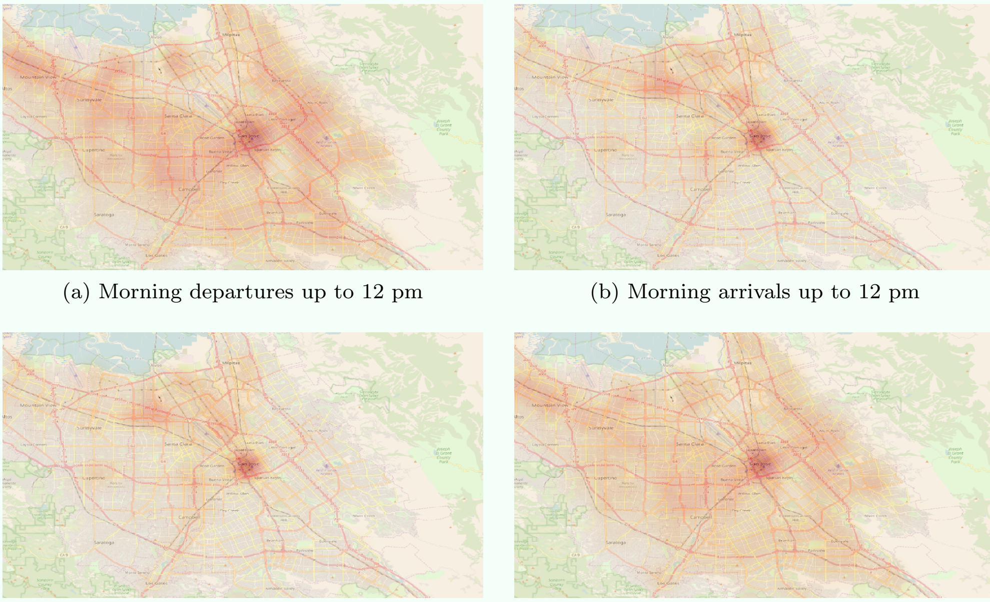
Analyzing the trip mode share, in the baseline scenario the vast majority of trips were performed with private cars (see Table 5.3). Trips involving Santa Clara County represent a higher share of car trips with respect to

	Bay Area	Involving SCC	Inside SCC
Number of trips	626,949	190,018	119,800
Total length [km]	17,583,796	5,190,390	1,951,887
Total time [h]	285,661	70,736	29,567
Total energy [GJ]	39,512,627	10,441,505	4,179,432

Table 5.1: Statistics of trips carried out in the baseline scenario

	From SCC	To SCC	From SCC [%]	To SCC [%]
Alameda	17,054	17,173	48.88	48.6
Contra Costa	3,012	3,179	8.63	9
Marin	225	242	0.64	0.68
Napa	61	71	0.17	0.2
San Francisco	1,989	1,937	5.7	5.48
San Mateo	12,039	12,187	34.51	34.49
Solano	313	326	0.9	0.92
Sonoma	193	217	0.55	0.61

Table 5.2: Statistics of trips departing from or arriving to Santa Clara County



(a) Morning departures up to 12 pm

(b) Morning arrivals up to 12 pm

(c) Afternoon departures after 12 pm

(d) Afternoon arrivals after 12 pm

Figure 5.2: Heat Map of departures and arrivals by time of day

the whole Bay Area, particularly for exchange trips with the other counties. Likewise, a lower share of trips involving Santa Clara County were made using sustainable means of transport: bike, transit, and walk. The addition of the automated shuttles did not produce a meaningful change of the modal share (see Table 5.4). Even though a total of 1,072 trips on the shuttles was registered, that theoretically corresponds to about 10,720 trips by considering the whole population. These trips replaced to a small extent all the other modes, but particularly the trips inside Santa Clara County performed by the existing ride hail system, some walking trips, and car trips that were intended to reach bus stops. Trips involving the use of the automated shuttles represent 0.5% of all the trips involving Santa Clara County.

Focusing only on the trips involving Santa Clara County that were most affected by the presence of the automated shuttles, Table 5.5 presents the

	Bay Area [%]	Involving SCC [%]	Inside SCC [%]	From SCC [%]	To SCC [%]
Bike	1.99	0.99	1.31	0.45	0.44
Car	77.1	88.79	87.06	91.81	91.68
Drive Transit	5.79	2.65	3.49	1.12	1.31
Ride Hail	3.34	3.03	3.07	3	2.92
Ride Hail Pooled	0.39	0.26	0.13	0.48	0.45
Ride Hail Transit	0.33	0.2	0.12	0.26	0.39
Walk	2.97	1.07	1.47	0.39	0.39
Walk Transit	8.07	3.02	3.35	2.49	2.41

Table 5.3: Mode share of trips involving Santa Clara County in the baseline scenario

	Bay Area [%]	Involving SCC [%]	Inside SCC [%]	From SCC [%]	To SCC [%]
Bike	2.05	1.09	1.46	0.45	0.44
Car	77.13	88.79	87.05	91.83	91.69
Drive Transit	5.77	2.61	3.43	1.09	1.34
Ride Hail	3.09	2.4	2.19	2.79	2.72
Ride Hail *AV*	0.08	0.27	0.4	0.03	0.03
Ride Hail Pooled	0.46	0.42	0.4	0.49	0.45
Ride Hail Pooled *AV*	0.01	0.03	0.05	0.01	0
Ride Hail Transit	0.3	0.06	0.03	0.09	0.12
Ride Hail Transit *AV*	0.07	0.23	0.2	0.2	0.35
Walk	2.9	0.98	1.34	0.39	0.36
Walk Transit	8.13	3.13	3.46	2.62	2.5

Table 5.4: Mode share of trips involving Santa Clara County in the *AV* Scenario

total length, duration, and energy for each different mode in the baseline scenario, and Table 5.6 reports on the automated shuttle scenario. The insertion of automated shuttles did not cause a large change in these values. The total length of trips including the automated shuttles was 15,800 km, for a cumulative trip duration of 347 hours and 32,404 GJ of energy used.

By analyzing the same attributes, this time separating out the trips by vehicle type, it is possible to observe that the use of both bus and tram increased in the automated shuttle scenario (see Table 5.8) with respect to the baseline scenario (see Table 5.7). In the automated shuttle scenario, the automated shuttles traveled for more than 6,000 km in 164 hours, registering an average speed of 39 km/h (24 mph), close to their maximum speed of 40 km/h (25 mph). The average trip distance on automated shuttles was almost 6 km, and the average trip duration was 10 minutes. For cars, the most widely used vehicle, the average speed was 84 km/h (52 mph), and the average trip duration was 20 minutes. Moreover, the multimodal trips consisting of automated shuttle plus either a bus or a tram were on average almost one hour long, suggesting that people are willing to accept a transfer between an automated shuttle and the public transport service only if the trip is long enough.

Funded by the U.S. Department of Energy, *BEAM* produces as one of its core outputs the energy use of the entire simulated transportation system, given inputs about fuel use by vehicle type. It reports energy in units of GJ, enabling comparison across fuel types; lower numbers indicate less energy used. Note the outputs do not compare elements of total life cycle energy cost such as impact of vehicle and battery manufacturing, fossil fuel extraction and processing, fuel mix for electricity generation, etc. A journalistic discussion of recent relevant studies about electric vehicle contributions to climate change reduction is presented by Carbon Brief[9] and an analysis of 59 world regions by Knobloch et al. (2020) indicates that emissions related to electric vehicles are less intensive than fossil fuel alternatives in 53 regions that represent 95% of global transport demand (see Knobloch et al. (2020)). In the United States, the U.S. Department of Energy Alternative Fuels Data Center provides a tool to compare well-to-wheel CO2-equivalent emissions across fuel types on a state-by-state basis, demonstrating that nationally and in all except a handful of states, emissions of all-electric vehicles are

[9]Carbon Brief - Factcheck: How electric vehicles help to tackle climate change: https://www.carbonbrief.org/factcheck-how-electric-vehicles-help-to-tackle-climate-change (accessed: February 16, 2022)

	Trips	Trips [%]	Length [km]	Length [%]	Time [h]	Time [%]	Energy [GJ]	Energy [%]
Bike	1,885	0.99	24,344	0.47	1,378	1.95	260	0
Car	168,724	88.79	4,775,980	92.02	57,516	81.31	8,557,083	81.95
Drive Transit	5,030	2.65	69,375	1.34	1,635	2.31	360,754	3.46
Ride Hail	5,756	3.03	132,236	2.55	1,619	2.29	233,767	2.24
Ride Hail Pooled	485	0.26	24,321	0.47	275	0.39	43,138	0.41
Ride Hail Transit	371	0.2	13,328	0.26	249	0.35	35,677	0.34
Walk	2,035	1.07	18,835	0.36	3,863	5.46	286	0
Walk Transit	5,732	3.02	131,972	2.54	4,200	5.94	1,210,541	11.59

Table 5.5: Statistics of trips involving Santa Clara County in the baseline scenario per transport mode used

	Trips	Trips [%]	Length [km]	Length [%]	Time [h]	Time [%]	Energy [GJ]	Energy [%]
Bike	2,066	1.09	25,741	0.5	1458	2.05	275	0
Car	168,837	88.79	4,779,468	92.01	57,787	81.29	8,564,277	81.87
Drive Transit	4,967	2.61	69,636	1.34	1,627	2.29	357,948	3.42
RH	4,565	2.4	119,953	2.31	1,490	2.1	211,789	2.02
RH *AV*	507	0.27	2,871	0.06	74	0.1	3,854	0.04
RH Pooled	806	0.42	26,165	0.5	364	0.51	46,129	0.44
RH Pooled *AV*	57	0.03	449	0.01	12	0.02	603	0.01
RH Transit	108	0.06	4,254	0.08	91	0.13	13,794	0.13
RH Transit *AV*	435	0.23	12,485	0.24	261	0.37	27,948	0.27
Walk	1867	0.98	17,426	0.34	3,569	5.02	243	0
Walk Transit	5,947	3.13	136,061	2.62	4,355	6.13	1,233,661	11.79

Table 5.6: Statistics of trips involving Santa Clara County in the *AV* scenario per transport mode used - RH = ride hail

	Trips	Trips [%]	Length [km]	Length [%]	Time [h]	Time [%]
AV shuttle	n/a					
Bike	1,949	0.92	24,563	0.48	1,385	1.98
Bus	10,328	4.89	78,518	1.52	2,708	3.87
Car	172,956	81.81	4,806,759	93.08	56,896	81.50
Rail	1,803	0.85	55,800	1.08	1,010	1.44
Ride Hail	6,539	3.09	159,357	3.09	1,841	2.63
Tram	1,856	0.88	13,296	0.26	495	0.71
Walk	15,979	7.56	25,601	0.50	5,488	7.85

