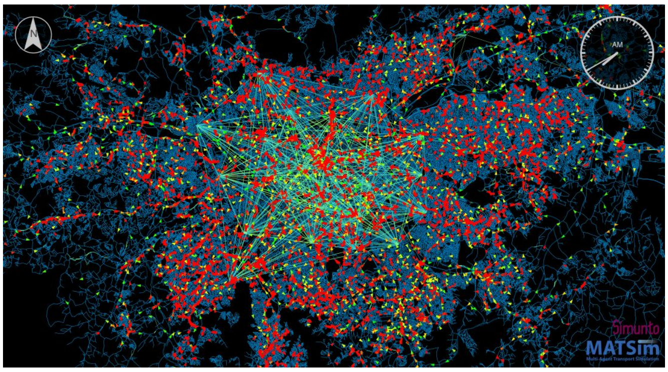


Analysis of urban air mobility's transport performance in São Paulo Metropolitan Region using MATSim

Simulation metamodel based on active learning to predict the simulations outputs

Sérgio Fonseca dos Santos

Thesis to obtain the Master of Science Degree in

Aerospace Engineering

Supervisors: Prof. João Pedro Faria Mendonça Barreto
Eng. Zafeiris Kokkinogenis

Examination Committee

Chairperson: Prof. Filipe Szolnoky Ramos Pinto Cunha
Supervisor: Prof. João Pedro Faria Mendonça Barreto
Members of the Committee: Prof. Frederico José Prata Rente Reis Afonso
Prof. Pedro Miranda de Andrade de Albuquerque d'Orey

November 2021

“We choose to go to the moon in this decade and do the other things, not because they are easy, but because they are hard”, John F. Kennedy

Dedicated to my parents and my sister.

Acknowledgments

The help and support of those below was an essential part for me to complete this thesis, and my gratefulness goes out to all of you.

My first acknowledgements go to my supervisors, to João Barreto and to Zafeiris Kokkinogenis who supported me throughout this journey and helped me grow as a learner at this time. The pleasure of working and learning with you is mine.

I also would like to express my gratitude to Instituto Superior Técnico, which was my home for the last few years, and to CEiiA, for the opportunity I was given to learn from the best.

I thank to Luís Rodrigues, a friend that this journey offered me, for all shared knowledge and his remarks, and for all the discussions we have. I thank to all my personal friends (you know who you are) for the partnership and for being there for what I need, especially to Alexandre Reis for all his help, that even though it is not his job, he took his time to be able to help me in this process.

Especially, to you, Marta, who were always love with open arms. Thank you for encouraging me to dream and never give up, and for being always by my side, no matter what. I promise to handle your heart with care, forever.

Last, but not least, to all my family, thank you for being wonderful persons and believe that I can achieve everything I want. In particular, to my mother, Cristina, to my father, José, and to my sister, Rita. I am grateful for having you in my life. Thanks for making this journey possible, to support my dreams and for being home, always. I just hope to make you proud.

I thank to all who can not be present now, but they guided me in this journey of life. “For even if they are far away, I hold them in my heart”.

I did my best efforts to make this thesis part of something and I wish that I could take all that I ever learn for the rest of my life. “To Infinity and beyond”

Thank you all!

Resumo

Devido ao rápido crescimento populacional, estima-se que o nível de congestionamento aumente nas cidades, sendo necessário integrar novos meios de transporte, de forma reduzir os tempos de viagem.

Recentes avanços na tecnologia aeronáutica, permitiram o desenvolvimento de um novo meio de transporte, Mobilidade Aérea Urbana (MAU). Este, quando implementado, permitirá deslocar pessoas, de um ponto para outro, através de veículos com capacidades descolagem e aterragem vertical. Enquanto estes veículos têm sido amplamente estudados, existe pouca investigação no que diz respeito ao impacto que a introdução destes veículos irá ter nas cidades. Torna-se assim fundamental prever o comportamento destas aquando da introdução deste novo meio de transporte, de forma a adotarem-se as melhores medidas.

Posto isto, este trabalho tem dois objetivos: prever o impacto que a introdução deste novo meio de transporte terá na cidade de São Paulo, Brasil, através de um Simulador de Mobilidade baseada em agentes (MATSim) e criar um metamodelo de simulação baseado em Processos Gausseanos e aprendizagem ativa, que permita prever o resultado das simulações, em vez de estas serem feitas.

Relativamente aos resultados das simulações obtidos, verifica-se que no cenário base simulado há um grande congestionamento aéreo. No entanto, a introdução deste novo modo de transporte pode servir para aliviar o congestionamento nas estradas.

Em relação ao metamodelo, este é capaz de prever, com alguma exatidão, resultados de simulações efetuadas. No entanto, é necessário aperfeiçoar a estratégia, de forma a aproximar o metamodelo ao simulador.

Palavras-chave: UAM, MATSim, Aprendizagem Computacional, Processos Gausseanos, Metamodelo, São Paulo.

Abstract

Due to the rapid population growth, it is estimated that the level of congestion increases in cities, making it necessary to integrate new transport modes, in order to reduce travel times.

Recent advances in aeronautical technology have allowed the development of a new transport mode, Urban Air Mobility (UAM). When implemented, it will make possible to move people from one point to another using vehicles with vertical take off and landing capabilities. While these vehicles have been extensively studied, there is little research regarding the impact the introduction of these vehicles will have on cities. It is, therefore, essential to predict the behavior of these when introducing this new transport mode, in order to adopt the best measures.

That said, this work has two objectives: to predict the impact that the introduction of this new transport mode will have in the city of São Paulo, Brazil, through an agent-based Mobility Simulator (MATSim) and to create a simulation metamodel based on Gaussian Processes (GP) and active learning, which allows to predict the result of simulations, instead of them being carried out.

Regarding the simulation results obtained, it is verified that in the simulated base scenario there is great air congestion. However, the introduction of this new transport mode can serve to alleviate congestion on the roads.

Regarding the metamodel, it can predict, with some accuracy, the results of simulations carried out. However, it is necessary to improve the strategy in order to bring the metamodel closer to the simulator.

Keywords: UAM, MATSim, Machine Learning, Gaussian Processes, Metamodel, São Paulo.

Contents

List of Tables

List of Figures

Nomenclature

GP Gaussian Process.

MATSim Multi-agent Transport Simulation.

PATS Personal Aerial Transportation Systems.

PAVs Personal Aerial Vehicles.

PT Public Transport.

UAMS Urban Aerial Mobility Systems.

UAM Urban Air Mobility.

VTOL Vertical Take-Off and Landing.

Greek symbols

α Mode parameter.

β Marginal utility parameters.

Roman symbols

$\mathbb{E}$ Expected value.

$\mathbb{P}$ Probability.

x' Input data point.

x Input data point.

c Travel costs.

k_f Kernel function.

m_f Mean function.

u Utility.

x Per-trip choice dimensions.

Chapter 1

Introduction

1.1 Motivation

Due to rapid population growth it is estimated that, by 2050, the population will reach 9.7 billion inhabitants, according to United ONU [1], an increase of about 2 billion, compared to the current number (equivalent to an increase of 25%). With this predicted increase, the level of congestion on the urban centers is expected to increase further. That said, the need arose to integrate new transport modes, in addition to the existing ones, in order to reduce travel times. In 2018, the average car journey time in São Paulo, Brazil, was 30 minutes [2].

A reduction in average travel times will, on one hand, reduce the average time spent on the road and, on the other, meet the standard proposed by the European Commission, for door-to-door travel in Europe, at a time maximum of 4 hours, until 2050, according to Raoul Rothfeld and Antoniou [3].

Recent advances in aeronautical technologies, have enabled the development of a new transport mode, Urban Air Mobility (UAM). This will allow people to move from point to point in cities using electric or hybrid vehicles with Vertical Take-Off and Landing (VTOL) capabilities. While these vehicles have been extensively studied, regarding their mains areas, like engines, aeroacoustics and aerodynamics, according to Christopher Silva and Patterson [4], there is little research on its introduction into cities and what influence it will have on them.

In view of this, there is a need to simulate the behavior that the city of São Paulo will have, in order to predict the number of people who can benefit from this new transport mode and to adopt the best policies for its introduction.

1.2 Topic overview

Personal Aerial Transportation Systems (PATS) or Urban Aerial Mobility Systems (UAMS) are an emerging form of transportation that promises to combine the best of ground-based and air-based transportation and to overcome the problems associated with either of these forms of transportation.

PATS are discussed as a mode to reduce urban congestion by making use of the free space in the air. There are a lot of barriers for the implementation of Personal Aerial Vehicles (PAVs). Firstly, there are aspects such as potential business cases (e.g. ownership or rental), usage scenarios (e.g. commuting or leisure) and technical requirements (electric or fossil fuel-based propulsion) for PAVs.

Moreover, due to the expected high interaction between existing transport systems and PATS it is fundamental to understand and evaluate the effects of UAM introduction. Therefore, it will be necessary to adapt all transport modes (stations and schedules, for example) so that maximum benefit can be obtained and, effectively, reducing travel times.

Analyzing PATS transport performance and potential modal shares is of particular importance. Hence, development of a PAV model and simulation environment is required. However, these simulations are not easy to carry out, as there are several obstacles. Firstly, because they require a great deal of preparation of the entire scenario. For the simulation to be valid and to obtain reliable results, the synthetic population and the network generated must be representative of the reality. To obtain this representative population, population specific data are needed, which is often not trivial to obtain. In addition, these simulations, the more detailed they are, the more expensive they are in terms of time and computer memory, which introduces the need to carry out a metamodel that can predict the results of the simulations.

1.3 Objectives and deliverables

This work has two main objectives:

- to instantiate an agent-based transport model of on-demand UAM services in São Paulo Metropolitan Region under the MATSim traffic simulator using the UAM extension. The analysis consists, studying the impact of this new urban transport paradigm on the overall transportation system performance, when compared with traditional transportation means. The study will try different selected UAM system properties, such as the number of stations and their locations, total process time, fleet size, vertical and cruising speed, vehicle capacity and costs. To reach this goal, it is necessary to create a synthetic population, a network, a model that allows the agents to do their choices and it is necessary to define the UAM service.
- create a metamodel based on Gaussian Processes with active learning, to predict the simulation outputs instead of doing the simulation itself, which can be a time-consuming process.

1.4 Thesis outline

This thesis is divided in 5 chapters, including this one, and are structured as follows:

- **Chapter 2:** presents a description of the UAM service and of the studies that have been carried out. Describes the different methods that allow carrying out transport simulations, and explains how the MATSim framework works together with its UAM extension. Furthermore, it describes simulation metamodels using Gaussian Processes and active learning.

- **Chapter 3:** describes the methodology used to carry out the simulations, how the scenario was prepared and presents the various results obtained.
- **Chapter 4:** describes the methodology used in the realization of the metamodel together with the creation of the Corsica scenario and presents the results obtained.
- **Chapter 5:** summarizes the work performed in this thesis with a conclusion and presents some recommendations for future work.

Chapter 2

Background and theoretical overview

It is essential, in any scientific work, to have knowledge of the state of the art. This allows you to see the work that has been developed in the area, as well as the possible improvements that can be made.

This chapter presents a brief description of the UAM service in general, the evolution of transport simulation methods, the MATSim framework, some UAM studies and simulation metamodels via machine learning. In addition, it also presents some basic concepts needed to understand this work.

2.1 UAM service – The future

UAM is one of the great current trends, with regard to urban mobility. With large cities reaching an extreme level of congestion, this new mode of transport has attracted the attention of the general public, as a way to substantially reduce the time spent in traffic jams. Thus, there are more than 100 vehicles in development worldwide, some still in the design phase, others already undergoing tests, according to Gollnick et al. [5], that will allow this new technology to advance. These vehicles can be divided according their morphology. Figure 2.1 presents the main differences between them.

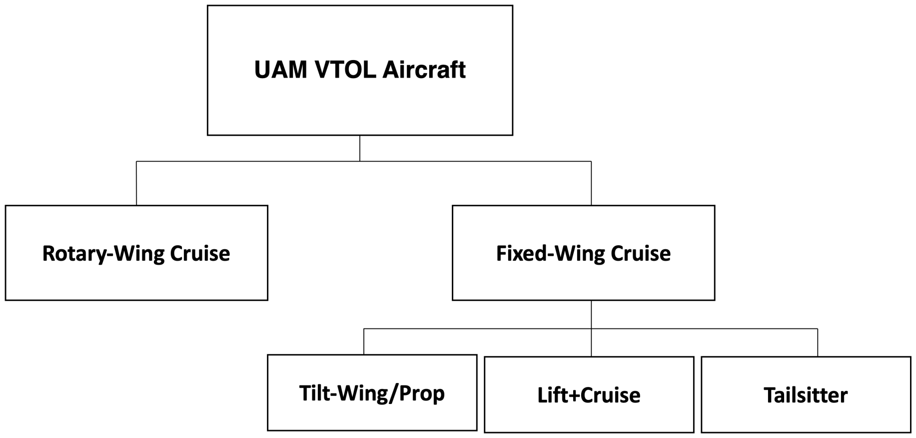


Figure 2.1: Vehicle morphologies for VTOL aircraft (adapted from Shamiyeh et al. [6]).

Although the new service can be used in many ways, there are 4 main ways that are usually foreseen, according to Straubinger et al. [7]: passenger transport, medical service, logistics and surveillance. The passenger transport service will allow people to travel from one point to another in the city quickly. This service can be scheduled, such as public transport or on demand, that is, the person requests the service and a vehicle is reserved for him. The medical service will be similar to what is currently provided by helicopters. The logistics will allow, for example, deliveries to be made in more remote areas of cities and in a faster way and the surveillance will allow to monitor the territory, whether at sea or on land, such as in zones where access is difficult.

In order for these vehicles to operate, infrastructures on land are needed, since these vehicles need specific infrastructures. The infrastructures can be called stations. As with other transports, this transport mode will have a preflight and a postflight time, which represent the time spent by people at the stations. There will also be a boarding and deboarding process, which is the time needed to get in or out of the vehicle. Due to advances in technology and since there is a great investment in electrification, a turn around time may be necessary, that is, a period in which the vehicle is inoperative to charge batteries. In addition to this, new regulations regarding airspace will be needed, which is not so trivial. These regulations will have as main objectives to define which airspace to use, maximum speeds, minimum distance between vehicles, maximum number of vehicles that can circulate, per hour, in a given area, as well as acoustic and environmental issues.

In conclusion, although this service still needs some advances, mainly with regard to the level of regulations, it is expected that this service will start operating within the next 5 years [8].

2.2 Transport modelling and simulation

Attempts to model transport go back over 140 years. According to Trattner [9], one of the first attempts to model transport systems, dates back to 1881, with the first isochoric map of travel times, by Galton [10] - Figure 2.2.

Figure 2.2: Isochronic map of travel times from London, United Kingdom, made in 1881 by Galton [10].

In 1955, the emergence of aggregate trip-based models, called four-step models, allowed for a breakthrough in transport modeling, by Weiner [11]. Four-step models are "the most common approach for travel demand modeling" according to Moeckel et al. [12] and, nowadays, are largely based on the framework proposed by Manheim [13]. They are a trip-based method and only consider network load, regarding different origin and destination pairs, according to Straubinger and Rothfeld [14]. This method consists of four sub-models that can be performed in different orders. These models can not capture the interactions individuals, and they ignore their decisions and behaviour, according to Hörl and Balac [15]. PTV Visium[1], for example, uses the four-step modeling methodology of (1) trip generation, (2) trip distribution, (3) mode choice and (4) assignment.

Subsequently, activity-based models emerged, in response to the failures of four-step models. These models are still in development today and some examples are presented in Moeckel et al. [12], Axhausen and Gärling [16] and Bowman and Ben-Akiva [17]. The individual's activities are limited by social and personal constraints and these models allow for scheduling activities and making mode and destination choices at the individuals' scale, within the household context. These models are based on the work developed by Chapin [18] and by Hägerstrand [19]. A disadvantage of these models is that they involve a range of econometric sub-models that need to be estimated, and later calibrated to fit the data. ActivitySim[2] is an example of a software that uses this methodology.

In addition to these methods, agent-based models were developed thanks to computational growth, according to Hennessy and Patterson [20], and are based on a synthetic population, described by Bonabeau [21].

These models have several applications, and they can be used to model complex adaptive systems, to make predictions in the area of biology, epidemiology, business, technology and network theory, in economics and social sciences and in water management.

Regarding the agent-based models that try to model complex adaptive systems, such as transit networks, the aim is to simulate the behavior of vehicles, the transport network and agents (persons), as well as their ability to adapt to the environment by maximizing the utility of each agent's plans, making sure that they choose the best route. Furthermore, these models allow modeling extremely dynamic services on a shorter timescale than activity-based models. These models are able to integrate the various phases of the four-step model into one coherent system and offer many advantages compared with conventional approaches in traffic simulation, according to Bazzan and Klügl [22], such as facilitating modelling data and behaviour heterogeneity. MATSim[3] and AnyLogic[4] are an example of softwares that uses this methodology. According to Chen and Cheng [23], agent-based models have been applied to several cases, which can be divided into five categories: traffic control and management system architecture and platforms, roadway transportation, air-traffic control and management, railway transportation and multiagent traffic modeling and simulation.

[1] https://www.ptvgroup.com/en/solutions/products/ptv-visum/
[2] https://activitysim.github.io
[3] https://www.matsim.org
[4] https://www.anylogic.com

2.2.1 Traffic simulations classification

The traffic simulations can be divided in four types: macroscopic, mesoscopic, microscopic and nanoscopic, according to Kokkinogenis et al. [24]. Macroscopic simulation models consider each vehicle in the same way and as a group and, therefore, are not good when the objective is to obtain simulations with great detail. Therefore, this model, which is based on high-level mathematical models, is useful when one intends to study a large area, such as motorway networks and interregional road networks.

On the contrary, microscopic simulators simulate individual entities, such as vehicles and drivers, with a high level of detail. These simulators, since they present a higher level of detail, are more realistic than the macroscopic simulators and, many times, are the emergent properties of the microscopic simulation. Therefore, microscopic simulators are useful for evaluating new traffic control and management measures.

The mesoscopic simulators aim to fill the gap between macro and micro simulators. In this case, the vehicles can be grouped and are treated as one entity. Finally, the nanoscopic model is a relatively new trend of traffic simulation and is mostly used in autonomous driving.

As the work was developed with the mesoscopic MATSim framework, its operation will be described in detail in Section 2.3.

2.2.2 Discrete choice model

As mentioned, in order to obtain results closer to reality, more complex models are needed. However, an increase in complexity also increases the time spent in each simulation, so it is essential to have computationally fast models to estimate and predict the behavior of people. Thus, there is a big interest in discrete choice modelling, which intends to predict the choice of an agent in relation to a particular mode of transport, route or destination. According to Sebastian Hörl and Axhausen [25], those are based on the expected travel characteristics in combination with socio-demographic factors. Equations 2.1 and 2.2 are used to compute the utility, u, and present an example that contains the mentioned parameters, where α is mode parameter, β is a marginal utility parameter, x represents the travel characteristics and c is a cost.

$$\begin{aligned} u_{car} = \alpha_{car} &+ \beta_{travelTime,car} * x_{travelTime,car} \\ &+ \beta_{cost} * c_{car} \end{aligned} \tag{2.1}$$

$$\begin{aligned} u_{pt} = \alpha_{pt} &+ \beta_{travelTime,pt} * x_{travelTime,pt} \\ &+ \beta_{cost} * c_{pt} \end{aligned} \tag{2.2}$$

The discrete choice models are considered statistical models that assign probabilities, $\mathbb{P}$, to specific alternatives in order to perform a decision task, using the utilities obtained. In the case of a simple Multinomial Logit Model, according to Sebastian Hörl and Axhausen [25], one would have:

$$\mathbb{P}(car) = \frac{exp(u_{car})}{exp(u_{car}) + exp(u_{pt})} \tag{2.3}$$

For the creation of these models, surveys of the population in question are necessary, so that the parameters presented in the equations can be estimated and the model has a good representation of the reality. Otherwise, global positioning system (GPS) analyses can be used. Fu et al. [26] is an example of a work aimed at defining the parameters of the city of Munich regarding the introduction of autonomous vehicles that used a survey.

