Technische Universität München
Master's Program in Transportation Systems

Master's Thesis

Total Cost Minimization Models for Shared Demand Responsive Transport Systems with Human Driven and Automated Vehicles

Author:
Aitan Militão

Supervised by:
Prof. Dr. Alejandro Tirachini

December 31, 2019

Abstract

Shared responsive transportation services are expected to have reduced operation cost with vehicle automation; this has raised questions about the feasibility of replacing fixed-route public transport by such services. The expected increase in competitiveness between the two services requires the development of frameworks that enable the analysis of the transportation costs. These tools can be used to guide the service selection process by designers, planners, and policymakers.

This research develops a total cost model for demand-responsive shared transportation services, including operator and user costs, for two different operational scenarios. A hybrid approach is used in which variables are analytically modeled and derived from multiple regression analyses using output data from agent-based simulations. A numerical application and fleet size optimization are performed using parameters from the city of Munich for different demand levels, and alternative scenarios for vehicle automation are analyzed.

The results indicate that the performance of on-demand shared systems depends on the operational scheme selected. If the system is provided to behave as public transport, without rejection of trips, economies of scale are not present, and the high user costs hinder its competitiveness even under the assumption of automated vehicles. In contrast, a system that allows for trip rejection based on waiting and travel times, economies of scale are present, and vehicle automation can reduce average operation costs at lower user cost levels, increasing the system competitiveness against other transportation modes.

Acknowledgments

I would like to express deep gratitude to Professor Alejandro Tirachini for supervising this thesis. His guidance throughout all the stages in the development of this work has enabled it to be successful and have made me broaden my knowledge and skills. Also, he has inspired me to do better and proper and meaningful research on transportation engineering.

I also like to thank Prof. Dr. Constantinos Antoniou and the staff of the Chair of Transportation Systems Engineering at TUM for providing the necessary resources that enabled the simulations and the research performed in this thesis.

I am also grateful to my family and all the close friends for their unceasing support during this work.

Table of contents

List of Figures

List of Tables

List of Abbreviations

ICT	Information and Communications Technology
AV	Automated Vehicle
DRT	Demand Responsive Transport -
MaaS	Mobility as a Service
UAM	Urban Air Mobility
VKT	Vehicle-Kilometers Travelled

Chapter 1

Introduction

The advances in Information and Communications Technology – ICT, together with the introduction of automated vehicles (AVs) and more efficient electric propulsion systems, are expected to change the mobility landscape. These technology advances have raised questions about new operational models for future transportation and their expected performance. Therefore, the development of frameworks that enable the analysis of these new transportation system's performance is needed.

Under this perspective, the analysis of expected upcoming new transportation models have gained focus on research in the last years, and many studies in this area have been developed. Services such as automated shared taxis and Demand Responsive Transport - DRT systems have been analyzed as alternatives to traditional line-based mass transit and private cars (Alonso-Mora et al., 2018; Badia & Jenelius, 2019; Bischoff et al., 2017; Bösch et al., 2018; Leich & Bischoff, 2019). The effects of automation in line-based public transportation was also performed recently and demonstrated the reduction of costs and operational perspectives (Tirachini & Antoniou, 2020). The impact of Mobility as a service (MaaS) including car-sharing, bike-sharing and ride-hailing on welfare have also been analyzed in city-scale simulations (Becker et al., 2019) and novel transportation modes, such as Urban Air Mobility (UAM) and its operation and adoption have been studied recently (Fu et al., 2019; Rothfeld et al., 2018b).

Although these scenarios and simulations depend on the accuracy of assumptions used, they can be used as estimates of the new transportation model's competitiveness. These studies allow policymakers, planners, and designers to better access the future transport demand, mobility patterns of people, and transportation externalities.

Among these new perspectives, a crucial case for transportation providers and city planners is the expected increase in competitiveness between conventional line-based public transport and shared DRT systems with the introduction of AVs. The usage and theoretical study of DRT systems have been limited to low demand density areas and as feeder transit services (Badia & Jenelius, 2019; Li & Quadrifoglio, 2010). However, the decrease in costs of such systems have enabled its provision as a mass transportation service (Jokinen et al., 2011), and recent focus has been given in analyzing the different operational schemes and its outcomes in different simulated scenarios (Bischoff et al., 2017; S. Hörl et al., 2019; Wang et al., 2019).

The analysis of the system design variables involved in the provision of DRT systems, such as fleet size, vehicle capacity, and dispatching strategies are of great importance in

determining the operational model's viability in terms of user acceptability and competitiveness of cost structures.

1.1. Research Motivation

The choice of technology for a transit system is a key strategic decision in urban transport policy. The comparison of the performance of different options for public transportation has been limited mostly to light and heavy rail, bus, and bus rapid transit (Tirachini et al., 2010a). The analysis of the performance of new transportation services that are being considered transit options, such as shared DRT, is therefore essential.

To the best of the author's knowledge, the related studies that analyze the performance of DRT systems rely mostly on operations characteristics (Alonso-Mora et al., 2018; Bischoff et al., 2017; Wang et al., 2019). Few are related to modeling the costs, and these latter are limited to feeder transit services involving only many-to-one demand patterns. (Badia & Jenelius, 2019; Kim & Schonfeld, 2014; Li & Quadrifoglio, 2010). Moreover, the traditional approach in such studies is to assume operational constraints based on maximum waiting time or travel time of customers, enabling the rejection of trips. If such systems are proposed to substitute traditional line-based public transport, no trip should be rejected.

Therefore, the development of a framework enabling the assessment of user and operator costs under different demand scenarios with human driven and automated vehicles is an essential tool for policymakers and transportation providers to better prepare cities for the upcoming future and of particular interest.

1.2. Research Questions and Objectives

The study has the following objective:

The development of a framework that enables modeling the total costs, including user cost and operator cost of shared DRT services for human-driven and automated vehicles.

Based on the aforementioned objective and issues discussed, the following research questions are proposed:

- Can total cost models be formulated with the utilization of analytical and simulation data output for shared DRT services?

- How do the operational scenarios analyzed for shared DRT perform under the minimization of total costs?

- What is the expected impact of automation on average costs of shared DRT systems for optimal provision?

1.3. Research Framework

This thesis is divided according to the research framework presented in Figure 1. The problem definition encompasses the motivation, objectives, research questions, and literature

review on related work. In the methodology, the theoretical total cost model is developed, data is collected and analyzed, and the simulation scenarios are set. Moreover, under results and discussion, the simulation results are analyzed, and the model formulated. A numerical application and optimization are performed, and results discussed. The conclusion is performed, followed by limitations and future work.

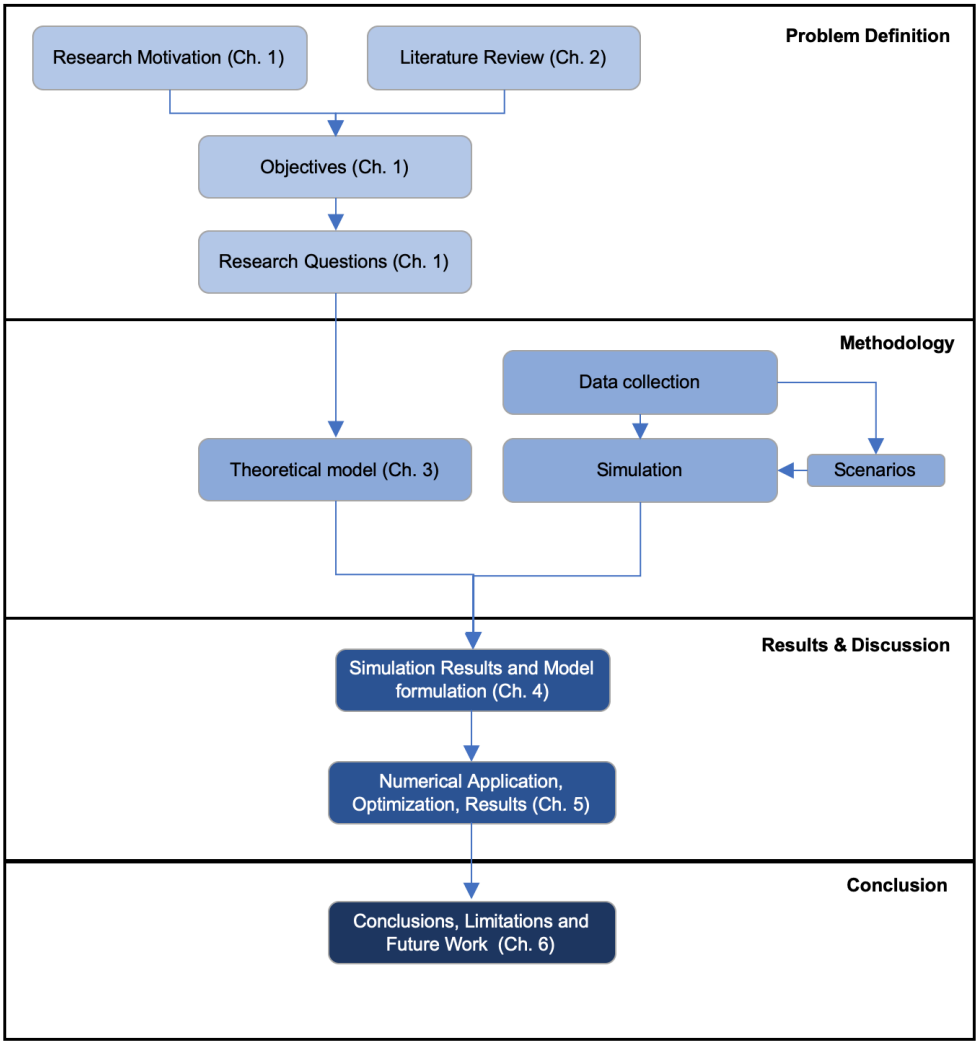

Figure 1 – Thesis framework

1.4. Thesis Contributions

The contributions of this thesis can be divided in two groups: Methodological contributions and Practical contributions.

1.4.1. Methodological contributions

This study is expected to contribute with the following methodological aspects:

- Development of a hybrid framework for modeling transportation total costs, including an analytical approach and functions derived from the modeling of shared DRT systems design variables, namely fleet size, vehicle capacity and demand, with system output variables: detour factor, average occupancy,

average waiting time and rejection rate using output agent-based simulation data; (Chapter 3 and 4).

- A comparison of operational scenarios with and without rejections for shared DRT systems. (Chapter 4)

1.4.2. Practical contributions

The practical relevance of this research is described below.

- Optimization of fleet size for different demand levels of two operational scenarios of shared DRT and estimation of average user, operator and total cost estimation based on parameters for the city of Munich. The values obtained can be used to compare the competitiveness of such system with different transportation modes from a provider perspective. (Chapter 5)
- Estimation of vehicle automation impact on costs and waiting time. The impact of automation on the costs and waiting time can be used as reference for future research and comparison with different transport modes in different scenarios. (Chapter 5)
- A comparison of the performance between scenarios with and without rejections for shared DRT systems. (Chapter 5).

Chapter 2

Literature Review

This chapter encompasses the literature review on DRT services and transportation total cost modeling and minimization. The overview of previous research includes simulation studies for different scenarios and cost estimations, including the assumption of vehicle automation.

2.1. Demand Responsive Transport

Demand Responsive Transport – DRT services are flexible shared transport services that adapt its route dynamically accordingly to the demand, without fixed routes or timetables (Jokinen et al., 2011). Although the concept of on-demand shared mobility has only become mainstream recently, its concept dates back to the 1970s, when DRT it was first introduced as a specific mode designed for people with mobility difficulties and had also the alternative name of Paratransit (Enoch et al., 2006). Because of the high costs involving operations and technology, this type of transportation has been mostly used and studied as a transit feeder service or in low demand density areas (Badia & Jenelius, 2019; Viergutz & Schmidt, 2019).

An important characteristic of DRT services is the trade-off observed between flexibility of a door-to-door service, and the increase in trip detour to serve passenger requests (Leich & Bischoff, 2019). From the operator perspective, the increase in shared rides reduce the system provision costs, on the other hand, the shared rides increase the waiting and travel times of users, making this system less attractive.

This transportation service has gained attention with the advances in ICT, introduction of AVs and more efficient electric propulsion systems, which are expected to reduce the operation costs and enable more efficient dispatching strategies. This recent attention focus on new implementations of DRT that propose large scale transportation solutions and are not restricted to the traditional service designed for disabled or elderly people. Therefore a distinction between these services is important (Jokinen et al., 2011).

The operation and performance of this type of service have been analyzed in different studies recently. Most of them rely on simulations of DRT services with different dispatching algorithms that assumes a fleet controlled by a central dispatcher responsible by assigning the requests to the vehicles and with introduction of AV. Systems with and without pooling of passengers are both analyzed. Different dispatching strategies have been developed and tested. These strategies have shown to be an important variable in the system design, being able influence directly on average waiting times and occupancy rates. (Alonso-Mora et al.,

2018; Bischoff et al., 2017; S. Hörl et al., 2019; Inturri et al., 2019; Moreno et al., 2018; Tsao et al., 2019; Wang et al., 2019).

Regarding potential benefits, the studies suggest that shared DRT with AV fleets can reduce potentially the number of vehicle in the roads, however it can also increase the Vehicle-Kilometers Travelled (VKT) (Dia & Javanshour, 2017; Javanshour et al., 2019; Moreno et al., 2018). In contrast to the traditional usage of DRT for low-density areas, Bischoff et al. (2017) simulated DRT services based on taxi demand for Berlin and found that sharing of rides works better in high demand areas.

Most of the studies use static demand for simulations and few research have been performed analyzing the system considering elastic demand, Sebastian Hörl et al. (2019) recently proposed a simulation framework including autonomous mobility on demand and a discrete choice model. The system performance in Paris is analyzed and found a dependency between fleet size and demand, showing the importance of the fleet size on the system design. Loeb and Kockelman (2019) simulated a fleet of shared autonomous electric vehicles in Austin, Texas, including a mode choice model in order to assess the vehicle factors that make the system more profitable. This study shows that not only the fleet size is important but also the vehicles characteristics.

Moreover, the majority of studies focus on simulation and operational aspects. Most of the operations allows for the rejection of requests and constraints for rejection based on maximum waiting times are common.

2.2. Transportation total cost modeling and minimization

The minimization of transportation total costs is a common method for public transport operations. Total costs C_t comprise operator cost C_o and user cost C_u, and the latter is a function of the times spent on transportation, such as access, waiting, and in-vehicle time. The objective function minimized usually has the form:

$$C_t = C_o + C_u(time)$$

When performed with inelastic demand models, the minimization of total cost is equivalent to the maximization of social welfare (Jara-Diaz, 1990).

The total cost minimization method has been used to model the system design variables that lead to the optimal total cost in transportation research. In the case of public transportation, the first studies focused on optimal values of service frequency, vehicle size, and stop spacing (Jansson, 1980; Kuah & Perl, 1988; Mohring, 1972). Subsequent new models and extensions were developed focusing on optimizing the number of lines, route density, spatial structure of bus services and short turn strategies (Delle Site & Filippi, 1998; Jara-Díaz & Gschwender, 2003; Shyue Koong & Schonfeld, 1991; Tirachini, 2014; Tirachini et al., 2010a).

Regarding its use with DRT services analysis, few studies have been performed. Focus is more given on the analysis of the substitution of fixed-route transit feeder services for demand-responsive systems (Badia & Jenelius, 2019; Li & Quadrifoglio, 2010). The optimization of system variables is mostly performed through simulations and iterative calculation of costs (Inturri et al., 2019). L. Wang et al. (2018) developed a model for minimizing the costs of a DRT system using zonal strategy, the area of service and bus headway are the system variables optimized. However, this latter study is also limited to a one-to-many routing, in which the DRT system connects a residential area to a terminal.

B. Wang et al. (2018) explored the optimal fleet size and different deployment strategies using agent-based simulation for station-based DRT services. A cost-based routing module is proposed, and the optimal fleet size is selected after simulation iterations.

To the best of the author's knowledge, there is no current literature on total costs minimization models of shared DRT systems in a network scale and with many-to-many routing that are able to analyze average total costs.

The comparison of different public transport modes has also been studied by minimizing total cost models. In these studies, the marginal costs performance is analyzed under different assumptions and scenarios in order to identify the conditions on which it makes more sense to provide a specific mode (Boyd et al., 1978; Bruun, 2005; Tirachini et al., 2010a, 2010b)

Regarding vehicle automation, Tirachini and Antoniou (2020) recently have demonstrated the effects of the reduction in operating costs of line based public transport. The same model proposes optimum service with smaller vehicles and shorter headways. Badia and Jenelius (2019) compared the substitution of a fixed-route system for a door-to-door transit service for last-mile solutions. They found that the demand threshold in which fixed routes are more competitive is higher under the automation scenario.

These studies have revealed interesting characteristics of public transport operations, such as the Mohring Effect, for example, in which the optimal bus frequency increases less proportionally than the increase in demand, diminishing average costs (Mohring, 1972). Such findings turn out to be handful tools for transportation planners and designers.

Chapter 3

Methodology

3.1. Total Cost Minimization Model

The performance of a shared Demand Responsive Transport - DRT service is analyzed analytically and numerically. A total cost model is developed and minimized.

The service operation is modelled under the assumption of a monopolistic market, in which the monopolist is able to change the supply of vehicles and select different input variables in the service.

Two operational scenarios analyzed are:

- Scenario 1: A zero-rejection operation scenario, in which every request is accepted.
- Scenario 2: A constrained operational scenario, in which there are rejections for requests based on waiting and in-vehicle time criteria.

The model considers temporal and spatial components of operator costs and user cost, following similar approach to Tirachini et al. (2010b).

The total cost is modelled for one hour of operation and is comprised of operator cost and user cost:

$$C_t = C_o + C_u \quad (1)$$

- C_o (€/hour): Operator cost
- C_u (€/hour): User cost

The operator cost is modelled as follows:

$$C_o = c_0^{rs} * B + c_1^{rs} * D_B \quad (2)$$

- c_0^{rs}(€/veh-hour): the cost associated to each vehicle in operation in the fleet B for every hour, including driver salaries, vehicle capital cost and charging infrastructure (in case of electric vehicles);
- c_1^{rs}(€/km-hour): variable cost associated to the number of kilometers operated by the fleet (energy consumption and maintenance);
- B: number of vehicles (fleet size)
- D_B (km): Total distance travelled by the vehicle fleet B in one hour of operation.

The total distance travelled by the vehicle fleet B (D_B) is given by the following expression:

$$D_B = \frac{Q * l * \rho(B, C, Q)}{o_B(B, C, Q)} \quad (3)$$

- l: average direct trip length when not considering shared trips (normally, it is the shortest path distance);
- Q: total demand of trips per hour (trips/h);

- ρ: average distance detour factor;
- O_B: average occupancy (pax/veh);
- C: vehicle capacity (number of seats);

Differently from line based public transportation, that follows a schedule with a defined frequency and headway, a shared DRT service has a dynamic operation and is able to generate different detour factors and occupancy rates depending on many factors, such as network characteristics and dispatching and rebalancing strategies. For the same network, dispatching strategy and same vehicle characteristics, the detour factor ρ and occupancy O_B are modelled as functions of fleet size, demand and vehicle seat capacity.

A seen in (3), the total distance travelled by the vehicles D_B depends on the ratio between detour factor and average occupancy, therefore the behavior of these two system output variables due to changes in the supply levels is the main factor in determining the fleet mileage contribution in the operator costs.

The expected relationship between the distance travelled by the fleet, and the fleet size is given by:

$$\frac{\partial D_B}{\partial B} = Q * l * \left| \frac{\frac{\partial \rho}{\partial B} * O_B - \rho * \frac{\partial O_B}{\partial B}}{{O_B}^2} \right| \quad (4)$$

Therefore:

If $(\frac{\partial \rho}{\partial B} * O_B - \rho * \frac{\partial O_B}{\partial B}) > 0$, then $\frac{\partial D_B}{\partial B} > 0$

If $(\frac{\partial \rho}{\partial B} * O_B - \rho * \frac{\partial O_B}{\partial B}) < 0$, then $\frac{\partial D_B}{\partial B} < 0$

This derivative shows the importance of the ratio between distance detour factor and average occupancy aforementioned.

Due to the dynamics of the on-demand shared service, the relationships $\frac{dD_B}{dB}$, $\frac{dD_B}{dQ}$ and $\frac{dD_B}{dC}$ depend on the detour factor and average vehicle occupancy, both of which are modelled as functions of B, C and Q.

The user cost is modelled as follows:

$$C_u = P_a * T_a + P_w * T_w + P_v * T_v \quad (5)$$

- P_a: Value of access time savings;
- P_w: Value of waiting time savings;
- P_v: Value of in-vehicle time savings;
- T_a: Total access time;
- T_w: Total waiting time;

- T_v:Total in-vehicle time;

Given the fact that the system analyzed is a door-to-door system, access time is assumed to be zero, therefore the user cost is:

$$C_u = P_w * T_w + P_v * T_v \quad (6)$$

Total waiting time is calculated by multiplying the hourly demand Q and the average waiting time t_w per trip. The average waiting time t_w is also a variable that depends on the dynamics of the operation, and it is modelled as a function of fleet size, demand and vehicle capacity.

$$T_w = Q * t_w \quad (7)$$

The total in-vehicle time is modelled as follows, in which v_m is the average speed of the vehicles:

$$T_v(Q, B, C) = \frac{Q * l * \rho(B, C, Q)}{v_m} \quad (8)$$

On scenario 2 the system allows for rejection of trips; therefore, the total cost functions have to be adapted to the presence of rejections. Realized demand Q_r is used instead of demand Q in equations (3), (7), and (8) . Realized demand is defined as:

$$Q_r = \big(1 - r(B, C, Q)\big) * Q \quad (9)$$

The rejection rate r is modelled as a function of fleet size, demand and vehicle capacity.

3.2. Waiting time – operation with no request rejections

The average waiting time for a passenger in a shared DRT service is the time between the request and the pick-up, it comprises of the time spent by the vehicle on route to pick the passenger. Therefore, average waiting time is modelled as a function of demand, fleet size and vehicle capacity.

For a fixed fleet size, an increase in the demand increases the probability of a new trip request after the current passenger request, and since there is no rejection this leads to an increase in the route length to the pick-up point and in the number of stops in between, increasing the average waiting time:

$$\frac{\partial t_w}{\partial Q} > 0 \quad (10)$$

This positive relationship between waiting time and demand is in line with previous research (Jokinen, 2016; Li & Quadrifoglio, 2010).

An increase in the fleet size increase the probability of having a vehicle available in the service area, therefore decreasing the expected distance between the vehicle at the time of request and the pick-up point:

$$\frac{\partial t_w}{\partial B} < 0 \quad (11)$$

This relationship between waiting time and fleet size had already been discussed by Jokinen (2016) and validated by simulation models, such as Alonso-Mora et al. (2018), Shen et al. (2018), Shen et al. (2018) and S. Hörl et al. (2019), for example.