Table 5.7: Statistics of vehicles involving Santa Clara County in the baseline scenario per vehicle

	Trips	Trips [%]	Length [km]	Length [%]	Time [h]	Time [%]
AV shuttle	1,072	0.51	6,381	0.12	164	0.23
Bike	2,138	1.01	25,987	0.50	1,465	2.08
Bus	10,525	4.97	79,743	1.54	2,772	3.95
Car	173.012	81.63	4,809,713	93.09	57,245	81.52
Rail	1,907	0.90	59,456	1.15	1,076	1.53
Ride Hail	5,310	2.51	146.985	2.84	1,691	2.4
Tram	1,978	0.93	14,052	0.27	523	0.74
Walk	16,008	7.55	24,575	0.48	5,289	7.53

Table 5.8: Statistics of vehicles involving Santa Clara County in the automated shuttle scenario per vehicle

lower than those of other fuel types: plug-in hybrid, hybrid, and gasoline[10]. Table 5.9 shows the total length, duration, and energy used for each type of fuel for the trips involving Santa Clara County. Gasoline, diesel, and biodiesel clearly dominate compared to electricity. The presence of automated shuttles replaced almost 1,000 gasoline trips (8,300 km and 15,600 GJ) by increasing at the same time the usage of diesel vehicles due to the passengers' shift to transit (see Table 5.10). Since the transit vehicles would have operated anyway, this increase pertains to the accounting only and is not related to an increase of pollutants.

Figure 5.4 represents the distribution of departures of the automated shuttles together with cars, other ride hails, and public transport services. It is evident that the automated shuttle service was used particularly during the night hours. In order to analyze the spatial distribution of the starting points of trips on the automated shuttles, Figure 5.5 represents the starting points superimposed on an *OSM* of Santa Clara County. The Figure shows that most trips are concentrated in downtown San José, but they are also distributed across the whole county. Regarding vehicle occupancy, there were only a few pooled trips on the automated shuttle. One limitation of *BEAM* is that it currently does not allow agents to choose a trip consisting of a pooled ride hail leg followed by a transit leg. (Agents can choose a trip consisting of a single-passenger ride hail leg connecting to a transit leg). If pooled trips to transit were allowed, we might see an increase in pooled trips and a decrease in single-passenger trips. Allowing pooled trips to transit would also effectively increase the capacity of the *AV* system without adding any vehicles. However, we don't believe the *AV* capacity was a limiting factor in the simulation because of the large number of AVs deployed.

5.3 Final Remarks

The main goal of the present study was to evaluate the impacts of a fleet of automated shuttles in Santa Clara County providing first and last mile connections to the current public transportation system. The secondary goal was to use and test the *BEAM* software developed at *LBNL*. *BEAM* was

[10] U.S. Department of Energy, Energy Efficiency Renewable Energy Program, Alternative Fuels Data Center. Emissions from Hybrid and Plug-In Electric Vehicles: https://afdc.energy.gov/vehicles/electric$_e$$missions.html(accessed: February16, 2022)$

Fuel Type	Trips	Length [km]	Duration [h]	Energy [GJ]	Trips [%]	Length [%]	Duration [%]	Energy [%]
Biodiesel	8,961	70,440	2,399	1,270,977	4.24	1.36	3.43	12.30
Diesel	5,035	117,109	1,954	231,157	2.38	2.27	2.79	2.24
Electricity	2,191	22,076	599	3,878	1.04	0.43	0.86	0.04
Food	15,979	25,601	5,488	632	7.56	0.50	7.85	0.01
Gasoline	179,244	4,928,668	59,474	8,824,334	84.79	95.44	85.07	85.42

Table 5.9: Statistics of trips involving Santa Clara County in the baseline scenario by fuel type

Fuel Type	Trips	Length [km]	Duration [h]	Energy [GJ]	Trips [%]	Length [%]	Duration [%]	Energy [%]
Biodiesel	9,142	71,185	2,446	1,284,419	4.31	1.38	3.48	12.41
Diesel	5,155	121,358	2,033	240,929	2.43	2.35	2.89	2.33
Electricity	3,390	29,397	795	12,537	1.60	0.57	1.13	0.12
Food	16,008	24,575	5,289	597	7.55	0.48	7.53	0.01
Gasoline	178,255	4,920,378	59,662	8,808,704	84.10	95.23	84.96	85.13

Table 5.10: Statistics of trips involving Santa Clara County in the automated shuttle scenario by Fuel Type

selected for use in this study because of its ability to simulate the combination of several critical features: defining a new vehicle type approximating the automated electric shuttle form factor and performance; assigning ride hail functions to the new shuttles; establishing geofenced areas throughout a broad area such as a county; assigning the shuttles to specific geofenced areas. *BEAM* demonstrated the capability to simulate the proposed automated shuttle service, and this report describes the lessons learned from using the model as well as suggestions for future users. Regarding the Santa Clara County scenario, the model suggests that automated shuttles would be chosen by some users for various trip types including first and last mile connections to transit. The model estimated 1,100 trips in one simulated day carried out by a 10% population sample. The trips were distributed across the county, though they were particularly concentrated in downtown San José. The automated shuttles were confined within circular catchment areas centered on major bus stops and transit stations. The average trip length was 6 km, with an average speed of 35 km/h, due to the low assumed maximum operating speed of 40 km/h. Users of the automated shuttles shifted mainly from gas trips — principally from other ride hail vehicles directed to bus stops — but also walking trips. The usage of public transport services increased with the presence of the automated shuttles. The choice to use an automated shuttle plus either a bus or tram was found to happen most frequently for trip duration of approximately one hour. Moreover, considering the higher usage of the automated shuttles during the night hours—mainly between midnight and 2am— as well as the increase of the usage of public transport and replacements of walking trips, it's possible that automated shuttles serve to increase the accessibility to users for public transport and offer a good solution in certain contexts. Some problems encountered in this study and possible solutions are discussed below. Future research could provide further insight into the potential use of the automated shuttles for connectivity.

- The *BEAM* model currently does not include an option for users to ride a pooled ride hail vehicle to connect to transit; instead only allowing single-passenger ride hail trips.The addition of this option could be crucial for the use of the automated shuttles, since they offer a capacity of 12 people.

- The center of each catchment area does not exactly align with the coordinates of the high-quality transit stops. The centers were placed to best approximate the 15-minute isochrones around the high-quality transit stops. This might have decreased the efficiency of the connections between the automated shuttles and the current public transport network.

- The coefficients of the utility function related to the automated shuttles could be adjusted to be different with respect to the already-calibrated coefficients associated with the standard ride hail system. Treating the automated shuttles as a separate mode would allow the more flexible use of coefficients to test modal assumptions and policy interventions of the shuttles, such as different pricing schemes, separately from the standard ride hail system.

- The use of a non-fixed radius could improve accuracy, based on the individual isochrones and tailored to the attractiveness of specific transit stops/stations.

- *BEAM* could impose a constraint requiring that both legs of a round trips should use the same modes. For example, if a user drives their car to a bus stop, they should arrive at this bus stop and reuse their car on the return trip. Conversely, if a user drives their car to their end destination, they should also reuse their car on the return trip. Imposing this constraint would make the individual agent-level modal choices more realistic.

- The number of automated shuttles supplied can be weighted based on sociodemographic attributes of catchment areas around high-quality stops, or otherwise optimized to match demand, which would give a better indication of fleet size requirements and cost implications.

- The model considers just 10% of the population. In future it will be possible to consider the whole population, since recently the entire Bay Area population has been made available for use with the model. Considering the whole population could allow a better estimate of pooled trips as well as reveal meaningful correlations between trip-making and land use characteristics such as density. Currently, there are too few simulated automated shuttle trips to draw a conclusion, and the sampled population changes between model runs. Clearly, simulating the whole population will require much greater computational resources for both the simulation and the output analysis.

- The model is calibrated to the relatively recent year 2015, but even since then, dramatic changes have occurred in observed travel behavior. These changes have included exponential increases in ride hail activity accompanied by modest declines in public transit use, the sudden introduction of completely new shared micro mobility services in many cities across the Bay Area, and most recently, in 2020, profound disruption of various transportation modes (particularly public transit) due to the COVID-19 pandemic. Calibration with sufficient data is a challenge for any model, and it is particularly important for the purpose of estimating new modes and scenarios.

The main benefit of this work has been demonstrating that the *BEAM* model, despite its beta status, limitations, and difficulty to use, allowed for the analysis of a large-scale scenario in a quite high level of detail. As the *BEAM* model continues to be developed, it will become more robust and capable in the future. The *BEAM* developers have been very open to inserting new tools, parameters, and functions to adapt *BEAM* to this case study. Moreover, the *BEAM* model is very flexible for analyzing different interventions on the baseline scenario and could clearly be adapted to a multitude of different studies.