When talking about discrete choice mode models, there are two different statistical models: trip-based or tour-based.

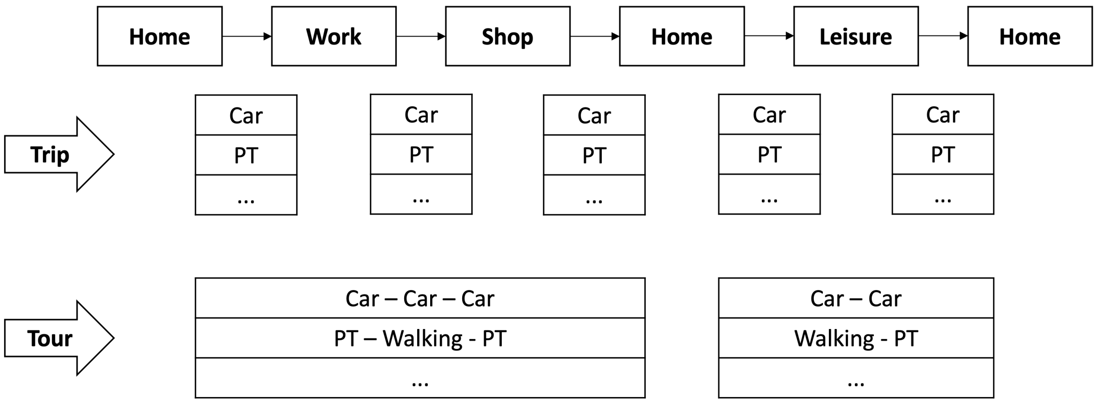


Figure 2.3: Difference between trip based and tour-based choice process (adapted from [25]).

The difference between the trip-based and the tour-based lies in the order of decisions, as Figure 2.3 illustrates. While, in trip-based, all the choices are made step by step, in the tour-based the choices are made keeping in mind the complete tour (a tour represents a set of trips that starts and ends in the same place, home for example). In other words, in the first case, the first choice influences all the others, in the sense that if the agent makes the first trip on foot to work and then has a distance of 50km to the gym, this will have to be done by public transport , while in the second case, the agent would no longer make the first trip to work on foot, knowing that he would then have a distance of 50km to cover.

One example of such a model is POLARIS [27]. In this case, the agents perform mode choices on a trip level and also consider vehicle availability constraints – an agent only has the vehicle available if he has traveled to the current location with it. L Miguel Martínez and Lopes [28] incorporate a mode choice model within an agent-based model to simulate carsharing operations in the city of Lisbon. Here, mode choice modelling is performed on a trip level, and vehicle constraints are not considered.

Sebastian Horl and Axhausen [29] and Sebastian Hörl and Axhausen [25] represent the first attempts to try to combine discrete choice mode models with MATSim, the simulation model that is used in this work.

2.3 MATSim framework

The Multi-agent Transport Simulation (MATSim) [30] is an open-source and java-based tool, developed by ETH Zürich and TU Berlin and which simulates the daily life of individuals. Is described as mesoscopic transport simulation framework as it is generally used for large scenarios and the smallest spatial units are road links [29], although it can also be considered microscopic transport simulation framework [31].

In order to understand how MATSim works, it is necessary to understand some concepts. First, the scenario that is simulated contains a synthetic population, consisting of agents (travel demand), a network, the location of the facilities and the vehicles. With the creation of the synthetic population, plans are assigned to agents, and plans are, as Dominik Ziemke and Moeckel [32] describes, a chain of activities (e.g. home–education–shop–home), which include their locations and end times. In turn, the network is represented through links, which simulate roads and these are linked by nodes and allow agents to be moved. In Figure 2.4 it is possible to see an example of a network, which is presented in the MATSim book, by Andreas Horni and Axhausen [30].

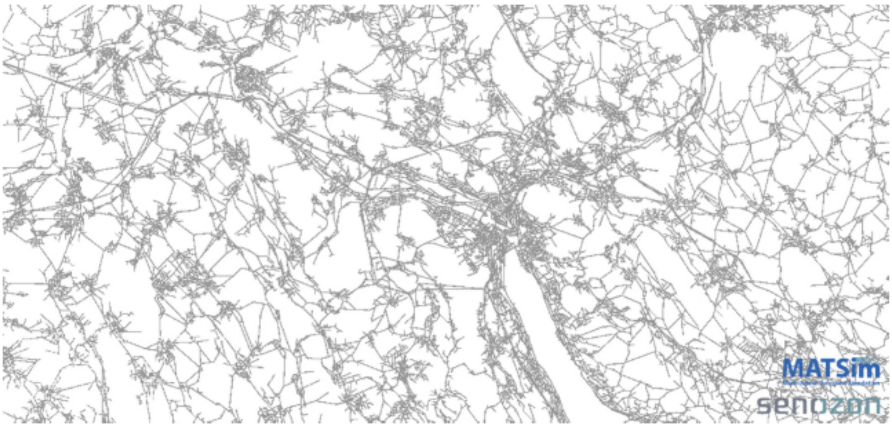


Figure 2.4: Zurich Network [30].

According to Felix Zwick and Axhausen [33], MATSim "utilizes an iterative, co-evolutionary learning approach" in which, the main goal of the agents is to optimise their daily plan of activities. The execution loop of MATSim, represented in Figure 2.5, is based on three major events:

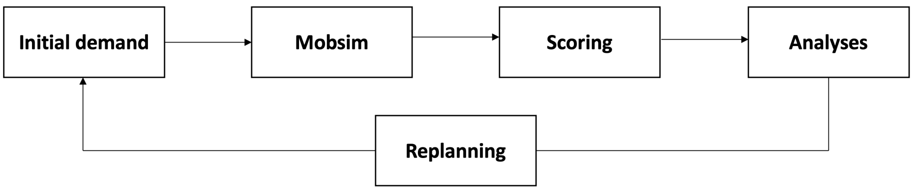


Figure 2.5: MATSim loop (adapted from [30]).

- **(1) Mobility Simulation** or mobsim, which executes the agents' plans. The available transport modes during the simulation, will allow the agents to perform their activities;

- **(2) Scoring**, a score is assigned to the plans executed by agents during the mobsim, through utility functions - as illustrated in Figure 2.6, present in the MATSim book by Andreas Horni and Axhausen [30]. Working, for example, is an activity that gives a positive score (since it generates a monetary income), while commuting to work gives a negative score;

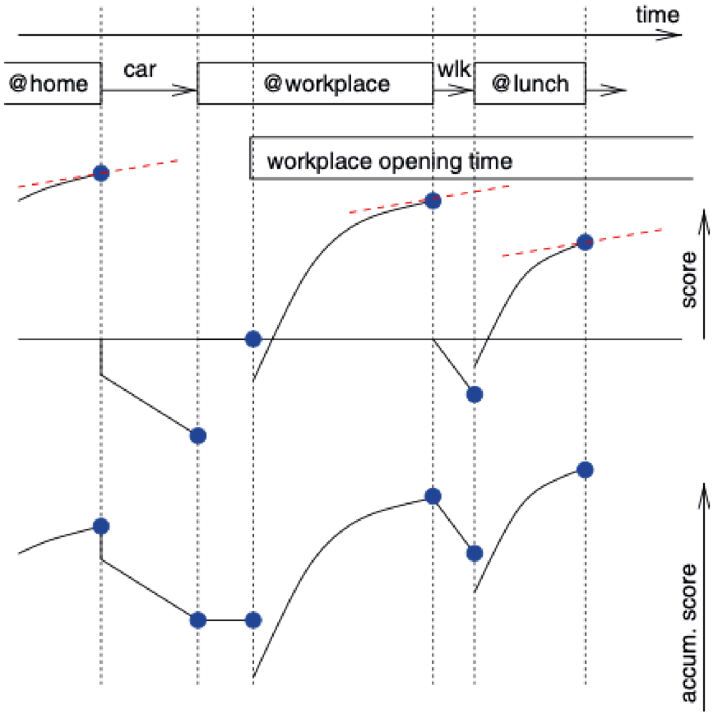


Figure 2.6: Illustration of the scoring function in MATSim [30].

- **(3) Re-planning** allows agents to change their plans, via innovative strategies, in order to improve their score. In this event, not all the agents are able to re-planning their plans and the ones that are able to change it, can modify the mode, departure time, route or trip of the previous plan.

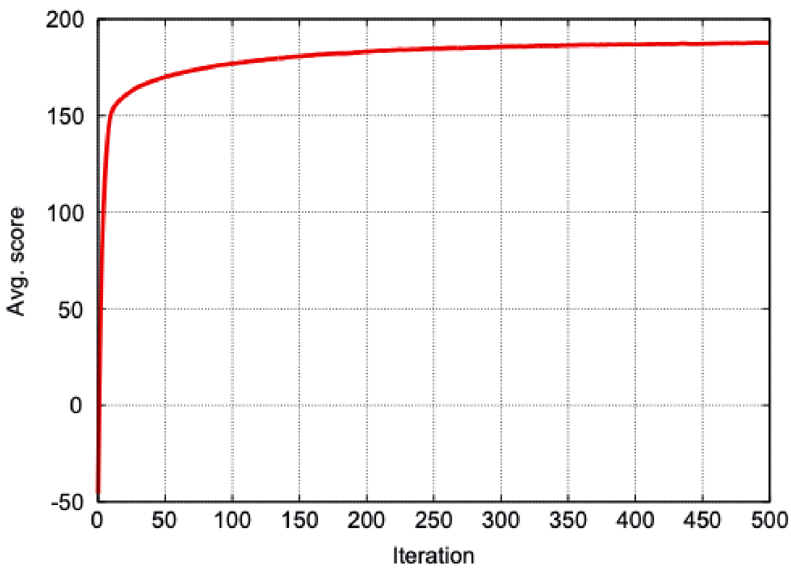


Figure 2.7: Typical score progress in MATSim [30].

The execution of all plans (mobsim), its scoring and re-planning is called an iteration. This process is repeated until the system reaches a Nash equilibrium, that is, as Dominik Ziemke and Moeckel [32] describes, the further development of agents' plan scores is "sufficiently relaxed", as illustrated in Figure

2.7. Through the example presented in figure, it is possible to see that the average score it is constant after the iteration 200, meaning that the agents are not more able to improve their plans and the system reached a Nah equilibrium.

Some of the parameters that allow the evolution of the score throughout the simulation are defined in the configuration file and can look like this in Figure 2.8, present in the MATSim book by Andreas Horni and Axhausen [30]. In addition to these parameters, it is also necessary to parameterize each transport mode, with, for example, a constant specific to each transport mode and a linear coefficient that translates all choice dimensions into a generalized utility.

```
<module name="planCalcScore" >
    <param name="performing" value="6.0" />
    <param name="waiting" value="-0.0" />
    <param name="earlyDeparture" value="-0.0" />
    <param name="lateArrival" value="-18.0" />
    <parameterset type="activityParams" >
            <param name="activityType" value="work" />
            <param name="typicalDuration" value="08:00:00" />
```

Figure 2.8: Some scoring parameters in MATSim.

Thanks to its versatility, MATSim has been continuously developed and extended, particularly with regard to the introduction of new transport modes, such as car-sharing, visible in Ayed et al. [34], Ciari et al. [35], Balac et al. [36], ride-hailing and sharing in Segui-Gasco et al. [37], Bischoff et al. [38], or Shared Autonomous Vehicle (SAV) services in Vosooghi et al. [39], Fagnant and Kockelman [40], Winter et al. [41]. Most of these studies, which simulate novel concepts of mobility, build upon the developments by Maciejewski [42] who introduced the ability to model dynamic vehicle routing with his Dynamic Vehicle Routing Problem (DVRP) extension for MATSim. The DVRP allows to assign a vehicle to an agent's travel request.

2.3.1 UAM extension

According to Balac et al. [43], MATSim has been "the only attempt to simulate commercial flights using an agent-based approach, where both aircraft and passengers are modeled on an agent level". This inclusion was done by Grether [44] with its focus on conventional commercial air transport with large passenger aircraft.

As mentioned in Section 2.3, MATSim has been used and extended to simulate the introduction of innovative modes of transport. Thus, since MATSim has all the conditions to extend its application to the UAM service, an extension was created in order to be able to simulate this transport mode together with the existing ones and thus predict the impact that this introduction will have. This one arose from a collaboration between Bauhaus Luftfahrt e.V., the Eidgenössische Technische Hochschule Zürich, and the Technical University of Munich and has since been published as an open-source project[5].

Under the assumption of on-demand operational models for UAM, the transport modeling of VTOL vehicles mirrors that of autonomous ground-based taxis, according to Rothfeld [45]. Thus, the UAM ex-

[5]https://github.com/BauhausLuftfahrt/MATSim-UAM

tension for MATSim is also based on the Dynamic Vehicle Routing Problem (DVRP) MATSim contribution by Maciejewski [42].

This extension allows to define parameters regarding the following aspects, inherent to this transport mode:

- **Stations:** are the necessary infrastructure for the take-off and landing of vehicles, and may be on the ground or on top of buildings. These should allow agents to change their mode of transport;
- **Flight paths:** link stations directly (euclidean distance) or via waypoints;
- **Vehicles:** circulate through flight paths, from station to station. Each vehicle must be capable of vertical take-off and landing and a cruise phase, with each phase having different speeds;
- **On-demand operation:** combines UAM stations, flight paths and vehicles to enable this mode of transportation.

2.4 UAM studies

As already mentioned, the UAM service is the future. Therefore, several studies have been carried out with regard to the vehicles that will be used, but few have focused on the changes that will occur in cities with the introduction of this new transport mode.

In relation to the different areas that need to be studied, mentioned in Section 2.1, Straubinger et al. [7] provides an overview of the studies that have been carried out regarding the vehicles. This article takes into account the different aspects related to vehicles, as well as their limitations and regulations. It also analyzes possible market structures and the integration of this new service in relation to other transport modes. In the end, they also analyze the potential users and public acceptance as well as UAM service modeling approaches.

Balac et al. [43] presents some research gaps that must be addressed before this new service is introduced into society. This paper also proposes some concepts relating to the various areas that must be studied, in order to be able to simulate and integrate this service in the future.

In turn, Straubinger and Rothfeld [14] evaluates relevant prerequisites for Personal Air Transport System introduction into the urban transport modelling environment. This article discusses different modes of operation as well as presents different demand drivers. In addition, it analyzes potential cities for the integration of the UAM service and the main requirements required. In the end, it explores different approaches to transport modeling.

Al Haddad et al. [46] presents an approach for the selection of indicators for a multi-criteria analysis for the assessment of UAM, in a case study of Upper Bavaria, Germany. The article presents a 5-stage approach with an expert assessment.

With regard to the vehicles to be used, Christopher Silva and Patterson [4] presents a pool of candidates for this service, as well as an overview of batteries, engines and gear box technologies.

In order to better understand the level of acceptance of the population to the UAM service, Fu et al. [26] demonstrates the results of a survey of the people of Munich. This article intends to understand the impact of autonomous transport modes, deriving the main characteristics of potential users who can adopt autonomous transport services.

Regarding the simulations that have been carried out to predict the impacts of the introduction of UAM and identify measures that need to be implemented, there are already several scenarios that have been simulated. Bulusu et al. [47] explores the addressable market for UAM as a multi-modal alternative in a community in the San Francisco Bay Area, using MATLAB. The authors develop a traffic demand analysis method, under two criteria of flexibility to travel time savings and three criteria of vertiport transfer times, to estimate the maximum number of people that can benefit from UAM.

Balac et al. [48] and Balac et al. [49] present a methodology for simulation and demand estimation for personal aerial vehicles in Zurich, Switzerland using MATSim with discrete mode choice models. Then, the impacts of PAVs with different vehicle and system parameters are analysed.

Rothfeld et al. [50] uses the MATSim UAM extension to simulate a test scenario in Sioux Falls, USA, in order to have an outlook on the transport performance of UAM with varying parameters.

Plötner et al. [51] define the required models and methods to analyze and quantify the potential demand for urban air mobility and possible impacts were defined and applied to the Munich Metropolitan region. The authors use an existing agent-based transport model of the study area to analyse three different UAM networks with different numbers of vertiports. To simulate the demand for this new mode they derive an incremental logit model for UAM. In the end, they perform sensitivity studies of ticket fare and structure, flying vehicle cruise speed, passenger process times at vertiports and different UAM networks sizes.

Rothfeld et al. [52] compares three pre-existent, and calibrated agent-based transport scenarios from three different cities, Munich, Paris and San Francisco, in terms of potential travel time savings, using the UAM-extension for MATSim. The authors, calculate congested trip travel times for each trip's original mode (car, public transport and UAM). In the end, the resulting travel times are compared and allow the deduction of potential UAM trip shares under varying UAM properties, such as the number of stations, total process time, and cruise flight speed.

Globally, several studies have been carried out, with regard to the most diverse areas that need to be studied, in order to be able to introduce UAM in the near future. However, there is still little progress regarding the laws that will govern this new transport mode and one of the possible facts for this is the difficult forecast of what will happen when the UAM service is introduced. That said, it is necessary to quickly understand what effects UAM will have in the different cities, taking into account that it will not be the same in all cities. In order to understand what will change and the level of acceptance by the population, two ways can be used. In the first, surveys can be made, where people are asked about whether they would be willing to use the UAM and, if so, under what conditions. If there is great reticence in acceptance, it is also important to understand the reasons, in order to adjust the service. Regarding the second form, it is possible to predict the changes in the societies based on simulations close to reality. In the end, the two forms can and should complement each other, since surveys can help to increase the level of reality in simulations.

2.5 Simulation metamodels via machine learning

The main objective in carrying out simulations is to study the performance of urban systems and to obtain a good approximation of the reality. In order to draw conclusions from these simulations, it is necessary that they present a great level of complexity. However, by increasing the level of complexity, they become computationally expensive and can exhibit prohibitive simulation runtimes.

In order to overcome this shortcoming, simulation metamodels are very often used to approximate the underlying simulation function. Therefore, these models can be considered functions that approximate the complex true function inherent to the real simulation. According to Kleijnen and Sargent [53], the use of simulation metamodels has four goals: real-world problem understanding, simulation output prediction, optimization, and verification/validation.

In more complex scenarios and where simulation data can be difficult to obtain, according to Antunes et al. [54], the use of simulation metamodels in conjunction with active learning proves to be an important tool. Despite being an old topic in the field of statistics and probabilities, the use of these processes in conjunction with machine learning only emerged in the last decade.

Among many significant machine learning tools, Gaussian Processes (GP) framework, by Rasmussen and Williams [55], for its Bayesian and non-linear modeling properties, allows to develop methodologies that combine active learning and metamodeling strategies.

In the end, if the simulation metamodel is very close to the simulation function, a reasonable number of simulations can be bypassed and, therefore, time and computing power are saved.