The capacity of the vehicles in the fleet are also an important input in determining the waiting times of passengers. For the scenario of no rejections, if vehicles are not full, waiting time is expected to be indifferent to increases in capacity for the same fleet size. However, if vehicles are full, an increase in capacity of all vehicles is expected to reduce waiting time.

Therefore, the relationship between waiting time and vehicle capacity depends on the system state regarding occupancy.

3.3. In-vehicle time – operation with no request rejections

Average in-vehicle time depends on the demand Q. The route length and number of stops is expected to increase as the demand increases (Jokinen, 2016). This relationship is already showed in the analytical equation (8). It also depends on the detour factor, which is a function of the decision variables. In an operation scenario without rejections, an increase in demand increases the probability of having requests satisfied before the passenger drop-off, therefore:

$$\frac{\partial \rho}{\partial Q} > 0 \; and \; \frac{\partial t_v}{\partial Q} > 0 \quad (12)$$

Results from Li and Quadrifoglio (2010) also have the same relationship for in-vehicle time.

For a fixed fleet size and demand, the increase in vehicle capacity increases the probability having empty seats near new requests, therefore the passengers that are in a trip would have an expected increase on the trip time due to the detour for the pick-up and the boarding time. Alonso-Mora et al. (2018) showed in their simulation model that the in-vehicle time does increase with the increase in vehicle capacity and that capacity is more important for smaller fleet sizes. It also happens because with increasing capacity, the detour factor is also expected to increase in a no-rejection scenario.

$$\frac{\partial \rho}{\partial C} > 0 \; and \; \frac{\partial t_v}{\partial C} > 0 \quad (13)$$

Fleet size is expected to reduce travel time, since given same demand level an increase in fleet size increase the probability of having less shared trips, hence, lower detour factor. This relationship between fleet sizes and shared rates is also observed on Alonso-Mora et al. (2018).

$$\frac{\partial \rho}{\partial B} < 0 \; and \; \frac{\partial t_v}{\partial B} < 0 \quad (14)$$

3.4. Variables under constrained operation

The analytic relationships between the cost function variables under constrained operations scenarios, in which trips can be rejected, such as waiting time and detour factor, are dependent on the operational constraints defined and the behavior of such constraints with the selected dispatching and pooling algorithm used. Therefore, it is not possible to define the expected relationship between those variables analytically.

3.5. Simulation Setup

The operation of shared DRT services includes dynamic decisions and vehicle routing problems. The values of variables such as average waiting time, average detour factor, and average occupancy depend not only on the system design variables but also on the operational decisions of dispatching algorithms and constraints. Therefore, the analytical solution for these variables is not feasible. To account for this, simulations are performed in order to analyze the system dynamics from the output data, and linear regression models are used to describe the output variables behavior. The objective of such approach is to analyze the data output and to model the theoretical causal relationship between the inputs and output variables. The input variables are the system supply design parameters, fleet size, demand and vehicle capacity. The output variables being modeled are distance detour factor, average occupancy, average waiting time, and for the case of constrained operations, rejection rate.

3.5.1. Simulation Framework – MATSim

The tool selected to perform the simulations is the simulation framework MATSim (Horni et al., 2016). MATSim is an open-source activity-based multi-agent transport simulation tool that is able to combine large scale transportation scenarios with microsimulation characteristics and has been used in research to model innovative and futuristic transportation modes; for example Rothfeld et al. (2018a), Balac et al. (2019), Michał Maciejewski and Nagel (2012), Hörl (2017) among others. The agent-based model uses a synthetic population where each agent follows a plan consisting of a chain of activities during the day.

In the MATSim framework the agents optimize their chain of activities repeatedly through simulation iterations by using a co-evolutionary algorithm, in which agents compete for space-time slots on the transportation infrastructure and have their daily activity schedule scored based on the executed plan's performance at every iteration, this process is repeated until the average population score stabilizes (Horni et al., 2016). During this process, at every iteration, the MATSim loop is executed (Figure 2). The loop consists of mobility simulation of an initial demand input (mobsim), where the agents try to execute their plan as scheduled in the environment simultaneously, the traffic is simulated based on a queue-based model. The

scoring step comes after execution, in which every plan is scored based on its performance using a utility function. In the replanning phase, a certain share of agents is allowed to modify their plans (e.g., change activity durations and mode choices). After the replanning phase, the loop starts again.

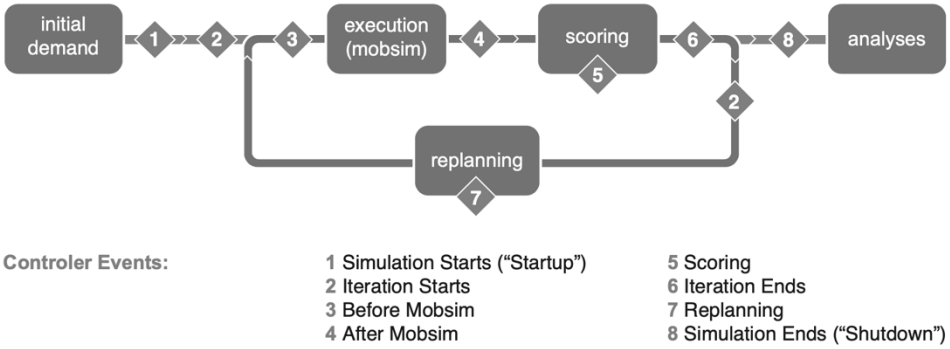

Figure 2 – The MATSim loop with controller events (Zilske, 2016)

In order to simulate the shared DRT service in MATSim, in which decisions are made during the simulation, the DRT (Demand-responsive Transport) extension is used together with the DVRP (Dynamic Vehicle Routing Problem) extension (Michal Maciejewski, 2016). The DVRP allows for the simulation of on-demand transport services by using the concept of dynamic agents, in which agents are able to perform dynamic activities and have trip legs that are created on-demand during the day and can be modified at any time depending on dispatching and optimizing algorithms. DVRP is deployed to model both demand and supply of the specific dynamic transport service, like a taxi service, ride-sharing, and autonomous taxis, whereas MATSim simulates the specific service with the overall transport system (Michal Maciejewski et al., 2017). The DRT extensions use DVRP and enable the pooling of passengers into a single-vehicle; the vehicle dispatching strategy can be constrained to maximum travel times and maximum waiting times, as explained further. These extensions have been used in different studies related to simulation of dynamic shared transportation services (Bischoff et al., 2018; Bischoff & Maciejewski, 2016; Bischoff et al., 2017; Leich & Bischoff, 2019; Michal Maciejewski et al., 2016; Moreno et al., 2018; Viergutz & Schmidt, 2019; B. Wang et al., 2018).

The vehicle dispatching strategy used by this framework is described by Bischoff et al. (2017) and have the following characteristics:

- A request is submitted at the moment of departure (no pre-booking of requests) and specify the exact pickup and drop-off locations (door-to-door service);
- The customer maximum waiting time for departure (waiting and boarding) is fixed as a simulation input;
- The time travelling (the sum of waiting time, boarding and riding) must not exceed $t_r = \alpha * t_r^{direct} + \beta$, in which t_r^{direct} is the direct ride time between origin

and destination of the request, while $\alpha \geq 1$ and $\beta \geq 0$ are input constants used to model the maximum time loss due to waiting, boarding and detours;

- Requests can be rejected based on the constraints only immediately after submission (cannot be rejected later);
- If scheduled, the request is guaranteed to be served, even if delays occur on the way and cause some constrain violation;
- Vehicles are managed by a central dispatch system responsible for scheduling or rejecting incoming requests;
- During the simulation at least one passenger gets in or out at each vehicle stop;
- Each stop is of a fixed duration t_r^{stop} (simulation input);
- Vehicles have a time window in which they operate consisting of initial and end time;
- The arrival at the last stop must be scheduled at latest for the difference of vehicle end time and t_r^{stop} (vehicles don't operate beyond time window);
- The vehicle capacity cannot be exceeded when driving between stops;
- After all scheduled stops have been visited, if the list of stops is empty, the vehicle remains idle until a new dispatch or the end of its time window.

The routing algorithm used in the DRT extension comprises an insertion heuristic with the objective of minimizing the total time spent on handling requests (excluding idle time). When a new request is submitted, it searches the routes of all vehicles for optimal insertions. The request insertion must satisfy the wait and travel time constraints both for the new and already inserted requests, and also, it must satisfy the vehicle time window, otherwise, the vehicle is rejected. The violations of wait and travel time constraints due to delays that vehicles might experience during the simulation affect only the scheduling of new requests, while the ones already accepted are not affected.

The simulations are performed using the Linux Clusters from the Leibniz Rechenzentrum (LRZ) via cloud computing services.

3.5.2. Input Data Analysis

The analysis in the following items contain data acquisition, filtering and exploration, and were conducted using R programming language (R Development Core Team, 2019), QGIS (QGIS Development Team, 2019), Microsoft Excel (Microsoft, 2019) and Java written scripts.

3.5.2.1. Trip Data

The base trip data used in this research to perform the simulations consists of a synthetic population provided by the research group Modeling Spatial Mobility group at Technical University of Munich **(**https://www.msm.bgu.tum.de/) free of charge. This data

consists of trips for one day in the year 2011 in the Munich metropolitan area**.** It contains approximately 8.7 million trips of a synthetic population in a typical working day. The trip data attributes are shown in Table 1. This data contains approximately 5300 zones for the whole region. Each row in the data corresponds to a trip made by an individual of the synthetic population. Each trip and each person are allocated a unique *trip ID* and *person ID*. Every trip is characterized by its origin/destination zones as well as geographical coordinates, mode of travel, travel distance, travel time, purpose of the trip, and departure time. Depending on the purpose, trips can be classified into two main categories viz. home-based and non-home-based trips. The home-based trips consist of two legs called departure and return legs, whereas non-home-based trips consist of only single legs i.e. departure leg. Furthermore, the data contains travel times by other non-chosen modes wherever applicable.

Table 1 – Attributes in the Trip data

Data Attribute	Description
'id'	Trip ID
'origin', 'destination'	Zone (origin/ destination)
'originX', 'originY' 'destinationX', 'destinationY'	Coordinates (origin/ destination) in reference system GK-4
'mode'	Chosen mode
'time_auto','time_bus', 'time_train', 'time_tram_metro',	Travel time by auto (Driver/ passenger), bus, train, tram/ metro (whichever applicable)
'purpose'	Purpose of the trip: Home-based work (HBW), home-based education (HBE), home-based shopping (HBS), home-based other (HBO), non-home-based work (NHBW), non-home-based other (NHBO)
'person'	Person id
'distance'	Distance from origin to destination with road network (Km)
'departure_time', 'departure_time_return'	Departure time for outbound and inbound/ return (if applicable) trip

3.5.2.2. Area of study

The operating area selected for the simulations is situated in the core of Munich city. A radius of 3 kilometers from the Munich Central Station is set and the data is filtered for the zones in within these boundaries, as shown in Figure 3 and Figure 4. The filtered data contains 349,585 trips for 427 zones.

The outbound trips (Internal- External) and inbound trips (External – Internal) from/to the study area are not considered in the analysis and subsequent modeling of simulation input data. The selection of the area in the city core, with a center in the Central Station and the exclusion of outbound and inbound trips is supported by the fact that such ride sharing services area likely to work best in dense city areas (Bischoff & Maciejewski, 2016).

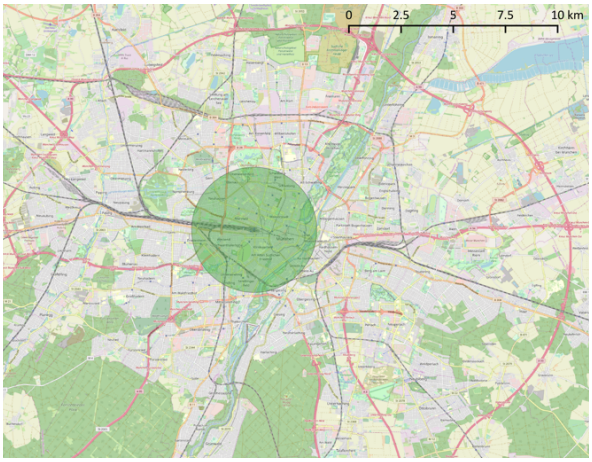

Figure 3- City of Munich with highlighted area of study

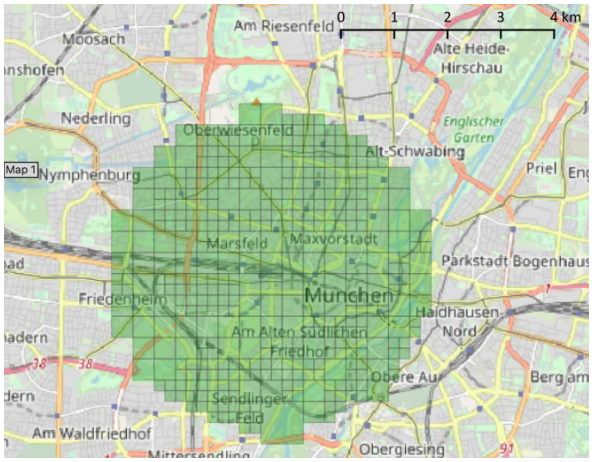

Figure 4 – Area of study divided in the original zones from the trip data.

3.5.2.3. Data exploration and adjustment

The mode-share of the filtered data to the specified area of study is shown in Figure 5, where it can be seen that private cars (Auto driver and Auto passenger) and public transport (train, tram, bus and metro) account respectively for 28% and 9% of the total trips while walk accounts for 36%. According to the available statistics, the mode share of cars and public transport for the city of Munich is 34 % and 24% respectively, and 24% for walking (Infrastruktur, 2019). Explanations for that are that are, firstly that the filtered area doesn't correspond exactly to the Munich city which has an area of 310.4 km², while the study area has a total area of 29.32 km²; secondly, this inconsistency could also be due to the calibration of the synthetic population data for the mode shares at the scale of Munich Metropolitan Area. And thirdly, the data was filtered to account only for internal trips, which excluded any trip from and to the study area. Therefore, the data is processed to account for the imbalance of the mode share and better represent the current mode share of Munich city.

To handle the mode imbalance an under-sampling technique (Chawla 2005) is used. This approach is applied when the majority class is overrepresented and it is used to balance it with the minority class, which in this case accounts for walk and the other modes. This can be done by randomly selecting a sample from the dataset, so that the mode share of the sampled trips corresponds to the statistical data by reducing the occurrences of the majority class.

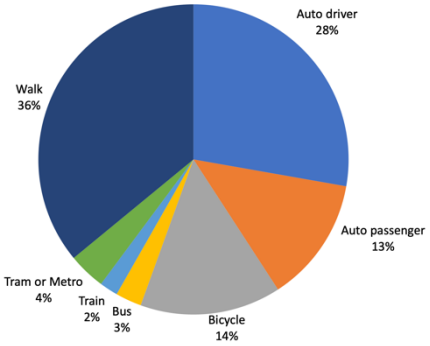

Figure 5 – Mode share in filtered synthetic population data

By using this technique, the total number of trips is reduced to 123,992, because its number is limited to the maximum value of the underrepresented class. Table 2 provides the number of trips per trips purpose, in which home-based trips accounts for 63.55 % of the total trips. Each trip has a number of legs depending on its purpose, in case of Home-based-work trips there are two legs, one from home to work and one from work to home, for the case of Non-home-based-trips there is only one leg. The legs have the same mode as the trip in which they are part. The final filtered data accounts for 202,787 legs.

The spatial distribution of the trips within the city of Munich and in the delimited area of study for the simulations is shown in Figure 6 and Figure 7. Trips are more concentrated near the city center, inside the selected area for simulation, where there is a high density of the residential, shopping and working places areas. As observed, origins and destinations are almost the same, the difference is due to the non-home-based trips, in which there is no returning leg, and accounts for 22% of the legs.

Table 2 – Trips by purpose

Trips purpose	Number of trips	%
Home-based-work (HBW)	6403	5.2
Home-based-education (HBE)	8328	6.7
Home-based-shopping (HBS)	23341	18.8
Home-based-other (HBO)	40723	32.8
Non-home-based-work (NHBW)	6737	5.4
Non-home-based-other (NHBO)	38460	31.0
Total	123992	100.0

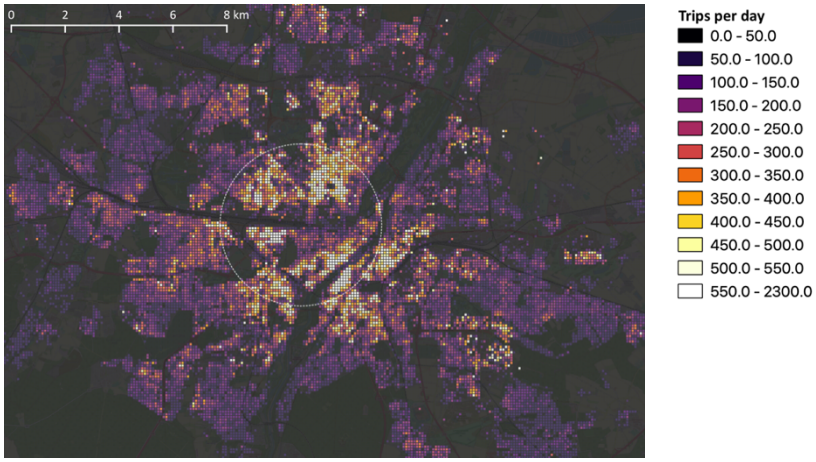

Figure 6 – Trip origins in 10,000 sqm area.

Figure 7 – Trip destinations in 10,000 sqm area.

The distribution of departure times for the trips is shown in Figure 8, where morning and evening peak hours can be clearly identified.

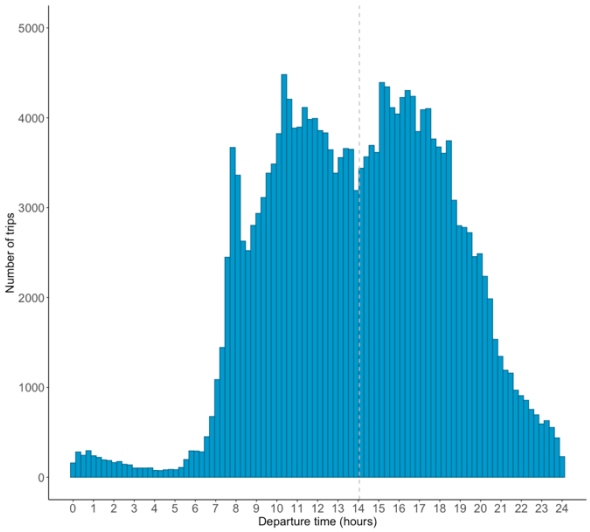

Figure 8 – Distribution of trip departure times within the study area. Dashed line represents mean value = 14.05.

The distribution of trip length has a mean and standard deviation of 2.7 km and 1.7 km respectively, as shown in Figure 9 , 75% of the trips have distances lower than 3.9 km and 90% of the trips cover a distance bellow 5.2 km, as shown in Table 3.

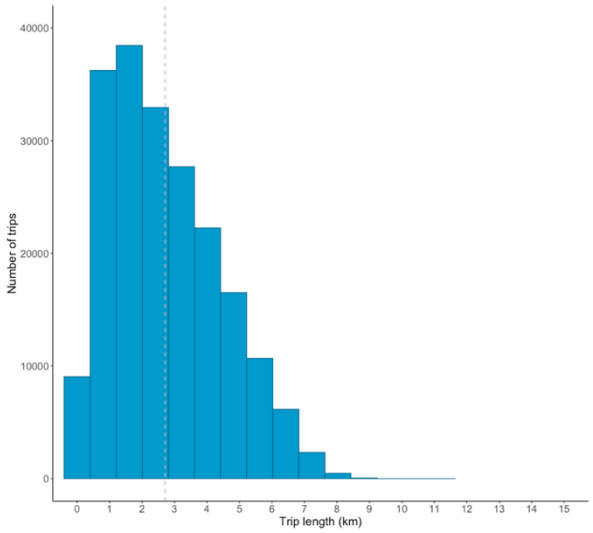

Figure 9 – Trip length distribution. Mean value = 2.71 km.

The characteristics and distribution of the trips per mode are described in Table 3. It is possible to see that regardless of mode, 90% of the trips have trips bellow 6 km.

Table 3 – Trip statistics per mode in km

	Car	Public Transport	Walk	Bicycle	All Modes
Mean	3.3	3.0	1.6	2.7	2.7
Median	3.1	2.5	1.3	2.4	2.4
Sd	1.7	1.7	1.2	1.6	1.7
Maximum	11.3	9.3	8.3	9.3	11.3
Minimum	0.0	0.0	0.0	0.0	0.0
75^{th} percentile	4.5	4.2	2.1	3.8	3.9
90^{th} percentile	5.6	5.4	3.2	5.0	5.2

The trip distribution per mode is depicted on Figure 10.

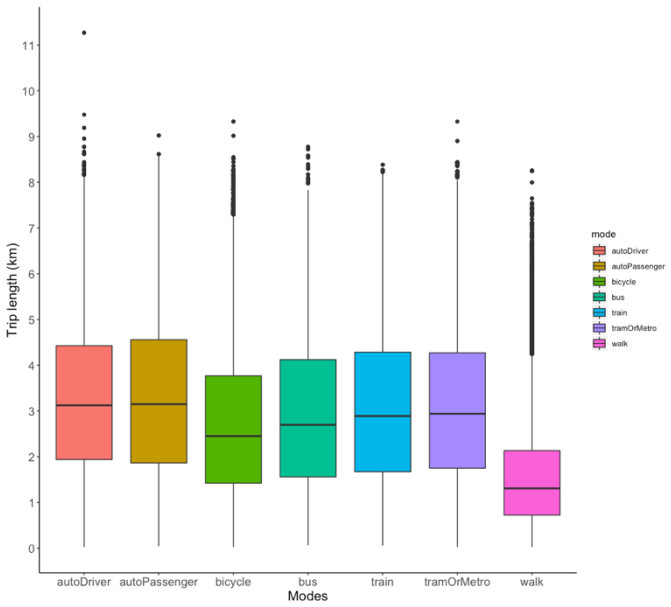

Figure 10 – Trip length distribution per mode.

A MATSim population input file on the XML format was created containing the filtered trips and all trip modes were converted to shared DRT.