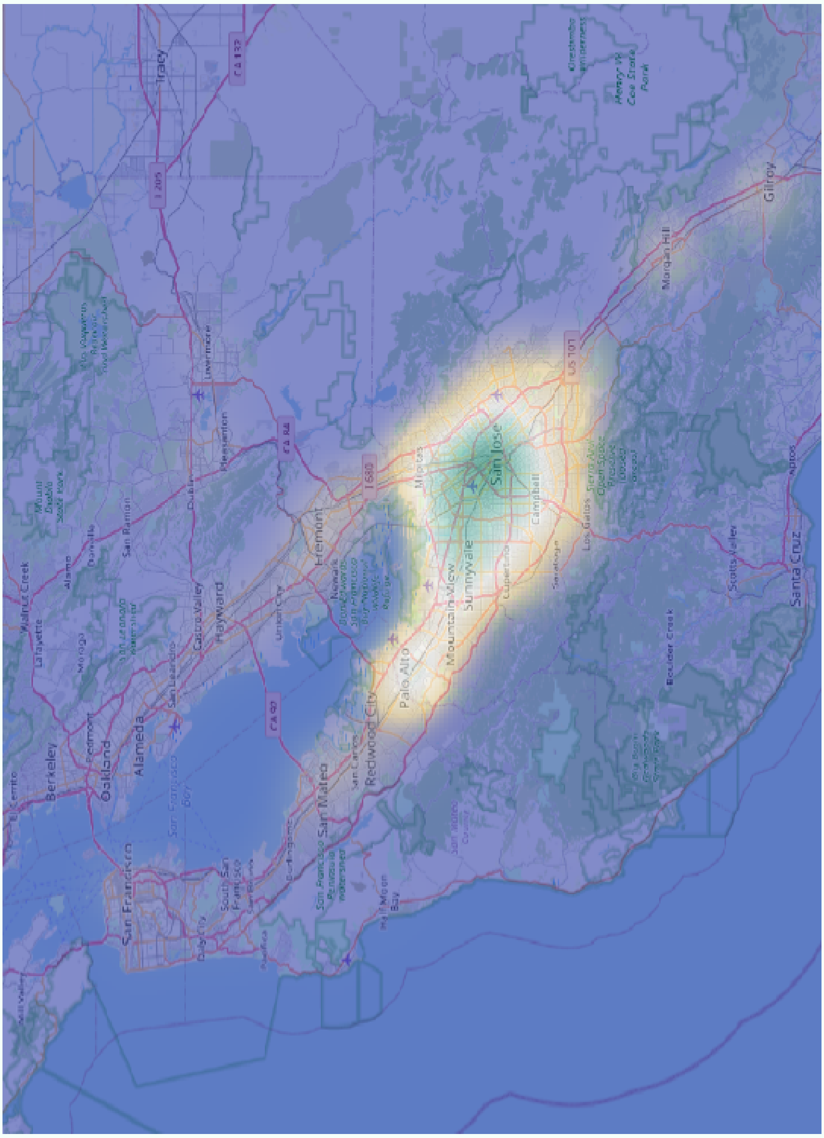


Figure 5.3: Heat map of trips involving Santa Clara County

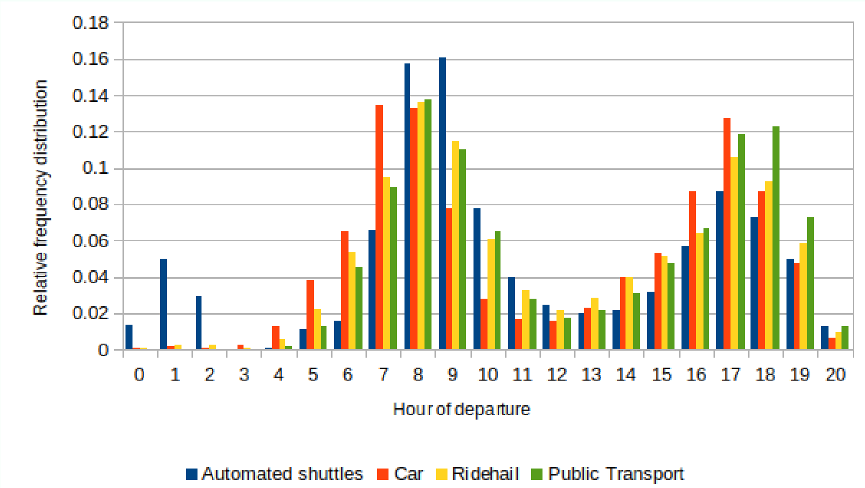


Figure 5.4: Relative frequency distribution of departure time of automated shuttles during the day, compared with the usage of cars, other ride hails, and Public transport service

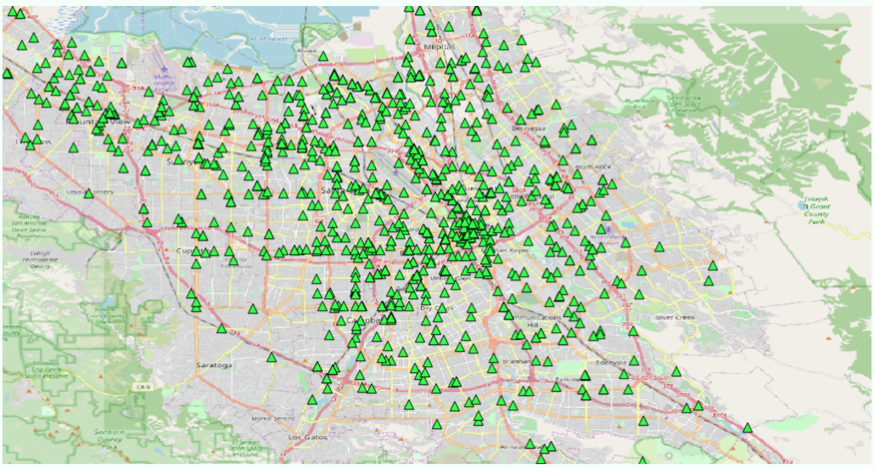


Figure 5.5: Spatial distribution of departures of automated shuttles' trips in Santa Clara County when using a 10% population sample

Chapter 6

Microscopic Transport Digital Twin of Bologna, IT

This application has the aim to, at least partially, fill the above-mentioned research gap (see 1.1), 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 online[1]. 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. This application has been published on the MDPI

[1]Github. SUMOPy: https://github.com/schwoz/sumopy/ (accessed: February 16, 2022)

journal (see Schweizer et al. (2021)).

6.1 The Model

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[2]. This Section explains how the data has been processed to represent the supply and demand of the transport systems using the *SUMOPy*[3]/*SUMO*[4] simulation suite. 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 (see Grimm et al. (2010)), 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 6.1.1 and 6.1.2 describe the transport supply of road and public transport, Sections 6.1.3 and 6.1.4 explain the data preparation of ODMs and *GPS* traces, Sections 6.1.5 and 6.1.6 explain the external and internal demand created from ODM and *GPS* data and Section 6.1.7 explain the calibration of te used plan choice model.

6.1.1 The Road Network

The road network of Bologna city has been converted from *OSM* in a *SUMO* XML format by SUMO's *NETCONVERT*[5] program and edited manually with SUMO's *NETEDIT*[6] 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,

[2]Dati ISTAT: http://dati.istat.it/ (accessed: February 16, 2022)

[3]SUMOPy: https://sumo.dlr.de/docs/Contributed/SUMOPy.html (accessed: February 16, 2022)

[4]SUMO: https://www.eclipse.org/sumo/ (accessed: February 16, 2022)

[5]SUMO's Netconvert: https://sumo.dlr.de/docs/netconvert.html (accessed: February 16, 2022)

[6]SUMO's Netedit: https://sumo.dlr.de/docs/netedit.html (accessed: February 16, 2022)

see Schweizer et al. (2016a) 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 6.1. The metropolitan area of Bologna covers a wider area of 3703 km^2 , see Figure 6.1. Figure 6.1 also shows the *TAZs* of the core area and the metropolitan area. The *TAZs* are derived from the 2001 national population census[7]. 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.

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 6.1.6. In addition, on-street parking lots have been created with some heuristics along roads with at least two lanes and road priority below eight.

[7]Censimento Popolazione e Abitazioni. 2001: https://www.istat.it/it/archivio/3847 (accessed: February 16, 2022)

6.1.2 The Public Transportation Service

The entire public transport (PT) provided by the local operator Trasporto Pubblico Emilia Romagna (*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 Schweizer et al. (2016a). 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 6.1.6. After this import procedure the ID, name, stop sequence, route, and service frequency of 234 bus lines are present in the scenario.

6.1.3 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 6.1.5 and 6.1.6. 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 6.1. 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[8]. The OD matrices have been updated to the year 2018 by considering the population increase

[8] Dati ISTAT: http://dati.istat.it/ (accessed: February 16, 2022)

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 6.1.6; 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 6.1.5.

6.1.4 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 geofenced 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 (see Rupi et al. (2019)) 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.

6.1.5 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 $f_o d$ from a zone of origin o to a zone of destination d has been used to generate $f_o d$ trips, between those zones; the first and last link of the $f_o d$ 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 6.1.4) 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.

6.1.6 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 (see Schweizer et al. (2018)). 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 6.1.7. 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 6.1.4) 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 6.1. Note that despite the different data sources, the mode share of the population is similar to the official statistics obtained from the *SUMP* of Bologna[9].

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

[9] Osservatotio PUMS: https://www.osservatoriopums.it/bologna (accessed: February 16, 2022)

Transport strategy	N° of people per strategy 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%	13,923.0	31,260
Bicycle	2,424	4.26%	7.20%	20,310.0	22,734
Bus	17,557	30.89%	27.70%	39,280.0	56,837
Scooter	6,199	10.91%	11.40%	5,168.0	11,367
Walking	13,320	23.44%	23.00%	12,467.0	25,787
Total	56,837	100.00%	100.00%	91,148.0	147,985

Table 6.1: Statistics of the synthetic population with share of preferred modes (M_s) and observed mode share O_s provided by the *SUMP* of Bologna[10]. 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 6.1.7

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[11] 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.

6.1.7 Calibration of a 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[12], as stated by Grimm et al. (2005): 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

[11] SUMO. Demand/Dynamic User Assigment: https://sumo.dlr.de/docs/Demand/Dynamic$_User_Assignm$
$February 16, 2022$)

[12] I Numeri di Bologna Metropolitana. Il Parco Veicolare di Bologna al 31.12.2017: http://inumeridibolognametropolitana.it/studi-e-ricerche/il-parco-veicolare-di-bologna-al-31122017 (accessed: February 16, 2022)

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 6.2; the iterations of plan (re)-selection and simulating are continued until all plans of all people have been simulated at least once.

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 6.2. 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.) - see Cascetta (2001). 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:

$$U_{s,i} = \alpha_s - \beta T_{s,i} \tag{6.1}$$

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 6.1), 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.

$$z = \sum_{s=1}^{5} \mid M_s - O_s \mid \tag{6.2}$$

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 Hansen (2011): 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 $p_j = [\alpha 2 \ldots \alpha 5]$ 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 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 6.2.

6.2 Results and Discussions

6.2.1 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 Section 6.1.7. For the present study, the value of time is assumed to be $\beta = 0.07$ €/min (see Cascetta (2001)) and the parameter space for all four parameters has been limited to the interval (-5, 5). Figure 6.3 shows that a good convergence has been achieved after 4,000 iterations, highlighting that the

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

Table 6.2: Calibrated parameters of utility function (Equation 6.1) of mode share model by minimizing objective function z

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).

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 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.

6.2.2 Microsimulation Results and and Model Validation

After each person has chosen the plan with the highest utility (with utility from Equation 6.1 and the calibrated parameters from Table 6.2), 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 6.4.