2.5.1 Gaussian processes

As described by Rasmussen and Williams [55], a GP is a stochastic process that can be defined by a mean and a covariance function (also known as kernel function), $m_f(\mathbf{x})$ and $k_f(\mathbf{x}, \mathbf{x'})$, respectively, being **x** and **x'** two input data points and they are represented in the follwing equations.

$$GP(m_f(\mathbf{x}), k_f(\mathbf{x}, \mathbf{x'})) \tag{2.4}$$

$$m_f(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \tag{2.5}$$

$$k_f(\mathbf{x}, \mathbf{x'}) = \mathbb{E}[(f(\mathbf{x}) - m_f(\mathbf{x}))(f(\mathbf{x'} - m_f(\mathbf{x'})] \tag{2.6}$$

One of the applications of GP is to perform regression via supervised learning, which is called Gaussian Process Regression (GPR). In this context, the GP framework places a prior over functions with continuous domain, i.e., $y = f(x) + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$ and $f(x) \sim GP(m_f(\mathbf{x}), k_f(\mathbf{x}, \mathbf{x'}))$

Regarding the kernel function, it measures the "similarities" between two different predictions. It means that a higher kernel function value implies that the two predictions are more similar. The kernel and the mean function have some free parameters – hyper-parameters of the GP – which can be optimized by marginal likelihood maximization subjected to the training data. After these parameters are obtained, the conditional distribution for any unlabeled test point x_* is given by $f_*|X, \mathbf{y}, \mathbf{x}_* \sim N^T(k_{f*}[K_y]^{-1}\mathbf{y}, (k_{f**} - K_{f*}^T[K_y]^{-1}k_{f*})$ where $k_{f*} = k_f(X, \mathbf{x}_*)$, $k_{f**} = k_f(\mathbf{x}_*, \mathbf{x}_*)$, $[K_y]$ is the covariance matrix, X the design matrix and $\mathbf{y}$ is the vector of the output values.

2.5.2 Active learning

Active learning allows any algorithm to actively choose the training data points during the learning process. This is extremely useful in scenarios where labeled data is expensive to obtain. Therefore, this paradigm choose the most informative points from which it learns.

An arbitrary active learning strategy includes five elements, (L, U, M, O, Q). The first element, L is the labeled training data set and U is the set of unlabeled data points. Usually, the number of unlabeled data points is much higher than the labeled set. M is the machine learning model, which can be a classification or a regression model, what makes the values in L being discrete or continuous, respectively. O is the oracle, which function is to provide labeled values. The last, Q, is the query function which contains the strategies and criteria for finding and selecting the most informative instances of U to be added to L. According to Settles [56], there are several query frameworks that can be adopted, such as error reduction by Roy and Mccallum [57], query-by-committee by Seung et al. [58], uncertainty sampling by Lewis and Gale [59], density-weighted methods by Settles and Craven [60], variance reduction by Geman et al. [61] and expected model change Settles et al. [62]. Related with the query function, this learning paradigm can be divided into pool-based or stream-based. In the pool-based, each individual data point is presented serially or in successive blocks, whereas in the stream-based all the unlabeled instances are available for querying.

Since this is an iterative process, there must be a stopping rule. In general, this criteria should take into account the performance of the model, is generalization capacity and the associated costs of acquiring new labeled data.

Although this strategy is used in a wide range of fields, in this work we focus mainly on applications of simulation metamodels.

2.5.3 Related works

The application of simulation metamodels based on GP is still rare, however, there are some works that have been developed. These works can be divided into traffic prediction or optimization of networks, according to Song et al. [63].

Ciuffo et al. [64] adopt a metamodel-based technique and conduct a sensitivity analysis using the mesoscopic traffic simulator AIMSUN. The authors also investigate the importance of selecting a proper sampling strategy based on low-discrepancy random number sequences and the importance of selecting a class of metamodels able to reproduce the inputs–outputs relationship in a robust and reliable way. The authors conclude that the estimated outputs from the metamodel and the real simulation outputs were very similar.

Chen et al. [65] also apply the metamodel for mesoscopic simulation to solve the bi-level Mixed Network Design Problem (MNDP) to enable dynamic traffic assignment (DTA) and the simulation-based optimization solution. The authors propose a surrogate-based optimization framework to solve the MNDP that is characterized by expensive-to-evaluate objective functions, using as a case study the large-scale Montgomery County network in Maryland.

Zhang et al. [66] utilizing a simulation-based dynamic traffic assignment model, propose a Bayesian stochastic Kriging metamodel to optimize integrated planning and operational ATM strategies for corridors using the real scenario of I-270 and MD355 in the state of Maryland, USA.

Antunes et al. [67] propose an active learning algorithm based on the GP framework. The methodology used gathers the most informative simulation data points in batches, according to both their predictive variances. The most informative data points are also selected according to the relative distance between them, in order to explore the simulators' input space with fewer data points and in a more efficient way. In this work, a simple traffic simulation and a Demand-Responsive Transportation (DRT) system simulator is used and the results show that the methodology used is able to improve the exploration efficiency of the simulation input space

Chen et al. [68] adopts a surrogate-based optimization approaches to approximate the response surface for the transportation simulation input–output mapping and search for the optimal toll charges in a transportation network. The authors use a simulation-based dynamic traffic assignment model DynusT (Dynamic Urban Systems in Transportation) to evaluate the system performance.

Chapter 3

UAM simulation in MATSim - São Paulo case study

3.1 Methodology

To carry out the simulations, the eqasim framework which is based on the MATSim framework with added components for simulation of discrete choice models was used together with the UAM extension, available on GitHub[1]. Therefore, it was necessary to create a synthetic population together with the network, to define the discrete mode choice model, that is, the equations that govern the model and the UAM service, which includes the designation of vehicles and stations.

3.1.1 Population and network

The creation of the synthetic population and the network of São Paulo Metropolitan Region was made based on open data and using the eqasim's repository available on GitHub[2]. The data required for its creation are listed below:

- **Census data 2010**, containing the socio-demographic information of people living in Brazil were available from Centro de Estudos da Metrópole[3];
- **São Paulo household travel survey 2017** was available from Transportes Metropolitanos[4];
- **OpenStreetMap data** which was available on Geofabrik[5];
- **Educational facilities** were available from Dados Abertos da Educação[6];
- **Census zoning system** was available from Centro de Estudos da Metrópole[7];

[1] https://github.com/eqasim-org/uam
[2] https://github.com/eqasim-org/sao_paulo
[3] https://centrodametropole.fflch.usp.br/pt-br/controle-acesso
[4] http://www.metro.sp.gov.br/pesquisa-od/
[5] http://download.geofabrik.de/south-america/brazil/sudeste.html
[6] https://dados.educacao.sp.gov.br/dataset/endereços-de-escolas
[7] https://centrodametropole.fflch.usp.br/pt-br/controle-acesso

- **Road network (OpenStreetMap)**, that was obtained thanks to the file obtained in the third item;
- **Public transit schedule (GTFS)** that was obtained from the OpenMobilityData[8].

Regarding the synthetic population, only about 1% of the total population residing in São Paulo was generated, in order to reduce the time of the simulations.

The generated network does not yet have the UAM service. In Figure 3.1, it is possible to see an overview of the created network (a) as well as a more detailed view (b). This network represents well the real network, since it was obtained using the Open Street Map. The Figures were obtained using Simunto Via[9], a tool that supports the loading of the typical MATSim input and output files and thus allows extracting a huge amount of data from the visualization simulation results such as tracking a single vehicle, activity times and link volumes.

3.1.2 Mode choice model

In order to use the discrete mode choice model, it was necessary to define the equations that will allow, during the simulation, to obtain a probability for each option, in order to allow the agent to make the best decision. The equations are already defined in the calibrated São Paulo scenario, in Eqasim's GitHub repository[10], so they have not been changed. That said, it was only necessary to define the equation of the UAM service. These are represented from 3.1 to 3.5

The options available are UAM, car, PT, taxi and walk. Bike is not considered because its modal share is zero in the real case [69], having in account the data of 2020.

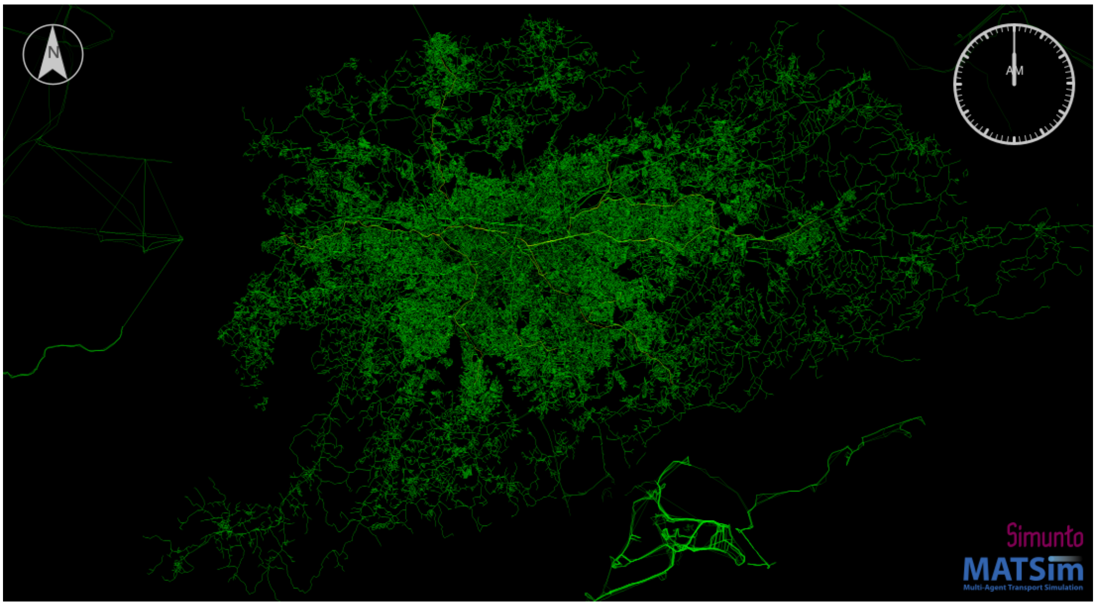
Looking at the equations, it is possible to see that there are several parameters that must be defined, with α being a constant specific to the transport mode, β a marginal utility parameter, C represents the cost of the trip and x is related with per-trip choice dimensions.

$$\begin{aligned} u_{UAM} = \alpha_{UAM} &+ \beta_{travelTime,UAM} * x_{travelTime,UAM} \\ &+ \beta_{cost} * c_{UAM} \end{aligned} \tag{3.1}$$

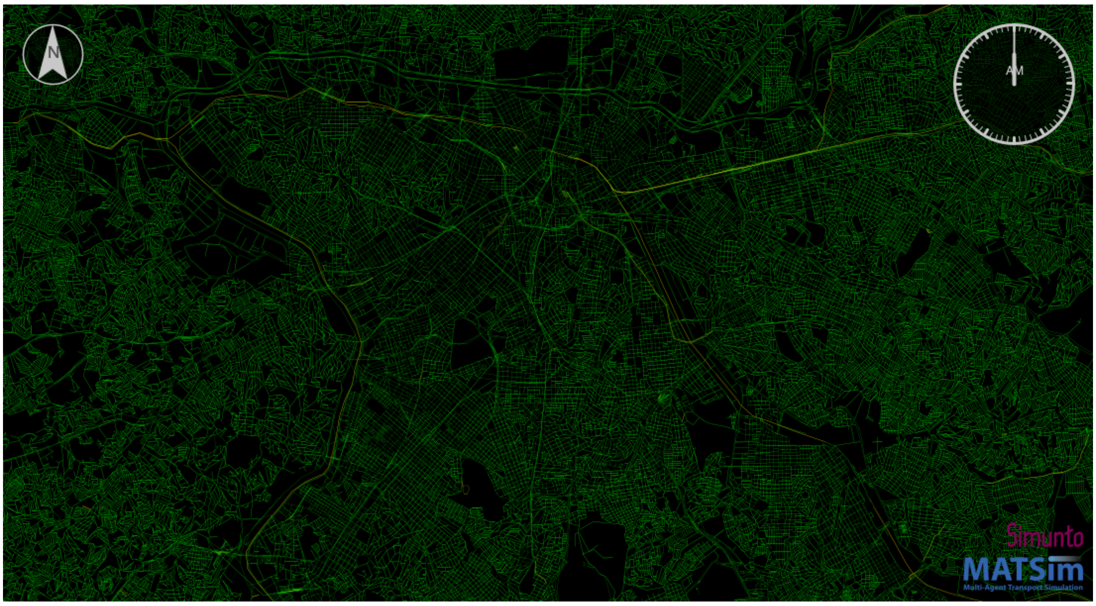
$$\begin{aligned} u_{car} = \alpha_{car} &+ \beta_{travelTime,car} * x_{travelTime,car} \\ &+ C_{trip} \\ &+ \beta_{accessEgressTime,car} * x_{accessEgressTime,car} \\ &+ \beta_{cost} * c_{car} \end{aligned} \tag{3.2}$$

$$\begin{aligned} u_{pt} = \alpha_{pt} &+ \beta_{travelTime,pt} * x_{travelTime,pt} \\ &+ \beta_{accessEgressTime,pt} * x_{accessEgressTime,pt} \\ &+ \beta_{waitTime,pt} * x_{waitTime,pt} \\ &+ \beta_{transfer,pt} * x_{transfer,pt} \\ &+ \beta_{cost} * c_{pt} \end{aligned} \tag{3.3}$$

[8]http://transitfeeds.com/p/sptrans/1049
[9]https://simunto.com/via/
[10]https://github.com/eqasim-org/eqasim-java



(a) General view.



(b) Detailed view.

Figure 3.1: Network of the city of São Paulo without UAM.

$$\begin{aligned} u_{taxi} = \alpha_{taxi} + \beta_{travelTime,taxi} * x_{travelTime,taxi} \\ + \beta_{accessEgressTime,taxi} * x_{accessEgressTime,taxi} \\ + \beta_{cost} * c_{taxi} \end{aligned} \tag{3.4}$$

$$u_{walk} = \alpha_{walk} + \beta_{travelTime,walk} * x_{travelTime,walk} \tag{3.5}$$

Having already defined the equations, it was necessary to assign values to the various variables. Table 3.1 presents the values of each variable for each transport mode. The parameters of car, public transport, taxi and walk were based on the calibrated scenario. Regarding the definition of the UAM service, the new service added, the parameter values were based on the values of public transport. With regard to the walk mode, there is also a penalty of -100 in the utility, if agents travel more than 20km on foot.

Table 3.1: Discrete mode choice parameters definition.

Parameter	UAM	Car	Public transport	Taxi	Walk
α	-0.2	0	-0.2	-3.0	2.2
$\beta_{travelTime}$	-0.0142	-0.0246	-0.0142	-0.15	-0.1257
β_{cost}	-0.0606	-0.0606	-0.0606	-0.0606	-
C_{trip}	-	-0.1597	-	-	-
$\beta_{accessEgressTime}$	-0.0142	-0.0246	-0.0142	0	-
$\beta_{waitTime}$	-0.0142	-	-0.0142	-	-
$\beta_{transfer}$	-	-	0	-	-

3.1.3 UAM service

Regarding the definition of the UAM Service, this was made based on the repository availabe on GitHub MATSim-UAM[11], which allows the creation of a file containing information about vehicles and stations, as well as updating the network, so that it includes the UAM service. It was then necessary to define the locations for placing the stations (x, y, and z coordinates), as well as the parameters listed below, which concern the stations and vehicles.

- **Stations:**
  - **Station ID and name**;
  - **VTOL distance:** the distance, in meters, that vehicles have to travel to go from the take off to the cruise phase or to land;
  - **Preflight time:** represents the time, in seconds, that an agent has to be at the station before being able to board;
  - **Postflight time:** represents the time, in seconds, that an agent has to be at the station before being able to leave;
  - **Default waiting time:** represents the average waiting time, in seconds, for passengers, when there is no information on the time of previous iterations (used in iteration zero);
  - **Ground access capacity and free-speed:** represent the capacity that a station has to receive agents in its base, in one hour, and the speed (m/s), respectively;

[11] https://github.com/BauhausLuftfahrt/MATSim-UAM

- **Flight access capacity and free-speed:** represent the maximum capacity of the links in one hour, as well as the maximum allowed speed (m/s), respectively.

- **Vehicles:**
  - **Type (name)**;
  - **Vehicles per stations:** how many vehicles are available per station at the beginning of the simulation;
  - **Start and end time:** represent the start and end time of vehicle operation, in hours, respectively;
  - **Range:** represents the maximum distance a vehicle can travel on a journey, in meters;
  - **Capacity:** represents the maximum number of people the vehicle can carry;
  - **Cruise and vertical speed:** represent the maximum speed of the vehicle, during the take-off/landing and cruise phase, in m/s;
  - **Boarding and deboarding time:** represent the time it takes an agent to board/deboard to or from a vehicle, in seconds;
  - **Turn around time:** represents the time, in seconds, that a vehicle has to wait before it can be used again.

The introduction of the UAM service in a city in general will involve analyzing the many factors that will influence the adoption of this service by the population and the UAM performance. That said, a set of parameters regarding stations and vehicles are constant, independently of the simulation and they are present in Table 3.2 and Table 3.3, respectively.

Table 3.2: Stations parameters.

Station ID / name	Stationn / n
VTOL distance (m)	600
Default wait time (s)	180
Ground access capacity / free speed (m/s)	9999 / 35
Flight access capacity / free speed (m/s)	9999 / 200

The others parameters vary according to the simulation, with the objective of being able to perceive the influence they have on the use of the UAM service. In order to assess the influence of each parameter, a base scenario was defined and the parameters were changed one by one, in relation to the base scenario. The parameters defined for the base scenario and their variations are in Table 3.4.

Regarding the values of the base scenario it is possible to see below a little explanation about how we got these values.

Table 3.3: Vehicles parameters.

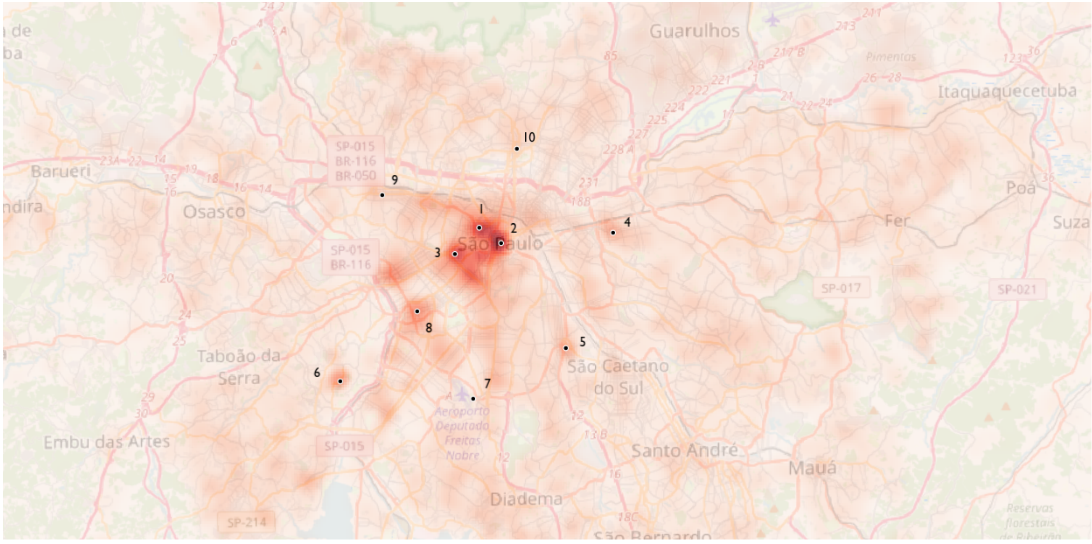
Name	1
Start / end time (h:min:s)	00:00:00 / 30:00:00
Range (m)	60000
Boarding / deboarding time (s)	30 / 30
Turn around time (s)	120

Table 3.4: Varied parameters in the simulations.