3.5.3. Network

The road network used for the simulation is extracted using MATSim JOSM plugin (Kühnel & Zilske, 2019) from the Munich metropolitan network provided by the research group Modeling Spatial Mobility – TUM. The network file is generated in XML format and contains a total of 7104 links and 3343 nodes. A snapshot is shown on Figure 11

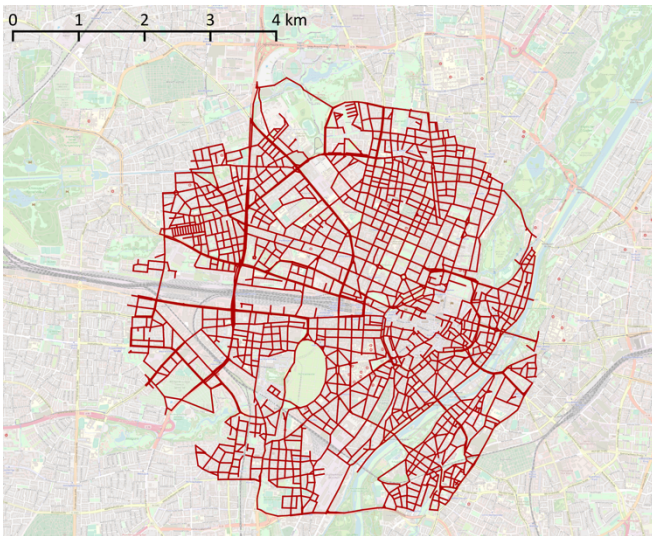

Figure 11 – Road network for the simulated area

3.5.4. Vehicles

The vehicles file used as input on the MATSim simulations were generated in XML format and contains each vehicle id, start link, initial operation time and end operation time (time window) and capacity (number of seats). Files with fleet sizes ranging from 50 to 4000

vehicles were created. Every fleet size was generated for three different vehicle capacities: 4, 6 and 8 seats. For every fleet, the starting links were randomly assigned to the vehicles. The operational time starts at the first simulation time step and the end time is after 30 hours. The selection of 30 hours is set to account for the trips that have drop-offs scheduled after the 24-hour time step.

3.5.5. Scenarios configuration

The scenario set up for the simulations is based on three main input variables: Demand, fleet size, and vehicle capacity. The simulation scenarios are designed in order to generate suitable data for modeling the variables. Different scenarios were designed with fleet sizes from 50 vehicles to 4000 vehicles. The different fleet sizes were tested together with defined percentage levels of the full demand for the selected area of study. All scenarios were tested for three different vehicle capacities: 4, 6 and 8 seats.

Regarding the operation without request rejections, for scenario cases with a high degree of undersupply (very small fleet size and very high demand), a fraction of trips would only end by the end of the next simulation day. Therefore, to rely on realistic results, for the different demand levels, the minimum suitable fleet size that would provide realistic outputs were selected. The criteria used for this selection is based on the simulation window time, and vehicle operational time, which are both set to 30h, the scenarios selected have trips starting and ending on this time window. Such behavior is not observed on constrained operations simulation because of the presence of rejections.

For each scenario defined, five different fleets sizes were tested together with the 3 different vehicle capacities. In summary, a total of 78 scenarios where simulated for the operations without request rejection and 90 scenarios for the constrained operation. Table 4 and Table 5 summarizes the scenarios built for the simulations.

Table 4 – Combinations of demand and fleet size for scenario 1: no rejections.

		Fleet Size (vehicles)									
		100	200	400	600	650	750	850	1000	2500	4000
Demand (%)	10	■	■		■				■	■	■
	20		■		■				■	■	■
	50			■	■				■	■	■
	80					■		■	■	■	■
	100						■	■	■	■	■

Table 5 – Combinations of demand and fleet size for scenario 2: with rejections.

		Fleet Size (vehicles)										
		50	100	200	400	600	650	750	850	1000	2500	4000
Demand (%)	10	■	■	■		■				■	■	■
	20			■		■				■	■	■
	50		■		■	■				■	■	■
	80		■				■		■	■	■	■
	100		■					■	■	■	■	■

As described previously on item 3.5.1, the DRT MATSim extension use constraints for the rejection of requests. The constraints are based on maximum permissible waiting time and maximum travel time for a direct ride, and are set by the two parameters α and β used in the following equation (Bischoff et al., 2017):

$$t_r = \alpha * t_r^{direct} + \beta$$

In which t_r is the maximum permissible travel time and t_r^{direct} is the direct ride time between origin and destination of the request.

The parameter values for the two operational scenarios studied are presented on Table 6.

Table 6 – Simulation constrains parameters

Scenario	Max Waiting Time (sec)	α	β
No-rejection scenario	1200	100.5	1200
Constrained scenario (with rejections)	600	1.5	240

Together with the parameters, the *requestRejection* option in the DRT module was also used. If set true, this option allows for the maximum travel and wait times of a submitted request to be considered hard constraints (the request gets rejected if one of the constraints is violated). If false, the max travel and wait times are considered soft constraints. The value is set false for the no-rejection scenario and true for the constrained scenario.

The parameter values selected for the no-rejection scenario allow for a no-rejection system, in which the high value of α allows for high trip detours when needed. Although this setup emulates the no-rejection scenario, due to the dispatching algorithm, the agents in the simulation who are already in the vehicle will not be delayed beyond their initial detour tolerance, defined by alpha and beta parameters.

The values for the parameters in the constrained scenario were selected based on the best results for the simulation without rejections and for the best results on waiting time from previous simulation studies using the same framework (Bischoff et al., 2017). The vehicle speed used in all simulations is set 30 km/h and is in accordance with values from literature (Bösch et al., 2018). The number of iterations was set for 400 to all scenarios.

Chapter 4

Simulation Results and Model Formulation

In this chapter the output data from the simulations performed is analyzed and linear regression models are developed for distance detour factor, average occupancy, average waiting time and for the case of constrained operations, rejection rate. The average distance detour factor is calculated using the output average trip distance. It is the ratio between the average trip distance actually performed by the average theoretical direct trip distance. Average occupancy is calculated using by the ratio between the total distance travelled by the passengers and the total distance travelled by the vehicles.

4.1. Scenario 1 – Operation with no request rejections

A total of 6 scenarios did not achieve convergence for the number of iterations set on MATSim simulations and were removed from the output data analysis. The dataset analyzed contains 72 data points. The following table and Figure 12 provide a summary of the dataset for scenario 1.

Table 7 – Summary of the Dataset

Variable	Min	Max	Mean	Median	Std	Percentiles			
						25%	50%	75%	90%
Distance detour factor	1.89	3.79	2.68	2.68	0.53	2.10	2.68	3.14	3.36
Avg waiting time – (sec)	561.8	1084.2	899.0	931.2	136.36	803.5	931.2	1000.5	1067.9
Vehicle occupancy	2.58	6.90	4.61	4.70	1.35	3.28	4.70	5.60	6.80

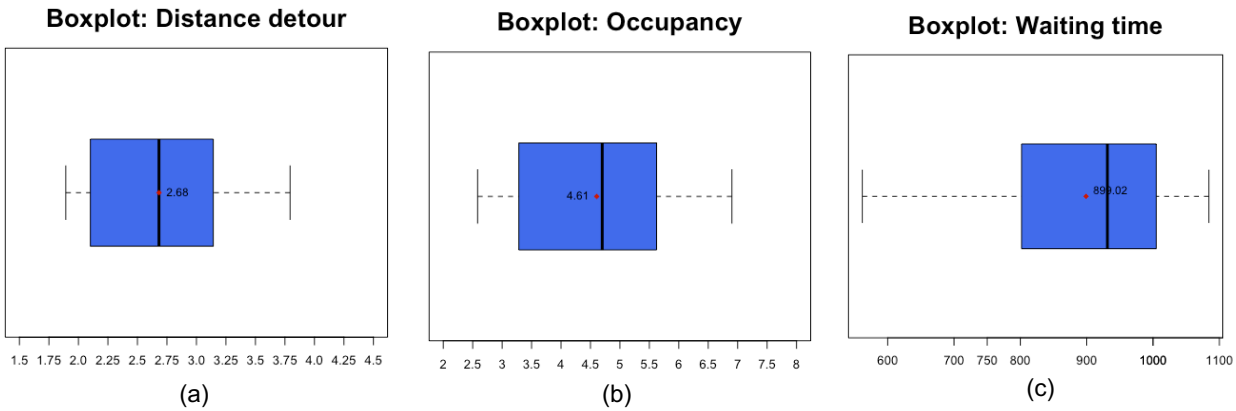

Figure 12 – Distance detour factor (a), Average occupancy (b) and Average waiting time (c) boxplots. Scenario 1.

The values obtained for the outputs are well distributed around mean and median values. The fact that there are no negative outliers and no zero values is consistent with the expected simulation outputs.

4.1.1.Detour Factor

4.1.1.1. Correlation and relationship between variables

Following is a correlation table between system input variables and detour factor.

Table 8 – Correlation between input variables and distance detour factor

Variable	Correlation
Demand	0.01
Fleet size	-0.25
Vehicle capacity	0.96

According to Table 8, the variable that has a higher linear correlation with the distance detour factor is vehicle capacity, with a correlation of 0.96. The high correlation can be explained by the fact that the system simulated is set for zero rejection rate, and based on the dispatching algorithm described on item 3.5.1, an increase in vehicle capacity will necessarily increase the probability of vehicle taking a detour for a new pick-up or drop-off, for the same demand level and fleet size, as long as there is an empty seat. There is negative relationship with fleet size, meaning that an increase in fleet size is related to a decrease in the detour factor, as theoretically expected on equation (14). Demand has shown a low positive correlation value with detour factor, as expected on equation (12). Although a very low correlation coefficient might suggest small influence of a variable in the detour factor, the relationship between these variables can also be not linear and must be studied in depth. Figure 13 contains the scatter plots of detour factor and the independent variables.

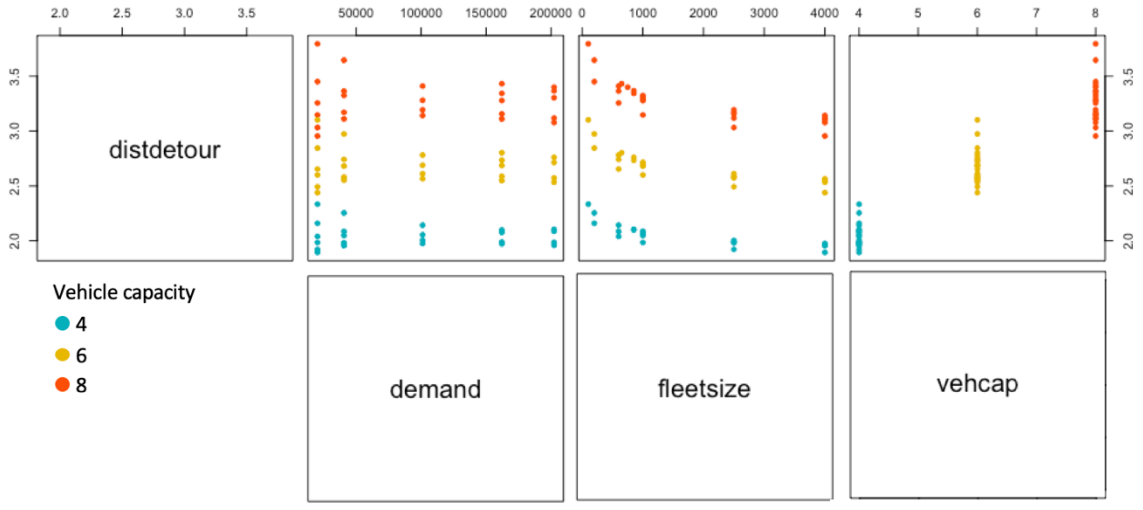

Figure 13 – Scatter plots of Detour factor against independent variables

It is possible to see that smaller fleet sizes accounts for higher distance detours and as the fleet size increases the detour decreases, for the same vehicle capacity. The relationship suggested is not linear, with a fleet size saturation effect: As the fleet size

increases and the system moves from a low supply level, passing through an equilibrium between supply and demand and reaching an oversupply configuration, the impact degree of an increase in fleet size in the detour decreases, that is, the slope in the curve decreases. This expected behavior, together with the data suggests an asymptotic function for fleet size in the modeling of detour factor.

Figure 14 depict the relationship between detour factor and vehicle capacity and detour factor and fleet size. As seen before, the capacity of the vehicles suggests a strong linear relationship for the range of the data analyzed. The change in detour together with changes in capacity and fleet size is depicted on Figure 15.

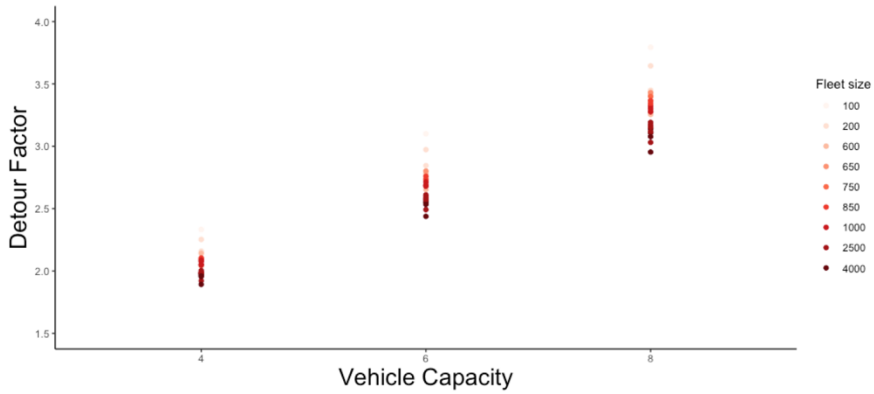

Figure 14 – Detour Factor according vehicle capacity.

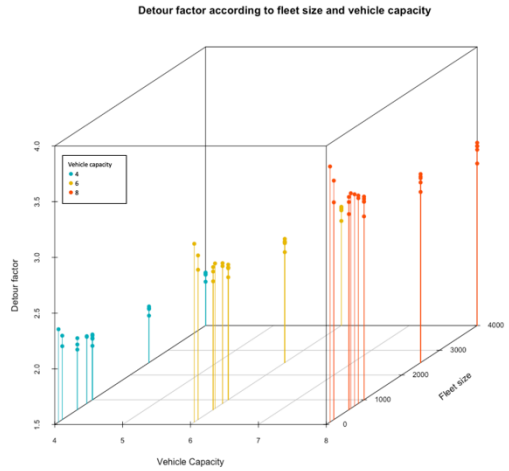

Figure 15 - 3D plot of Detour factor, vehicle capacity and Fleet Size

The correlation between demand and distance detour factor is small for scenario 1 and Figure 16 and Figure 17 depicts how detour changes with demand. For the cases when demand is low, there is a positive relationship, as demand increases its influence on detour diminishes. The positive relationship at low demand levels is explained by the fact that there is no request rejection in the system, and it is not saturated yet: for each new request there will be an increase in the detour factor in order to fill the empty seats available. After enough demand, its impact on detour is almost none.

Even though there is a theoretical relationship between the variables, the overall impact of demand in the detour factor is very low for the data range analyzed.

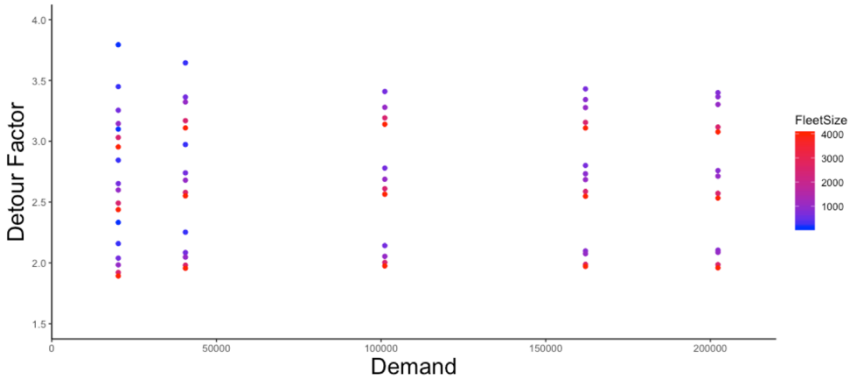

Figure 16 - Detour Factor according number of trips (A) colored by fleet size (B) colored by vehicle capacity.

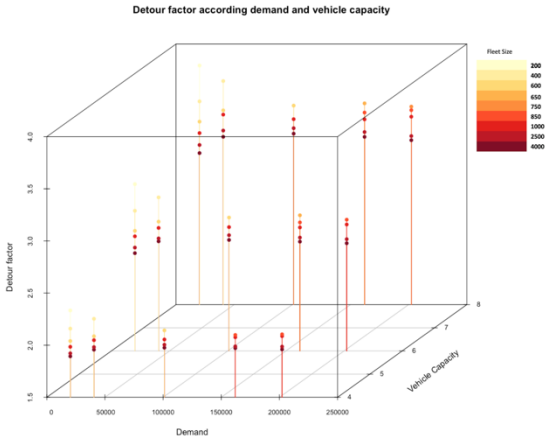

Figure 17 – 3D plot with Detour Factor, demand and vehicle capacity, colored by fleet size.

4.1.1.2. Multiple regression models

Given the non-linear relationships observed between distance detour factor with fleet size and demand, different types of relationship were analyzed and used as regressors for the detour function.

The following Figure 18, Figure 19 and Figure 20 show the correlation of distance detour factor with transformed variables, this exploratory measure helps to identify better the type of non-linear relationship that accounts for the curvilinear existing relationship and to eliminate the nonlinearity.

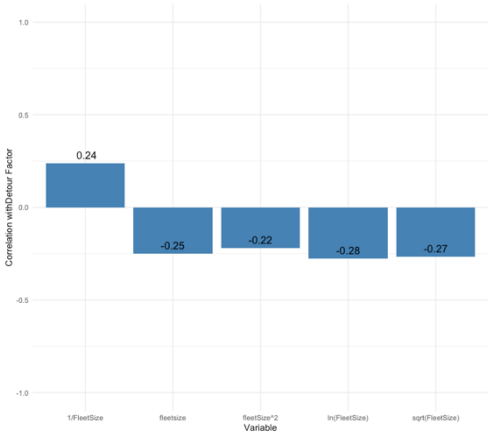

Figure 18 – Correlation between detour factor and non-linear functions of fleet size

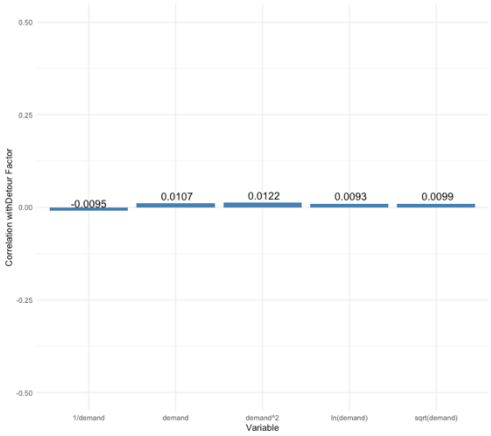

Figure 19 – Correlation between detour factor and non-linear functions of demand

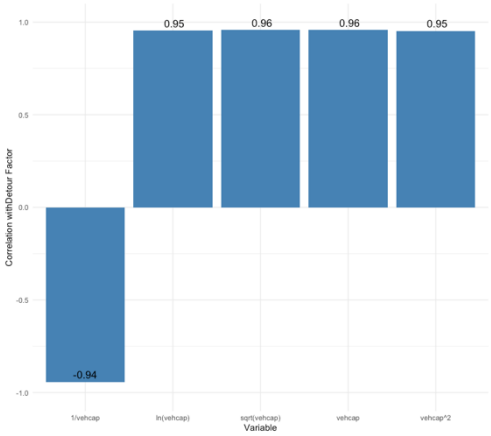

Figure 20 – Correlation between detour factor and non-linear functions of vehicle capacity

The correlations indicate the strong linearity between vehicle capacity and distance detour factor. For fleet size, a slight increase for logarithmic and square root functions. For demand, the reciprocal and log functions appear to improve the correlation, however the value is still very low.

Based on the previous exploratory data analysis and theoretical explanations, multivariate regression models were built and are summarized in Table 9. The models presented account for the best three model fittings taking into account demand and the three best fittings without using demand as an independent variable. For modeling demand the potential demand was divided by the hours in the day, therefore potential demand used for the models are in rides/h.

Table 9 – Best fitted models Detour Factor regression – scenario 1.

Number	**Model Form**	R^2	Adjusted R^2	Residual SE	F-statistic (p-values on joint Hypotheses)	Lower Coef. Signif.
1	$\beta_0 + \beta_1 * C - \beta_2 * \ln B - \beta_3 * \frac{1}{Q}$	0.989	0.989	0.055	< 2.2e-16	***
2	$\beta_0 + \beta_1 * C - \beta_2 * \ln B + \beta_3 * Q - \beta_4 * Q^2$	0.991	0.989	0.058	< 2.2e-16	**
3	$\beta_0 + \beta_1 * C - \beta_2 * \ln B + \beta_3 * \ln Q$	0.988	0.988	0.058	< 2.2e-16	***
4	$\beta_0 + \beta_1 * C - \beta_{2*} \ln B$	0.985	0.984	0.066	< 2.2e-16	***
5	$\beta_0 + \beta_1 * C - \beta_2 * B + \beta_3 * B^2$	0.979	0.978	0.078	< 2.2e-16	***
6	$\beta_0 + \beta_1 * C - \beta_{2*} \sqrt{B}$	0.976	0.975	0.082	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

All models provide a good fit with satisfactory R^2 and statistical significance for the coefficients, with model 1 being able to explain 99% of data variance.

The normality of residuals assumption is analyzed graphically and shown in Figure 21 and by applying statistical tests, with results shown in Table 10. The graphical analysis provides satisfactory results for all models. According to the Kolmogorov-Smirnov test, the null hypothesis of normality cannot be rejected in any model, however for Shapiro-Wil this only happens for model 4. One reason for that are the values for simulated scenario number 17 and 18, in which fall at a higher distance from the Q-Q plot curve and confidence level. The sample size might cause discontinuity in the data and cause this.

Table 10 – Tests for OLS regression residuals normality

Number	Kolmogorov-Smirnov test (p-value)	Shapiro-Wilk normality test (p-value)
1	0.7719	0.0015
2	0.8466	0.0185
3	0.8512	0.0297
4	0.2827	0.1102
5	0.2593	0.0085
6	0.3594	0.0039

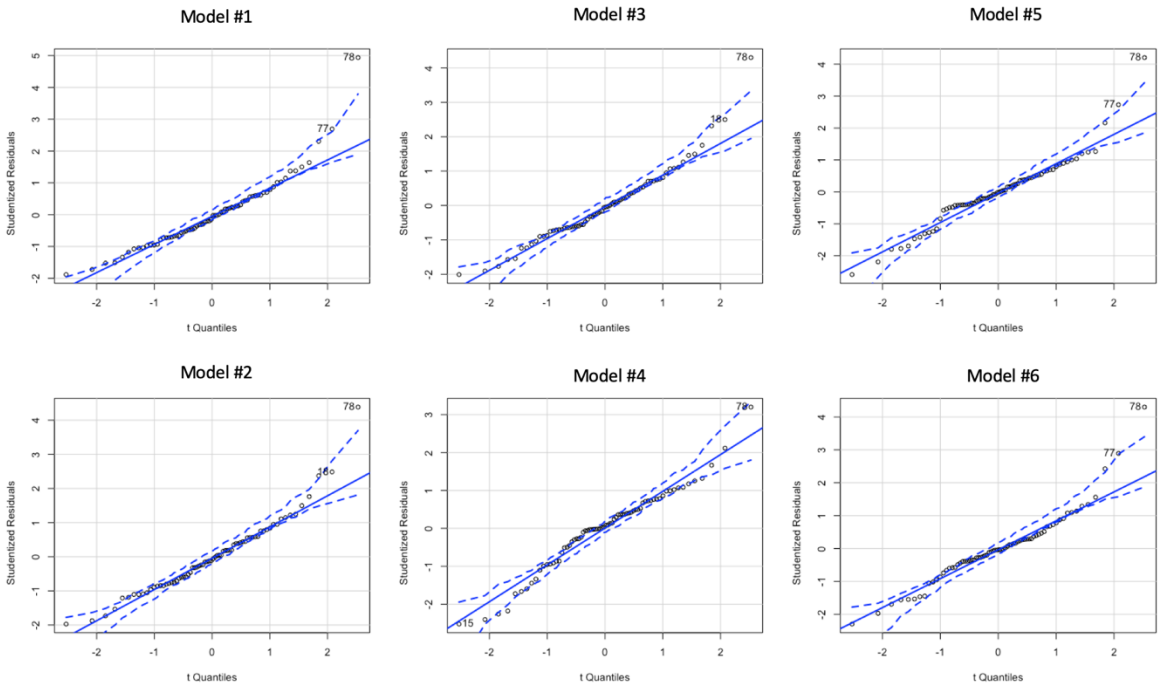

Figure 21 – Normal Q-Q plots for residuals of regression models for detour factor including 0.95 confidence level– scenario 1.