The maximum flows of 2,500 vehicles/h per lane on Figure 6.4 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 1,000 to 1,500 vehicles/h are also realistic. In order to validate the simulation, the simulated flows shown in Figure 6.4 are compared to the flows measured by induction loop based detectors, scattered around the city, as shown in Figure 6.5. 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. The plot of link-flows from the microsimulation run over the link-flows from the detectors is shown in Figure 6.6, 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 (see Kang and Aldstadt (2018)). 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: 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 6.6. 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 $1.4710^{-}91$ for the slope and $8.5310^{-}7$ for the intercept. According to literature (see Cascetta (2001)), 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 Cascetta (2001) 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 (see Tawfeek et al. (2018)). In brief, the GEH of link i is determined by

$$GEH_i = \sqrt{\frac{(f_i - f_i^*)^2}{0.5(f_i - f_i^*)}} \tag{6.3}$$

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 6.3 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.

6.3 Final Remarks

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 (see Roulland et al. (2015), Wilson (2018), Cuauhtemoc et al. (2017) and Croce et al. (2019)) 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 (see

Road Link Type	# Links	Slope m	Intercept (veh/h)	R2	$GEH < 5$	$5 < GEH < 10$	$10 < GEH$
$0m < width < 5m$	128	0.87	157.22	0.34	29%	30%	41%
$5m < width < 7m$	203	0.83	150.94	0.51	33%	27%	40%
$width > 7m$	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%

Table 6.3: Flow comparison data and indicators for different road widths and number of lanes

Maciejewski (2010)). 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 Traffic Control Interface (*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.

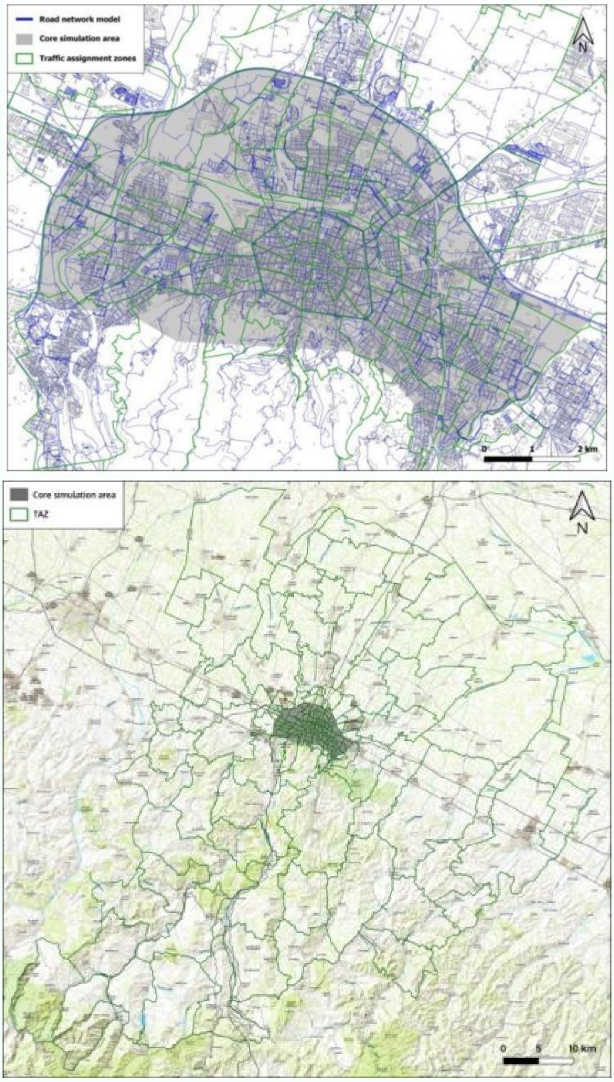


Figure 6.1: (up) Core simulation area: city of Bologna, 50 km^2; (down) Entire simulation area with extra-urban areas, 3700 km^2. *TAZs* in green

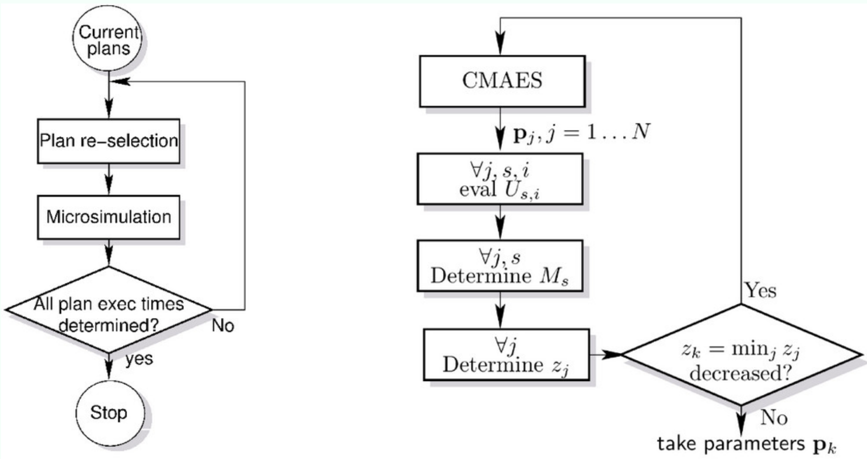


Figure 6.2: Application of the plan choice calibration model: (left) Scheme of algorithm that calculates the plan execution time of all plans while maintaining the official modal split. (right) Iteration of CMASE minimization algorithm; determination of objective function z_j and choice of final parameter vector p_k.

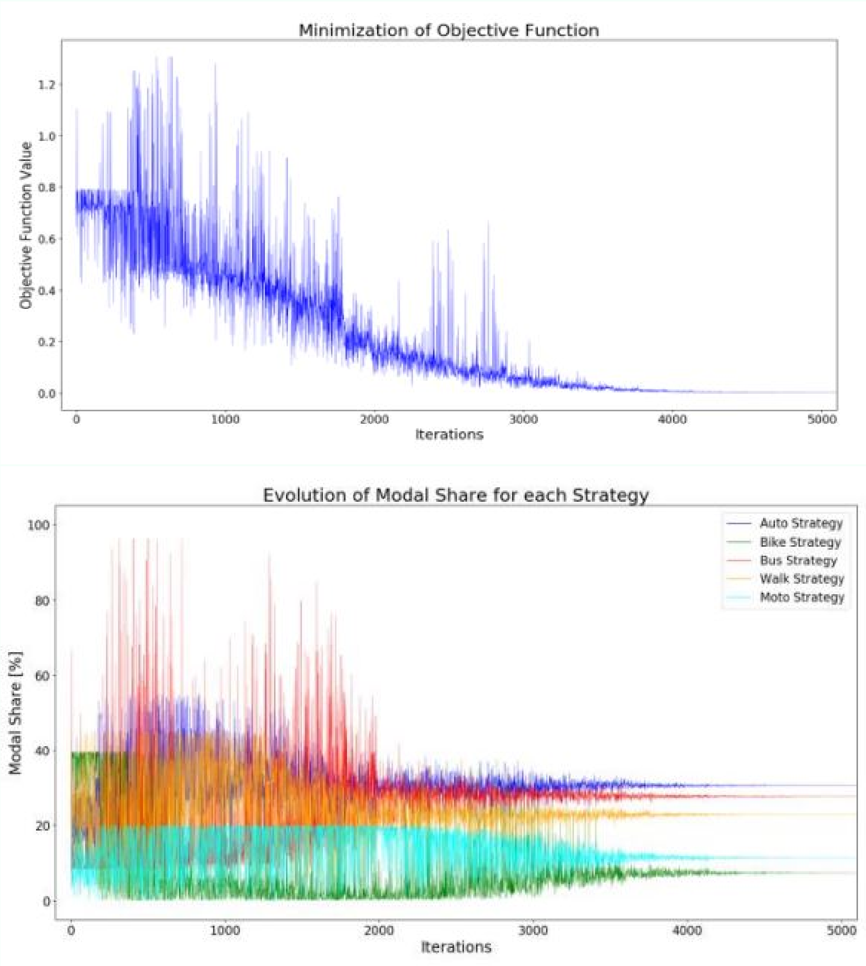


Figure 6.3: Convergence of calibration process: (up) Convergence of objective function z over iterations. (down) Calibrated mode share M_s over iterations, converging to the official mode share, as shown in Table 6.2.

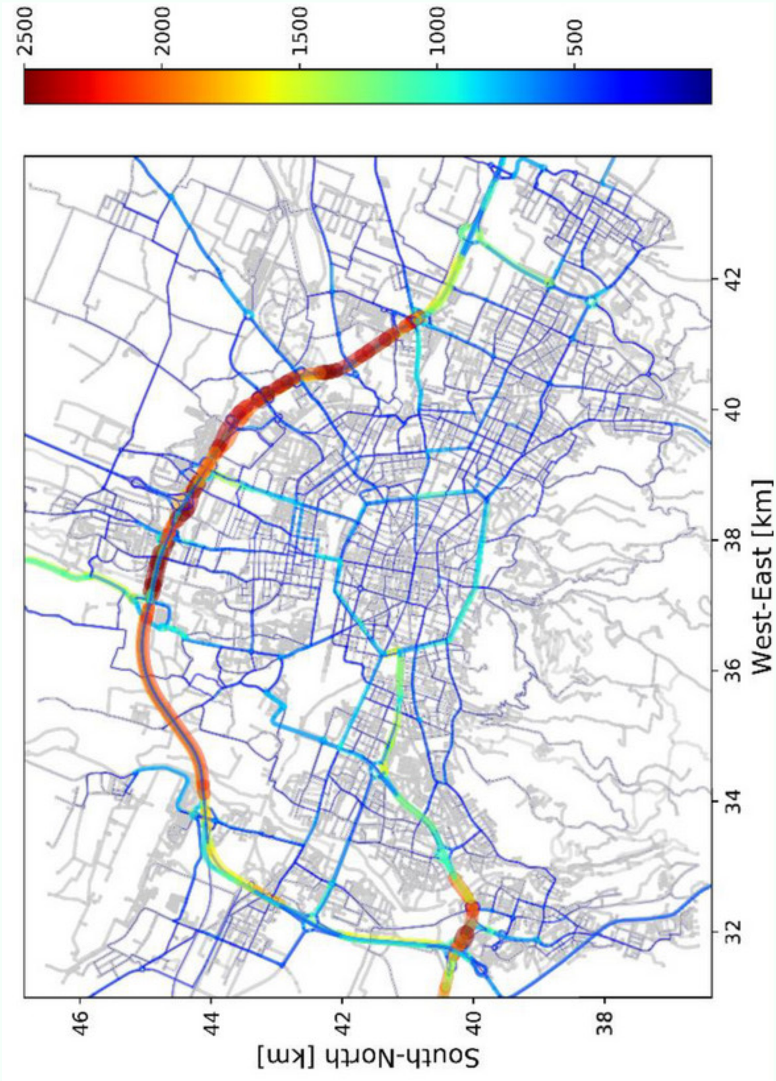


Figure 6.4: Measured simulated flows in the core simulation area as number of vehicles entered into a link in 30 min.