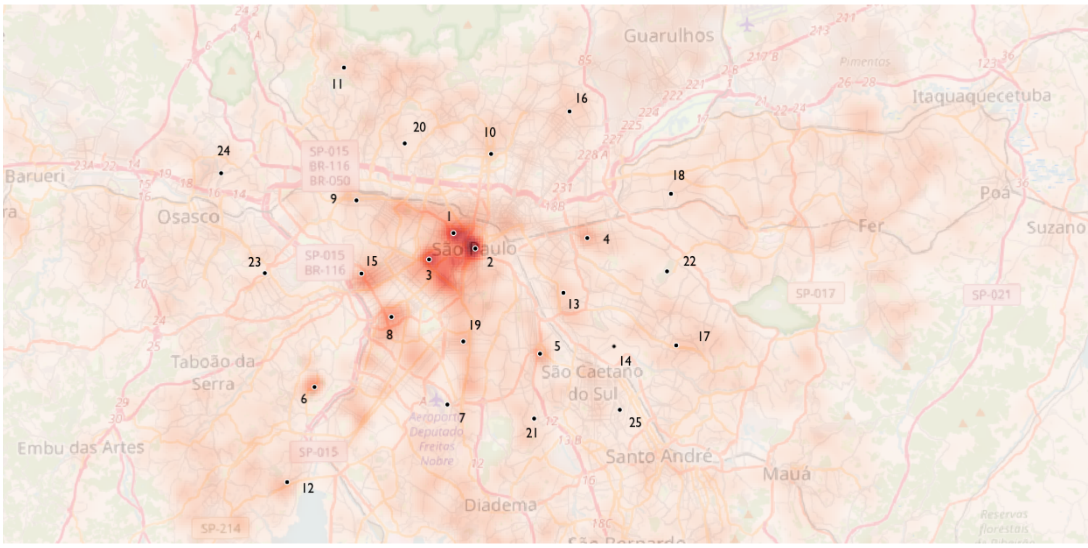
Parameter	Base value	Variation
Cruising speed (km/h)	150	50, 250, 350
Vertical speed (m/s)	10	5, 20
Ground Process Time (min)	3	0.5, 5, 15
Vehicle Capacity	1	2, 6
Fleet Size (uniform distributed)	1000	100, 2000
Number of stations	25	10
Fare per km (BRL)	2	0, 5
Fare per min (BRL)	0.5	0, 2
Base fare (BRL)	3	0, 10

- **Cruising and vertical speed values:** based on the vehicles being developed;
- **Ground process time:** or the check-in and check-out time, is 3 minutes because we are assuming that as this is a transport mode whose objective is to complement and alleviate existing ones, this process needs to be agile, unlike what happens at airports where other controls are needed;
- **Vehicle capacity:** based on the 1% of the population of São Paulo that is used in the synthetic population;
- **Number of Vehicles:** based on the number of agents in the simulation and they were uniformly distributed across the available stations;
- **Number of stations:** based on the travels that are made and in the number of agents used during the simulations;
- **Fares:** similar to public transport prince in the city of São Paulo.

It was also necessary to introduce the stations in the network and, for this, an analysis was carried out, based on the origin-destination pairs of all trips, of all agents. In this way, a file was created with all the coordinates of the places where a trip starts or ends and all these points have been marked on a map using QGIS[12]. Later, in order to make the busiest places noticeable, a heat map was created where it was possible to identify the regions of the city where the most departures or arrivals of trips are made. Since two different scenarios were simulated for the parameter ”No. of stations”, it was necessary to create two different networks, one with 10 stations and the other with 25. In Figure 3.2, it is possible to observe the locations chosen for the 10 stations (a) and for the 25 (b). For both scenario with 10 and the scenario with 25 stations, the stations were placed in the areas with more movement (hottest spots), while in the scenario with 10 stations they were placed only in the city center, in the scenario with 25 these are more dispersed, thus occupying a larger area.



(a) Location of the 10 UAM stations.



(b) Location of the 25 UAM stations.

Figure 3.2: Location of UAM stations in the São Paulo scenario.

[12] A free and open source geographic information system, https://www.qgis.org/en/site/

In Figure 3.3, it is possible to see the complete network with 25 stations. These stations were directly connected to each other without any deviation, in both cases.

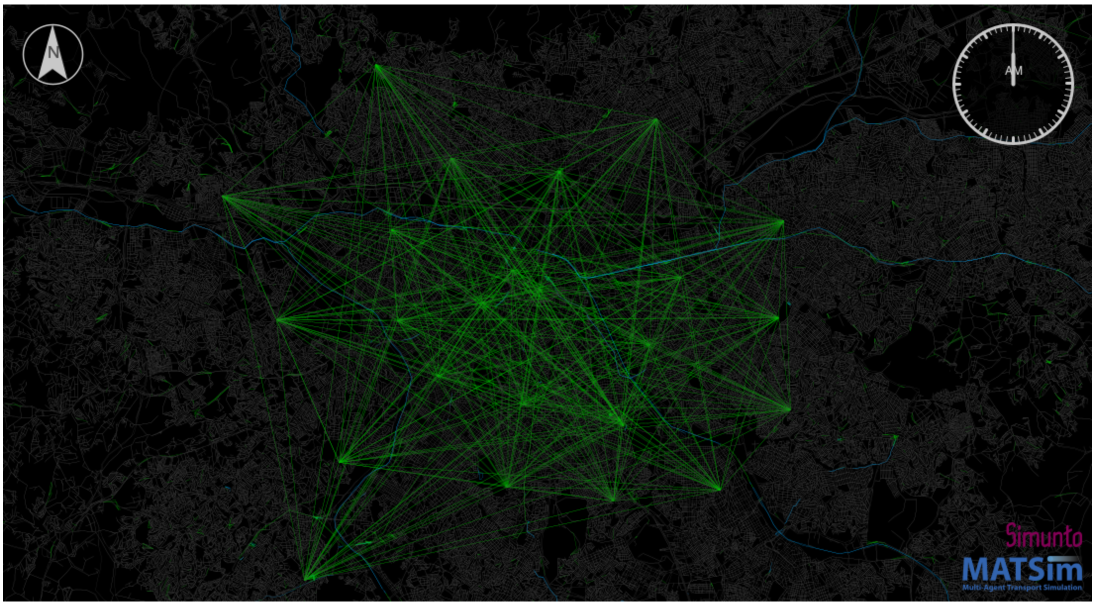


Figure 3.3: Network with 25 stations linked directly to each other.

3.2 Synthetic population validation

In order to prove that the population used in the simulations is effectively representative of the city of São Paulo and that it will reflect their choices, a comparative analysis was made to the data obtained with the real values, taken from the results of the 2010 census[13].

With regard to the number of men and women present in the synthetic population (SynPp) and in the census, it is possible to see in Figure 3.4 that there is a good fit between the reality and what was generated.

Going deeper and dividing the population by gender and age, it is possible to see in Figure 3.5 that there is a good fit not only between gender but also by age group.

Comparing the percentage of the population residing in São Paulo that is employed, unemployed or is a student with the household travel survey, it is possible to verify in Figure 3.6 that there is a good matching between these categories.

Analyzing now only the synthetic population, in order to understand its characteristics, and since no information was found about these topics, in Figure 3.7 it is possible to see the percentage of people who have a driving license and also the percentage of people who are subscribed to the public transport service.

The population that was generated includes the agents' plans, with the activities they have to perform throughout the simulation (home, work, education, leisure, shop or other), therefore, these were also analyzed. First, in Figure 3.8 it is possible to observe, in percentage, the number of trips that are carried

[13] https://www.ibge.gov.br/censo2010/apps/sinopse/

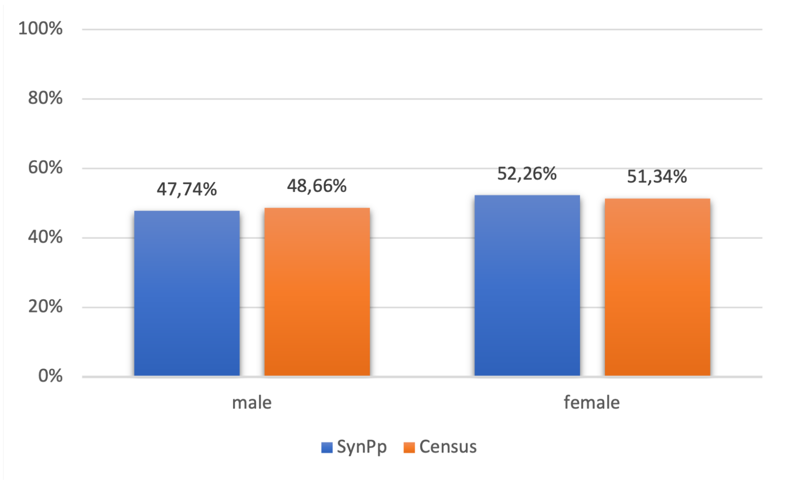


Figure 3.4: Gender comparison between real and synthetic population.

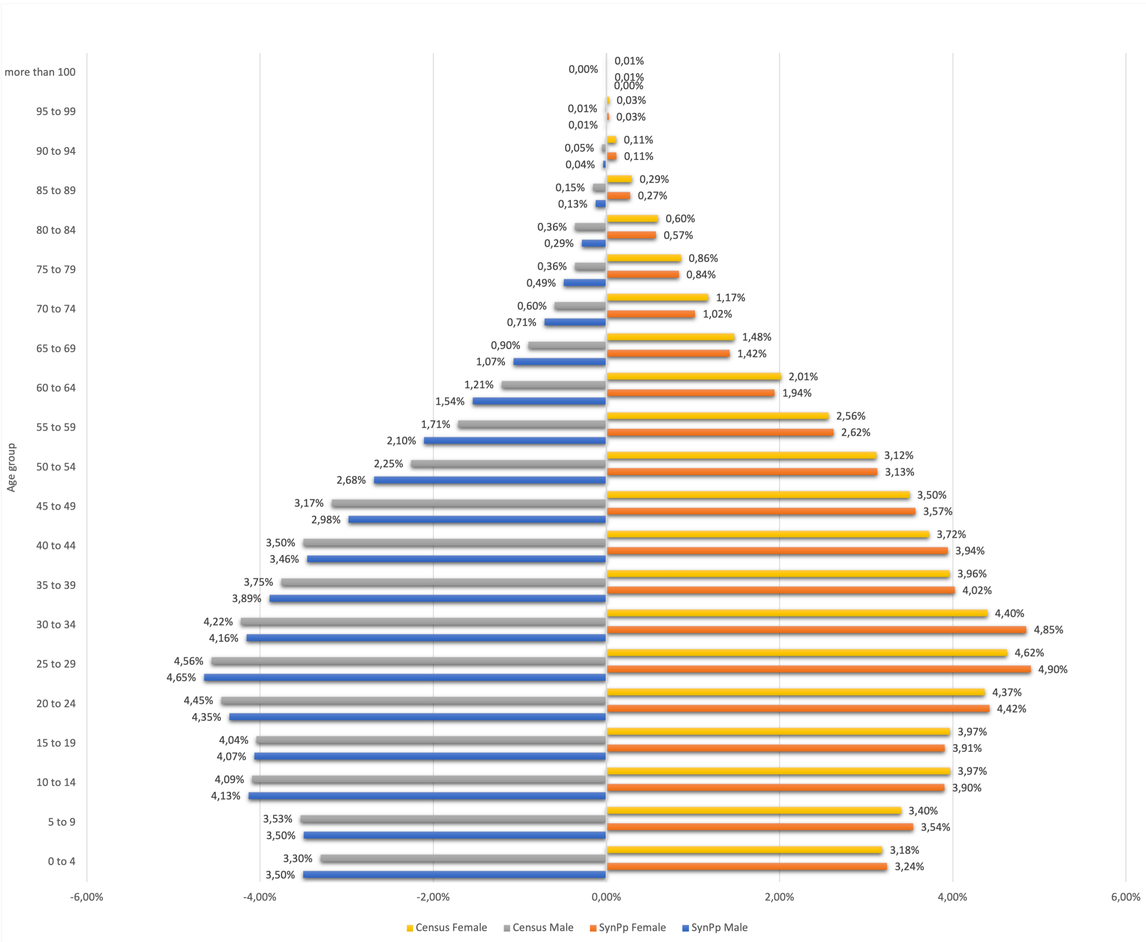


Figure 3.5: Comparison of genders and ages between real and synthetic population.

out by each chain of activities. For example, the home-work-home activity chain has 2 trips, one from home to work and another from work to home.

Now analyzing the chains of agents' activities, in Figure 3.9, it appears that, effectively, chains with two trips are the most common.

Regarding the chains of activities that specifically involve work (a), education (b), leisure (c), shop

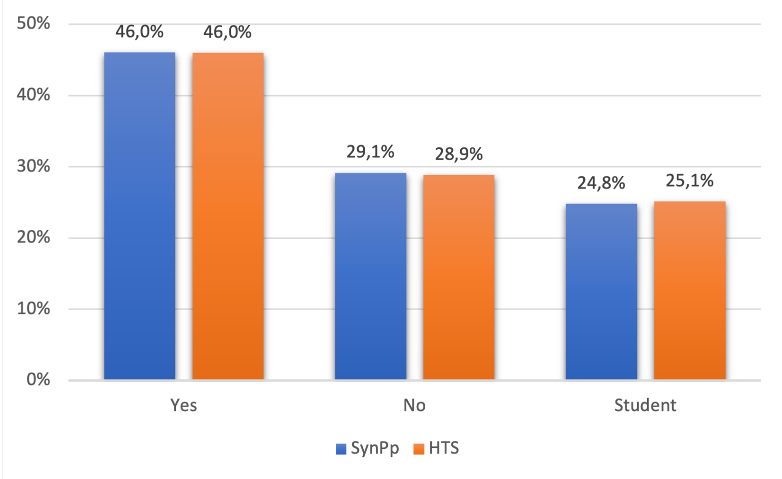
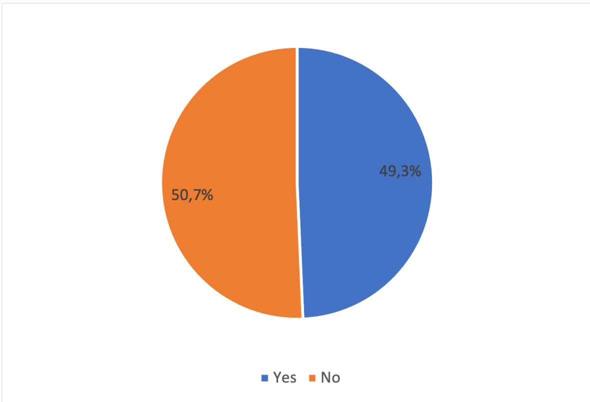
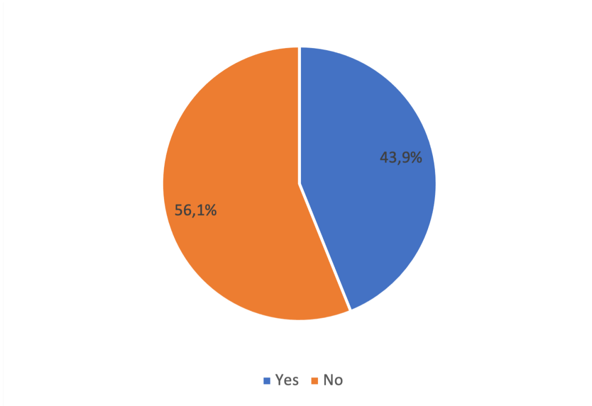


Figure 3.6: Comparison of the percentage of employed, unemployed and students between the synthetic population and the household travel survey.



(a) Public transport subscription



(b) License

Figure 3.7: Percentage of public transport subscription and persons with license in the synthetic population.

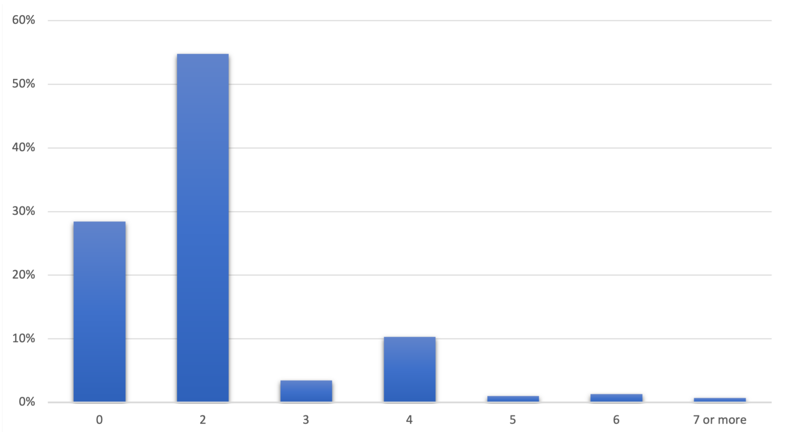


Figure 3.8: Number of trips that are performed by each chain of activities.

(d) and other (e) activities, the most accomplished plans are represented in Figure 3.10.

Finally, the average income per person was analyzed, based on the income of each family. This average income is shown in Figure 3.11.

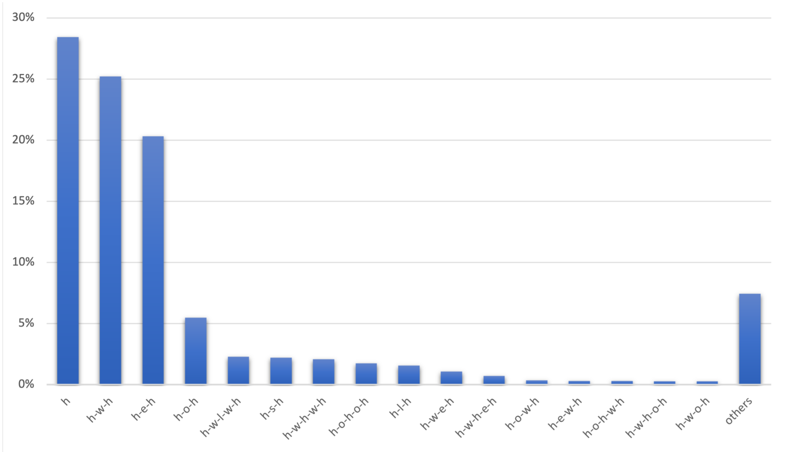


Figure 3.9: Effective frequency of each activity chain in a simulation.

3.3 UAM MATSim results

This Section describes the results obtained, taking into account the methodology described before. The results are divided into two parts. The first analyzes the base scenario and makes a comparison with the scenario without UAM. The second part analyzes the results obtained, taking into account the variation of the different input parameters.

3.3.1 Overall evaluation

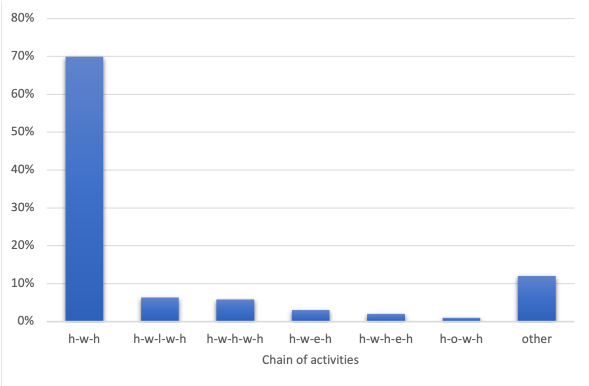
First, in order to verify the impact that the introduction of UAM has on the city, a simulation without UAM was carried out, in order to verify if it represents well the reality. Figure 3.12 presents the differences between the scenario without UAM and the real values.