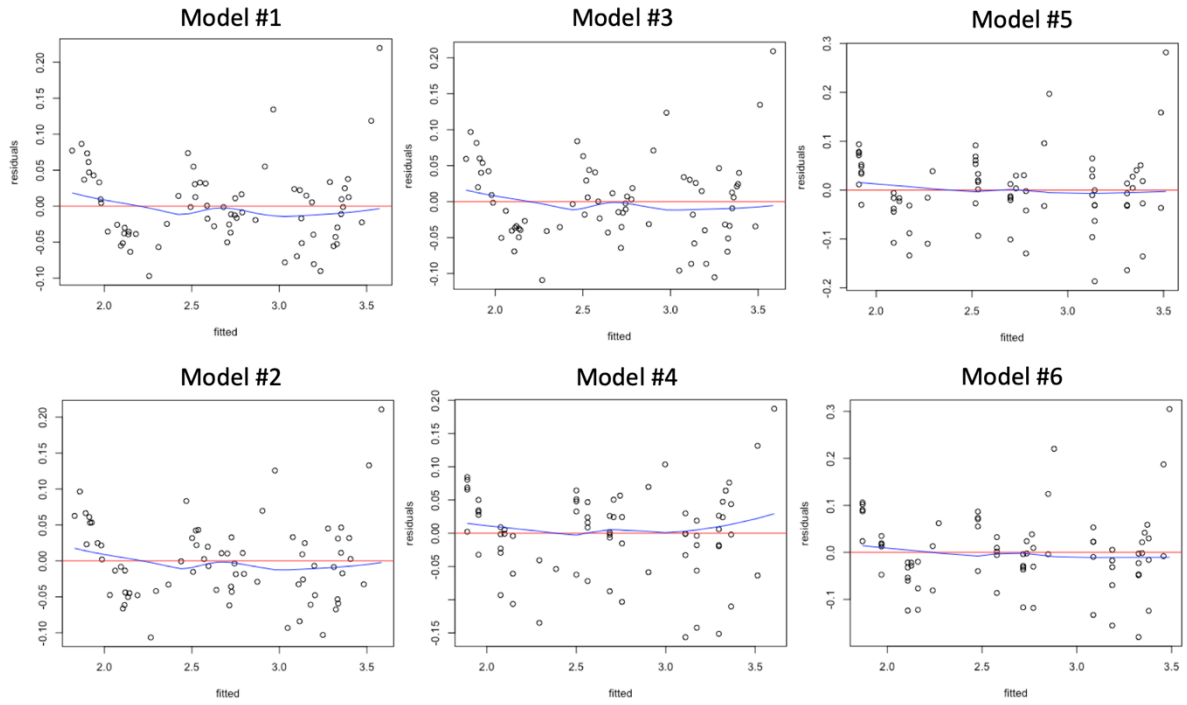

Figure 22 – Residual plots against fitted values for regression models of detour factor – scenario 1.

Residuals variance independence is also analyzed graphically in Figure 22. All models show a good distribution of residual variance. To analyze better the variance of residuals, the residuals of each model are ordered according to their values and divided in sets of 3 groups, with approximately equal number of elements, the conditional variance of each group is

calculated and the ratio to the largest to the smallest variance of groups is analyzed (Cohen et al., 2003), results are in the following table.

Table 11 – Ratio between maximum and minimum conditional variance of residual groups for detour factor regression – scenario 1.

Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
2.63	2.48	2.88	13.26	11.91	11.51

Model 4, 5 and 6 have the highest ratio values; all the other models have values lower than 3, showing a good level of homoscedasticity.

Given the results presented, model number 1 is chosen, equation model parameters are shown in the following table.

Table 12 - Equation model for detour factor – scenario 1

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	1.925753	0.057673	33.391	< 2e-16 ***
C	0.304043	0.003986	76.272	< 2e-16 ***
$\ln B$	-0.146934	0.006781	-21.667	< 2e-16 ***
$\frac{1}{Q}$	-90.450886	16.132082	-5.607	4.08e-07***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.05516 on 68 degrees of freedom
Multiple R-squared: 0.9895, Adjusted R-squared: 0.989
F-statistic: 2129 on 3 and 68 DF, p-value: < 2.2e-16

The model selected is able to take into account the non-linear relationship between the detour factor and fleet size. It worth to mention that based on the demand coefficient, it will be responsible for low increase on detour values. It is also worth to mention that the linear model with the form $\beta_0 + \beta_1 * C - \beta_2 * B$ provides also good results, being able to explain 97% of data variance.

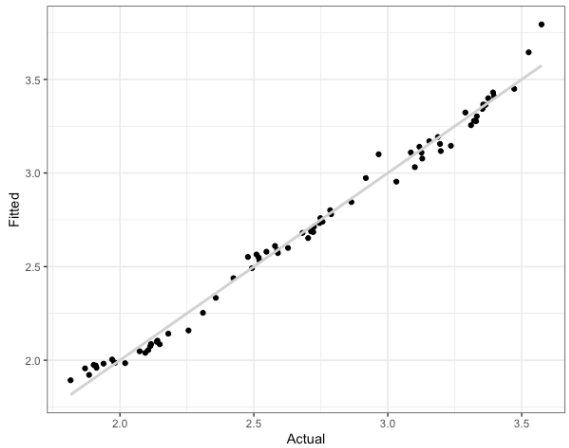

Figure 23 – Actual vs Fitted values for the regression model selected for detour factor.

Some of the key findings are:

- The distance detour factor for this scenario is mainly influenced by the vehicle capacity, the linear relationship between these variables appear to be strong and is able to explain more than 92% of the sample variance.
- The log function for the fleet size predictor is able to explain the diminishing rate of influence in which the fleet size influence the detour factor, although the function ln(x) does not have the asymptotic property that the reciprocal function 1/x has, it explains better the variance in the data, one reason might be that the asymptotic behavior would only be observed for very large fleet sizes, which might not be realistic for a real scenario. The negative reciprocal function of demand is able to explain the saturation effect observed. In general, the analysis presented is satisfactory.

4.1.2. Average Occupancy

4.1.2.1. Correlation and relationship between variables

Following is a correlation table for the input variables and occupancy.

Table 13 – Correlation between input variables and occupancy

Variable	Correlation
Demand	0.35
Fleet size	- 0.07
Vehicle capacity	0.93

According to Table 13, the variable that has a higher linear correlation with the occupancy is vehicle capacity, with a correlation of 0.92. The explanation is the same as for the detour factor, the fact that the system simulated is set for zero rejection rate, an increase in vehicle capacity will necessarily increase occupancy. There is positive relationship with demand, meaning that an increase in demand is related to an increase in the occupancy, what is also expected: given the same area and network, if fleet size and number of seats are the same, the increase in demand will increase the probability of shared rides. Fleet size has shown a very low negative correlation with detour factor. Although a very low correlation coefficient might suggest no influence of the variable in the occupancy, the relationship between these variables can also not be linear and must be studied in depth. Figure 24 contains the scatter plots of detour factor and the independent variables.

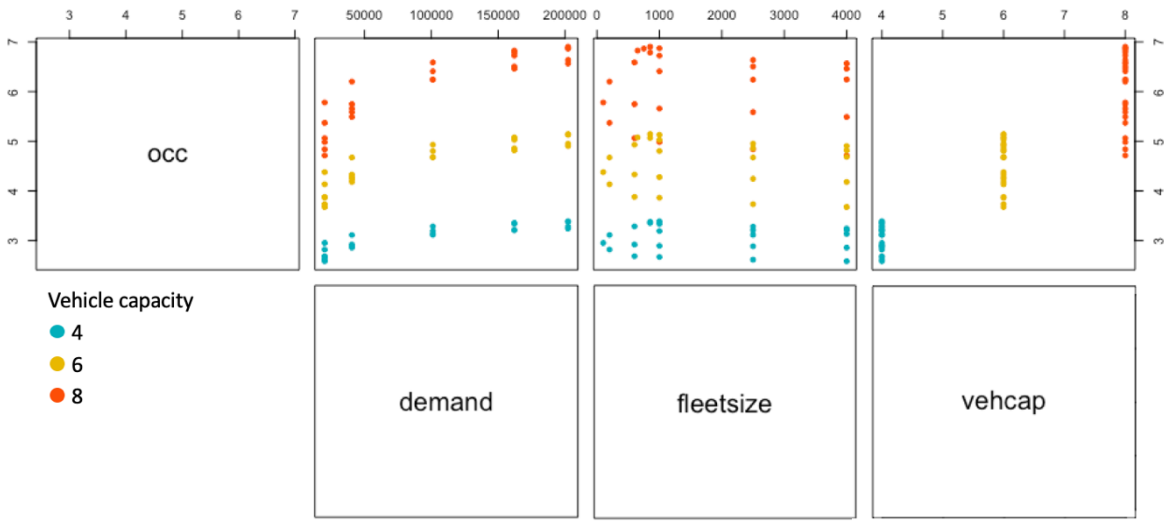

Figure 24 – Occupancy x independent variables

Figure 25, Figure 26 depict the relationship between occupancy and vehicle capacity and fleet size. Vehicle capacity suggests a strong linear relationship, it is also possible to note that the curve inclination seems to increase as demand increases, meaning that as higher the demand, higher the influence of vehicle capacity on the average occupancy. For the same demand levels and small fleet sizes, the increasing in the number of vehicles decreases occupancy, and as fleet size gets larger the rate of change decreases and the curve gets flatter.

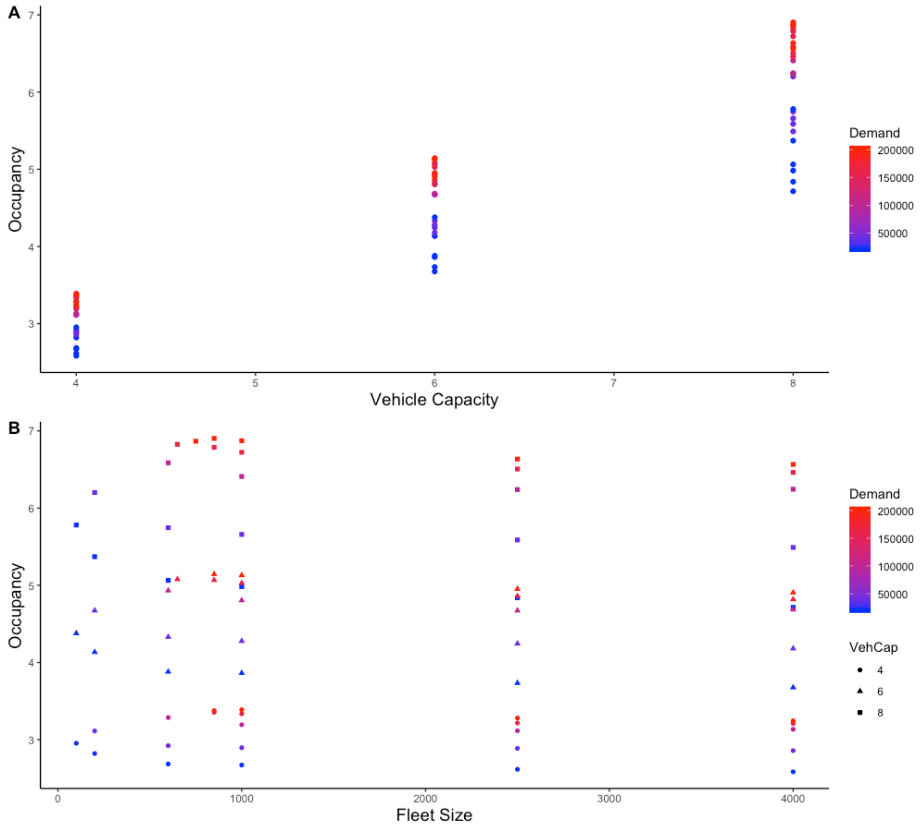

Figure 25 - Occupancy according vehicle capacity (A) and Occupancy according fleet size (B).

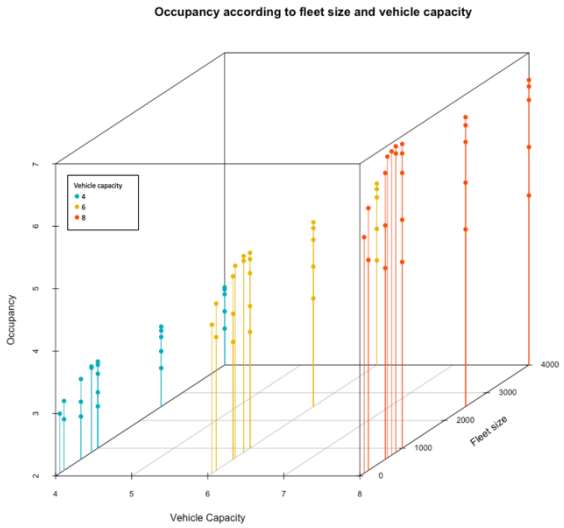

Figure 26 - 3D plot of Occupancy, vehicle capacity and Fleet Size

Regarding demand, a saturation effect is observed, the increase in demand increases occupancy in a higher rate for low levels of demand and as it increases, the degree of influence diminishes, depicting a non-linear relationship. Theoretically, this curve has to be asymptotic to the vehicle capacity of the fleet (occupancy cannot be higher than the number of vehicle seats available).

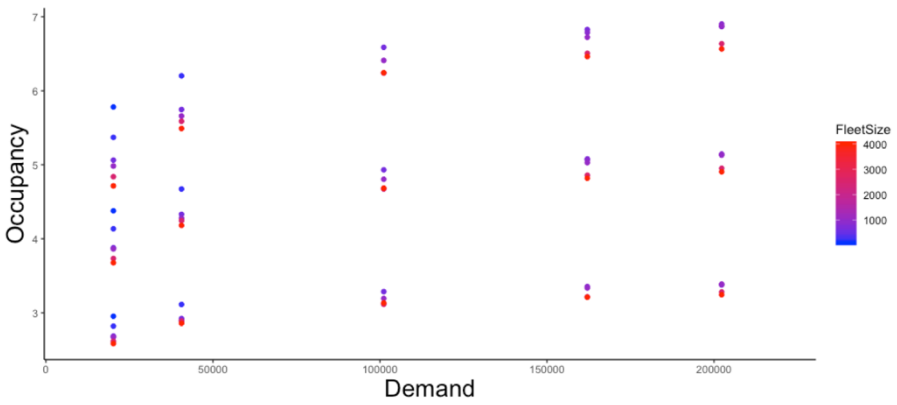

Figure 27 - Occupancy according number of trips (A) colored by fleet size (B) colored by vehicle capacity.

4.1.2.2. Multiple Regression Analysis

Given the non-linear relationships observed between occupancy with fleet size and demand, different types of relationship were analyzed and used as regressors for the average occupancy function.

The following Figure 28, Figure 29 and Figure 30 show the correlation of occupancy with transformed variables, this exploratory measure helps to identify better the type of non-linear relationship that accounts for the curvilinear existing relationship and to eliminate the nonlinearity.

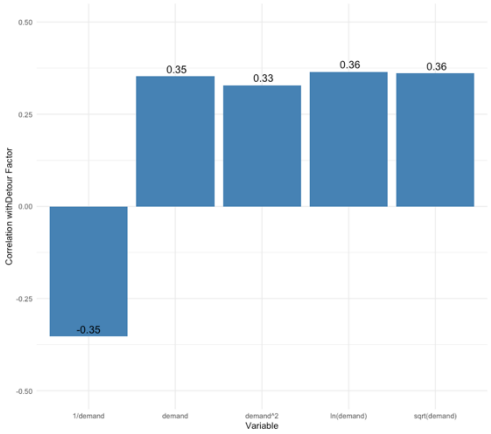

Figure 28 - Correlation between occupancy and non-linear functions of demand – occupancy regression analysis scenario 1.

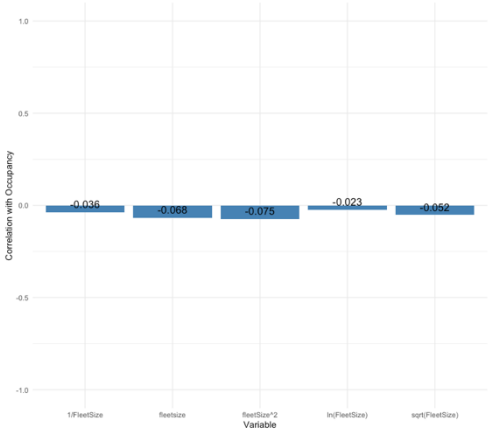

Figure 29 - Correlation between occupancy and non-linear functions of fleet size occupancy - regression analysis scenario 1.

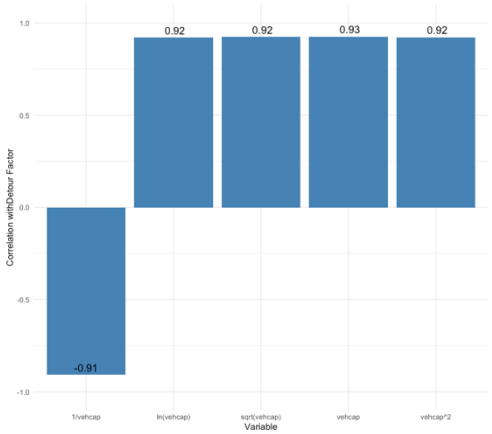

Figure 30 - Correlation between occupancy and non-linear functions of vehicle capacity - occupancy regression analysis scenario 1.

The correlations indicate the strong linearity between vehicle capacity and occupancy. For fleet size there was a change in the values, but the correlation still very low. For demand, the reciprocal and log functions appear to improve the correlation.

Based on the previous exploratory data analysis and theoretical explanations, multivariate regression models were performed and are summarized on Table 14. Because of the relationship observed between capacity and demand, the best two models that account for the interaction are presented, the other models are the best that do not include the interaction term. For modeling demand the potential demand was divided by the hours in the day, therefore potential demand used for the models are in rides/h.

Table 14 – Best fitted models average occupancy regression – scenario 1.

Number	**Model Form**	R^2	Adjusted R^2	Residual SE	F-statistic(p-values on joint Hypotheses)	Lower Coef. Signif.
1	$\beta_0 - \beta_1 * C - \beta_2 * \ln B - \beta_3 * \ln Q + \beta_4 * C * \ln Q$	0.995	0.995	0.093	< 2.2e-16	*
2	$\beta_0 + \beta_1 * C - \beta_2 * \ln B + \beta_3 * Q - \beta_4 * Q^2 + \beta_5 * C * Q$	0.992	0.991	0.124	< 2.2e-16	***
3	$-\beta_0 + \beta_1 * C - \beta_2 * \ln B + \beta_3 * \ln Q$	0.981	0.980	0.190	< 2.2e-16	***
4	$-\beta_0 + \beta_1 * C - \beta_2 * B + \beta_3 * B^2 + \beta_4 * \ln Q$	0.98	0.978	0.198	< 2.2e-16	***
5	$-\beta_0 + \beta_1 * C + \beta_{2*} \ln Q$	0.97	0.966	0.249	< 2.2e-16	***
6	$-\beta_0 + \beta_1 * C + \beta_{2*} Q - \beta_3 * Q^2$	0.965	0.964	0.256	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

All models provide a good fit with satisfactory R^2 and statistical significance for the coefficients, with model 1 being able to explain 99% of data variance.

The normality of residuals assumption is analyzed graphically and shown in Figure 31 and by applying statistical tests for the null hypothesis of normality, with results shown in Table 15. Based on the statistical tests performed, the null hypothesis of normality distribution of residuals cannot be rejected at 0.05 significance level for all models, with exception for model 1 and shapiro test. The graphical analysis provides further insights for all models. Overall all models have a good fit with normal distribution.

Table 15 – Tests for OLS regression residuals normality – occupancy (scenario1)

Number	Kolmogorov-Smirnov test (p-value)	Shapiro-Wilk normality test (p-value)
1	0.1399	0.0253
2	0.7787	0.2527
3	0.4425	0.0593
4	0.7331	0.0700
5	0.9817	0.3821
6	0.9669	0.2088

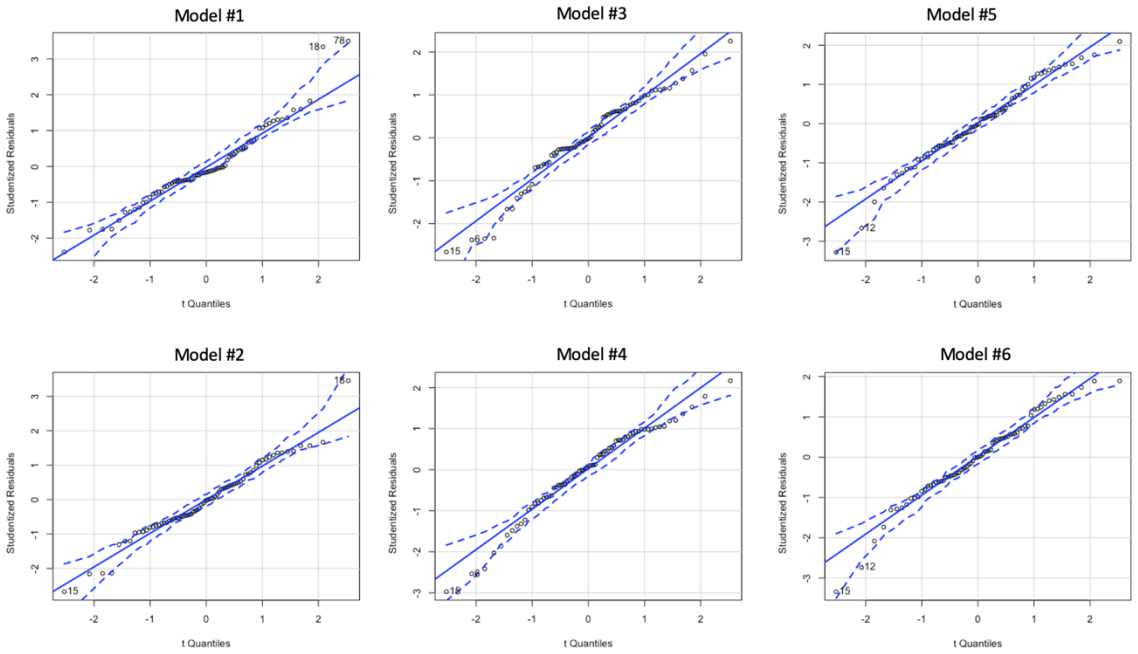

Figure 31 – Normal Q-Q plots for residuals of regression models for average occupancy including 0.95 confidence level– scenario 1.