Figure 6.5: Location of flow detectors (cyan circles).

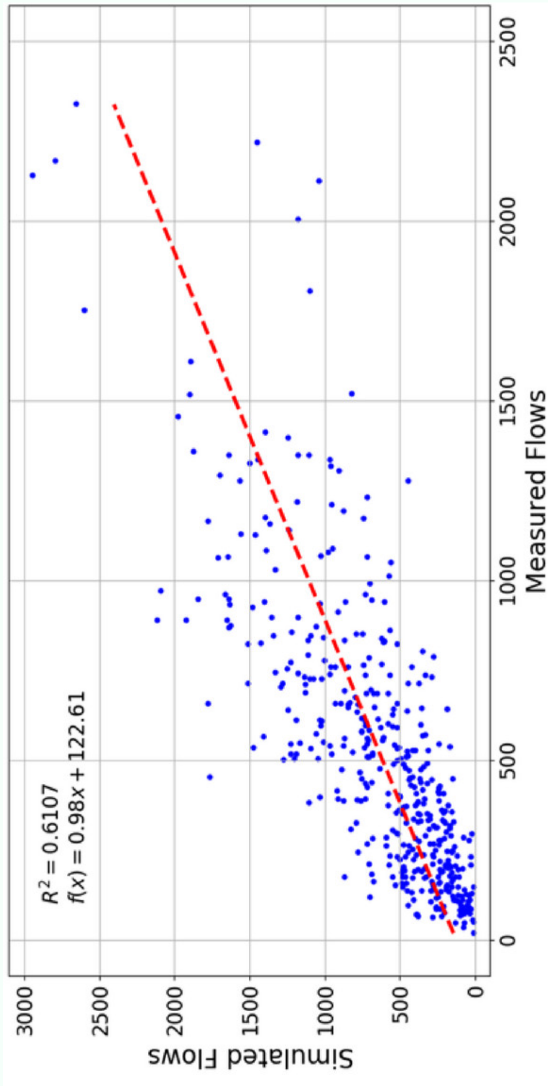


Figure 6.6: Simulated flows over measured (observed) flows in vehicles per hour.

Chapter 7

Conclusions

The three years passed as a PhD student has been plenty of research activities together with Prof. Federico Rupi, Prof. Joerg Schweizer, and external collaborators, without forgetting the beautiful research experience in California, which fortunately was not stopped by the COVID-19 pandemic, during which I could enrich my research activity as well as experiment American work and living culture.

The main intellectual contributes are reported on Section 7.1, while policy and research implications are showed on Section 7.2, and Section 7.3 illustrates future research.

7.1 Intellectual Contributes

The main intellectual contributes have been the development and calibration of new models and algorithms to build and analyze large scale and time-dependent digital twins exploiting different type of data and perform two practical case studies: the application of the *BEAM* model to the Bay Area scenario, California (see Chapter 5), with a mesoscopic approach, and the creation of a scenario for the city of Bologna, Italy (see Chapter 6), with a microscopic approach using the SUMOPy software. In particular, I actively participated on the SUMOPy software developement; also, the second scenario has been licensed through University of Bologna and some professionals are already using it for their purposes. During my PhD studies I also developed a critical thinking on why we need advanced models to analyze transportation systems, what are the main benefits, what are the main challenges and what are the differences between a macroscopic, a mesososcopic and a microscopic approach (see Chapter 3).

My main contributes related to the field of the microscopic approach for the analysis of a transportation system and implemented through the *SUMOPy* environment have been:

- Importing, checking, correcting and developing transportation infrastructure from *OSM*, with particular attention to soft mobility
- Developing and coding a script to create agents from big data related to real *GPS* traces
- Exploiting *OD* matrices to create both a virtual population resident in the study area and external traffic
- Developing and coding a new strategy-choice model
- Building public transportation services from *GTFS* files big data
- Calibrating a dynamic user equilibrium to fix vehicle routes
- Implementing a new algorithm to elaborate the results and plot graphical illustrations

It worth noting saying that during all the three years I've been glad to contribute on the SUMOPY's python-software developement including new classes, new processes and new functions aimed to model, analyze and divulge digital twins.

From the other hand, my main contributions in the field of a mesoscopic approach from the analysis of a transportation system and implemented with *BEAM* environment have been:

- Contributing on inserting a service of on-demand shuttles for first and last-mile connection on the *BEAM* software
- Performing a sensitivity analysis of *BEAM* models on test scenarios
- Report pros and cons for further developments of *BEAM* models and for future users.

- Coding of a *BEAM* output analyser to extrapolate information of trips and vehicles after the simulation of a scenario

My main contributes in terms of case study and practical applications have been:

- Build a state-of-the-art microscopic digital twin which is large-scale model including all main modes of transportation, has been validated. This model represents an innovation in terms of existent microscopic digital twin, in fact, it has been licensed through University of Bologna.

- Demonstrate with a practical application the potentialities of the innovative software *BEAM* while analysing impacts of fixed-area shuttle shared service in a Bay Area's mesoscopic digital twin built by the *BEAM* developers from *LBNL*.

  Regarding the study I conducted with University of Bologna related to microscopic digital twin of the city of Bologna, I'm working on it from my first year of my Master's Course in Civil Engineering and I started working on the transportation network with other interns in order to check and correct *OSM* data and importation errors. During my Master's thesis I developed a new strategy-choice model for the microscopic digital twin, I have coded it in python and I added it to the software SUMOPy: this was my first programming application after a year of independent study on python language and programming basis. After this first programming application I become more and more experienced with python and I continued for the whole PhD Course developing the software when needed. Regarding the large-scale microscopic case study, I actually started with small and artificial areas, passing from small and real areas and arriving to big areas with artificial demand. Only developing new models to exploit big data we were able with the help of my supervisors to reproduce a real transport demand in a large and real context. Regarding the application of the Bay Area scenario, I immediately started using the already calibrated scenario, therefore I only needed to learn how to run the scenario, how to change configuration parameters and how to launch the process on a remote computer - we have used *AWS* to satisfy the scenario requirements. A first analysis consisted on running several time the scenario with a reduced demand - to reduce simulation time - just to test the sensibility of some configuration parameters. Successively, after been configured and simulated the case studies, my work focused on learning R software programming language to develop a new R-code to analyse, aggregate and elaborate the output table.

7.2 Policy/research Implications

The developed models, scripts and software applications try to enrich the literature of large scale and time-dependent digital twins construction and analysis, providing new tools and case studies that will hopefully help future researchers on their studies and to better understand the potentiality of advanced approaches for the analysis of a transportation system and at the same time be aware of the related challenges.

The two case studies built during my research activities represent an application of new models and software implementations but not only:

- The Bologna's Case Study can be used by transportation planners as a test platform to build scenarios for evaluating and optimizing the evolution of the transportation system on a real and large-scale context, optimizing effects on people and the collectivity, to realize a resilient, safe, accessible, equitable, green, clean and climate neutral mobility. Moreover, industries can use the scenario to quantify the effects of the developed technologies in the system: intelligent transportation system, fleets of autonomous shared vehicles and so on. The created scenario has been licensed through University of Bologna and some professionals in the field of transportation engineering are already using it.

- The *BEAM* model development and application has been useful for the public authorities to better understand the potentialities of a fixed-area shared shuttles in the territory and to direct future investments; also, this application served to demonstrate *BEAM* capabilities and comprehend possible future improvements and developments of the software.

7.3 Future Research

Even if new models have been calibrated and digital twins have been built and tested (see Chapters 5 and 6), future research should focus on the one hand on enriching the literature with other models able to exploit even more data source as well as to model other advanced technologies and services in the system. On the other hand, further validation and some more application of the digital twins are necessary to respectively confirm the model validity and to exploit its potentialities for practical case studies. Also, future research can aim at expanding the digital twin in terms of both size and interval time.

In particular, some further applications may include different 'what if' scenarios: 1) transform the vehicle fleet either partially or totally in autonomous or electric vehicles, modeling in this last case the electric grid in the city; 2) test the platooning technologies on autonomous vehicles; 3) test synchronized traffic light; 4) study the resiliency of a study area from small to catastrophic events, and so on.

Big data sets can be available in different types and formats and can be more or less representative of the population; therefore, each dataset should be accurately pre-screened before applying or building some models on it. Beside being expensive in terms of time and resources, big data are not always available, as explained on Section 3.3. For these reasons future research should focus on the generalization of the digital twin construction process, with less efforts and using only data available for all the study area - e.g. starting directly from the population density.

Finally, as a consequence of the detailed comparison between the mesoscopic and microscopic approach (see Chapter 3), future research should focus on their combination to take advantages of the benefit of their respective advantages, while eclipsing their biggest challenges.

After concluding my PhD in transportation engineering at University of Bologna, I have been accepted for a post doctoral position at EA-Energy Analysis Environmental Impacts Division at *LBNL*, California. My employment will start right after the conclusion of my PhD studies and will allow me to continue my research on the advanced approaches for transportation system analysis.

Dissemination

Licenses:

1. Microscopic scenario for the simulation of traffic in the city of Bologna, deposited by University of Bologna, invented by: Joerg Schweizer and **Cristian Poliziani**
Microscopic transport scenarios of entire cities are an editable database that contains all information to simulate the traffic in an unprecedented detail. They can represent the actual traffic or include all present and potential future transport modes, services and technologies, allowing a precise quantification of environmental impacts and energy consumption and high sensibility to the system change.