Analyzing the mode shares of the simulation without UAM, it can be concluded that they represent the reality relatively well, comparing to the real data[14]. However, there is a difference with regard to the use of the car, since in 2020 the real value was 36% and in the simulation result is, approximately, 21%. The values regarding the mode share of taxi and car passenger were not found. Then, the results of this simulation were compared with the results of the simulation of the base scenario.

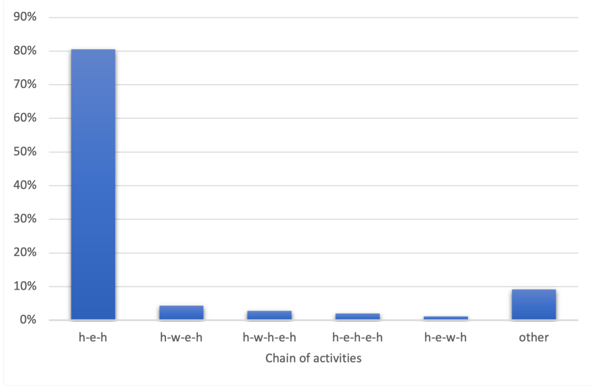
Regarding the result of the base scenario, 41128 UAM legs were performed and they represent 8.28% of the total number of legs. The mode shares of each transport mode (including walk) were analyzed, and, organized in a decreasing manner, the result is as follows: walk (30.34%), public transport (28.71%) , UAM (18.94%), car (12.49%), car passenger (8.7%) and taxi (0.82%). Figure 3.13 presents the differences in the mode share of each mode of transport, compared to the scenario without UAM.

It is possible to see that, the UAM introduction will have a negative impact on all transport modes and this impact will have a greater effect on the reduction of the car's mode share. One of the reasons that may justify the drop in the use of public transport is the fact that UAM stations are not placed in strategic places, where a change from train to UAM can be made, for example. These strategic places would be next to bus and train terminals. In this way, a better complement between all public transport could be

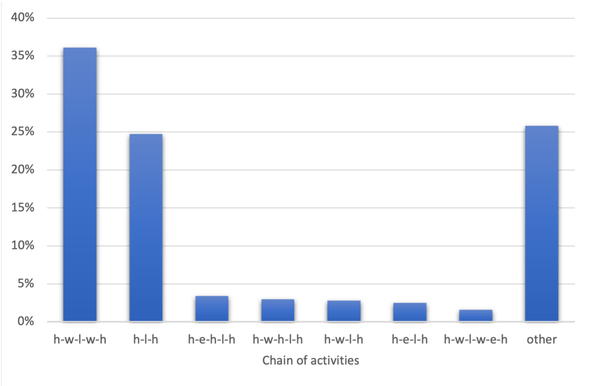
[14] https://www2.deloitte.com/content/dam/insights/us/articles/4331_Deloitte-City-Mobility-Index/SaoPaulo_GlobalCityMobility_WEB.pdf



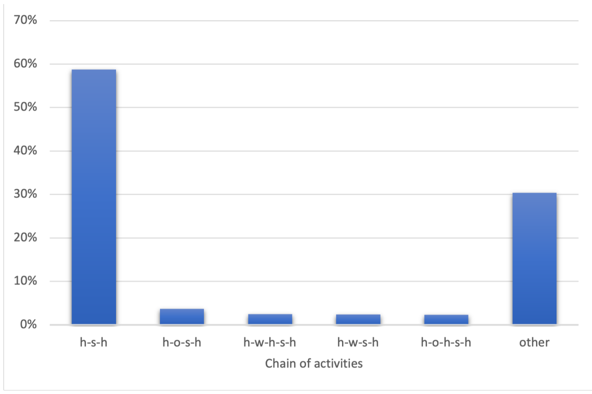
(a) Work.



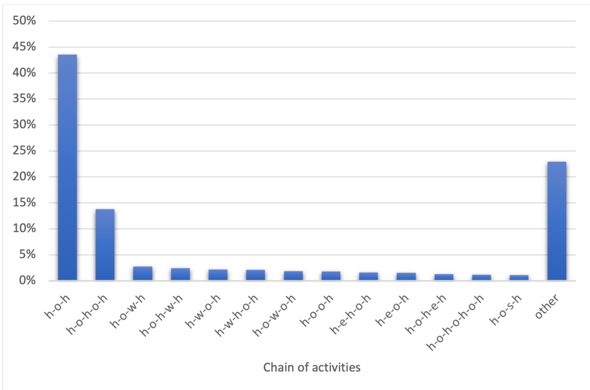
(b) Education.



(c) Leisure.



(d) Shop.



(e) Other.

Figure 3.10: Plans most carried out depending on the activity and regarding the synthetic population.

achieved.

Then, the average time of all journeys was calculated, which is approximately 28 minutes. In order to check whether this was a reasonable value for the trips that were being made, the average time of all trips was calculated assuming that all UAM vehicles would travel at the maximum speeds allowed, that is, a cruising speed of 150km/h and a vertical speed of 10m/s. That said, the time each vehicle needs to take off and landing (2*(600/10)=120s) and the time that the cruise phase would take at maximum speed (distance/cruising speed) were calculated and these were added together. The final result shows that the average time of all trips at maximum speed should be about 5 minutes, which shows that 23 minutes are being spent in traffic jam. Therefore, we can conclude that with the parameters defined for this

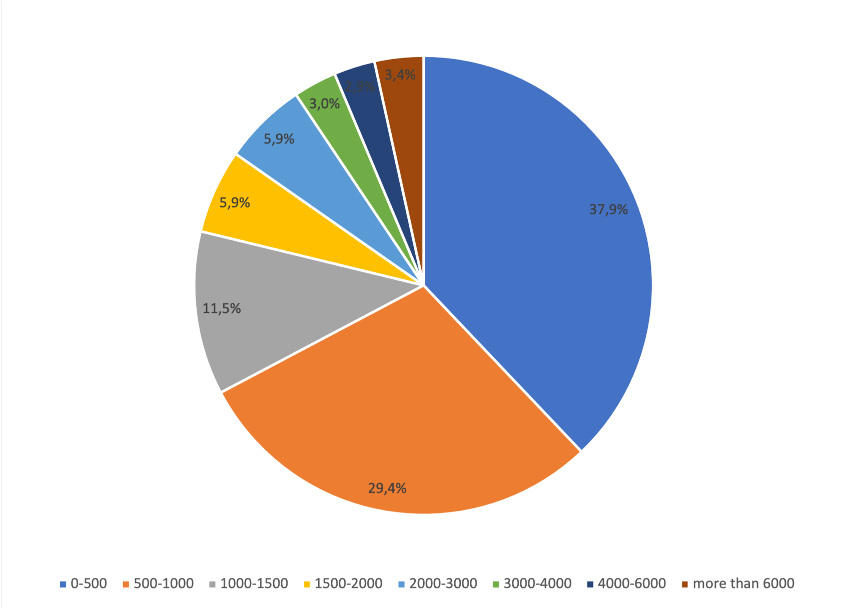


Figure 3.11: Mean income per person in the synthetic population.

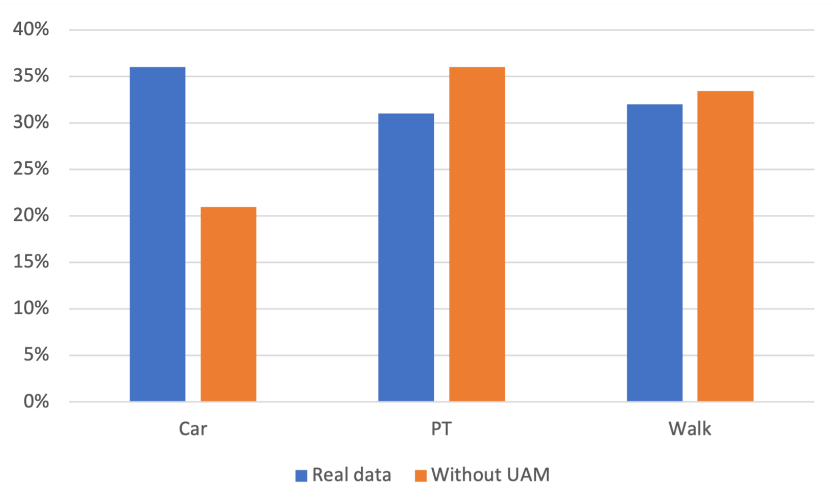


Figure 3.12: Mode shares comparison between scenario without UAM and real values.

scenario, there is strong UAM congestion, which can be due to several factors, including the unrealistic value of the ground and flight access capacity of 9999 persons per hour, which hardly puts a limit on the number of people at stations and the number of vehicles per link and the low preflight/postflight time. Another reason could be the low number of stations available.

The average distance covered was 7.4km and, taking into account the average time of all trips, the average speed was 20km/h, a value far below which the vehicle can travel. Regarding the distances covered, a histogram was created that allows to understand which trips were made the most. This histogram of the base scenario is shown in Figure 3.14.

Analyzing the histogram of this scenario, it appears that the minimum distance traveled in a trip is in the range of 1.5km-2.5km meters and the maximum distance in the range of 27.5km-28.5km. However, most trips are in the range of 4.5km-5.5km meters.

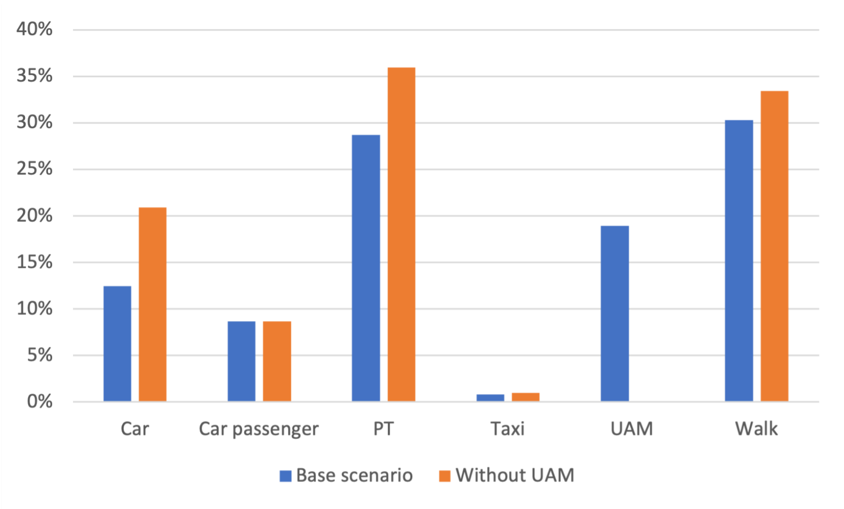


Figure 3.13: Mode shares comparison between scenario base and scenario without UAM.

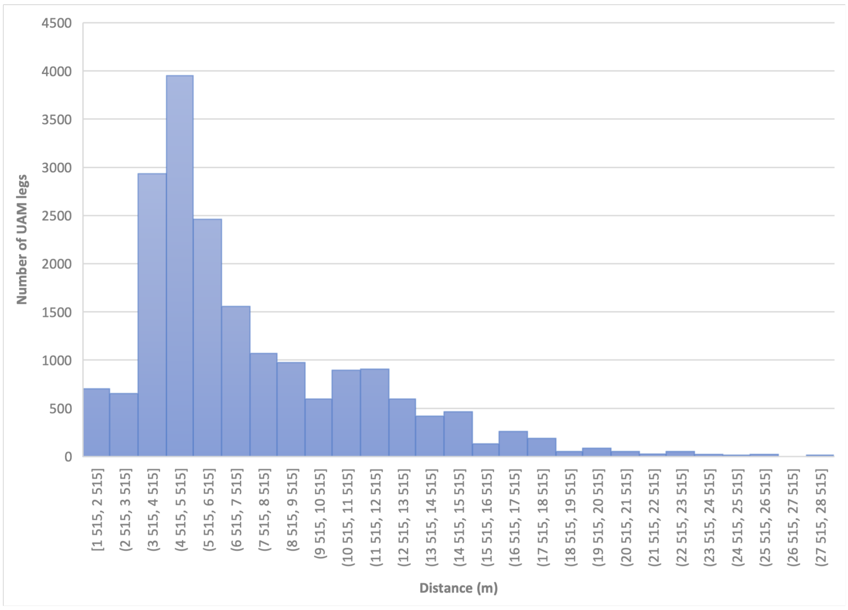


Figure 3.14: Histogram of distances from the base scenario.

The stations most and least used by agents were also analyzed, in order to understand whether the fact that they were placed in a more or less busy area (darker or lighter on the heat map, respectively) had an influence. In Figure 3.15, the 3 busiest stations are shown in green and the 3 least busy in red.

We can conclude that there is an agreement with the analysis made for the placement of stations taking into account the departures and arrivals of trips. Regarding the least used stations, it can be seen that they are found in areas with a light spot and on the outskirts, while the 1st and 2nd most used are found in the darkest spots, in the center of the city. There is a surprise, if we look at the 3rd most used station, in the sense that it is located on the outskirts and in a light spot. Therefore, this should be further studied, to understand the main reasons.

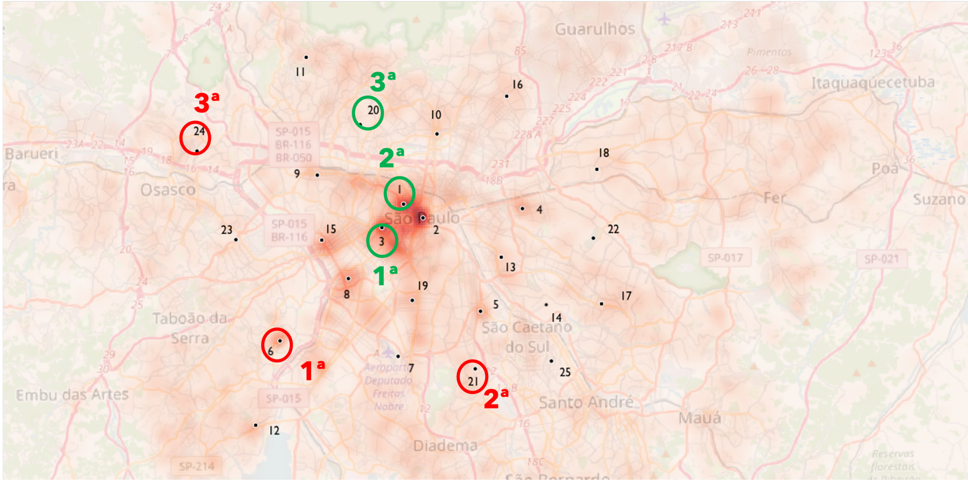
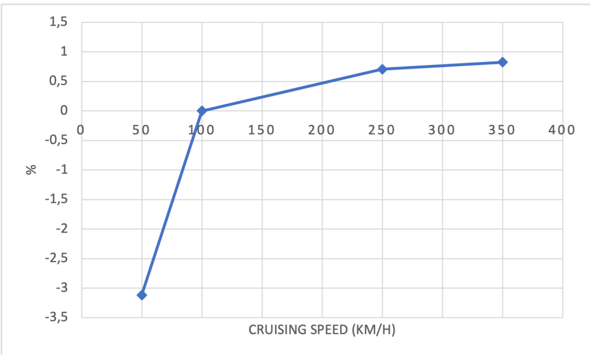


Figure 3.15: Most used stations (green) and least used stations (red) in the simulation.

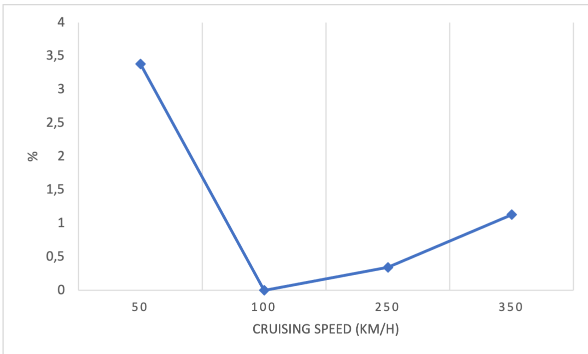
3.3.2 Sensitivity to main UAM parameters

Cruising speed variation

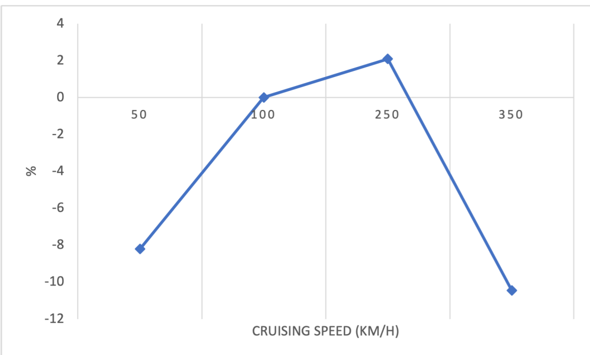
Analyzing the cruising speed variation, Figure 3.16 presents the influence that this change had on the number of UAM legs (a), the average travel time (b), the average distance traveled (c) and the average time spent in traffic (d).



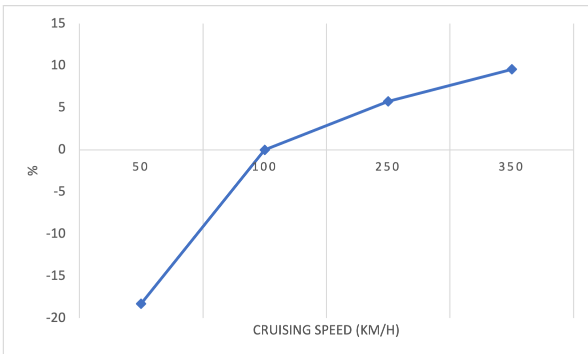
(a) Variation of UAM legs.



(b) Average travel time variation.



(c) Average distance variation.

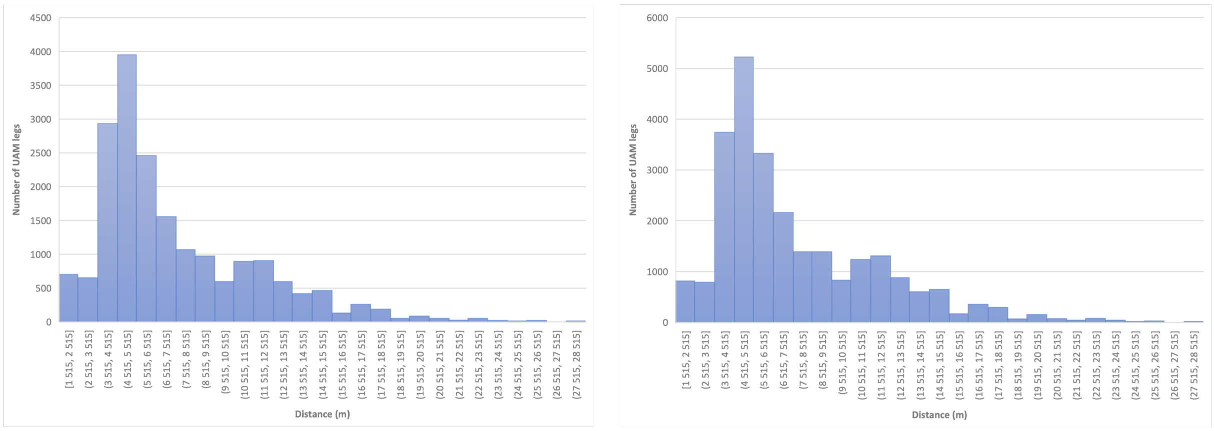


(d) Variation of average time in traffic jam.

Figure 3.16: Cruising speed variation (km/h).