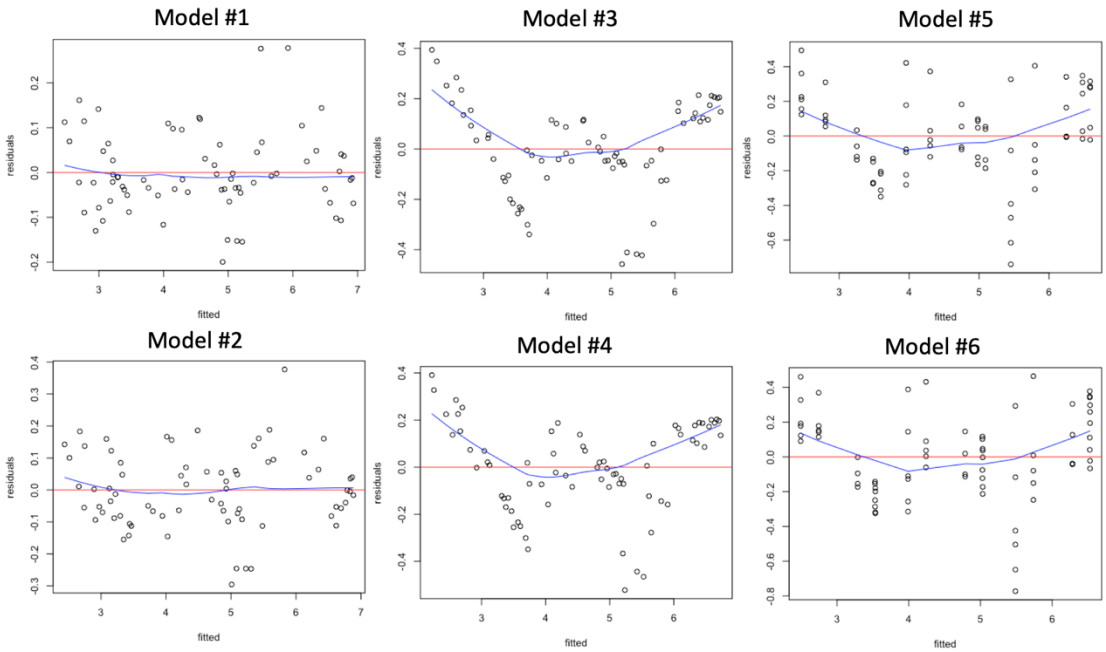

Figure 32 – Residual plots against fitted values for regression models of occupancy– scenario 1.

Residuals variance independence is also analyzed graphically in Figure 32. The models show a good a similar distribution, with residuals variance being more homogenous in values around the mean. The conditional variance of each group is calculated and the ratio to the largest to the smallest variance of groups is analyzed (Cohen et al., 2003), results are in Table 16.

Table 16 – Ratio between maximum and minimum conditional variance of residual groups for average .occupancy regression – scenario 1.

Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
14.42	4.83	10.41	10.02	8.12	7.94

All models have ratios bellow 10, with only model 1 having a ratio of 14.42.

Therefore, given the results presented, model number 2 is chosen, equation model parameters are shown in the following table.

Table 17 - Equation model for occupancy – scenario 1

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	1.016	0.1385	7.337	4.06e-10 ***
C	0.6155	0.01534	40.137	< 2e-16 ***
$\ln B$	-0.1578	0.01519	-10.389	1.59e-15 ***
Q	2.200e-04	3.126e-05	7.037	1.39e-09 ***
Q^2	-2.831e-08	2.752e-09	-10.285	2.41e-15 ***
Q*C	3.180e-05	3.038e-06	10.466	1.18e-15 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.1238 on 66 degrees of freedom
Multiple R-squared: 0.9921 Adjusted R-squared: 0.9915
F-statistic: 1665 on 5 and 66 DF, p-value: < 2.2e-16

The model selected is able to take into account the linear relationship between occupancy and vehicle capacity and the non-linear relationship with demand and fleet size, including the increase in occupancy for lower fleet sizes. The interaction term between demand and capacity is also able to explain the behavior observed with satisfactory statistical significance. The logarithmic functions are able to accrue for the positive saturation effect of demand.

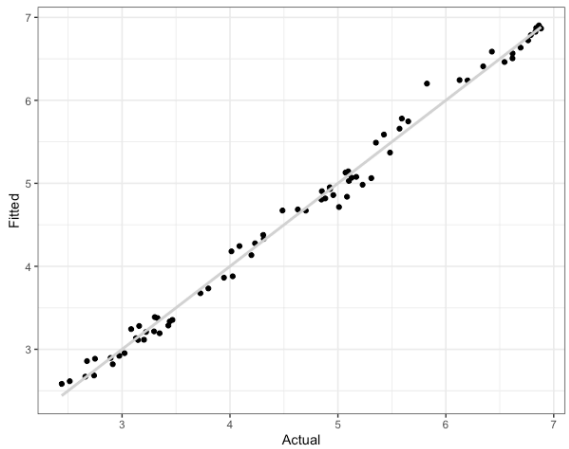

Figure 33 – Actual vs Fitted values for the regression model of average occupancy. Scenario 1.

Some of the key findings are:

- As the detour, the average occupancy for this scenario is mainly influenced by the vehicle capacity, the linear relationship between these variables appear to be strong and the vehicle capacity itself is able to explain 86% of the sample variance.
- For higher demand levels the vehicle capacity tends to have a stronger impact on occupancy.

4.1.3. Average Waiting Time

4.1.3.1. Correlation and relationship between variables

Following is a correlation table for the input variables and waiting time.

Table 18 – Correlation between input variables and waiting time

Variable	Correlation
Demand	0.786
Fleet size	-0.332
Vehicle capacity	0.190

According to Table 18 the variable that has a higher linear correlation with the waiting time is the demand, with a correlation of 0.79. This relationship is expected for a system without rejections. An increase in the demand increases the probability of a new trip request after the current passenger request, this leads to an increase in the route length to the pick-up point and in the number of stops in between, increasing the average waiting time.

There is negative relationship with fleet size, the explanation is that an increase in the fleet size increase the probability of having a vehicle available in the service area, and since there are no rejections, it leads to a decrease in the expected distance between the vehicle at the time of request and the pick-up point. Correlation with vehicle capacity is positive but rather low. Figure 34 contains the scatter plots of waiting time against the independent variables.

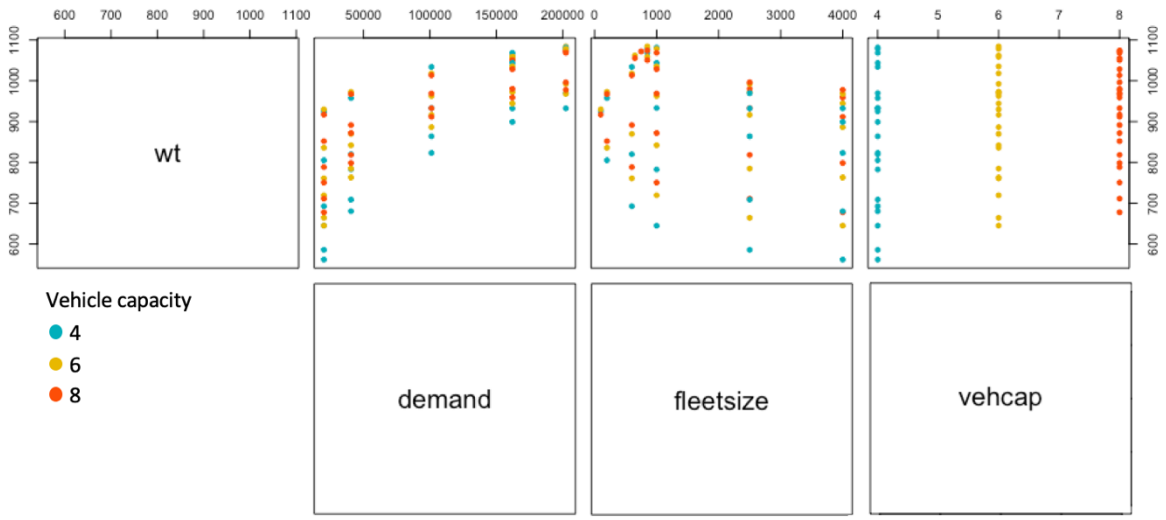

Figure 34 – Waiting time x independent variables

Error! Reference source not found. and **Error! Reference source not found.** show the relationship between waiting time and vehicle capacity and fleet size. Waiting time seems to only change with increases with vehicle capacity for low levels of demand, as demand increases, vehicle capacity has almost no influence on the waiting time. This behavior is depicted on Figure 35. The reason for this lies on the routing algorithm described on 3.5.1. and the fact that there are no rejections. When the demand is very low compared to the fleet size, the drop-off location of an in-route vehicle that has already a good occupancy can be close enough to a new request that the increase on its workload is lower than the waiting time necessary for a closer vehicle to satisfy this request. By increasing vehicle capacity, the probability of having such behavior increases. In fact, Bischoff et al. (2017) explains that the use of insertion without re-ordering of stops or moving requests between vehicles in the heuristic algorithm is justified by the waiting and travel times constraints. Since for this scenario the constraints are relaxed enough to not allow for no rejections, it explains the data behavior. This depicts the importance of the dispatching and routing algorithms on the performance of shared DRT systems.

For the fleet size relationship with waiting time, data suggests a non-linear relationship, in which at low supply levels of fleet, an increase in number of vehicles has a higher decrease impact on waiting time, as the fleets gets larger and the demand is satisfied by the supply, the rate of decrease in waiting time due to the increase in fleet size diminishes.

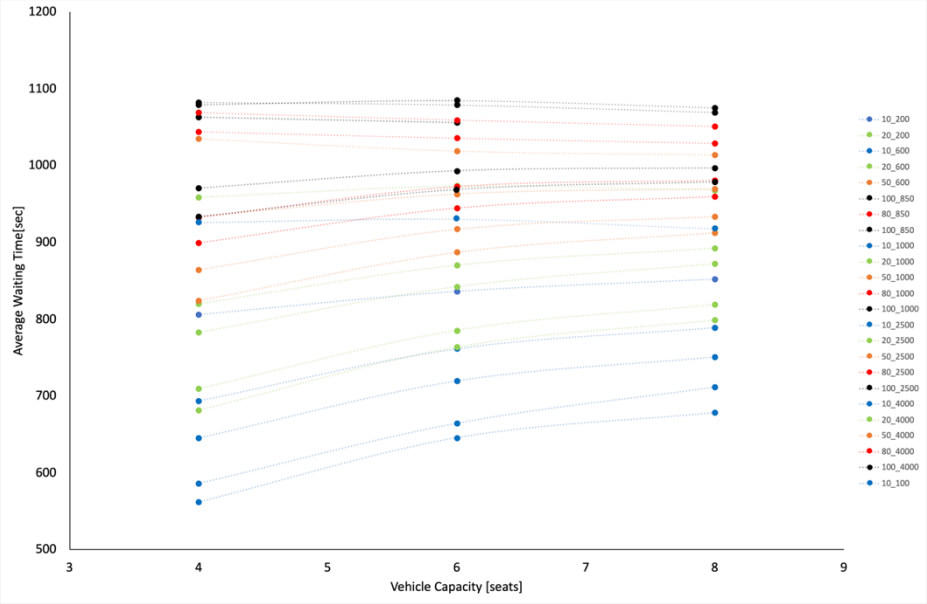

Figure 35 – Waiting time and vehicle capacity. The impact of vehicle capacity in increasing the waiting time is higher for low values of demand, as it increases the vehicle capacity has a lower influence on waiting time.

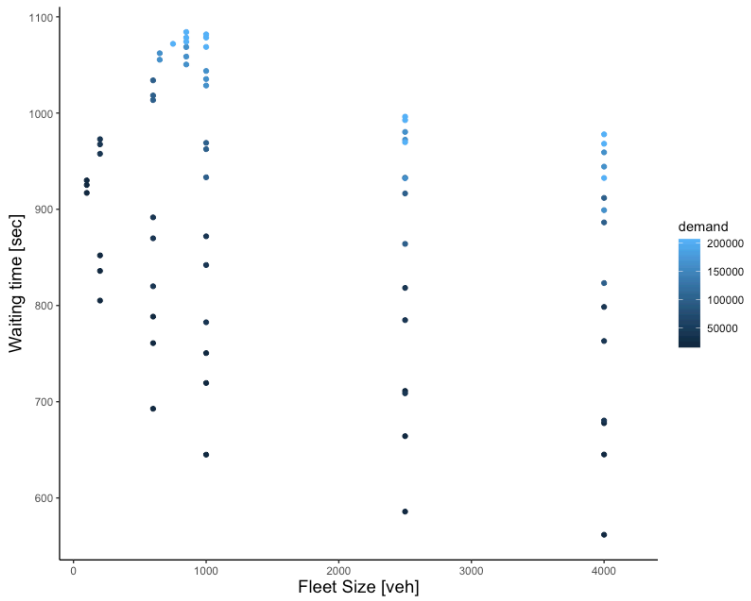

Figure 36 - Waiting time according to fleet size

The demand has the highest correlation with waiting time, for the same fleet size and vehicle capacity, an increase in demand leads to an increase in waiting time, as stated in equation (10). The data suggests a non-linear relationship, in which as demand increases, its influence on waiting time decreases.

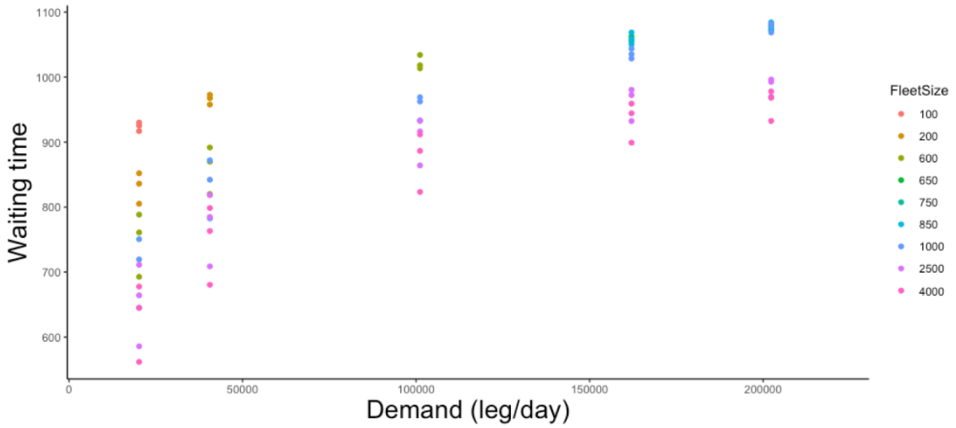

Figure 37 – Waiting time according number of trips (A) colored by fleet size (B) colored by vehicle capacity.

4.1.3.2. Multiple Regression models

Given the non-linear behavior observed for the waiting time with different scenario configurations, different types of non-linear functions were analyzed and used as regressors for the detour function.

The following Figure 38, Figure 39 and Figure 40 show the correlation of waiting time with transformed variables, this exploratory measure helps to identify better the type of non-linear relationship that accounts for the curvilinear existing relationship and to eliminate the nonlinearity.

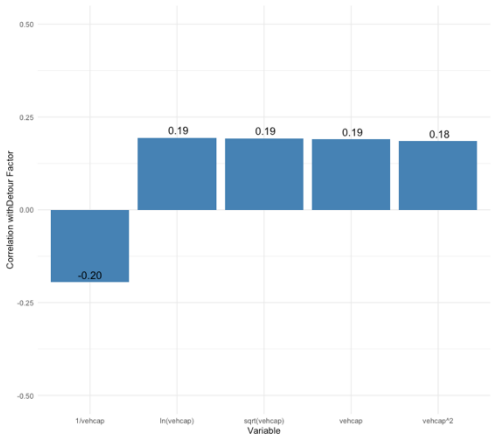

Figure 38 - Correlation between waiting time and non-linear functions of vehicle capacity

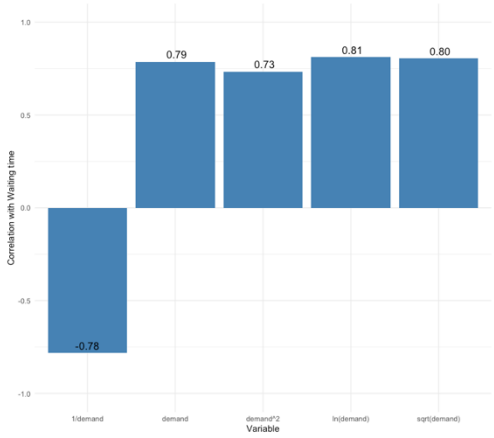

Figure 39 - Correlation between waiting time and non-linear functions of demand

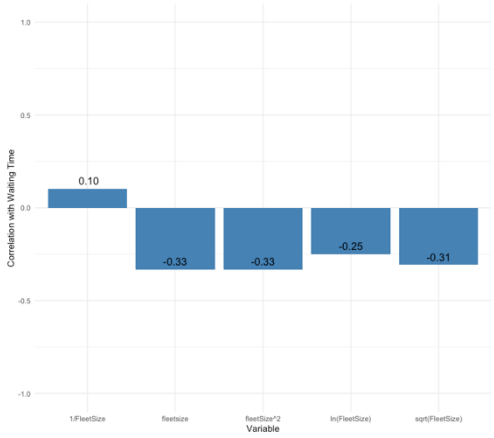

Figure 40 - Correlation between waiting time and non-linear functions of fleet size

There was no significant change in the correlation with vehicle capacity and fleet size. However, there was a improve in the correlation for square and logarithmic functions of demand the variables transformation, further research has to be performed in the regression analysis.

For modeling demand the potential demand was divided by the hours in the day, therefore potential demand used for the models are in rides/h. In all models an interaction term between vehicle capacity and demand was added in order to account for the curvilinearity change seen on Figure 35, the addition of the interaction term improved the results quality for all models presented when compared without it.

Table 19 – Best fitted models average waiting time regression – scenario 1.

Number	**Model Form**	R^2	Adj. R^2	Residual SE	F-statistic (p-values on joint Hypotheses)	Lower Coef. Signif.
1	$-\beta_0 + \beta_1 * \ln Q - \beta_2 * \ln B + \beta_3 * C - \beta_4 * C^2 - \beta_5 * \ln Q * C$	0.977	0.975	21.62	< 2.2e-16	*
2	$\beta_0 + \beta_1 * \ln Q - \beta_2 * \ln B - \beta_3 * \frac{1}{C} + \beta_4 * \ln Q * \frac{1}{C}$	0.976	0.975	21.54	< 2.2e-16	***
3	$-\beta_0 + \beta_1 * \ln Q - \beta_2 * \ln B + \beta_3 * \ln C * -\beta_4 * \ln Q * \ln C$	0.977	0.974	21.82	< 2.2e-16	***
4	$-\beta_0 + \beta_1 * \ln Q - \beta_2 * \ln B + \beta_3 * \sqrt{C} - \beta_4 * \ln Q * \sqrt{C}$	0.975	0.974	22.06	< 2.2e-16	***
5	$-\beta_0 + \beta_1 * \ln Q - \beta_2 * \ln B + \beta_3 * C - \beta_4 * C * \ln Q$	0.975	0.973	22.35	< 2.2e-16	*

6	$\beta_0 + \beta_1 * \ln Q - \beta_2 * \ln B + \beta_3 * \sqrt{C}$	0.967	0.965	25.32	< 2.2e-16	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

All models provide a good fit with satisfactory R^2 and statistical significance for the coefficients, with all models explaining between 97 and 98% of data variance.

The normality of residuals assumption is analyzed graphically and shown in Figure 41 and by applying statistical tests for the null hypothesis of normality, with results shown in Table 20. Based on the statistical tests performed, the null hypothesis of normality distribution of residuals can only be rejected at 0.05 significance level only based on shapiro tests, with exception to model 6. The graphical analysis provides further insights for all models, and overall there is a good fit with normal distribution, with models 1, 2 having one extreme value outside the confidence interval, which might contribute for the low p-value of Shapiro-Wilk tests on these models.

Table 20 – Tests for OLS regression residuals normality – waiting time (scenario1)

Number	Kolmogorov-Smirnov test (p-value)	Shapiro-Wilk normality test (p-value)
1	0.1625	0.0006
2	0.1066	0.002
3	0.2564	0.0116
4	0.3117	0.0305
5	0.5118	0.0791
6	0.7807	0.7617

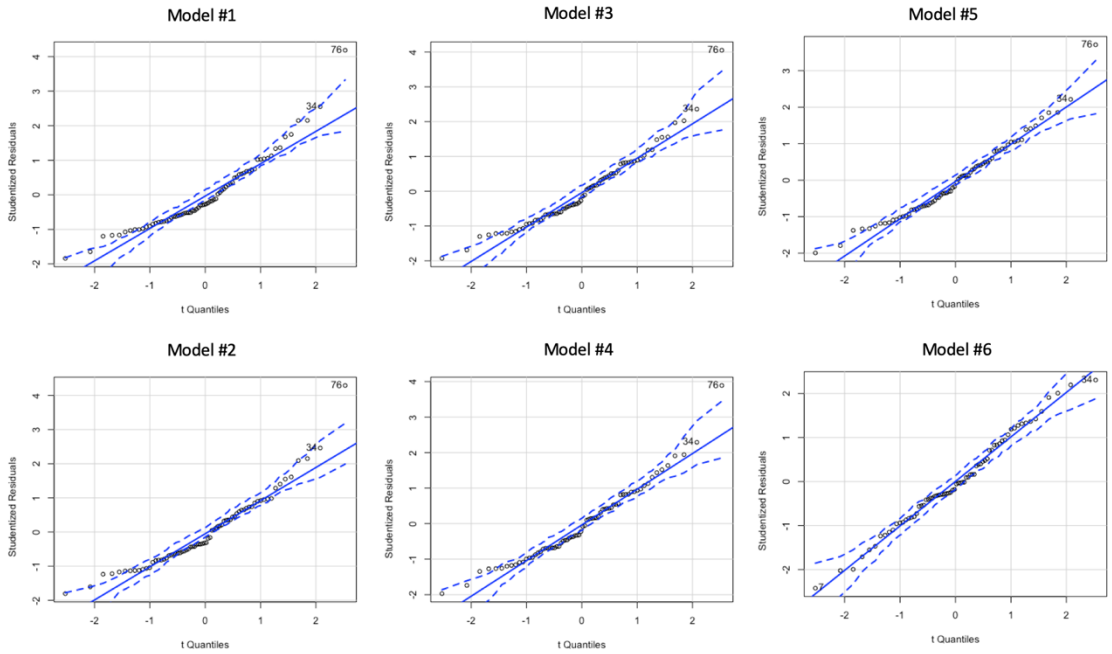

Figure 41 – Normal Q-Q plots for residuals of regression models for average waiting time including 0.95 confidence level– scenario 1.