Main Articles:

1. Schweizer, J., **Poliziani, C.**, Rupi, F., Morgano, D., Magi, M., Building a Large-Scale Micro-Simulation Transport Scenario Using Big Data. ISPRS International Journal of Geo-Information. 2021; 10(3):165. https://doi.org/10.3390/ijgi10030165 [Journal paper]

2. 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, Mineta Transportation Institute Publications, 2021. doi:10.31979/mti.2021.1822 [Research Project]

Other Articles:

3. **Poliziani, C.**, Schweizer, J., Rupi, F., Supply and demand analysis of a Free Floating bike Sharing System. Komunicàcie. 24(2), A53-A65, 2022 [Journal paper]

4. Schweizer, J., Rupi, F., **Poliziani, C.**, Simulating automated vehicles in high capacity networks. SUMO2021. 2021 [Conference paper]

5. Velasquez-Martìnez, R., Torres-Bohòrquez, C., Baza-Solares, N., Martinez-Estupinan, Y., **Poliziani, C.**, Comparative analysis of open source and commercial traffic micro simulation tools. Komunicàcie. 24(2), E49-E62, 2021 [Journal paper]

6. **Poliziani, C.**, Rupi, F., Schweizer, J., Traffic surveys and GPS traces to explore patterns in cyclist's in-motion speeds. Transportation Research Procedia 2021. 60, 410-417, 2022 [Journal paper]

7. **Poliziani, C.**, Rupi, F., Schweizer, J., Saracco, M., Capuano, D., Cyclist's waiting time estimation at intersections, a case study with GPS traces from Bologna. Transportation Research Procedia 2021. 2021 – in press [Journal paper]

8. **Poliziani, C.**, Rupi, F., Schweizer, J., Exploring Patterns in the Free-Floating Bike Sharing Usage by Analysing GPS Traces. European Transport Conference 2020. 2020 [Conference paper]

9. **Poliziani, C.**, Rupi, F., Mbuga, F., Schweizer, J., Tortora, C., Categorizing three active cyclist typologies by exploring patterns on a multitude of GPS crowdsourced data attributes. RTBM – Research in Transportation Business & Management. 40(1), 100572, 2021 [Journal paper]

10. Schweizer, J., Rupi, F., **Poliziani, C.**, Estimation of link-cost function for cyclists based on stochastic optimization and GPS traces. IET – Intelligent Transport System. 14(13), 1810-1814, 2020 [Journal paper]

11. Rupi, F., Schweizer, J., **Poliziani, C.**, Analysing the dynamic performances of a bicycle network with a temporal analysis of GPS traces. Case Studies on Transport Policy. 8(3), 770-777, 2020. [Journal paper]

12. Rupi, F., **Poliziani, C.** Schweizer, J. Data-driven bicycle network analysis based on traditional counting methods and GPS traces from smartphone. ISPRS International Journal of Geo-Information. 8(8), 322, 2019 [Journal paper]

13. Danesi, A., Ongari, D., **Poliziani, C.**, Rupi, F., Evolution of the road and rail transport of goods in European countries before and after the financial crises. Komunicàcie. 21(4), 3-12, 2019 [Journal paper]

14. Schweizer, J., Rupi, F., **Poliziani, C.**, Filippi, F. - Generating activity based, multi-modal travel demand for SUMO. Epic Series in Engineering. 2, 118 – 133, 2018 [Conference paper]

Presentation at Conferences

1. **TRB2021** Washington, D.C., 100th Transportation Research Board Annual Meeting
   Hsueh, G., Czerwinski, D., **Poliziani, C.**, Using *BEAM* Software to Simulate the Introduction of On-Demand, Automated, and Electric Shuttles for Last Mile Connectivity in Santa Clara County, presented by Gary Hsueh, David Czerwinski and **Cristian Poliziani**

2. **SETT2021** Irvine, CA Conference on Sustainability and Emerging Transportation Technology
   Hsueh, G., Czerwinski, D., **Poliziani, C.**, Using *BEAM* Software to Simulate the Introduction of On-Demand, Automated, and Electric Shuttles for Last Mile Connectivity in Santa Clara County, presented by Gary Hsueh

3. **SUMO2021** Berlin - Germany, *SUMO* user Conference
   Schweizer, J., Rupi, F., **Poliziani, C.**, Simulating automated vehicles in high capacity networks, presented by Joerg Schweizer

4. **EWGT2021** Aveiro – Portugal, EURO Working Group on Transportation Meeting
   Poliziani, C., Rupi, F., Schweizer, J., Saracco, M., Capuano, D., Cyclist's waiting time estimation at intersections, a case study with *GPS* traces from Bologna, presented by **Cristian Poliziani**

5. **LWC2021** Brescia, Living and Walking in Cities XXV International Conference

Poliziani, C., Rupi, F., Schweizer, J., Traffic surveys and *GPS* traces to explore patterns in cyclist's in-motion speeds, presented by **Cristian Poliziani**

6. **ETC2020** Milan, Italy - European Transport Conference Online
Poliziani, C., Rupi, F., Schweizer, J., Prediction of the Free Floating Bike Sharing System demand from *GPS* traces, presented by **Cristian Poliziani**

7. **SIDT2019** Salerno, Italy - Scientific seminar of Società Italiana Docenti di Trasporti (SIDT)

    - **Poliziani, C.**, Rupi, F., Schweizer, J., An analysis of dynamic demand patterns in free-floating bike sharing schemes based on *GPS* traces, presented by **Cristian Poliziani**
    - Rupi, F., Schweizer, J., **Poliziani, C.**, Analyzing an urban bike network by comparing traditional counting methods with *GPS* traces from Smartphone, presented by Federico Rupi
    - Schweizer, J., Rupi, F., **Poliziani, C.** Calibration of bicycle link-cost functions using *GPS* traces and stochastic optimization methods, presented by Joerg Schweizer

8. **SUMO2018** Berlin - Germany, *SUMO* user Conference
Schweizer, J., Rupi, F., Filippi, F., **Poliziani, C.**, Generating activity based, multi-modal travel demand for SUMO, presented by Joerg Schweizer

Acknowledgments

My first acknowledgment goes to my PhD Supervisors Prof. Federico Rupi and Prof. Joerg Schweizer, and to the whole transportation engineer team at University of Bologna, who has always been comprehensive, welcoming, and competent people (see 7.1); I would like to thank them for making me grow from research, didactic aspects and not only. I really hope to have the opportunity to continue collaborating with the team in the future.

I would like to thank the PhD program coordinators Prof. Luca Vittuari and Prof. Alessandro Tugnoli for their guidance. It has been nice relating with you as a PhD student representative.

I would like to thank my family, my friends and my partner Chiara.

I would like to thank Eng. Gary Hsueh and Prof. David Czerwinski, their guidance and for hosting me respectively at Prospect Silicon Valley and San José State University (see Figure 7.2). What an amazing experience in California from both research and cultural point of view.

I would like to thank the *BEAM* team at *LBNL* who supported me using the software during my research period in California, and for offering me a post doctoral position after the conclusion of my PhD studies. Look forward to collaborate with them.

I would like to thank some special collaborators from outside University of Bologna: Prof. Silvio Nocera from University Iuav of Venezia, Italy; Prof. Cristina Tortora from San José State University, California and Eng. Nelson Baza-Solares from University of Santander, Colombia.

I would like to thank all the 45 Master's and Bachelor's graduated students from all over the world I had the pleasure to co-supervise for the final dissertation and all students of Bachelor's and Master's Courses in Transportation Engineering at University of Bologna, with whom I had the pleasure to share my research during didactic activities.

Figure 7.1: From left to right: Prof. Joerg Schweizer, Prof. Federico Rupi and myself. Conference trip to Milan

Figure 7.2: March 19, 2020. A rare sunny day in San Francisco. View from Twin Peaks without tourists and with social distancing. Photo taken during my research period in San José, in the middle of the first wave of COVID-19 pandemic

Bibliography

M. J. Alam and M. A. Habib. Investigation of the impacts of shared autonomous vehicle operation in halifax, canada using a dynamic traffic microsimulation model. *Procedia Comput. Sci.*, 130:496–503, 2018. doi: http://doi.org/10.1016/j.procs.2018.04.066.

B. Bartin, K. Ozbay, J. Gao, and A. Kurkcu. Calibration and validation of large-scale traffic simulation networks: A case study. *Procedia Comput. Sci.*, 130:844–849, 2018. doi: http://doi.org/10.1016/j.procs.2018.04.076.

M. Behrisch, L. Bieker-Walz, J. Erdmann, and D. Krajzewicz. Sumo—simulation of urban mobility: An overview. *Proceedings of the SIMUL 2011, Barcelona, Spain*, pages 23–28, 2011.

S. Bernardi, K. J. Krizek, and F. Rupi. Quantifying the role of disturbances and speeds on separated bicycle facilities. *J. Transp. Land Use*, 2015. doi: http://doi.org/10.5198/jtlu.2015.715.

S. Biba, K. M. Curtin, and G. Manca. Who can access transit? reviewing methods for determining population access to bus transit. *Cityscape*, 16 (2):193–203, 2014. doi: https://doi.org/10.1080/13658810802646679.

J. Bowman and M. Ben-Akiva. Activity-based disaggregate travel demand model system with activity schedules. *Transp. Res. Part A Policy Pract.*, 35:1–28, 2001. doi: http://doi.org/10.1016/S0965-8564(99)00043-9.

S. Calvert, W. Schakel, and J. V. Lint. Will automated vehicles negatively impact traffic flow? *J. Adv. Transp.*, 2017:1–17, 2017. doi: http://doi.org/10.1155/2017/3082781.

E. Cascetta. Transportation systems engineering: Theory and methods. *Springer Science and Business Media LLC: Berlin/Heidelberg, Germany*, 49, 2001.

E. Cascetta. Transportation systems analysis: Models and applications, 2nd edition. *Springer*, 2009.

D. Charypar, K. W. Axhausen, and K. Nagel. Event-driven queue-based traffic flow microsimulation. *Transp. Res. Rec. J. Transp. Res. Board*, 2003:35–40, 2007. doi: http://doi.org/10.3141/2003-05.

P. N. E. Chee, Y. O. Susilo, and Y. D. Wong. Determinants of intention-to-use first-/last-mile automated bus service. *Transportation Research Part A: Policy and Practice*, 139:350–375, 2020. doi: https://doi.org/10.1016/j.tra.2020.06.001.

S. Childress, B. Nichols, B. Charlton, and S. Coe. Using an activity-based model to explore the potential impacts of automated vehicles. *Transp. Res. Rec. J. Transp. Res. Board*, 2493:99–106, 2015. doi: http://doi.org/10.3141/2493-11.

L. Codeca, R. Frank, and S. F. T. Engel. Luxembourg sumo traffic (lust) scenario: Traffic demand evaluation. *EEE Intell. Transp. Syst. Mag.*, 9: 52–63, 2017. doi: http://doi.org/10.1109/MITS.2017.2666585.

. G. Corey, Canapary. Vta on-board survey 2013: Management summary report. *Management Summary Report, San José: Santa Clara Valley Transit Authority*, pages 1–76, 2014. URL `http://openbook.sfgov.org/webreports/details3.aspx?id=2705`.