An increase in this speed in relation to the baseline scenario translates into an increase in the number of UAM legs which, in turn, implies an increase in the average travel time and in the congestion as there are more vehicles circulating. By reducing the value of this parameter, there is an increase in travel time, despite the decrease in the number of UAM legs, due to the speed being lower. In turn, reducing the number of UAM legs in this case reduces congestion time. Analyzing the average distance covered, it is possible to observe that a reduction in cruising speed to 50km/h implies a reduction of approximately 8% in this, that is, as the speed is lower, agents prefer to make shorter trips. Increasing the value of cruising speed to 250km/h, there is a slight increase in the average distance covered, however, when it increases to 350km/h, there is a decrease of almost 12% in the average distance.

Histograms were also analyzed, which show the number of trips made for each distance interval, and the only histogram that presents a significant variation in relation to the base scenario histogram is the one in which the cruising speed was 350km/h. Thus, the two histograms are found in Figure 3.17.



(a) Base scenario.

(b) 350km/h cruising speed scenario.

Figure 3.17: Histograms of cruising speed variation.

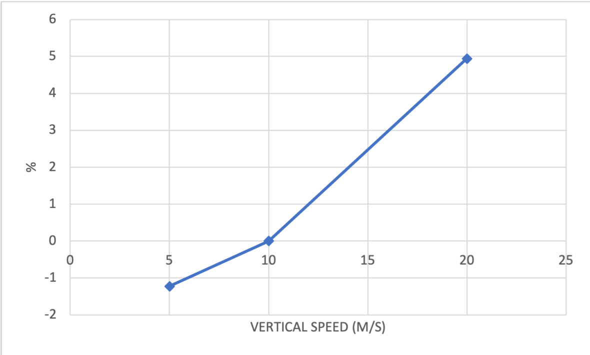
The reduction in the average distance of almost 12% in the scenario with 350km/h is visible in the comparison of histograms, that is, in the histogram of the cruising speed of 350km/h, there is a substantial increase in trips between 4.5km and 5.5km, compared to the base scenario.

Vertical speed variation

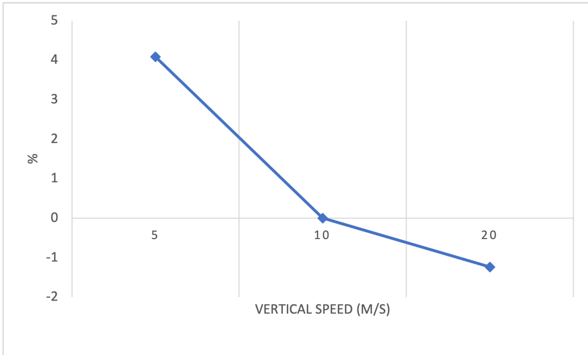
Regarding the variation of vertical speed, Figure 3.18 presents the influence that this change had on the number of UAM legs (a), the average travel time (b), the average distance traveled (c) and the average time spent in traffic (d).

With regard to this variation, an increase in relation to the baseline scenario, to 20m/s, translates into an increase in the number of UAM legs and, therefore, an increase in congestion. With this increase in vertical speed, agents reduced the average travel distance, which translates into a reduction in average travel time. Reducing this parameter to 5m/s, there is a slight reduction in the number of UAM legs and, consequently, in congestion. In relation to the average travel time, this increases by about 4%, although the average distance remains practically the same.

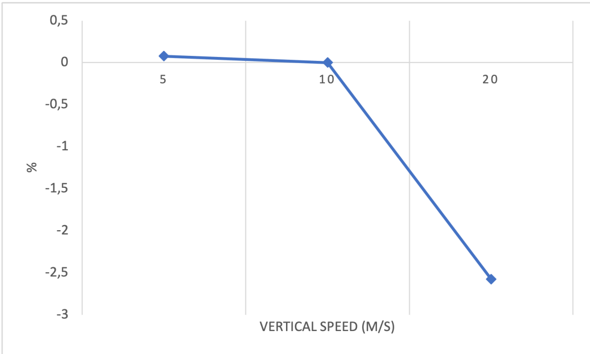
Regarding the histograms for each simulation performed as a function of the distance covered, the



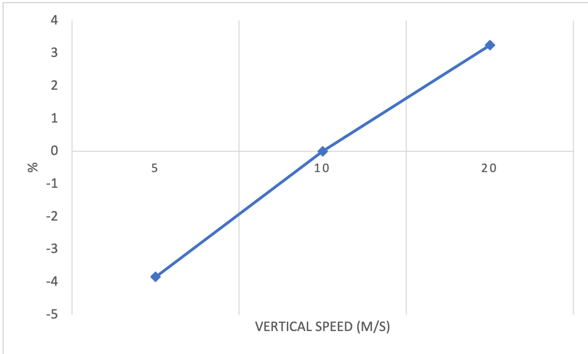
(a) Variation of UAM legs.



(b) Average travel time variation.



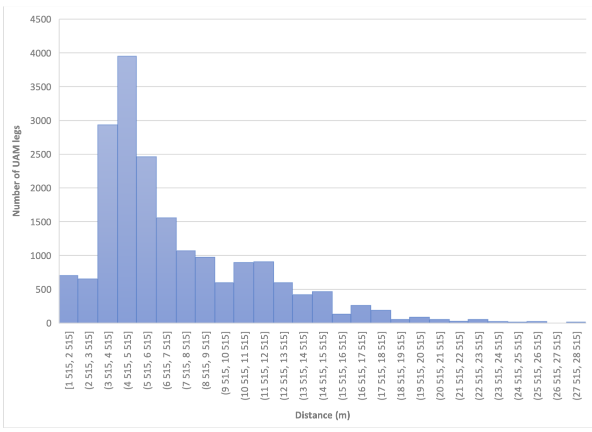
(c) Average distance variation.



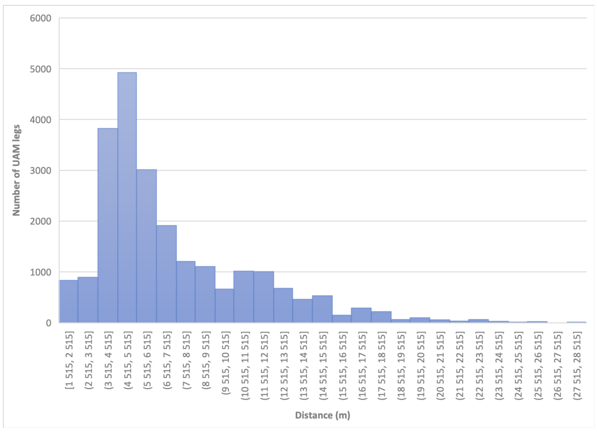
(d) Variation of average time in traffic jam.

Figure 3.18: Vertical speed variation (m/s).

only histogram that presents a significant variation in relation to the histogram of the base scenario is the one in which the vertical speed was 20m/s. Thus, the two histograms are found in Figure 3.19.



(a) Base scenario.



(b) 20m/s vertical speed scenario.

Figure 3.19: Histograms of vertical speed variation.

As already mentioned that an increase from 10m/s to 20m/s implied a reduction in the average distance, it can be seen through the histograms that there was a significant increase in trips between 4.5km and 5.5km, compared to the base scenario.

Ground process time variation

Regarding the variation of the ground process time, that is, the time an agent spends at the station before being able to enter the vehicle and the time they have to stay at the station after finishing a trip, the only parameter that had a significant variation, was the variation of the average time in traffic jam. This can be found in Figure 3.20.

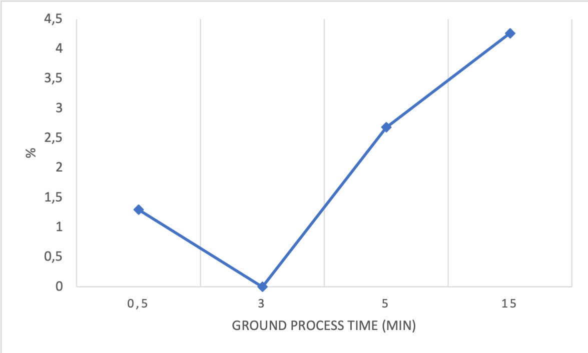


Figure 3.20: Variation of the average time in traffic jam varying the ground process time (min).

This may be due to the fact that, instead of spending time in the air traffic jam, agents now spend at the station. Analyzing this variation, it appears that an increase in GPT implies an increase in traffic and, on the other hand, a reduction also means an increase in congestion, although not so sharply. This is due to the fact that, although slight and not represented, a decrease in GPT, translated into an increase in UAM legs.

In the histograms, a significant change in the distribution of trips as a function of distance is not visible.

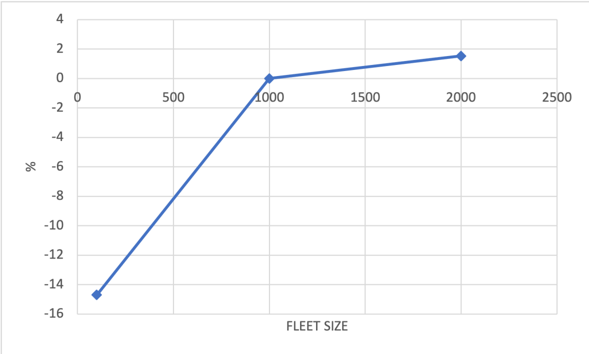
Vehicle capacity variation

The change in vehicle capacity did not show significant differences in any of the analyzed values, so no graphs are presented for this parameter. This is due to the fact that a vehicle, at this time, having an agent ready, leaves immediately to carry out the trip, instead of waiting, for example, 1 or 2 minutes to be able to transport more agents to the same location.

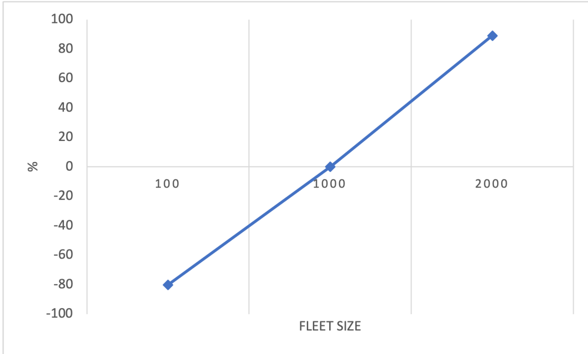
Fleet size variation

Analyzing the fleet size variation, Figure 3.21 presents the influence that this change had on the number of UAM legs (a), the average travel time (b), the average distance traveled (c) and the average time spent in traffic (d).

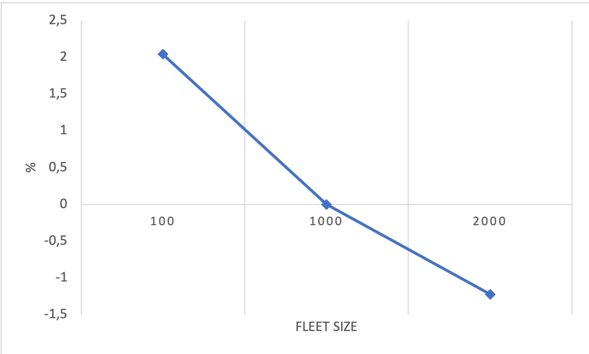
Doubling the fleet size resulted in an increase in the number of UAM legs, average travel time and average time in congestion. With an increase in congestion, agents chose to take shorter trips, thus reducing the average distance traveled. On the other hand, a reduction of the fleet to 100 vehicles caused the number of UAM legs to be reduced by approximately 15%. With fewer vehicles circulating, congestion has decreased and therefore the average travel time has also decreased.



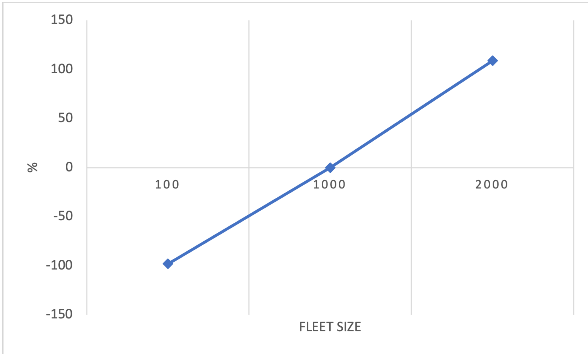
(a) Variation of UAM legs.



(b) Average travel time variation.



(c) Average distance variation.



(d) Variation of average time in traffic jam.

Figure 3.21: Fleet size variation.

Regarding the histograms for each simulation performed as a function of the distance covered, both scenarios in which the fleet size was varied present variations in relation to the base scenario. So it is presented in Figure 3.22 the 3 histograms. It is visible that with 2000 vehicles, the number of short trips increased.

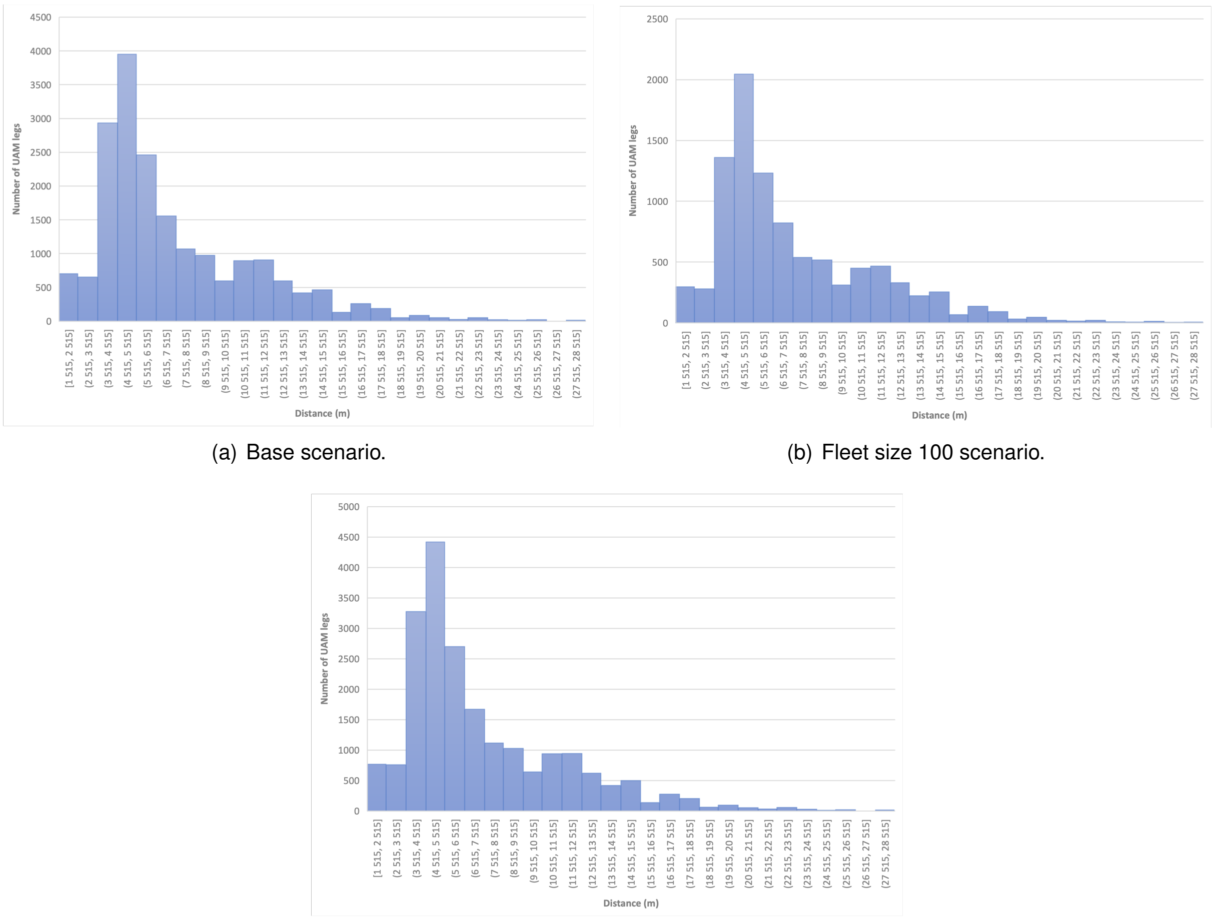
Number of stations variation

Regarding the variation in the number of stations, Figure 3.23 presents the influence that this change had on the number of UAM legs (a), the average travel time (b), the average distance traveled (c) and the average time spent in traffic (d).

The reduction in the number of stations to 10 resulted in a reduction of almost 60% in the number of UAM legs. With fewer stations, but with the same number of vehicles circulating (1000), this reduction implied an increase of approximately 130% in the average travel time and 160% in the average time in congestion. The average distance reduced by 16%, as the 10 stations are more concentrated in the city center, compared to the 25 stations.

Figure 3.24 presents the histogram for the simulation carried out with 10 stations, depending on the distance covered.

It is possible to conclude that with fewer stations, the trips that were made the most are in the range of values between 5.5km and 6.5km.



(a) Base scenario.

(b) Fleet size 100 scenario.

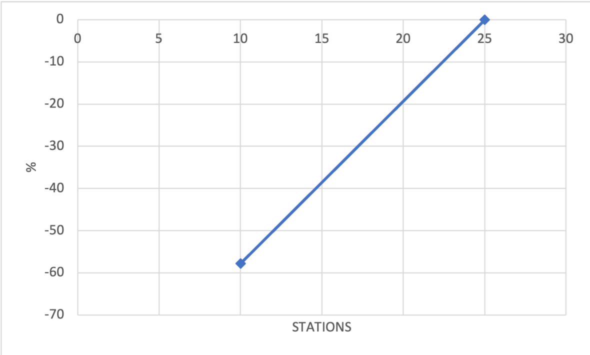
(c) Fleet size 2000 scenario.

Figure 3.22: Histograms of fleet size variation.

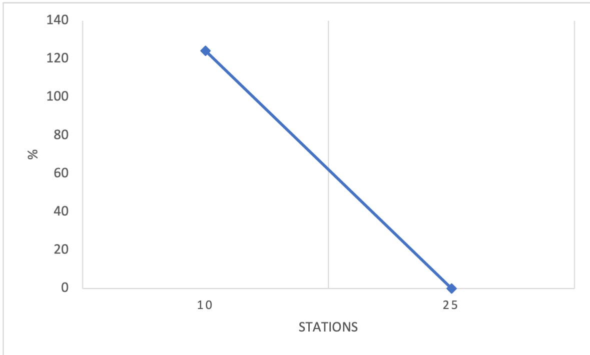
Fare per km, per minute and base fare variation

In the simulations carried out in which the fare per minute, per kilometer and base fare were changed, the only significant variation in the performance metrics analyzed is related to the scenario in which the fare per minute was changed. Even so, the only changes are related to the variation in the average travel time and the variation in the average time spent in traffic jam. These two graphics are shown in Figure 3.25.

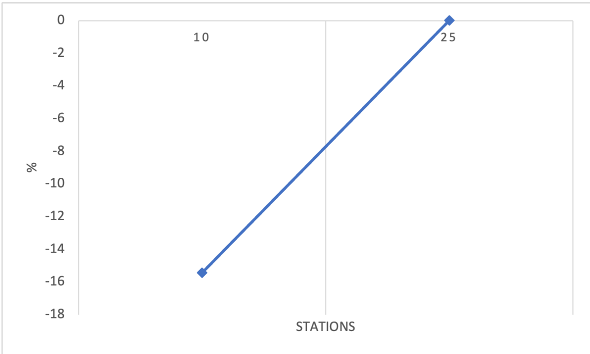
With the variation of the different fares, the only significant change in the performance metrics is found in the variation in average travel time and congestion with the variation in fare per minute. The fact that the fare per kilometer and base fare do not have much influence is due to the fact that agents who are already using the UAM service under the base scenario conditions with high congestion do not mind paying more to carry out the trip, as this continues to be advantageous for them. Analyzing the variation of the average travel time and congestion, both have a similar behavior, in the sense that, with a reduction in fare per minute, the average time increases and, therefore, so does the congestion.