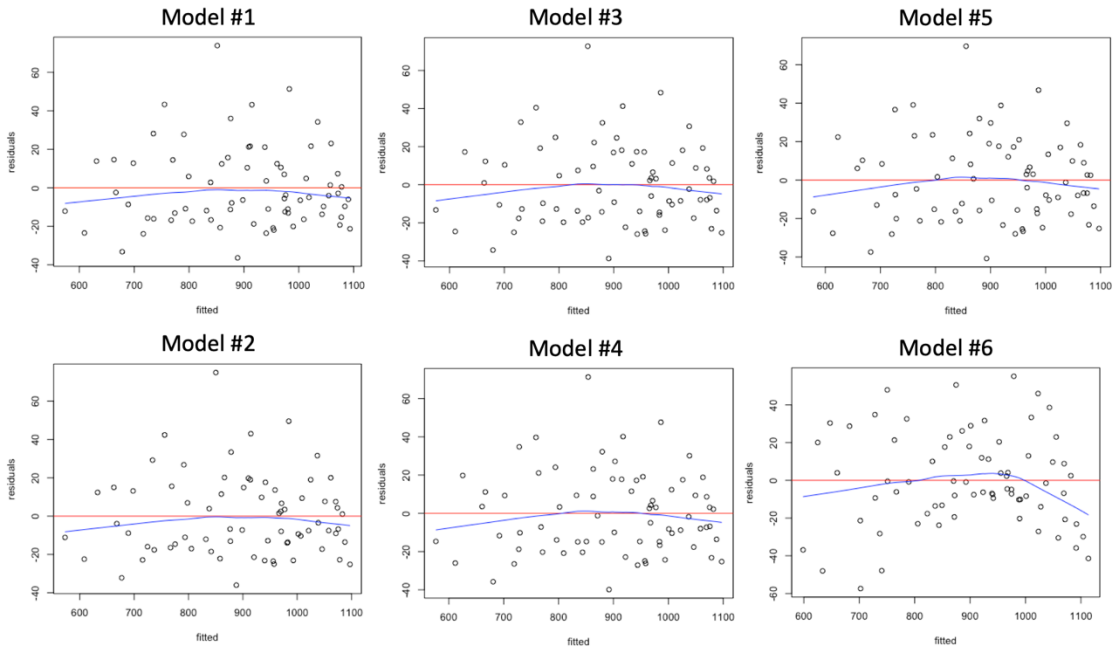

Figure 42 – Residual plots against fitted values for regression models of waiting time– scenario 1.

Residuals variance independence is also analyzed graphically in Figure 42. The models show a similar distribution, with a good independence of residuals. The conditional variance of each group is calculated and the ratio to the largest to the smallest variance of groups is analyzed (Cohen et al., 2003), results are in Table 21.

Table 21 – Ratio between maximum and minimum conditional variance of residual groups for average .waiting time regression – scenario 1.

Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
5.16	2.19	1.84	1.79	1.88	8.7

Models 1 and 6 have the higher ratio values, while all other models having ratio values between bellow 3, showing a better variance independence of residuals.

Although other models 1, 2 and 3 have a slight better coefficient of determination, model 5 accounts for the linear influence of vehicle capacity and is able to explain 98% of data variance while having the better normality and independence of residuals. Therefore, given the results presented, model number 5 is chosen, equation model parameters are shown in the following table.

Table 22 - Equation model for waiting time – scenario 1

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	-227.075	90.458	-2.510	0.0145 *
$\ln Q$	198.928	11.324	17.567	< 2e-16 ***
$\ln B$	-75.238	2.741	-27.448	< 2e-16 ***
C	78.829	14.366	5.487	6.75e-07 ***
$\ln Q * C$	-8.387	1.791	-4.682	1.43e-05 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Residual standard error: 22.35 on 63 degrees of freedom
Multiple R-squared: 0.9746 Adjusted R-squared0.9731
F-statistic: 643.8 on 4 and 63 DF, p-value: < 2.2e-16

The model selected is able to take into account the non-linear relationship of average waiting time with demand and fleet size. The logarithmic functions are able to accrue for the saturation effects of demand and fleet size. The linear predictor of vehicle occupancy fits well the data. The interaction term included moderates the steepness of the relationship between waiting time and vehicle occupancy.

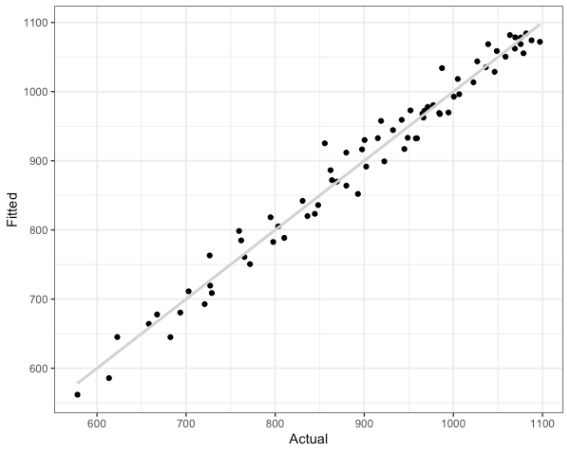

Figure 43 – Actual vs Fitted values for the average waiting time regression model. Scenario 1.

Some of the key findings are:

- As the detour, the average waiting time for this scenario is mainly influenced by the demand, the linear relationship between these variables appear to be strong and is able to explain more than 62% of the sample variance.
- Fleet size and demand have a saturation effect on waiting time, this effect is well explained by the logarithmic functions in the model.

4.2. Scenario 2 – Operation with rejections

A total of 5 scenarios did not achieve convergence for the number of iterations set on MATSim simulations and were removed from the output data analysis. The dataset analyzed contains 85 data points. The following table and Figure 44 provide a summary of the dataset for scenario 1.

The following table summarize the dataset output from the simulations used in the multiple regression analysis used for scenario 2.

Table 23 – Summary of the Dataset Scenario 2

Variable	Min	Max	Mean	Median	Std	Percentiles			
						25%	50%	75%	90%

Distance detour factor	1.06	1.22	1.16	1.16	0.05	1.12	1.16	1.19	1.22
Avg waiting time (sec)	104.7	308.8	187.4	168.0	64.66	133.7	168.0	243.0	287.0
Vehicle occupancy	0.86	1.43	1.21	1.21	0.15	1.09	1.21	1.32	1.42
Rejection Rate	0.00	0.90	0.19	0.01	0.28	0.00	0.01	0.31	0.88

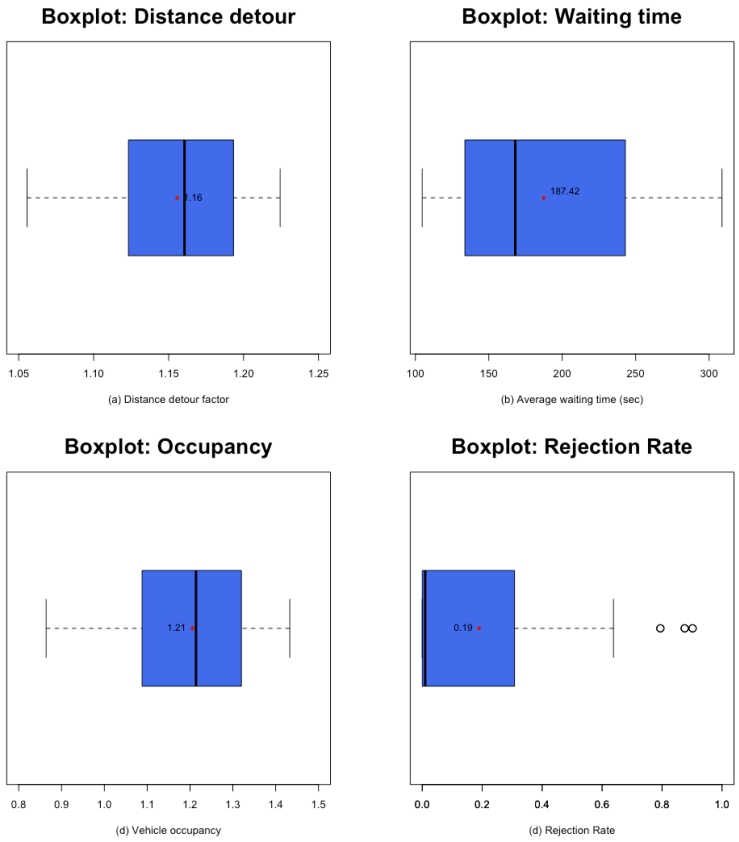

Figure 44 – Boxplots for scenario 2 (a) distance detour factor, (b) Average waiting time, (c) Average vehicle occupancy, (d) Rejection rate. The mean of the variables and its value is also depicted in each boxplot.

The values obtained for the outputs show values well distributed around mean and median values for all variables with the exception of rejection rate values, the reason is because 75% of the scenarios simulated resulted in rejection rates below 31%. Therefore, the lower number of datapoints with higher values depicted in the boxplot for rejection rate are not outliers and are included in the dataset. The fact that there are no negative outliers is consistent with the expected simulation outputs. It possible to note that for this case the range values of the variables are lower when comparing to values of scenario 1. Detour factor for example ranges from ranges from 1.06 to 1.22, while in the no rejection operation scenario it ranges from 1.89 to 3.79. This difference in the range of values is a direct result of the constrained operation.

4.2.1.Detour Factor

4.2.1.1. Correlation and relationship between variables

Following is a correlation table for the input variables and detour factor for scenario 2.

Table 24 – Correlation between input variables and distance detour factor

Variable	Correlation
Potential demand	0.12
Fleet size	0.76
Vehicle capacity	-0.03

According to Table 24, the variable that has a higher linear correlation with the distance detour factor is the fleet size, with a correlation of 0.76. In contrast with scenario 1, where vehicle capacity had the main influence on detour, here there is almost no influence of vehicle capacity on detour, with a correlation of -0.03.

The high correlation with fleet size is due to the constrained operation case, in which the relationship between supply and demand now plays an important role in the rejection of trips. Although linear correlation is positive, is shown further that this relationship is not linear and depends on the ratio between demand and supply.

There is also a positive linear correlation with potential demand, which is also expected for the same reasons as for the scenario 1 and described on equation (12). The increase in number of vehicles increase the chances to match requests to closer vehicles. Figure 45 contains the scatter plots of detour factor and the independent variables.

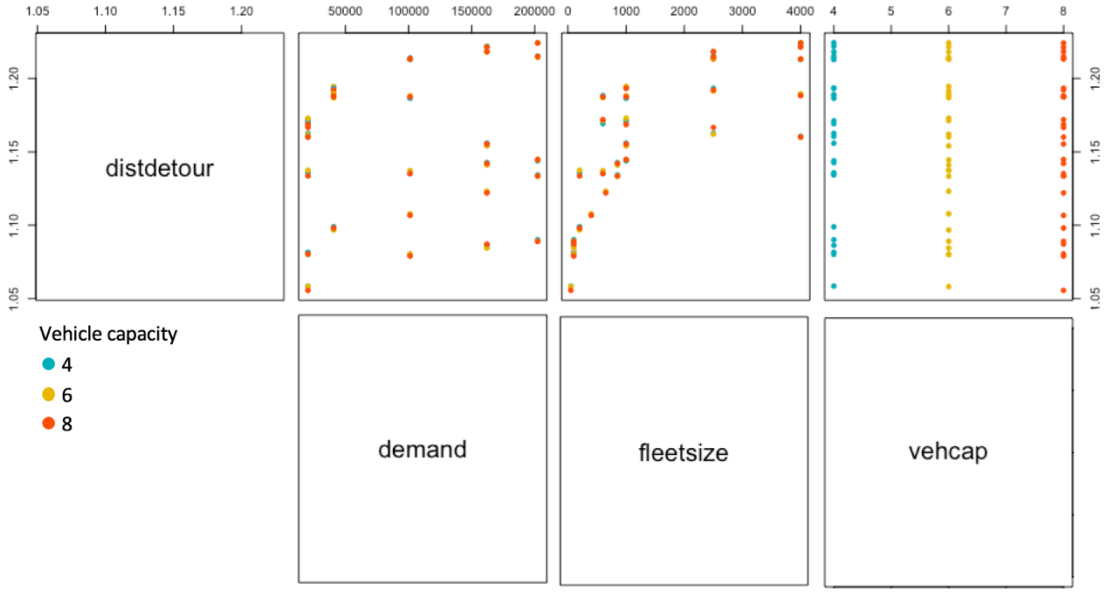

Figure 45 – Scatter plots of Detour factor against independent variables for scenario 2.

Figure 46 and Figure 47 depict the relationship between detour factor and vehicle capacity and detour factor and fleet size for scenario 2. The vehicle capacity has almost no influence on determining the detour, with almost a constant detour with capacity changes.

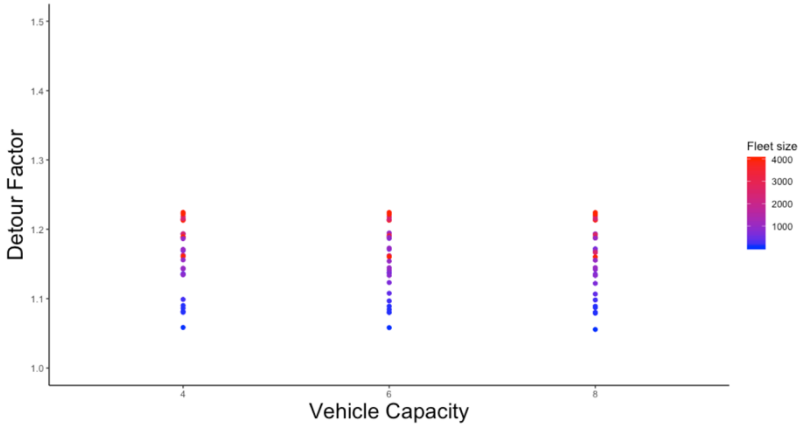

Figure 46 – Detour Factor according vehicle capacity (A) and Detour Factor according fleet size (B) for scenario 2.

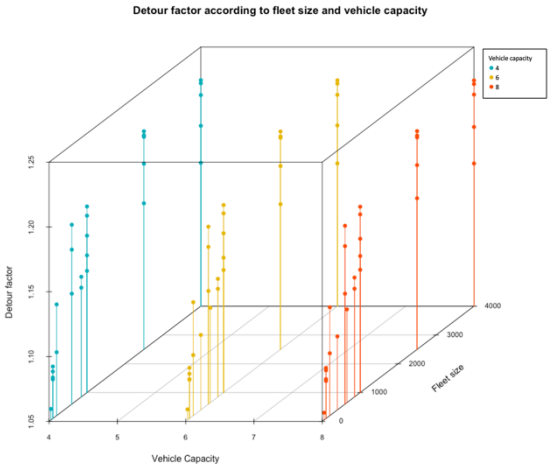

Figure 47 - 3D plot of Detour factor, vehicle capacity and Fleet Size. Scenario 2.

Figure 48 details the scenarios and the change in detour factor with changes in fleet size. It is possible to note that due to the constrained operation, the ration between demand and supply (fleet size) plays an important role. As bigger the ration between demand and fleet size (undersupply), the increase in fleet size increases the detour, after some level in which increases in fleet seems to be indifferent. Therefore, fleet size and detour factor have shown a non-linear relationship dependent also on the interaction between fleet supply and demand.

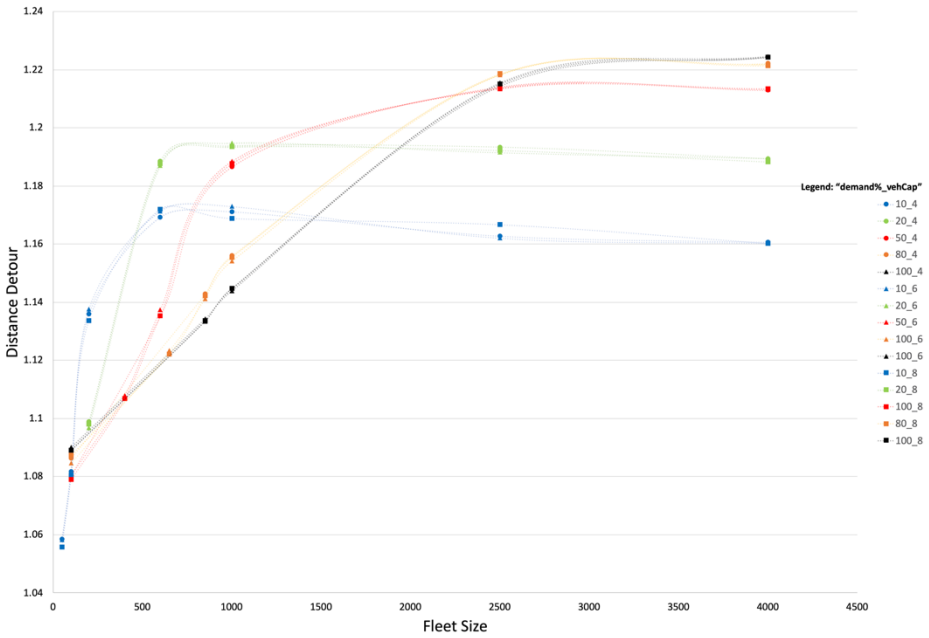

Figure 48 – Detour factor and fleet size. Scenario 2.

The correlation between potential demand and distance detour factor is higher for scenario 2 when compared to scenario 1. Figure 49 and **Error! Reference source not found.** depicts how detour changes with demand. As already described, due to the constrained operation, the detour factor depends on the ratio between demand and supply. The data shows that the detour change with demand is very low, and it tends to increase slightly for increases of demand at low levels, having less influence as demand increases.

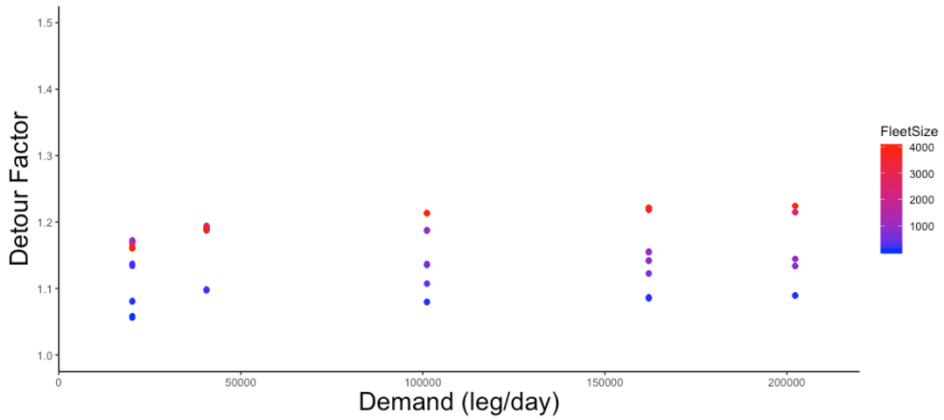

Figure 49 - Detour Factor according number of trips (A) colored by fleet size (B) colored by vehicle capacity. Scenario 2.

4.2.1.2. Multiple Regression

Given the non-linear relationships observed between distance detour factor with fleet size and demand, different types of relationship were analyzed and used as regressors for the detour function.

The following Figure 50, Figure 51 and Figure 52 show the correlation of distance detour factor with transformed variables, this exploratory measure helps to identify better the type of non-linear relationship that accounts for the curvilinear existing relationship and to eliminate the nonlinearity.

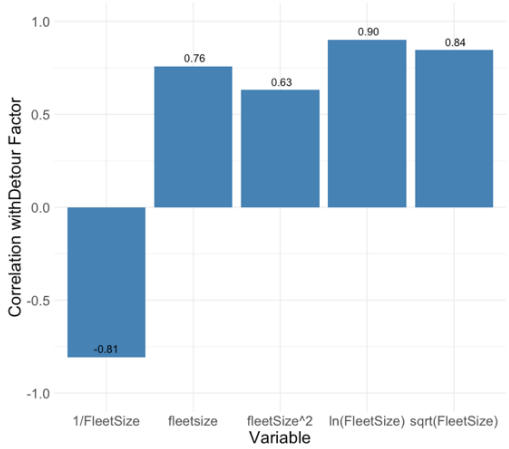

Figure 50 – Correlation between detour factor and non-linear functions of fleet size

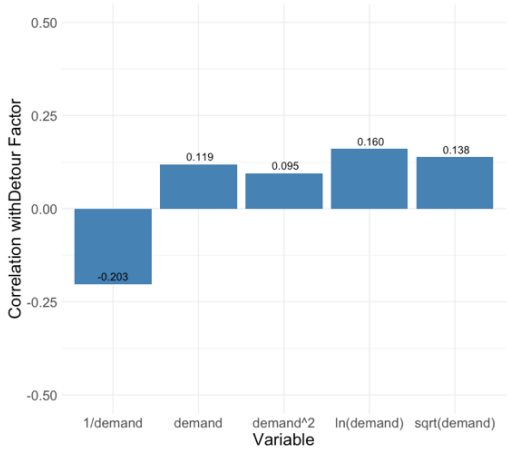

Figure 51 – Correlation between detour factor and non-linear functions of demand

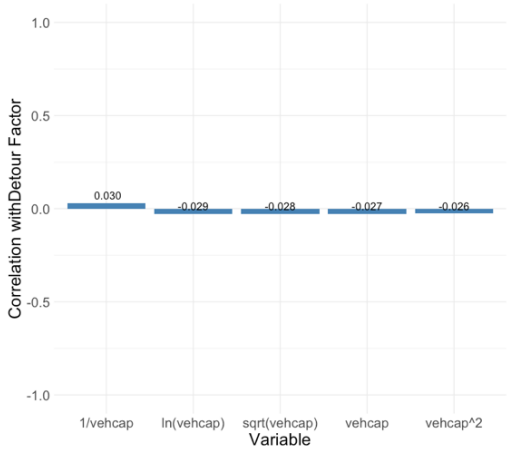

Figure 52 – Correlation between detour factor and non-linear functions of vehicle capacity

The correlations indicate almost no correlation of vehicle capacity and distance detour factor. For fleet size and demand, there is an increase on the correlation for logarithmic and square root functions. For demand, there is no significant change on the variable's transformation.

Based on the previous exploratory data analysis and theoretical explanations, multivariate regression models were performed, and best performing model forms are summarized on Table 25. Only models that have all coefficients statistically significant for at least 0.1 significance level are selected for analysis. For modeling demand, the potential demand was divided by the hours in the day, therefore potential demand used for the models are in rides/h.

Table 25 – Best fitted models Detour Factor regression – scenario 2.