A. I. Croce, G. Musolino, C. Rindone, and A. Vitetta. Transport system models and big data: Zoning and graph building with traditional surveys, fcd and gis. *ISPRS Int. J. Geo-Inf.*, 8:187, 2019. doi: http://doi.org/10.3390/ijgi8040187.

A. Cuauhtemoc, E. Alexander, and F. P. Jacobus. Transport modelling in the age of big data. *Int. J. Urban Sci.*, 21:19–42, 2017. doi: http://doi.org/10.1080/12265934.2017.1281150.

W. Do, O. M. Rouhani, and L. Miranda-Moreno. Simulation-based connected and automated vehicle models on highway sections: A literature review. *J. Adv. Transp.*, 2019:1–14, 2019. doi: http://doi.org/10.1155/2019/9343705.

A. El-Geneidy, M. Grimsrud, R. Wasfi, P. Tétreault, and J. Surprenant-Legault. New evidence on walking distances to transit stops: Identifying redundancies and gaps using variable service areas. *Transportation*, 41 (1):193–210, 2013. doi: https://doi.org/10.1007/s11116-013-9508-z.

J. K. Eom, J. Choi, M. S. Park, and T.-Y. Heo. Exploring the catchment area of an urban railway station by using transit card data: Case study in seoul. *Cities*, 95:102364, 2019. doi: https://doi.org/10.1016/j.cities.2019.05.033.

S. Farber and M. Grandez. Transit accessibility, land development and socioeconomic priority: A typology of planned station catchment areas in the greater toronto and hamilton area. *Journal of Transport and Land Use*, 10(1):879–902, 2017. doi: https://doi.org/10.5198/jtlu.2017.980.

M. Fellendorf and P. Vortisch. Microscopic traffic flow simulator vissim. *Harvey J. Greenberg*, 145:63–93, 2010.

P. Fernandes and U. Nunes. Platooning of autonomous vehicles with intervehicle communications in sumo traffic simulator. *Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems, Funchal, Portugal, 19–22 September 2010; IEEE: New York, NY, USA*, 2010.

G. Flötteröd, Y. Chen, and K. Nagel. Behavioral calibration and analysis of a large-scale travel microsimulation. *Networks Spat. Econ.*, 12:481–502, 2011. doi: http://doi.org/10.1007/s11067-011-9164-9.

M. Frank and P. Wolfe. An algorithm for quadratic programming. *Nav. Res. Logist. Q.*, 3:95–110, 1956. doi: http://doi.org/10.1002/nav.3800030109.

V. Grimm, E. Revilla, U. Berger, F. Jeltsch, W. M. Mooij, S. F. Railsback, H.-H. Thulke, J. Weiner, T. Wiegand, and D. L. DeAngelis. Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*, 310:987–991, 2005. doi: http://doi.org/10.1126/science.1116681.

V. Grimm, U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback. The odd protocol: A review and first update. *Ecol. Model*, 221: 2760–2768, 2010. doi: http://doi.org/10.1016/j.ecolmodel.2010.08.019.

E. Guerra, R. Cervero, and D. Tischler. Half-mile circle. *Transportation Research Record: Journal of the Transportation Research Board*, 2276(1): 101–109, 2012.

I. Haas and B. Friedrich. Developing a micro-simulation tool for autonomous connected vehicle platoons used in city logistics. *Transp. Res. Procedia*, 27:1203–1210, 2017. doi: http://doi.org/10.1016/j.trpro.2017.12.084.

N. Hansen. The cma evolution strategy: A tutorial. *Inria*, pages 1–34, 2011. URL `https://hal.inria.fr/hal-01297037/document`.

D. B. Hess. Access to public transit and its influence on ridership for older adults in two u.s. cities. *Journal of Transport and Land Use*, 2(1):3–27, 2009. doi: https://doi.org/10.5198/jtlu.v2i1.11.

G. Hsueh, D. Czerwinski, C. Poliziani, T. Becker, A. Hughes, P. Chen, and M. Benn. 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*, page 343, 2021. doi: http://doi.org/10.31979/mti.2021.1822.

C. juan Liu, Z. Liu, Y. jie Chai, and T. ting Liu. Review of virtual traffic simulation and its applications. *J. Adv. Transp.*, 2020:1–9, 2020. doi: http://doi.org/10.1155/2020/8237649.

J. Kamel, R. Vosooghi, J. Puchinger, F. Ksontini, and G. Sirin. Exploring the impact of user preferences on shared autonomous vehicle modal split: A multi-agent simulation approach. *Transp. Res. Procedia*, 37:115–122, 2019. doi: http://doi.org/10.1016/j.trpro.2018.12.173.

J.-Y. Kang and J. Aldstadt. Using multiple scale spatio-temporal patterns for validating spatially explicit agent-based models. *nt. J. Geogr. Inf. Sci.*, 33:193–213, 2018. doi: http://doi.org/10.1080/13658816.2018.1535121.

C. Kanuri, K. Venkat, S. Maiti, and P. Mulukutla. Leveraging innovation for last mile connectivity to mass transit. *Transportation Research Procedia*, 41:655–669, 2019. doi: https://doi.org/10.1016/j.trpro.2019.09.114.

B. Kloostra and M. J. Roorda. Fully autonomous vehicles: Analyzing transportation network performance and operating scenarios in the greater toronto area, canada. *Transp. Plan. Technol.*, 42:99–112, 2019. doi: http://doi.org/10.1080/03081060.2019.1565159.

F. Knobloch, S. V. Hanssen, A. Lam, H. Pollitt, P. Salas, U. Chewpreecha, M. A. J. Huijbregts, and J.-F. Mercure. Net emission reductions from electric cars and heat pumps in 59 world regions over time. *Nat Sustain*, 3:437–447, 2020. doi: https://doi.org/10.1038/s41893-020-0488-7.

S. Lee, B. G. Heydecker, J. Kim, and S. Park. Stability analysis on a dynamical model of route choice in a connected vehicle environment. *Transp. Res. Procedia*, 23:720–737, 2017. doi: http://doi.org/10.1016/j.trpro.2017.05.040.

D. Lin, Y. Zhang, R. Zhu, and L. Meng. The analysis of catchment areas of metro stations using trajectory data generated by dockless shared bikes. *Sustainable Cities and Society*, 49, 2019. doi: https://doi.org/10.1016/j.scs.2019.101598.

R. Liu. The dracula dynamic network microsimulation model. *Human Resource Development and Information Technology; Springer International Publishing: Berlin/Heidelberg*, pages 23–56, 2006.

R. Liu, D. V. Vliet, and D. Watling. Microsimulation models incorporating both demand and supply dynamics. *Transp. Res. Part A Policy Pract.*, 40:125–150, 2006. doi: http://doi.org/10.1016/j.tra.2005.05.003.

Q. Lu, T. Tettamanti, D. Hörcher, and I. Varga. The impact of autonomous vehicles on urban traffic network capacity: An experimental analysis by microscopic traffic simulation. *Transp. Lett.*, 12:540–549, 2019. doi: http://doi.org/10.1080/19427867.2019.1662561.

Z. Luo, Y. Liu, and C. Guo. Operational characteristics of mixed traffic flow under bi-directional environment using cellular automaton. *J. Traffic Transp. Eng.*, 1:383–392, 2014. doi: http://doi.org/10.1016/S2095-7564(15)30288-9.

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

M. Martínez-Díaz and F. Soriguera. Autonomous vehicles: Theoretical and practical challenges. *ransp. Res. Procedia*, 33:275–282, 2018. doi: http://doi.org/10.1016/j.trpro.2018.10.103.

I. Mavromatis, A. Tassi, R. J. Piechocki, and M. Sooriyabandara. On urban traffic flow benefits of connected and automated vehicles. *Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 25–28 May 2020; IEEE: New York, NY, USA, 2020*, pages 1–7.

K. N. Michael Balmer, Kay W. Axhausen. Agent-based demand-modeling framework for large-scale microsimulations. *Transp. Res. Rec. J. Transp. Res. Board*, 1985:125–134, 2006. doi: http://doi.org/10.1177/0361198106198500114.

V. Milanés, S. E. Shladover, J. Spring, C. Nowakowski, H. Kawazoe, and M. Nakamura. Cooperative adaptive cruise control in real traffic situations. *IEEE Trans. Intell. Transp. Syst.*, 15:296–305, 2013. doi: http://doi.org/10.1109/TITS.2013.2278494.

A. Moorthy, R. de Kleine, G. Keoleian, and J. Good. Shared autonomous vehicles as a sustainable solution to the last mile problem: A case study of ann arbor-detroit area. *SAE International Journal of Passenger Cars—Electronic and Electrical Systems*, 10(2), 2017. doi: https://doi.org/10.4271/2017-01-1276.

E. T. Mtoi, R. Moses, and E. E. Ozguven. An alternative approach to network demand estimation: Implementation and application in multi-agent transport simulation (matsim). *Procedia Comput. Sci.*, 37:382–389, 2014. doi: http://doi.org/10.1016/j.procs.2014.08.057.

A. Papadoulis, M. Quddus, and M. Imprialou. Evaluating the safety impact of connected and autonomous vehicles on motorways. *Accid. Anal. Prev.*, 124:12–22, 2019. doi: http://doi.org/10.1016/j.aap.2018.12.019.

M. Patriksson. The traffic assignment problem: Models and methods. *Topics in Transportation; Grafton, J., Ed.; VSP, Linköping Institute of Technology: Gothenburg, Sweden*, 1994.

X. Pi, W. Ma, and Z. S. Qian. general formulation for multi-modal dynamic traffic assignment considering multi-class vehicles, public transit and parking. *Transp. Res. Procedia*, 38:914–934, 2019. doi: http://doi.org/10.1016/j.trpro.2019.05.047.

L. Po, F. Rollo, C. Bachechi, and A. Corni. From sensors data to urban traffic flow analysis. *roceedings of the 2019 IEEE International Smart Cities Conference (ISC2), Casablanca, Morocco, 14–17 October 2019; IEEE: New York, NY, USA*, pages 478–485, 2019. doi: http://doi.org/10.1109/ISC246665.2019.9071639.

C. Poliziani, F. Rupi, and J. Schweizer. Traffic surveys and gps traces to explore patterns in cyclist's in-motion speeds. *Transportation Research Procedia*, 60:410–417, 2022a. doi: https://doi.org/10.1016/j.trpro.2021.12.053.

C. Poliziani, F. Rupi, J. Schweizer, M. Saracco, and D. Capuano. Cyclist's waiting time estimation at intersections, a case study with gps traces from bologna. *Transportation Research Procedia*, in press, 2022b. doi: https://doi.org/10.1016/j.trpro.2021.12.053.