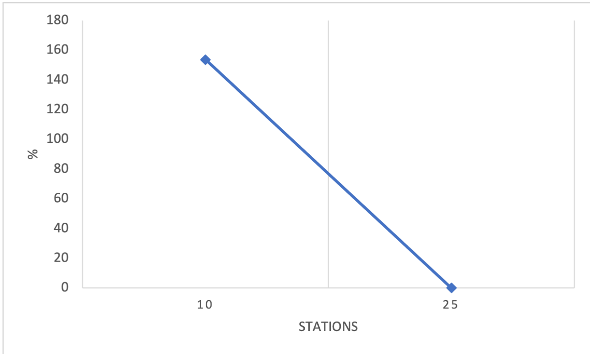
(a) Variation of UAM legs.



(b) Average travel time variation.

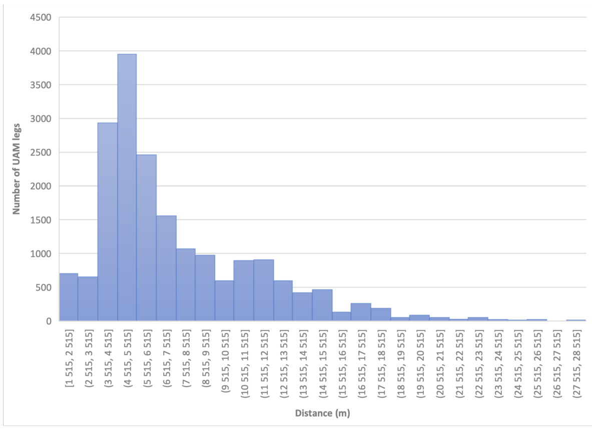


(c) Average distance variation.

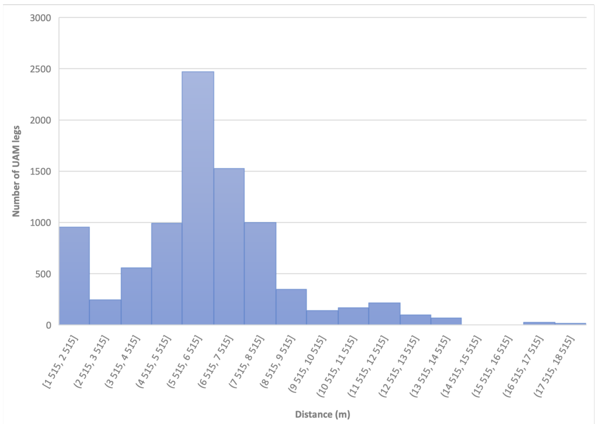


(d) Variation of average time in traffic jam.

Figure 3.23: Variation in the number of stations.



(a) Base scenario.



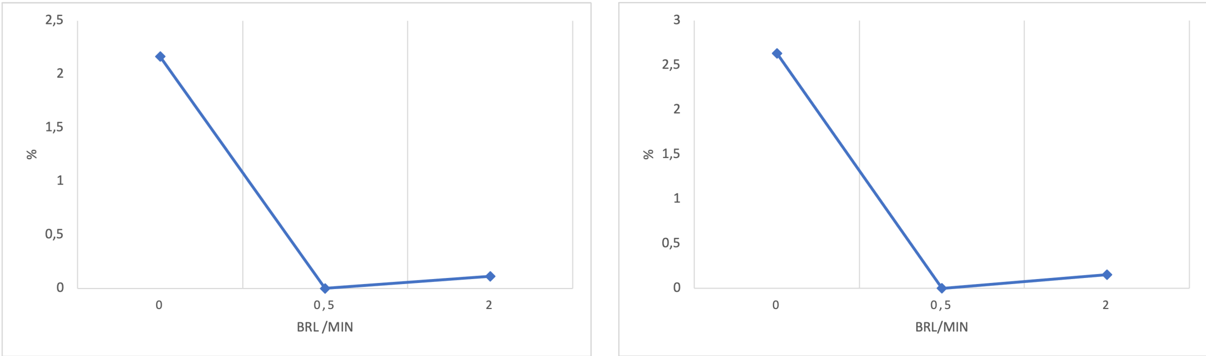
(b) 10 stations scenario.

Figure 3.24: Histograms of the variation in the number of stations.

3.4 Summary

- **Base scenario:** in general, although the UAM appears to have acceptable levels of utilization, the parameters defined for the base scenario cause a high level of air congestion. It is important to adjust these parameters in order to optimize the service, and this optimization can range from updating the values chosen for the input parameters to the strategic placement of UAM stations, along with other public transport modes.

- **Cruising speed variation:** an increase in the cruising speed translates into an increase in the



(a) Average travel time variation.

(b) Variation of average time in traffic jam.

Figure 3.25: Fare per minute variation.

number of UAM legs and, consequently, an increase in congestion. On the other hand, a reduction of this parameter results in a reduction in the number of UAM legs and also in the congestion time. Due to a lower speed, it results in an increase in the average travel time.

- **Vertical speed variation:** an increase in the vertical speed translates into an increase in the number of UAM legs and, consequently, an increase in congestion. By reducing this parameter, there is a slight reduction in the number of UAM legs.
- **Ground process time variation:** in this case, the only variation that occurred was the increase in congestion, either with the decrease or increase in the value of this parameter.
- **Vehicle capacity variation:** there were no significant changes with the change in vehicle capacity, due to the fact that the vehicles, in the simulation, departure as soon as they have an agent, rather than waiting for more.
- **Fleet size variation:** an increase in fleet size translates into an increase in the number of UAM legs, as well as in the average travel time and the level of congestion. On the other hand, a reduction also causes a reduction in the number of UAM legs and in the level of congestion.
- **Number of stations variation:** the reduction in the number of stations translated into a drastic reduction in the number of UAM legs. However, with the same number of vehicles circulating, the average travel time and the level of congestion increased dramatically.
- **Fare per km, per minute and base fare variation:** a change in fares did not cause a major change in the results, due to the fact that those who use the UAM service in this condition are willing to pay more to use it, as it is still worth it. However, with a reduction in fare per minute, an increase in the average travel time can be verified.
- **Vehicles analysis:** analyzing the results obtained, it can be concluded that fully electric vehicles seem to be a good choice for the city of São Paulo, assuming these conditions. This choice is related to the fact that the trips are relatively short, on average, with the longest trip being around 28km. With regard to vehicle capacity, no comment can be made, as the vehicle capacity parameter did not produce desirable results.

Chapter 4

Simulation metamodel methodology applied on UAM simulation

4.1 Methodology

This work presents an active learning metamodeling methodology based on the work developed by Antunes et al. [70]. Therefore, a pool-based active learning strategy is adopted and it is presented in Figure 4.1. As mentioned in Section 2.5.2, in a pool-based active learning strategy, each individual data point is presented serially or in successive blocks. In this case, the algorithm will choose the 3 most informative points and will ask to the oracle to perform the simulations of these input points. The most informative points means the point with the highest predictive variance.

The algorithm was developed using the high-level programming language python and the scikit learn library.

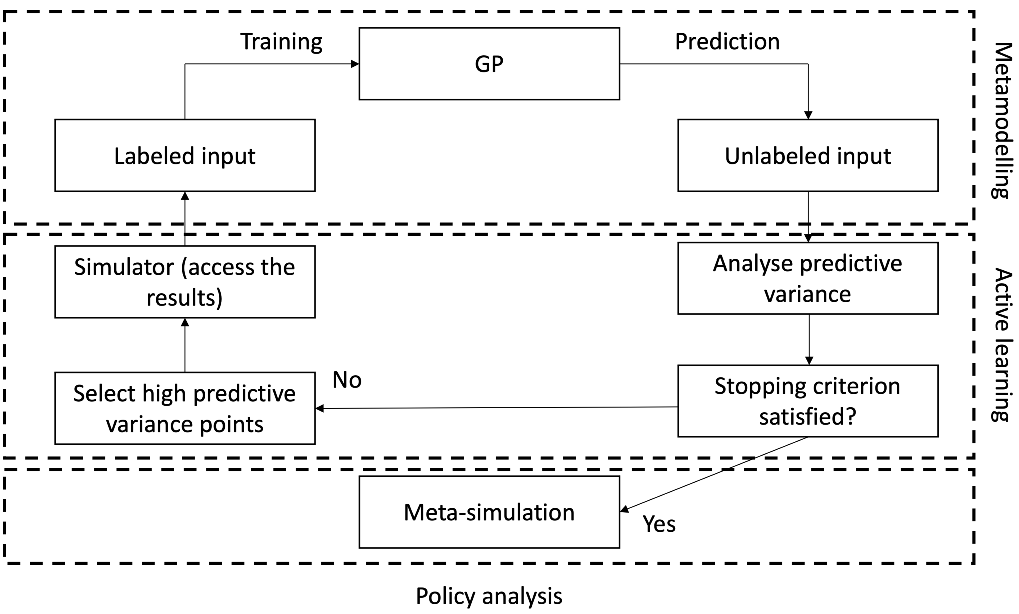


Figure 4.1: Active learning metamodeling methodology (adapted from [70]).

The methodology adopted, described below, can be divided into 3 main blocks: the simulation metamodeling, the active learning strategy and the policy analysis.

- **Simulation metamodeling:** approaches the simulation model using a GP. In this sense, the GP (M) is provided with a set of results already obtained through simulations (labeled input L), which allow training the model. After this, the model makes several predictions about a dataset that the metamodel does not yet know the true value of the simulations output (unlabeled input U).

- **Active learning strategy:** increases the quality of the GP, iteratively. In this block, the algorithm analyzes the predictive variance of each prediction that was made and selects the 3 most informative points. The 3 points with the highest predictive variance are provided to the simulator (MATSim O), which will give the real output of the 3 unlabeled input. After the simulator obtains the true results of these points, they are added to the L set. The stopping criterion, in our case, is the main difference, comparing to Antunes et al. [70]. Instead of using the total predictive variance, we use the mean variance across the unlabeled input simulation region. Therefore, the stopping criterion is the Current Mean Variance (CMV) divided by the Initial Mean Variance (IMV) and this should be bigger or equal to $1 - \alpha$, being α a parameter defined by the user, depending on the level of accuracy that he want. Based on that, the Q is the query function and is based on the analysis of the mean variance.

- **Policy analysis:** Lastly, when the stopping criterion is satisfied, meaning that the metamodel fits well the reality, according to our expectations, previsions can be done and the best policies can be adopted.

4.2 Corsica case study

Due to the long time that simulations of the São Paulo scenario took to be completed, and due to the high amount of results needed to carry out the metamodel, it became impracticable to apply this strategy to the case studied above. Therefore, the scenario of the island of Corsica, available in the MATSim-UAM repository, mentioned above, was chosen.

This scenario has a smaller population and a smaller network, therefore 500 different simulations were performed. However, the available scenario only features 3 UAM stations, so it was necessary to add more stations. These were added randomly, as the objective here is not to assess the impact of the introduction of UAM on the island of Corsica, but rather to assess the usefulness of creating a metamodel for predicting the output of simulations. Figure 4.2 presents the networked created, with the 10 UAM stations.

To carry out these simulations, a javascript was created that changes the input parameters between each simulation, in order to have 500 different simulations. Therefore, it was necessary to define a range of values for each input parameter. Table 4.1 presents which parameters were changed and their range of values.

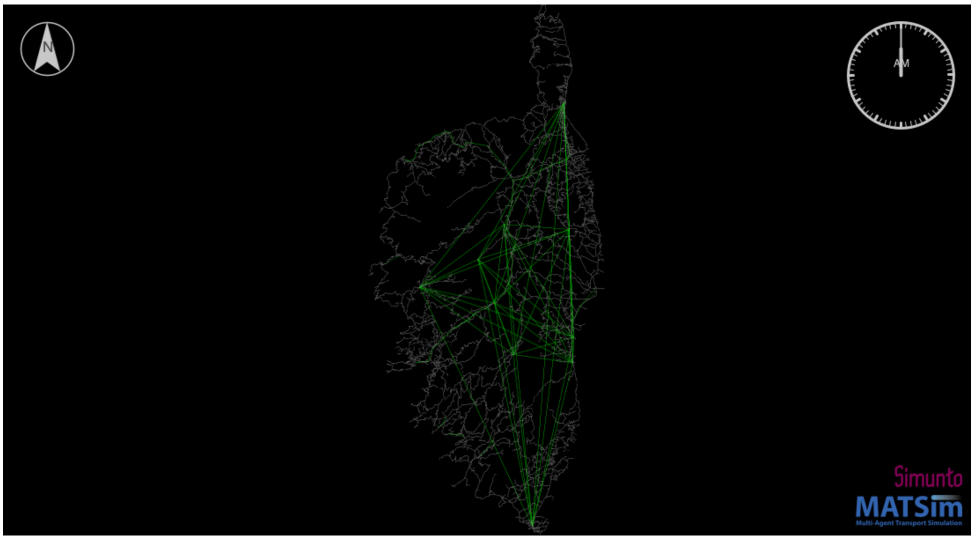


Figure 4.2: Corsica network without UAM.

Table 4.1: Corsica scenario parameters variation.

Preflight time (s)	1-900
Postflight time (s)	1-900
Default wait time (s)	1-500
Vehicle capacity	1-6
Cruising speed (m/s)	1-35
Vertical speed (m/s)	1-35
Boarding time (s)	1-120
Deboarding time (s)	1-120
Turn around time (s)	1-120

Despite having used a scenario with 10 stations, these were assumed to be uniform, that is, they present the same values in each simulation, in order to reduce the number of inputs of the metamodel. However, this assumption is not realistic, as each station will have its own characteristics.

After the 500 simulations had been carried out, another 35 different simulations were carried out. Unlike the 500 simulations where all parameters could be chosen within the maximum range of values defined in Table 4.1, these 35 simulations focused on reducing the range of each parameter in order to obtain simulations results in a smaller input space. Listed below, is presented the sequence performed by the javascript.

- **1 –** Access the configuration file of the Corsica scenario and change the input parameters randomly, according to their range;
- **2 –** Perform the simulation, based on the generated configuration file;

- **3 –** A file containing the input parameters of all simulations is created and the results are saved. This process is repeated until the end of the 500 simulations.

After all the simulations performed, it was necessary to implement the metamodel based on the active learning strategy defined before. In order to the model be trained with the labeled inputs and be able to make predictions of the unlabeled inputs, the 500 simulations were divided randomly.

In this case, we chose to use 67% of the 500 simulations carried out to train the metamodel and, thus, form the set of labeled input. This set is divided in the various input parameters obtained with javascript, called x_{train}, as well as the respective results of the simulations carried out with these parameters, called y_{train}. The set y_{train} is constituted by the number of UAM legs of each simulation.

The remaining 33% were used to enable the metamodel to make predictions and therefore form the set of unlabeled input. This set contains only input parameters without simulation results and is called x_{test}. The remaining 35 simulations were added separately to the respective sets in the labeled group, in order to see if, reducing the input space, it was possible to improve the metamodel. In the end, basically, 2 metamodels were created, based on the same strategy. The first one starts with 67% of the 500 simulations performed in the labeled input set and contains the input parameters chosen randomly within the maximum possible range. The second starts with 67% of the 500 simulations performed plus the 35 in the labeled input set although, in this case, there is an attempt to refine the input space, as the input parameters of the 35 simulations restrict the random value of the input parameters.

Then, it was necessary to define the kernel function and the alpha parameter. Equation 4.1 presents the kernel function, chosen randomly. Alpha was set to 0.3, meaning that the iterations stop when the actual mean predictive variance is less than 70% of the initial mean predictive variance.

$$kernel = Lin() + WhiteKernel() \tag{4.1}$$

Listed below, is presented the sequence performed by the algorithm to obtain the first metamodel, although this sequence is the same for the second metamodel, since the only thing that changes is the set L.

- **1 –** The set L is defined being $L = 500 \times 0.67$ and is divided in the x_train and in the y_train, containing the input parameters and the outputs, respectively.

- **2 –** The GP is trained based on the set L. Then it predicts the 165 unlabeled inputs ($U = 500 \times 0.33$) and obtain their corresponding predictive variances;

- **3 –** The algorithm finds the top 3 highest predictive variance points in U (most informative points);

- **4 –** The true values for the top 3 highest predictive variance points are obtained;

- **5 –** L is updated with the new 3 observations;

- **6 –** The process finishes when the stopping criterion is satisfied, meaning that in each iteration, the mean variance is computed.

It is important to note that, in this case, when the metamodel asks for the results of the top 3 highest predictive variance points in U, these values, although not being used, are already available and stored, due to the fact that the 500 simulations have already been carried out and they correspond to the outputs of the x_{test}. Therefore, the algorithm just finds the true values and add them to y_{train}.

4.3 Simulation metamodel Results

The results obtained can be divided into two parts, according the 2 metamodels. The first part is the one in which the metamodel is only based on the 500 simulations, while the second part takes into account the 500 simulations plus the 35 simulations carried out in a more restricted input space. Figure 4.3 presents the mean variance variation in function of the number of iterations of the metamodel.

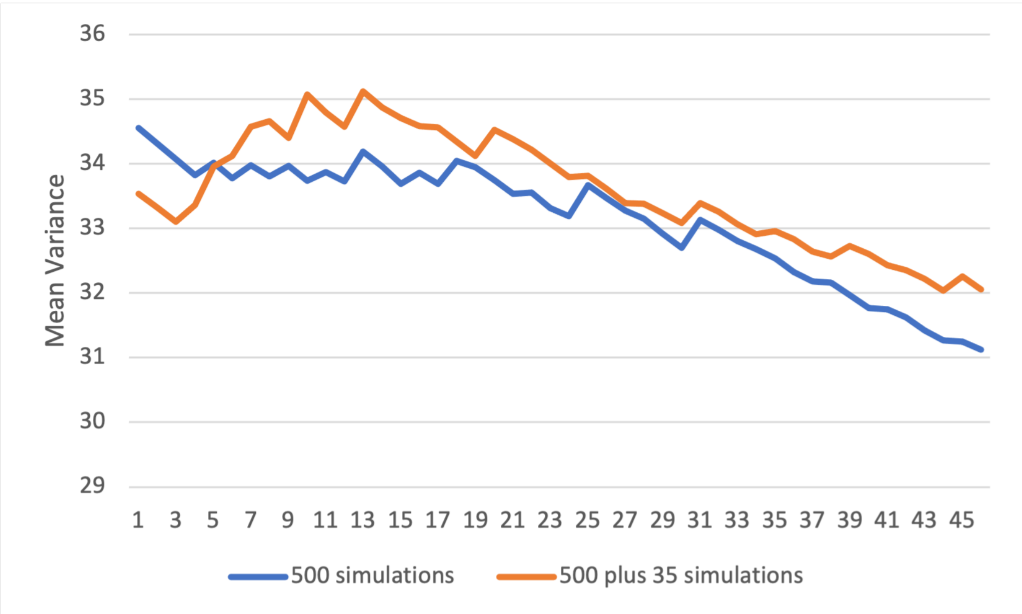


Figure 4.3: Comparison of the mean variation as a function of iterations between the metamodel with 500 simulations and the metamodel with 535 simulations.

Regarding the results, the most important thing to note is that, until iteration number 47, the average mean variance decreased compared to the first iteration. The metamodel was stopped at iteration 47, as if it continued, it would have no more results to predict, since at the end of each iteration, 3 results are added to the training set. It should be noted that, due to this, the stopping criteria did not have time to be satisfied.