Number	**Model Form**	R^2	Adjusted R^2	Residual SE	F-statistic(p-values on joint Hypotheses)	Lower Coef. Signif.
1	$\beta_0 + \beta_1 * \ln B + \beta_2 * \frac{1}{Q} - \beta_3 * \ln B * \frac{1}{Q}$	0.840	0.835	0.0110	< 2.2e-16	***
2	$\beta_0 + \beta_1 * \ln B - \beta_2 * \sqrt{Q} - \beta_3 * \ln B * \sqrt{Q}$	0.834	0.828	0.020	< 2.2e-16	**
3	$\beta_0 + \beta_1 * B - \beta_2 * B^2 - \beta_3 * \sqrt{Q} + \beta_4 * B * \sqrt{Q}$	0.833	0.825	0.0205	< 2.2e-16	***
4	$\beta_0 + \beta_1 * \ln B - \beta_2 * Q + \beta_3 * \ln B * Q$	0.831	0.825	0.0205	< 2.2e-16	**
5	$\beta_0 + \beta_1 * B - \beta_2 * B^2 - \beta_3 * \frac{1}{Q} - \beta_4 * B * \frac{1}{Q}$	0.831	0.822	0.0206	< 2.2e-16	**
6	$\beta_0 + \beta_1 * B - \beta_2 * B^2 - \beta_3 * Q + \beta_4 * B * Q$	0.830	0.821	0.0207	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

All models provide similar fit with satisfactory R^2 and statistical significance for the coefficients, with model 1 being able to explain 84% of data variance.

The normality of residuals assumption is analyzed graphically and shown in Figure 53 and by applying statistical tests, with results shown in Table 26. The graphical analysis provides satisfactory results for all models. It is only possible reject the null hypothesis of normality distribution of residuals at 0.05 significance level for Shapiro-Wilks tests only.

Table 26 – Tests for OLS regression residuals normality

Number	Kolmogorov-Smirnov test (p-value)	Shapiro-Wilk normality test (p-value)
1	0.2403	0.0041
2	0.2486	0.0161
3	0.1006	0.0126
4	0.2338	0.017
5	0.084	0.004
6	0.1857	0.0338

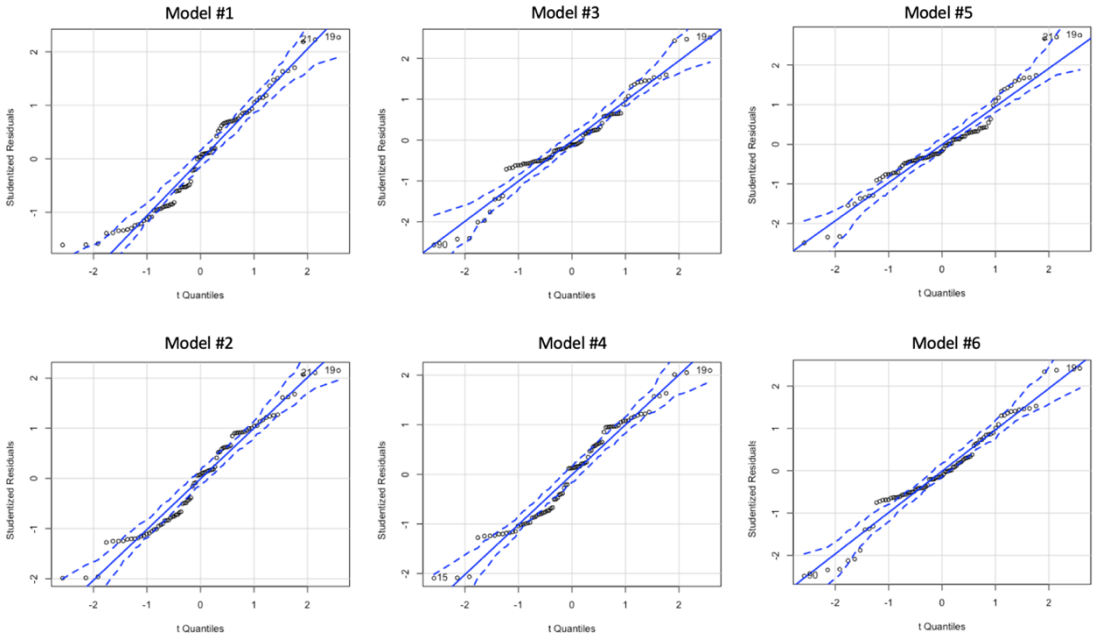

Figure 53 – Normal Q-Q plots for residuals of regression models for detour factor including 0.95 confidence level– scenario 2.

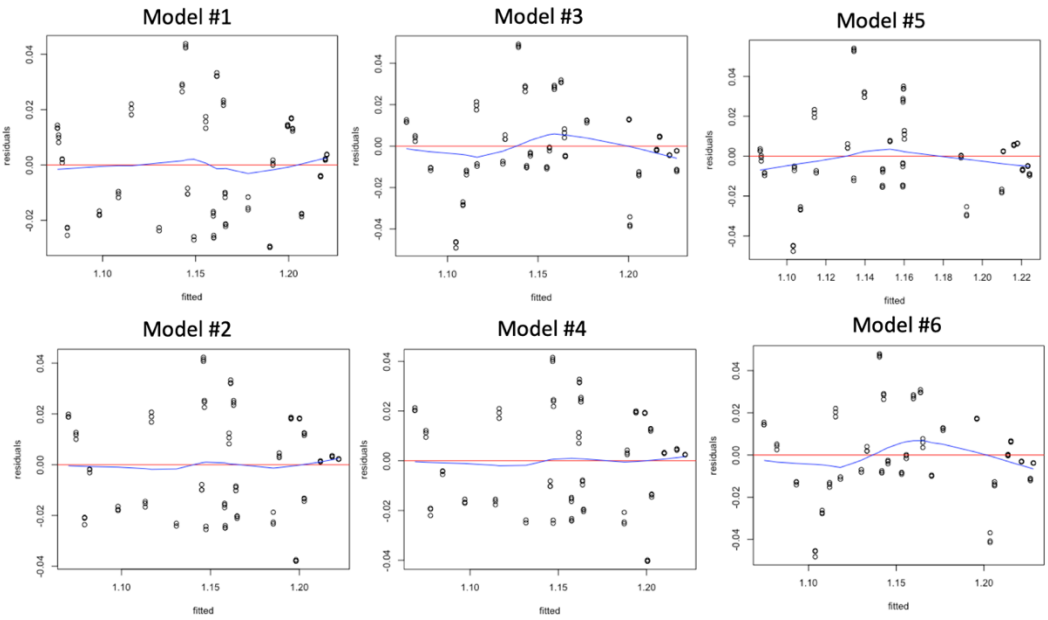

Figure 54 – Residual plots against fitted values for regression models of detour factor – scenario 2.

Residuals variance independence is also analyzed graphically in Figure 54. The conditional variance of each group is calculated and the ratio to the largest to the smallest variance of groups is analyzed (Cohen et al., 2003), results are in Table 27.

Table 27 – Ratio between maximum and minimum conditional variance of residual groups for detour factor regression – scenario 2.

Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
3.95	1.82	11.78	1.34	15.74	10.3

Model 3, 5 and 6 have the highest ratio values; all the other models have values lower than 10.75. Therefore, given the results presented, model number 1 is chosen, equation model parameters are shown in the following table.

Table 28 - Equation model for detour factor – scenario 2

Parameter (coefficient)	Estimate	Std. Error	t value	Pr(>|t|)
Intercept (β_0)	0.885	0.0183	48.341	< 2e-16 ***
$\ln B$ (β_1)	0.040	0.003	15.394	< 2e-16 ***
$\frac{1}{Q}$ (β_2)	83.093	23.599	3.521	0.000709 ***
$\ln B * \frac{1}{Q}$ (β_3)	-13.478	3.561	-3.775	0.000304 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.01992 on 81 degrees of freedom
Multiple R-squared: 0.8405, Adjusted R-squared: 0.8346
F-statistic: 142.3 on 3 and 81 DF, p-value: < 2.2e-16

The model selected is able to take into account the non-linear relationship between the detour factor and fleet size and detour factor and demand. The interaction term included represents the system state regarding supply and demand, which influences directly the operation in a constrained operation scenario, this interaction term moderates the curvilinearity of the relationships between detour and independent variables. The reciprocal function for demand and logarithmic for fleet size are in line with the results and the saturation effect observed.

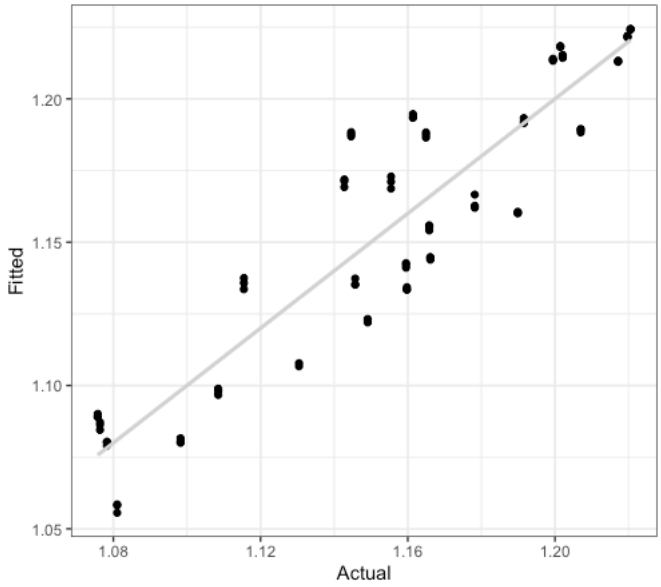

Figure 55 – Fitted vs Actual values for the model selected for distance detour factor. Scenario 2.

4.2.2.Average Occupancy

4.2.2.1. Correlation and relationship between variables

Following is a correlation table for the input variables and average occupancy for scenario 2.

Table 29 – Correlation between input variables and occupancy. Scenario 2.

Variable	Correlation
Potential Demand	0.33
Fleet size	0.84
Vehicle capacity	-0.01

According to Table 29, the variable that has a higher linear correlation with the occupancy is fleet size, with a correlation of 0.84. Contrasting to scenario 1, vehicle capacity has almost no influence on occupancy levels for this case. The high correlation with fleet size is expected for the constrained operation scenario. There is also a positive correlation with potential demand. Figure 56 contains the scatter plots of occupancy and the other variables.

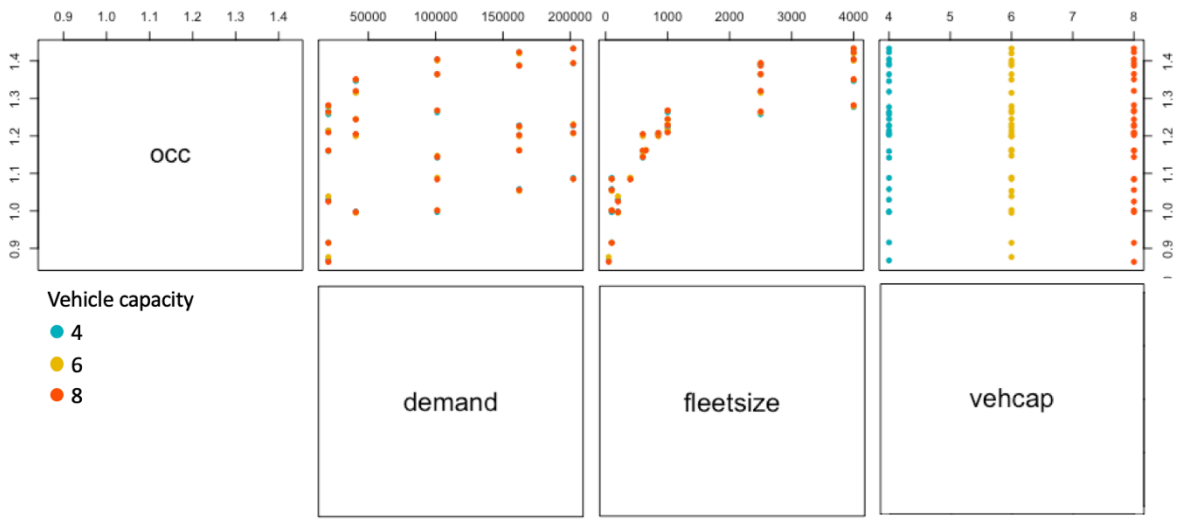

Figure 56 – Occupancy x independent variables for scenario 2.

Figure 57 depicts the relationship between occupancy and vehicle capacity and fleet size for scenario 2. As the correlation already shown, the number of seats has almost no influence on the occupancy levels. The fleet size increase leads to an increase in occupancy for the same demand levels, the data suggests a non-linear relationship accounting for the saturation effect of fleet size. For the same demand level, a lower fleet size has a higher rejection rate and lower realized demand, in these cases the probability of shared rides are lower and occupancy tends to be small, as fleet size increases and the rejection rate approaches zero, the occupancy is also increased, a saturation effect is also observed for this relationship, in which the rate of change in fleet size decreases. Therefore, this relationship is highly related to the operational scheme used and the presence of rejections.

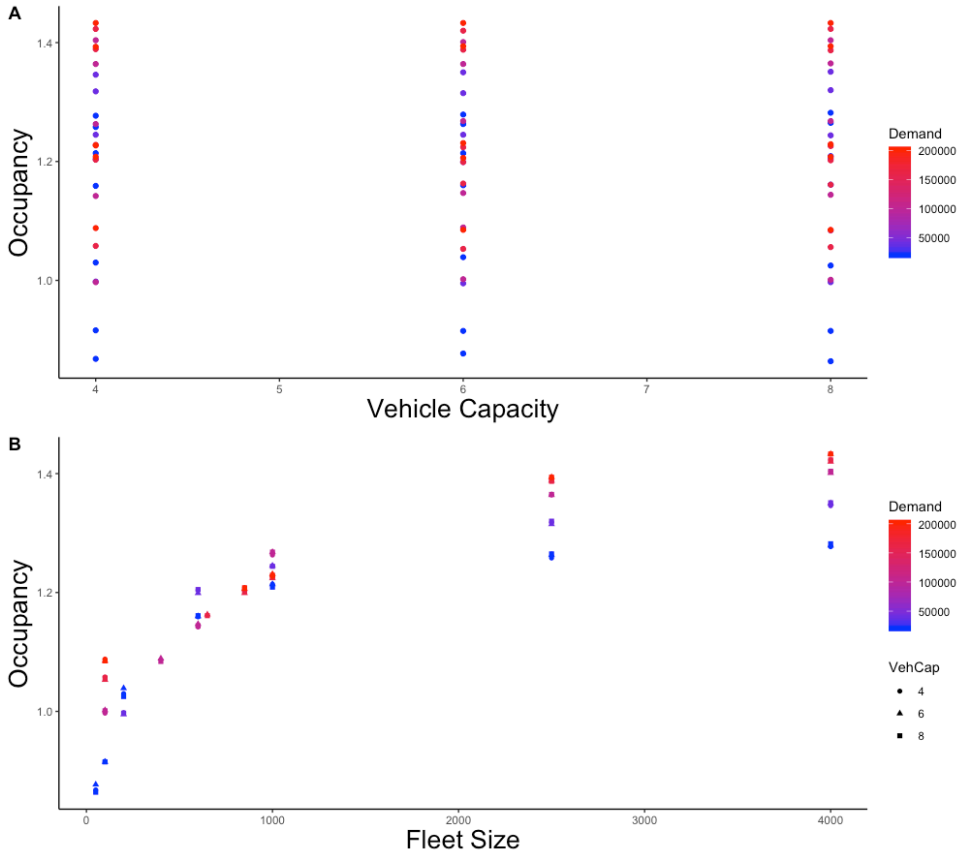

Figure 57 - Occupancy according vehicle capacity (A) and Occupancy according fleet size (B).

Regarding demand, the same saturation effect observed in scenario 1 is present, as can be observed in Figure 58. For this case, the theoretical asymptote seems to be lower than the vehicle capacity value, a direct effect of the constrained operation.

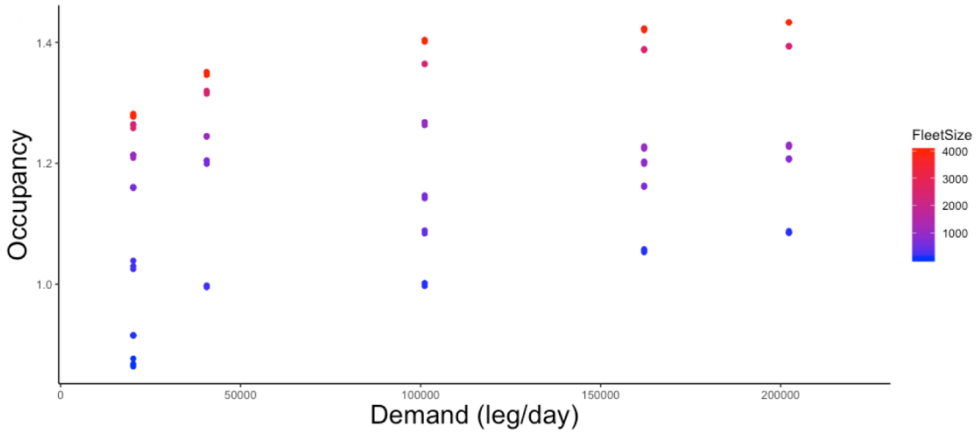

Figure 58 - Occupancy according number of trips colored by fleet size. Scenario 2.

4.2.2.2. Multiple Regression

Given the non-linear relationships observed between average occupancy with fleet size and demand, different types of non-linear relationships were analyzed and used as regressors for the average occupancy function.

The following Figure 59, Figure 60 and Figure 61 show the correlation of average occupancy with transformed variables, this exploratory measure helps to identify better the type of non-linear relationship that accounts for the curvilinear existing relationship and to eliminate the nonlinearity.

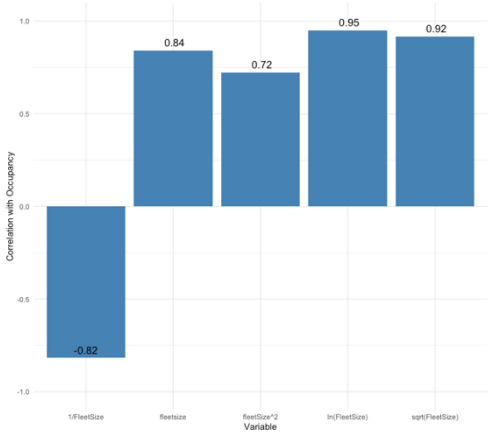

Figure 59 – Correlation between occupancy and non-linear functions of fleet size – scenario 2.

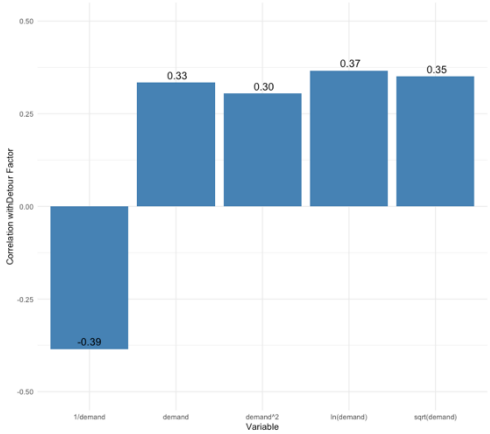

Figure 60 – Correlation between occupancy and non-linear functions of demand – scenario 2.

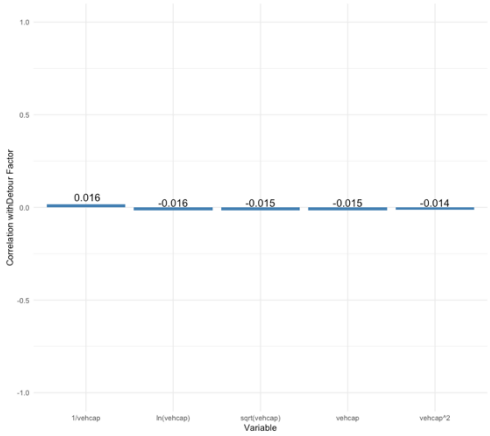

Figure 61 – Correlation between occupancy and non-linear functions of vehicle capacity – scenario 2.

The correlations indicate an almost null linearity between vehicle capacity and occupancy. For fleet size an increase in correlation was observed for logarithmic and square root functions. For demand, there is an increase for logarithmic and reciprocal functions.

Based on the previous exploratory data analysis and theoretical explanations, multivariate regression models were performed and are summarized on Table 30. For modeling demand the potential demand was divided by the hours in the day, therefore potential demand used for the models are in rides/h.

Table 30 – Best fitted models average occupancy regression – scenario 2.

Number	**Model Form**	R^2	Adjusted R^2	Residual SE	F-statistic(p-values on joint Hypotheses)	Lower Coef. Signif.
1	$\beta_0 + \beta_1 * \ln B + \beta_2 * \ln Q$	0.949	0.948	0.0346	< 2.2e-16	***
2	$\beta_0 + \beta_1 * \ln B + \beta_2 * \sqrt{Q}$	0.948	0.947	0.0349	< 2.2e-16	***
3	$\beta_0 + \beta_1 * \ln B + \beta_2 * \frac{1}{Q}$	0.948	0.946	0.035	< 2.2e-16	***
4	$\beta_0 + \beta_1 * \ln B + \beta_2 * Q$	0.946	0.945	0.0356	< 2.2e-16	***
5	$\beta_0 + \beta_1 * B - \beta_2 * B^2 - \beta_3 * \frac{1}{Q} + \beta_4 * B * \frac{1}{Q} + \beta_5 * B^2 * \frac{1}{Q}$	0.938	0.934	0.0391	< 2.2e-16	*
6	$\beta_0 + \beta_1 * B - \beta_2 * B^2 - \beta_3 * \ln Q + \beta_4 * B * \ln Q + \beta_5 * B^2 * \ln Q$	0.936	0.932	0.0396	< 2.2e-16	**

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

All models provide a good fit with satisfactory R^2 and statistical significance for the coefficients, with model 1 being able to explain 95%.

The normality of residuals assumption is analyzed graphically and shown in Figure 62 and by applying statistical tests, with results shown in Table 31. Based on the statistical tests performed, it is only possible reject the null hypothesis of normality distribution of residuals at 0.05 significance level for model number 2 and 3 according to Kolmogorov-Smirnov, this happens for all models with Shapiro-Wilk test. The graphical analysis provides further insights for all models. Models 1 and 2 have a good normal distribution, with higher discrepancy in values around the mean.