C. Poliziani, J. Schweizer, and F. Rupi. Supply and demand analysis of a free floating bike sharing system. *Komunicàcie*, 24(2):A53–A65, 2022c. doi: https://doi.org/10.26552/com.C.2022.2.A53-A65.

M. Ramezani, J. A. Machado, A. Skabardonis, and N. Geroliminis. Capacity and delay analysis of arterials with mixed autonomous and human-driven vehicles. *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*, pages 280–284, 2017. doi: http://doi.org/10.1109/MTITS.2017.8005680.

J.-P. e. a. Rodrigue. The geography of transport systems. fifth edition. *New York, NY: Routledge*, 2020. URL `https://transportgeography.org`.

F. Roulland, C. R. D. Souza, L. Ulloa, A. Mondragón, M. Niemaz, and V. Ciriza. Towards data-driven simulations in urban mobility analytics. *Proceedings of the 14th ITS Asia Pacific Forum, Nanjing, China*, pages 27–29, 2015.

S. Roy and D. Basu. Selection of intervention areas for improving travel condition of walk-accessed bus users with a focus on their accessibility: An experience in bhubaneswar. *Transport Policy*, 96:29–39, 2020. doi: https://doi.org/10.1016/j.tranpol.2020.06.004.

F. Rupi, C. Poliziani, and J. Schweizer. Data-driven bicycle network analysis based on traditional counting methods and gps traces from smartphone. *ISPRS Int. J. Geo-Inf.*, 8:322, 2019. doi: http://doi.org/10.3390/ijgi8080322.

F. Rupi, J. Schweizer, and C. Poliziani. Analysing the dynamic performances of a bicycle network with a temporal analysis of gps traces. *Case Studies on Transport Policy*, 8(3):770–777, 2020. doi: https://doi.org/10.1016/j.cstp.2020.05.007.

M. Savrasovs, I. Pticina, and V. Zemlyanikin. Wide-scale transport network microscopic simulation using dynamic assignment approach. *Comput. Devices Commun.*, 36:241–251, 2018.

J. Schweizer, S. Bernardi, and F. Rupi. Map-matching algorithm applied to bicycle global positioning system traces in Bologna. *ET Intell. Transp*, 10:244–250, 2016a. doi: http://doi.org/10.1049/iet-its.2015.0135.

J. Schweizer, T. Parriani, E. Traversi, and F. Rupi. Optimum vehicle flows in a fully automated vehicle network. *Proceedings of the International Conference on Vehicle Technology and Intelligent Transport Systems; SCITEPRESS—Science and Technology Publication: Rome, Italy*, page 195–202, 2016b. URL `https://www.scitepress.org/PublicationsDetail.aspx?ID=WOrOwi3OENM=&t=1`.

J. Schweizer, F. Rupi, and C. Poliziani. Generating activity based, multimodal travel demand for sumo. *Proceedings of the SUMO User Conference 2018, Berlin, Germany*, 2018.

J. Schweizer, F. Rupi, and C. Poliziani. Estimation of link-cost function for cyclists based on stochastic optimization and gps traces. *ITE Intelligent Transport Systems*, 14(13):1810–1814, 2020. doi: https://doi.org/10.1049/iet-its.2019.0683.

J. Schweizer, C. Poliziani, F. Rupi, D. Morgano, and M. Magi. Building a large-scale micro-simulation transport scenario using big data. *ISPRS International Journal of Geo-Information*, 10(3), 2021. ISSN 2220-9964. doi: 10.3390/ijgi10030165. URL `https://www.mdpi.com/2220-9964/10/3/165`.

S. E. Shladover, D. Su, and X.-Y. Lu. Impacts of cooperative adaptive cruise control on freeway traffic flow. *ransp. Res. Rec. J. Transp. Res. Board*, 2324:63–70, 2012. doi: http://doi.org/10.3141/2324-08.

M. H. Tawfeek, M. Elesawey, and K. El-Araby. Calibration and validation of micro-simulation models using measurable variables. *Proceedings of the Canadian Society for Civil Engineering 2018 Annual Conference, Fredericton, NB, Canada, 13–16 June*, 2018.

S. Uppoor and M. Fiore. Large-scale urban vehicular mobility for networking research. *Proceedings of the 2011 IEEE Vehicular Networking Conference (VNC), Amsterdam, The Netherlands, 14–16 November 2011; IEEE: New York, NY, USA*, pages 62–69, 2011.

R. Velasquez-Martínez, C. Torres-Bohórquez, N. Baza, Y. Martinez-Estupiñan, and C. Poliziani. Traffic simulation with open-source and commercial traffic microsimulators: a case study. *Komunicàcie*, 24(2): E49–E62, 2022. doi: https://doi.org/10.26552/com.C.2022.2.E49-E62.

A. K. W., M. Konrad, B. Michael, C. Francesco, H. Andreas, R. Marcel, and W. R. A. Axhausen, k. large-scale agent-based travel demand optimization applied to switzerland, including mode choice. *roceedings of the 12th World Conference on Transportation Research, Lisboa, Portugal*, pages 11–15, 2010.

M. Wang and L. Mu. Spatial disparities of uber accessibility: An exploratory analysis in atlanta, usa. *omputers, Environment and Urban Systems*, 67: 169–175, 2018.

J. G. Wardrop. Road paper. some theoretical aspects of road traffic research. *Proc. Inst. Civ. Eng.*, 1:325–362, 1952. doi: http://doi.org/10.1680/ipeds.1952.11259.

M. J. Wierbos, V. L. Knoop, F. S. Hänseler, and S. P. Hoogendoorn. A macroscopic flow model for mixed bicycle–car traffic. *Transp. A Transp. Sci. 202*, 17:340–355, 2021. doi: http://doi.org/10.1080/23249935.2019.1708512.

A. Wilson. The future of urban modelling. *Appl. Spat. Anal. Policy*, 11: 647–655, 2018. doi: http://doi.org/10.1007/s12061-018-9258-6.

R. Woodman, K. Lu, M. D. Higgins, S. Brewerton, P. A. Jennings, and S. Birrell. Gap acceptance study of pedestrians crossing between platooning autonomous vehicles in a virtual environment. *Transp. Res. Part F Traffic Psychol. Behav.*, 67:1–14, 2019. doi: http://doi.org/10.1016/j.trf.2019.09.017.

Y. Zhao and A. W. Sadek. Large-scale agent-based traffic micro-simulation: Experiences with model refinement, calibration, validation and application. *Procedia Comput. Sci.*, 10:815–820, 2012. doi: http://doi.org/10.1016/j.procs.2012.06.105.

T. Zuo, H. Wei, and N. Chen. romote transit via hardening first-and-last-mile accessibility: Learned from modeling commuters' transit use. *Transportation Research Part D: Transport and Environment*, 86:102446, 2020a. doi: https://doi.org/10.1016/j.trd.2020.102446.

T. Zuo, H. Wei, N. Chen, and C. Zhang. irst-and-last mile solution via bicycling to improving transit accessibility and advancing transportation equity. *Cities*, 99:102614, 2020b. doi: https://doi.org/10.1016/j.cities.2020.102614.

Ömer Verbas, H. S. Mahmassani, M. F. Hyland, and H. Halat. 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*, 2564:78–88, 2016. doi: http://doi.org/10.3141/2564-09.

Acronyms

AV Automated Vehicles. 27, 59, 63, 66–68, 129

AWS Amazon Web Service. 52, 54, 103

Aimsun Advanced Interactive Microscopic Simulator for Urban and Non-Urban Network. 44

ArcGIS Aeronautical Reconnaissance Coverage Geographic Information System. 54

BEAM Behavior, Energy, Autonomy, and Mobility. VIII, 23, 35–37, 47, 48, 50–57, 64, 68, 71–73, 101–104, 107, 109, 111, 131, 132

CAV Connected Automated Vehicles. 27, 43

DOE U.S. Department of Energy. 51

DRT Demand Responsive Transportation. 44

EEA European Environment Agency. 1

EV Electric Vehicles. 27

GPS Global Positioning System. IX, 2, 7, 8, 23, 24, 29, 78, 81, 82, 84, 92, 102, 109, 110, 131

GRT Group Rapid Transit. 27

GTFS General Transit Feed Specification. IX, 8, 24, 53, 54, 80, 102

GUI Graphical User Interface. 32, 37, 40, 132

ISTAT Italian institute for statistics. 80

ITS Intelligent Transport System. 16, 30, 132

LANL Los Alamos National Lab. 45

LBNL Lawrence Berkeley National Lab. VIII, 35, 51, 68, 103, 105, 111

MATSim Multi-Agent Transport Simulation. 36, 44, 51

Maas Mobility as a service. 2, 43, 44

NASA United States National Aeronautics and Space Administration. 45

NETCONVERT NETwork CONVERTer. 16, 78

NETEDIT NETwork EDITor. 16, 78

NumPy = Numerical Python library. 37

OD Origin/Destination. IX, 13, 14, 22, 23, 42, 102

OSM Open Street Map. IX, 8, 16, 20, 23, 33, 36, 38, 45, 53, 68, 78, 79, 102, 103, 132

PRT Personal Rapid Transit. 27

PTV Planung Transport Verkehr AG - Planning Transportation Traffic. 43

Paramics Paramics Discovery - originally S-Paramics. 43, 44

PyOpenGL Python Open Graphics Library. 37

SAGA SUMO Activity GenerAtion. 45

SDGs Sustainable Development Goals. 2

SMART Systems and Modeling for Accelerated Research in Transportation. 51

SUMOPy Simulation of Urban MObility - python interface. VIII, 8, 16, 18, 35, 37–40, 43, 46, 78, 80, 84, 94, 102, 131, 132

SUMO Simulation of Urban MObility. 6–8, 16–19, 21, 35–37, 39, 44, 45, 78, 79, 94, 109, 110, 131, 132

SUMP Sustainable Urban Mobility Plan. 2, 84, 85

TAZs Traffic Assignment Zones. 13, 79–82, 95, 133

TLS Traffic Light System. 16

TPER Trasporto Pubblico Emilia Romagna. 80

TraCI Traffic Control Interface. 94

Transims TRansportation ANalysis SIMulation System. 45

V2X Vehicle-to-Everything. 20, 27, 94

VTA Vally Transportation Authority. 54

Vissim Verkehr In Städten - SIMulationsmodell. 7, 31, 43, 132

List of Tables

List of Figures