Analyzing the figure, it can be concluded that, although slightly, the scenario with only 500 simulations at the beginning, due to having fewer observations, presents a higher average variance, however, at the end of the 47 iterations, it managed to reduce more in percentage than the scenario with 535 simulations.

In order to evaluate the obtained metamodel with the 500 simulations, it was used to predict the results of the 35 simulations performed later. Table 4.2 presents the results obtained.

Observing the results, it is possible to see that the metamodel is still not perfect, especially when the number of UAM legs is reduced. However, for high UAM legs values, the metamodel presents a good

Table 4.2: Predictions over the 35 simulations with the metamodel of the 500 simulations.

Simulation	**Predict**	**Simulation**	**Predict**	**Simulation**	**Predict**	**Simulation**	**Predict**
42	45.4	2	19.8	28	33.1	42	40.4
42	44.8	28	31.6	34	38.1	42	38.8
42	47.1	28	30.2	28	41.4	40	38.2
42	46.1	28	29.6	28	40.6	26	24.9
42	46.2	26	28.8	40	42.5	22	23.2
2	20.1	26	28.7	40	43.7	32	26.5
11	22.9	34	37.3	28	41.5	26	25.4
2	19.6	32	34.1	42	41.6	22	23.3
2	19.9	34	38.0	42	41.8		

estimate.

In conclusion, due to the reduced test group, it was not possible to reduce the desired mean variance value and, thus, complete the metamodel, in order to have good predictions. Therefore, according to the results, it is important to have more simulations to add to the training group, mainly simulations with low number of UAM legs in the result.

Chapter 5

Conclusions

The objectives of this thesis were to instantiate an agent-based transport model of on-demand UAM services in São Paulo Metropolitan Region under the MATSim traffic simulator using the UAM extension in order to analyse the impact of this new transport mode and to create a metamodel based on Gaussian Processes with active learning, to predict the simulation outputs. Therefore, in this Section, the main results achieved will be explored as well as future work to be developed.

5.1 Achievements

The creation of the synthetic population and the network and the UAM service definition were properly done. This allowed to run several simulations in the MATSim framework, with the UAM extension and to evaluate the impact that the introduction of UAM will have in the city of São Paulo, which is the first objective of the thesis.

First, regarding the simulation model, it is possible to conclude that the introduction of this new transport mode can be a great driver for the reduction of congestion on the roads. Furthermore, based on the results obtained and taking into account that the synthetic population and the network generated are very close to the reality, it can be concluded that the introduction of the UAM service will be well accepted by society, however, it is important to remember that there are other aspects that can affect people's opinion, such as noise.

Regarding the results of the simulations of the base scenario, the parameters defined cause a high level of air congestion. However, analyzing the variation of the input parameters, it is possible to conclude that the reduction in the number of stations and the number of UAM vehicles has a very negative effect on the number of trips made by the UAM service. Therefore, having a high number of stations well distributed throughout the city is essential. On the other hand, the increase in cruising speed and vertical speed have a positive influence. However, there is no big advantage in increasing the cruising speed too much beyond 150km/h, in this case. Regarding the fares and the ground process time, due to the high congestion, no major conclusions can be drawn. With regard to vehicle capacity, it is not possible to determine anything, as the vehicle does not wait for more agents and leaves with just one.

Using the updated scenario of the Corsica island with the 10 stations, it was possible to create the metamodel based on Gaussian Processes with active learning, however it needs some improvements. Although the metamodel did not achieve the desired results, the adopted strategy allows to reduce the average variance over the iterations of the metamodel. Furthermore, it is possible to conclude that the metamodel created allows to predict with some quality the results of the simulations, when the number of UAM legs is high. On the contrary, when the simulations results are a reduced value, the metamodel has some difficulty in being able to predict, due to the reduced number of training cases with a reduced output value.

5.2 Future work

In the future, regarding the simulation model, it will be important to adapt the base scenario, in order to adjust it even more to reality. For this, it will be important to obtain information from residents in São Paulo, and adapt the parameters accordingly. It will also be interesting to know which zones will be prohibited for vehicles to circulate and, thus, adjust the network. It will also be important to understand where the best places for the placement of stations will be, verifying in this sense if there are already vertiports in the city and trying to adjust the stations according to other transport modes. It will also be interesting to allow vehicles to wait at stations, in order to be able to carry more than one passenger on their trips and thus reduce air congestion. Another important improvement would be the individual characterization of each station, since they don't have all the same parameters.

With regard to the metamodel, it will be important to adapt the strategy so that the most informative points that are selected are distanced from each other, a certain value, in order to better explore the input space and speed up the process. In relation to the Corsica scenario, it is necessary to increase the test group so that the metamodel can progress and then be able to improve the predictions in relation to simulations that have a lower number of UAM legs. It will also be necessary to create new strategies, if one intends to apply the metamodel to the São Paulo scenario, as it is impossible to carry out so many simulations in this scenario, due to the long time it takes for each simulation to run.

Bibliography

[1] ONU. World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100. https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html, 2017. Accessed: 19/06/2021.

[2] Globo1. Average car travel time has dropped in sp in recent years, but average slowness has risen at peak times. https://g1.globo.com/sp/sao-paulo/noticia/2019/07/12/anda-sp-tempo-medio-de-viagens-de-carro-caiu-em-sp-nos-ultimos-anos-mas-lentidao-\media-subiu-nos-horarios-de-pico.ghtml, 2019. Accessed: 11/04/2021.

[3] A. P. Raoul Rothfeld, Anna Straubinger and C. Antoniou. Analysis of european airports' access and egress travel times using google maps. *Transport Policy*, 81:148–162, January 2019.

[4] K. R. A. Christopher Silva, Wayne Johnson and M. D. Patterson. Vtol urban air mobility concept vehicles for technology development. June 2018.

[5] V. Gollnick, M. Niklaß, M. Swaid, J. Berling, and N. Dzikus. A methodology and first results to assess the potential of urban air mobility concepts. 02 2020.

[6] M. Shamiyeh, R. Rothfeld, and M. Hornung. A performance benchmark of recent personal air vehicle concepts. 09 2018.

[7] A. Straubinger, R. Rothfeld, M. Shamiyeh, K.-D. Büchter, J. Kaiser, and K. Plötner. An overview of current research and developments in urban air mobility -setting the scene for uam introduction. 07 2019.

[8] ESA. Urban air mobility (uam). https://www.easa.europa.eu/domains/urban-air-mobility-uam, 2020. Accessed: 11/04/2021.

[9] Z. Trattner. Modeling future mobility. https://www.iaacblog.com/programs/modeling-future-mobility/, 2016. Accessed: 13/04/2021.

[10] F. Galton. On the construction of isochronic passage charts. *Proceedings of the Royal Geographical Society*, 3:657–658, January 1881.

[11] E. Weiner. Urban transportation planning in the united states: an historical overview fifth edition, 1997.

[12] R. Moeckel, N. Kuehnel, C. Llorca, A. T. Moreno, and H. Rayaprolu. Agent-based simulation to improve policy sensitivity of trip-based models. *Journal of Advanced Transportation*, 2020, 2020.

[13] M. L. Manheim. The challenge of transportation systems analysis, 1979.

[14] A. Straubinger and R. Rothfeld. Identification of relevant aspects for personal air transport system integration in urban mobility modelling. *Transport Research Arena TRA*, 2018.

[15] S. Hörl and M. Balac. Open data travel demand synthesis for agent-based transport simulation: A case study of paris and Île-de-france. *Arbeitsberichte Verkehrs- und Raumplanung*, 1499, November 2020.

[16] K. W. Axhausen and T. Gärling. Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transport Reviews*, 12:323–341, 1992.

[17] J. Bowman and M. Ben-Akiva. Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice*, 35:1–18, 2001.

[18] F. S. Chapin. Human activity patterns in the city: Things people do in time and in space, 1974.

[19] T. Hägerstrand. What about people in regional science? *Papers of the Regional Science Association*, 24:6–21, 1970.

[20] J. L. Hennessy and D. A. Patterson. A new golden age for computer architecture. *Commun. ACM*, 62(2):48–60, 2019. doi: 10.1145/3282307. URL https://doi.org/10.1145/3282307.

[21] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedingsof the National Academy of Sciences*, pages 7280–7287, 2002.

[22] A. Bazzan and F. Klügl. A review on agent-based technology for traffic and transportation. *The Knowledge Engineering Review*, 29:375–403, 06 2013. doi: 10.1017/S0269888913000118.

[23] B. Chen and H. Cheng. A review of the applications of agent technology in traffic and transportation systems. *Intelligent Transportation Systems, IEEE Transactions on*, 11:485 – 497, 07 2010. doi: 10.1109/TITS.2010.2048313.

[24] Z. Kokkinogenis, L. S. Passos, R. Rossetti, and J. Gabriel. Towards the next-generation traffic simulation tools: a first evaluation. In *Proceedings of the 6th Doctoral Symposium In Informatics Engineering. FEUP, University of Porto, Portugal*, pages 77–90. Citeseer, 2011.

[25] M. B. Sebastian Hörl and K. W. Axhausen. Pairing discrete mode choice models and agent-based transport simulation with matsim. *98th Annual Meeting of the Transportation Research Board (TRB 2019)*, July 2019.

[26] M. Fu, R. Rothfeld, and C. Antoniou. Exploring preferences for transportation modes in an urban air mobility environment: Munich case study. *Transportation Research Record: Journal of the Transportation Research Board*, 2673:036119811984385, 05 2019. doi: 10.1177/0361198119843858.

[27] J. Auld, M. Hope, H. Ley, V. Sokolov, B. Xu, and K. Zhang. Polaris: Agent-based modeling framework development and implementation for integrated travel demand and network and operations simulations. *Transportation Research Part C: Emerging Technologies*, 64:101–116, 2016.

[28] F. M. L Miguel Martínez, Gonçalo Homem de Almeida Correia and M. M. Lopes. Insights into carsharing demand dynamics: Outputs of an agent-based model application to lisbon, portugal. *International Journal of Sustainable Transportation*, 11:148–159, 2017.

[29] M. B. Sebastian Horl and K. W. Axhausen. A first look at bridging discrete choice modeling and agent-based. *Procedia Computer Science*, 130, 2018.

[30] K. N. Andreas Horni and K. W. Axhausen. *The Multi-Agent Transport Simulation MATSim*. London: Ubiquity Press, August 2016.

[31] A. Horni, K. Nagel, and K. W. Axhausen. *Introducing MATSim*, pages 3–8. Ubiquity Press, 2016. ISBN 9781909188754. URL `http://www.jstor.org/stable/j.ctv3t5r7p.8`.

[32] K. N. Dominik Ziemke and R. Moeckel. Towards an agent-based, integrated land-use transport modeling system. *Procedia Computer Science*, 83, 2016.

[33] R. M. Felix Zwick, Nico Kuehnel and K. Axhausen. Agent-based simulation of city-wide autonomous ride-pooling and the impact on traffic noise. *Transportation Research Part D: Transport and Environment*, 90, 2021.

[34] H. Ayed, D. Khadraoui, and R. Aggoune. Using matsim to simulate carpooling and car-sharing trips. In *2015 World Congress on Information Technology and Computer Applications (WCITCA)*, pages 1–5, 2015. doi: 10.1109/WCITCA.2015.7367046.

[35] F. Ciari, M. Balac, and K. Axhausen. Modeling carsharing with the agent-based simulation matsim: State of the art, applications, and future developments. *Transportation Research Record: Journal of the Transportation Research Board*, 2564:14–20, 01 2016. doi: 10.3141/2564-02.

[36] M. Balac, H. Becker, F. Ciari, and K. Axhausen. Modeling competing free-floating carsharing operators – a case study for zurich, switzerland. *Transportation Research Part C Emerging Technologies*, 98:101–117, 11 2018. doi: 10.1016/j.trc.2018.11.011.

[37] P. Segui-Gasco, H. Ballis, V. Parisi, D. Kelsall, R. North, and D. Busquets. Simulating a rich ride-share mobility service using agent-based models. *Transportation*, 46, 12 2019. doi: 10.1007/s11116-019-10012-y.

[38] J. Bischoff, I. Kaddoura, M. Maciejewski, and K. Nagel. Simulation-based optimization of service areas for pooled ride-hailing operators. *Procedia Computer Science*, 130:816–823, 01 2018. doi: 10.1016/j.procs.2018.04.069.

[39] R. Vosooghi, J. Kamel, J. Puchinger, V. Leblond, and M. Jankovic. Robo-taxi service fleet sizing: assessing the impact of user trust and willingness-to-use. *Transportation*, 46, 12 2019. doi: 10.1007/s11116-019-10013-x.

[40] D. Fagnant and K. Kockelman. Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in austin, texas. *Transportation*, 45, 01 2018. doi: 10.1007/s11116-016-9729-z.

[41] K. Winter, O. Cats, K. Martens, and B. Arem. Relocating shared automated vehicles under parking constraints: assessing the impact of different strategies for on-street parking. *Transportation*, 48, 08 2021. doi: 10.1007/s11116-020-10116-w.

[42] M. Maciejewski. *Dynamic Transport Services*, pages 145–152. Ubiquity Press, 2016. ISBN 9781909188754. URL `http://www.jstor.org/stable/j.ctv3t5r7p.30`.

[43] M. Balac, A. R. Vetrella, and K. Axhausen. Towards the integration of aerial transportation in urban settings. *97th Annual Meeting ofthe Transportation Research Board*, January 2018. doi: 10.3929/ethz-b-000193150.

[44] D. Grether. Extension of a multi-agent transport simulation for traffic signal control and air transport systems. Technical University of Berlin, 2014.

[45] R. Rothfeld. *Agent-based Modelling and Simulation of Urban Air Mobility Operation - An Evaluation of Travel Times and Transport Performance*. PhD thesis, 07 2021.

[46] C. Al Haddad, M. Fu, A. Straubinger, K. Plötner, and C. Antoniou. Choosing suitable indicators for the assessment of urban air mobility: A case study of upper bavaria, germany. *European Journal of Transport and Infrastructure Research*, 20:214–232, 11 2020. doi: 10.18757/ejtir.2020.20.4.4464.

[47] V. Bulusu, E. Onat, R. Sengupta, P. Yedavalli, and J. Macfarlane. A traffic demand analysis method for urban air mobility. *IEEE Transactions on Intelligent Transportation Systems*, PP:1–9, 01 2021. doi: 10.1109/TITS.2021.3052229.

[48] M. Balac, A. R. Vetrella, R. Rothfeld, and B. Schmid. Demand estimation for aerial vehicles in urban settings. *IEEE Intelligent Transportation Systems Magazine*, 11:105–116, 06 2019. doi: 10.1109/MITS.2019.2919500.

[49] M. Balac, R. Rothfeld, and S. Hörl. The prospects of on-demand urban air mobility in zurich, switzerland. pages 906–913, 10 2019. doi: 10.1109/ITSC.2019.8916972.

[50] R. Rothfeld, M. Balac, K. Plötner, and C. Antoniou. Initial analysis of urban air mobility's transport performance in sioux falls. 06 2018. doi: 10.2514/6.2018-2886.

[51] K. Plötner, C. Al Haddad, C. Antoniou, F. Frank, M. Fu, S. Kabel, C. Llorca, R. Moeckel, A. Moreno, A. Pukhova, R. Rothfeld, M. Shamiyeh, A. Straubinger, H. Wagner, and Q. Zhang. Long-term application potential of urban air mobility complementing public transport: an upper bavaria example. *CEAS Aeronautical Journal*, 11, 08 2020. doi: 10.1007/s13272-020-00468-5.

[52] R. Rothfeld, M. Fu, M. Balac, and C. Antoniou. Potential urban air mobility travel time savings: An exploratory analysis of munich, paris, and san francisco, 02 2021.

[53] J. Kleijnen and R. Sargent. A methodology for fitting and validating metamodels in simulation. Workingpaper, Operations research, 1997. Pagination: 36.

[54] F. Antunes, B. Ribeiro, F. Pereira, and R. Gomes. Efficient transport simulation with restricted batch-mode active learning. *IEEE Transactions on Intelligent Transportation Systems*, PP:1–10, 07 2018. doi: 10.1109/TITS.2018.2842695.

[55] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN 026218253X.

[56] B. Settles. Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison, 2009. URL `http://axon.cs.byu.edu/~martinez/classes/778/Papers/settles.activelearning.pdf`.

[57] N. Roy and A. Mccallum. Toward optimal active learning through monte carlo estimation of error reduction. 01 2001.

[58] H. Seung, M. Opper, and H. Sompolinsky. Query by committee. In *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory, pages 287–294. Publ by ACM, Jan. 1992. ISBN 089791497X. doi: 10.1145/130385.130417. Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory ; Conference date: 27-07-1992 Through 29-07-1992.

[59] D. Lewis and W. Gale. A sequential algorithm for training text classifiers. *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in Information Retrieval*, 29, 08 2001. doi: 10.1145/219587.219592.

[60] B. Settles and M. Craven. An analysis of active learning strategies for sequence labeling tasks. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 1070–1079, Honolulu, Hawaii, Oct. 2008. Association for Computational Linguistics. URL `https://aclanthology.org/D08-1112`.

[61] S. Geman, E. Bienenstock, and R. Doursat. Neural networks and the bias/variance dilemma. *Neural Computation*, 4:1–58, 01 1992. doi: 10.1162/neco.1992.4.1.1.

[62] B. Settles, M. W. Craven, and S. Ray. Multiple-instance active learning. In *NIPS*, 2007.

[63] W. Song, K. Han, Y. Wang, T. Friesz, and E. Del Castillo. Statistical metamodeling of dynamic network loading. *Transportation Research Procedia*, 23:263–282, 2017. ISSN 2352-1457. doi: 10.1016/j.trpro.2017.05.016. Publisher Copyright: © 2017 The Authors. Elsevier B.V. All rights reserved.

[64] B. Ciuffo, J. Casas, M. Montanino, J. Perarnau, and V. Punzo. Gaussian process metamodels for sensitivity analysis of traffic simulation models. case study of aimsun mesoscopic model. *Transportation Research Record Journal of the Transportation Research Board*, 2390:87–98, 12 2013. doi: 10.3141/2390-10.

[65] X. M. Chen, Z. Zhu, X. He, and L. Zhang. Surrogate-based optimization for solving a mixed integer network design problem. *Transportation Research Record Journal of the Transportation Research Board*, 2497, 03 2015. doi: 10.3141/2497-13.

[66] L. Zhang, X. Chen, X. He, C. Xiong, and Z. Zhu. Bayesian stochastic kriging metamodel for active traffic management of corridors. *IIE Annual Conference and Expo 2014*, pages 1790–1799, 01 2014.

[67] F. Antunes, B. Ribeiro, F. C. Pereira, and R. Gomes. Efficient transport simulation with restricted batch-mode active learning. *IEEE Transactions on Intelligent Transportation Systems*, 19(11): 3642–3651, 2018. doi: 10.1109/TITS.2018.2842695.

[68] X. M. Chen, L. Zhang, X. He, C. Xiong, and Z. Li. Surrogate-based optimization of expensive-to-evaluate objective for optimal highway toll charges in transportation network. *Computer-Aided Civil and Infrastructure Engineering*, 29, 12 2013. doi: 10.1111/mice.12058.

[69] Deloitte. Deloitte city mobility index 2020. `https://www2.deloitte.com/content/dam/insights/us/articles/4331_Deloitte-City-Mobility-Index/SaoPaulo_GlobalCityMobility_WEB.pdf`, 2020. Accessed: 23/08/2021.

[70] F. Antunes, M. Amorim, F. Pereira, and B. Ribeiro. Active learning metamodeling for policy analysis: application to an emergency medical service simulator. *Simulation Modelling Practice and Theory*, 97:101947, 06 2019. doi: 10.1016/j.simpat.2019.101947.