Table 31 – Tests for OLS regression residuals normality – occupancy (scenario2)

Number	Kolmogorov-Smirnov test (p-value)	Shapiro-Wilk normality test (p-value)
1	0.0530	0.0001
2	0.0330	0.0000
3	0.0427	0.0001
4	0.0778	0.0124
5	0.0826	0.0169
6	0.0765	0.0032

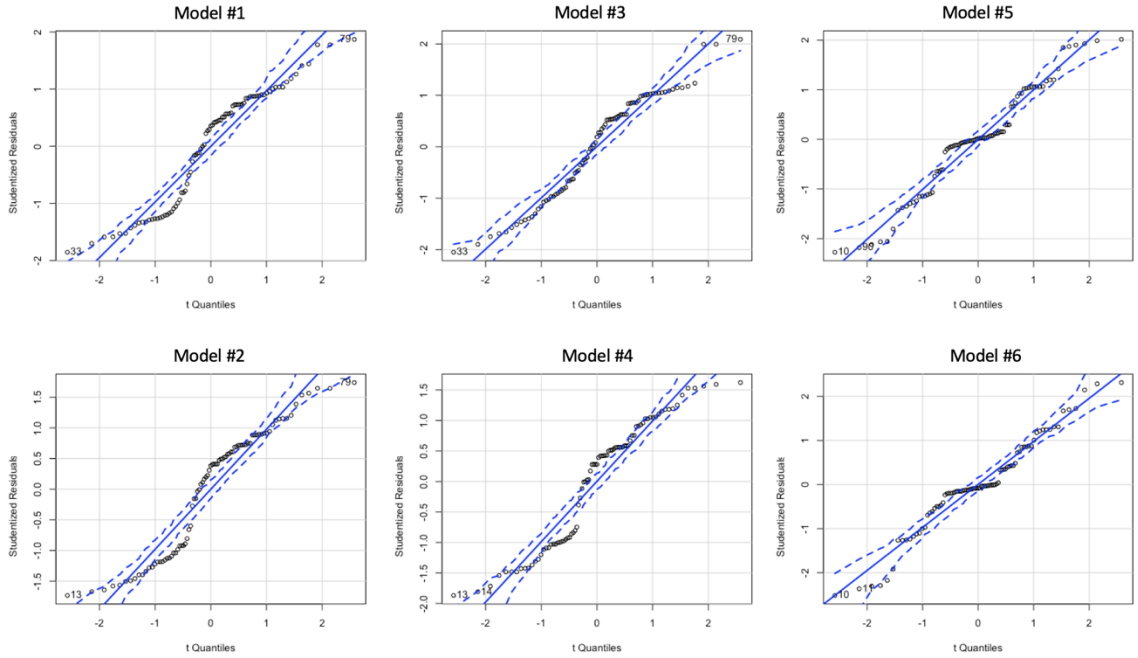

Figure 62 – Normal Q-Q plots for residuals of regression models for average occupancy including 0.95 confidence level– scenario 2.

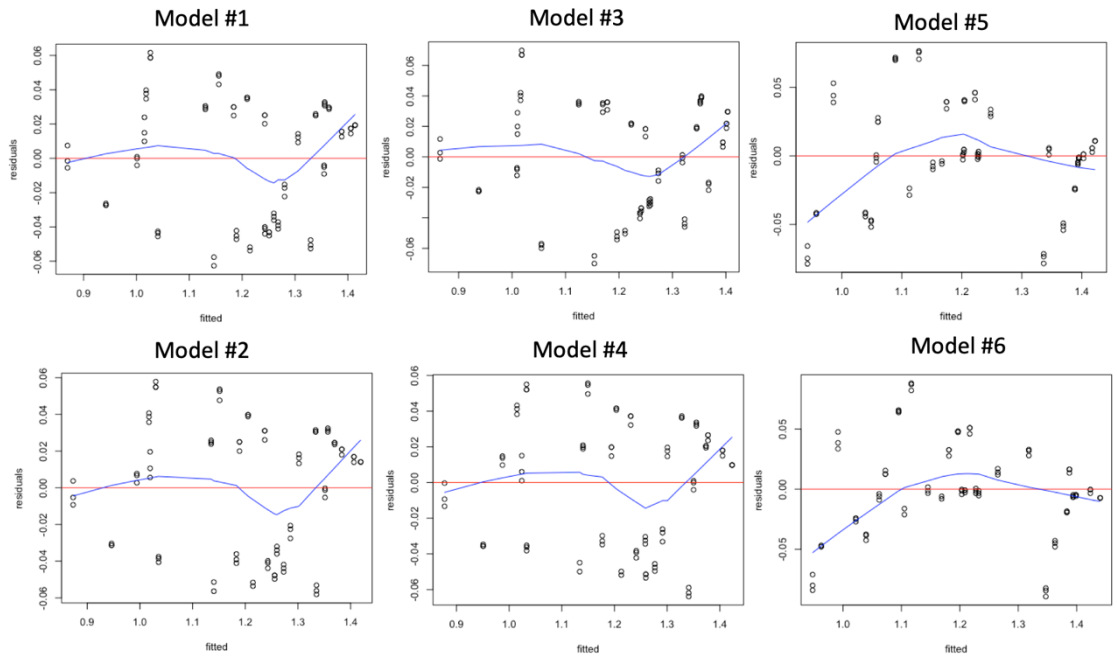

Figure 63 – Residual plots against fitted values for regression models of occupancy– scenario 2.

Residuals variance independence is also analyzed graphically in Figure 63. The models present a good degree of homoscedasticity. The conditional variance of each group is calculated and the ratio to the largest to the smallest variance of groups is analyzed (Cohen et al., 2003), results are in Table 32.

Table 32 – Ratio between maximum and minimum conditional variance of residual groups for average occupancy regression – scenario 2.

Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
2.12	2.71	2.38	61.14	53.07	93.32

Model 4, 5 and 6 have the highest ratio values. The other models all have values bellow 3, presenting a good independence of residuals. Therefore, given the results presented, model number 1 is chosen, equation model parameters are shown in the following table.

Table 33 - Equation model for occupancy – scenario 2

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
Intercept (β_0)	0.213098	0.036439	5.848	9.78e-08***
$\ln B$ (β_1)	0.104994	0.002888	36.359	< 2e-16 ***
$\ln Q$ (β_2)	0.036480	0.004250	8.583	4.86e-13***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.03458on 82 degrees of freedom
Multiple R-squared: 0.9494, Adjusted R-squared: 0.9482
F-statistic: 769.6 on 2 and 82 DF, p-value: < 2.2e-16

The model selected is able to take into account the non-linear relationship between the occupancy and fleet size and demand with the logarithmic functions. The fleet size is the main variable defining occupancy, $\ln B$ by itself is able to explain 90% of data variance.

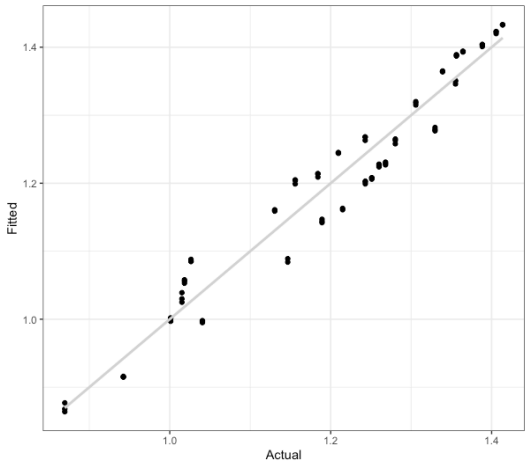

Figure 64 - Fitted vs Actual values for the model selected for average occupancy. Scenario 2.

4.2.3. Average Waiting time

4.2.3.1. Correlation and relationship between variables

Following is a correlation table for the input variables and waiting time for scenario 2.

Table 34 – Correlation between input variables and waiting time

Variable	Correlation
Potential Demand	0.30
Fleet size	-0.75
Vehicle capacity	0.03

According to Table 34, the variable that has a higher linear correlation with the waiting time is the fleet size, with a correlation of -0.75.

The negative relationship with fleet size is expected and has same the same explanation as for scenario 1. However, the higher degree of correlation for this scenario is due to the operational constraints. Vehicle capacity has almost null correlation with waiting time for this scenario. Figure 65 contains the scatter plots of waiting time against the independent variables.

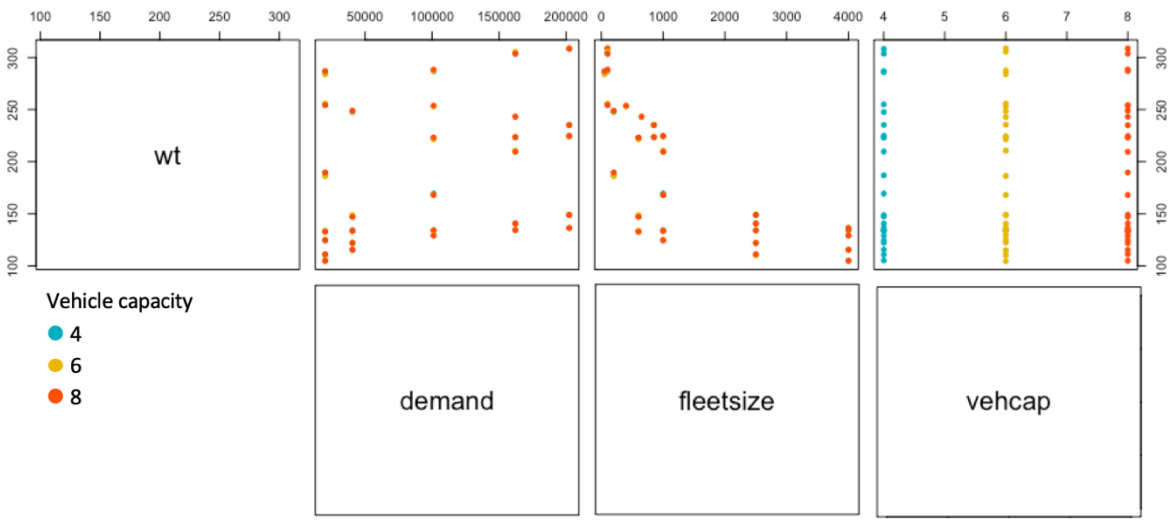

Figure 65 – Waiting time x independent variables, scenario 2.

Figure 66, Figure 67 and Figure 68 show the relationship between waiting time and vehicle capacity and fleet size. Vehicle capacity doesn't influence the waiting time in the scenario with rejections. Fleet size and demand are the most important variables under constrained operation.

The fleet size relationship with waiting time data suggests a non-linear function with a saturation behavior: as the number of vehicles increase, the rate of decrease in waiting time due to the increase in fleet size diminishes.

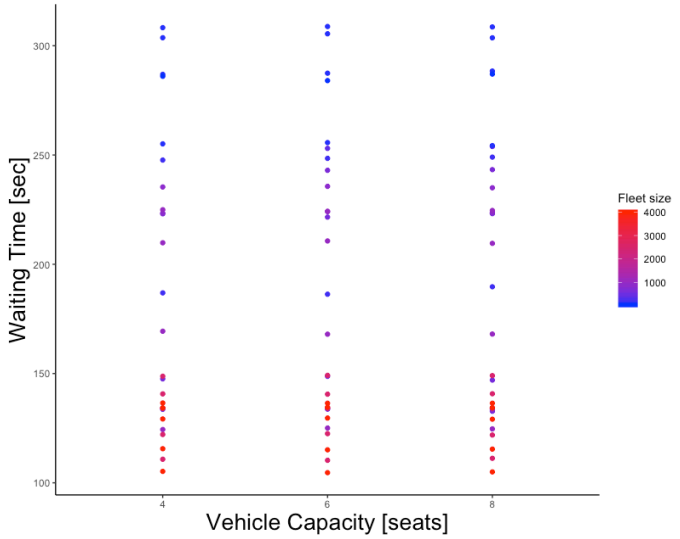

Figure 66 – Waiting time according vehicle capacity. Scenario 2.

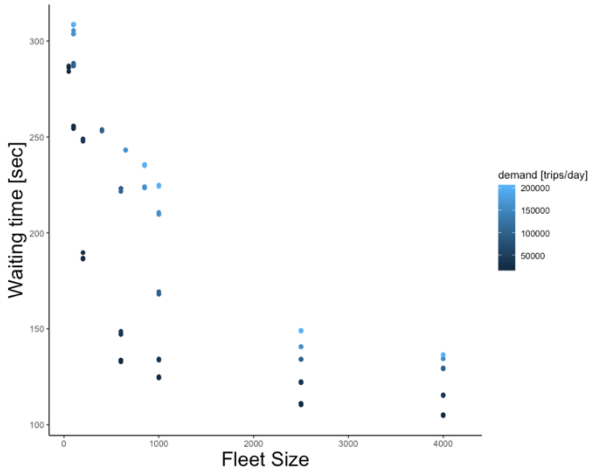

Figure 67 - Waiting time according to fleet size, scenario 2.

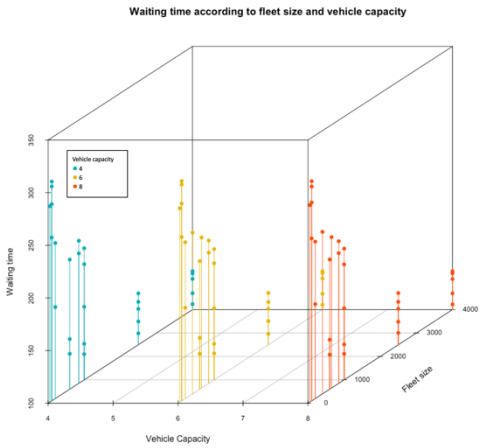

Figure 68 - 3D plot of waiting time [sec], vehicle capacity[seat] and Fleet Size[veh].

Potential demand has a positive correlation with waiting time; however, the data suggests a non-linear relationship depending on the ratio between demand and supply. Figure 69 presents this relationship. For lower ratios (higher fleet sizes) the behavior is similar to scenario 1, which is expected since for higher supply the rejections are almost null. For higher ratios (high demand and low fleet supply) the demand has a stronger influence on waiting time. This is explained by the increase in number of rejections in these specific scenarios, therefore suggesting an interaction between demand and supply (fleet size) that accounts for that effect and exists mainly because of the use of constraints allowing for request rejections.

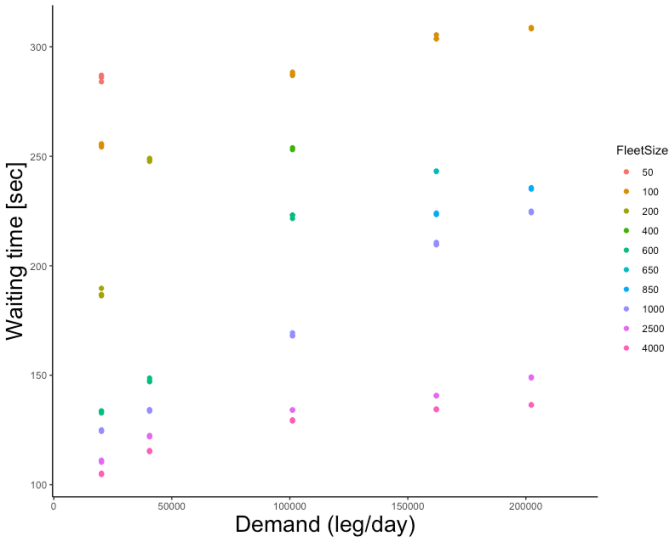

Figure 69 – Waiting time according number of trips colored by fleet size. Scenario 2.

4.2.3.2. Multiple Regression

Different types of non-linear relationships are analyzed and used as regressors for the waiting time function in order to account for the non-linear behavior observed.

The following Figure 70, Figure 71 and Figure 72 show the correlation of average waiting time with transformed variables, this exploratory measure helps to identify better the type of non-linear relationship that accounts for the curvilinear existing relationship and to eliminate the nonlinearity.

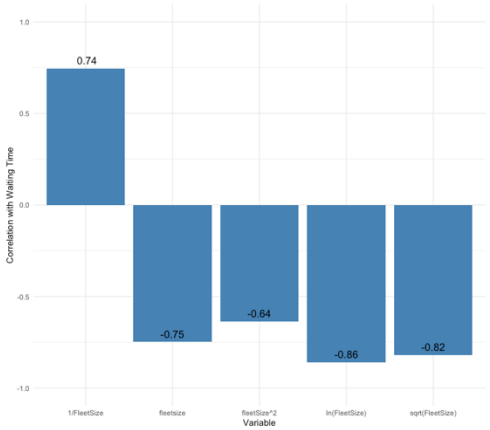

Figure 70 – Correlation between waiting time and non-linear functions of fleet size – scenario 2.

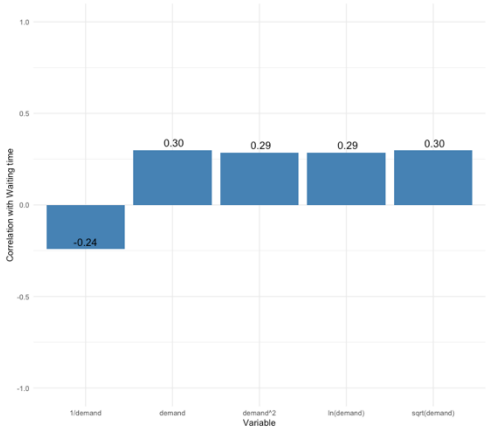

Figure 71 – Correlation between waiting time and non-linear functions of demand – scenario 2.

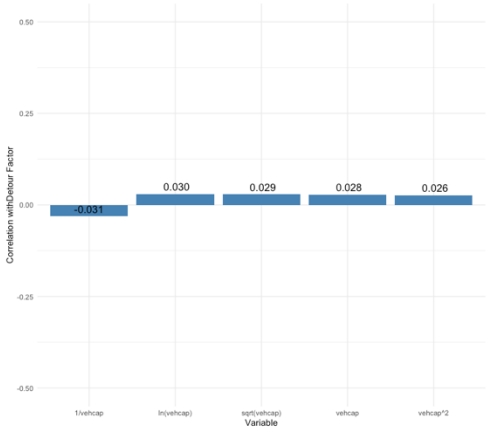

Figure 72 – Correlation between waiting time and non-linear functions of vehicle capacity – scenario 2.

As for all other variables in the constrained operational scenario, the correlations indicate an almost null correlation between vehicle capacity and average waiting time. For fleet size there is a slight increase in correlation was observed for logarithmic and square root functions. No particular improvements on correlation of demand.

Based on the previous exploratory data analysis and theoretical explanations, multivariate regression models were performed and are summarized on Table 35. For modeling demand the potential demand was divided by the hours in the day, therefore potential demand used for the models are in rides/h.

Table 35 – Best fitted models average waiting time regression – scenario 2.

Number	**Model Form**	R^2	Adjusted R^2	Residual SE	F-statistic(p-values on joint Hypotheses)	Lower Coef. Signif.
1	$\beta_0 - \beta_1 * \ln B + \beta_2 * \sqrt{Q} - \beta_3 * \ln B * \sqrt{Q}$	0.929	0.927	17.50	< 2.2e-16	*
2	$\beta_0 - \beta_1 * \ln B + \beta_2 * Q - \beta_3 * Q^2$	0.928	0.925	17.68	< 2.2e-16	**
3	$\beta_0 - \beta_1 * \ln B + \beta_2 * \ln Q$	0.927	0.925	17.65	< 2.2e-16	***
4	$\beta_0 - \beta_1 * \ln B + \beta_2 * \sqrt{Q}$	0.926	0.924	17.81	< 2.2e-16	***

5	$\beta_0 - \beta_1 * \ln B + \beta_2 * Q - \beta_3 * \ln B * Q$	0.922	0.919	18.36	< 2.2e-16	*
6	$\beta_0 - \beta_1 * \ln B + \beta_2 * Q$	0.918	0.916	18.77	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

All models presented provide a good fit with satisfactory R^2, with model 1 being able to explain 93% of data variance. Model 1 and 5 have the interaction term between fleet size and demand, however, the estimated coefficients for the interaction terms have a lower value in both cases, indicating that these two variables independently are able to explain the variance in the waiting time well, as seen in the other models. In the models presented, the main estimator is fleet size. In model 2 and 6, for example, an increase in 200 vehicles decreases the waiting time in 239 and 238 seconds, respectively. In order to increase the same amount of time, the necessary increase in the demand for these models are 11936 and 11937 pax/hour. This fact shows the important role of fleet size in determining the average waiting time and hence the performance of the system analyzed in scenario 2.

The normality of residuals assumption is analyzed graphically and shown in Figure 73, all models follows the normal distribution curve with some deviations, higher for model 5 and 6. From statistical tests of Table 36 all models can't reject the null hypothesis of normality at 0.05 significance level from Kolmogorov-Smirnov test, however the results from the Shapiro-Wilk test allows this null hypotheses to be rejected for all models but 2 and 6. The sample size is one of the reasons to this discrepancy between both test results and graphical analysis

Table 36 – Tests for OLS regression residuals normality – waiting time (scenario2)

Number	Kolmogorov-Smirnov test (p-value)	Shapiro-Wilk normality test (p-value)
1	0.1582	0.0064
2	0.6192	0.0566
3	0.2460	0.0295
4	0.2841	0.0292
5	0.2302	0.0027
6	0.7284	0.0558

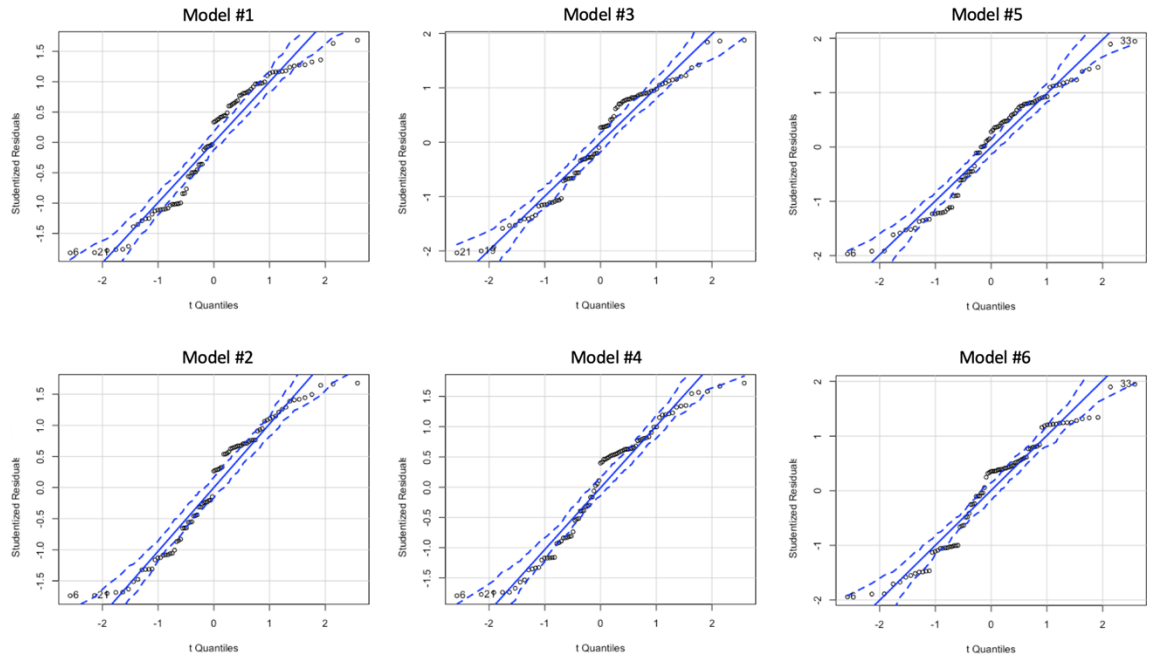

Figure 73 – Normal Q-Q plots for residuals of regression models for average waiting time including 0.95 confidence level– scenario 2.

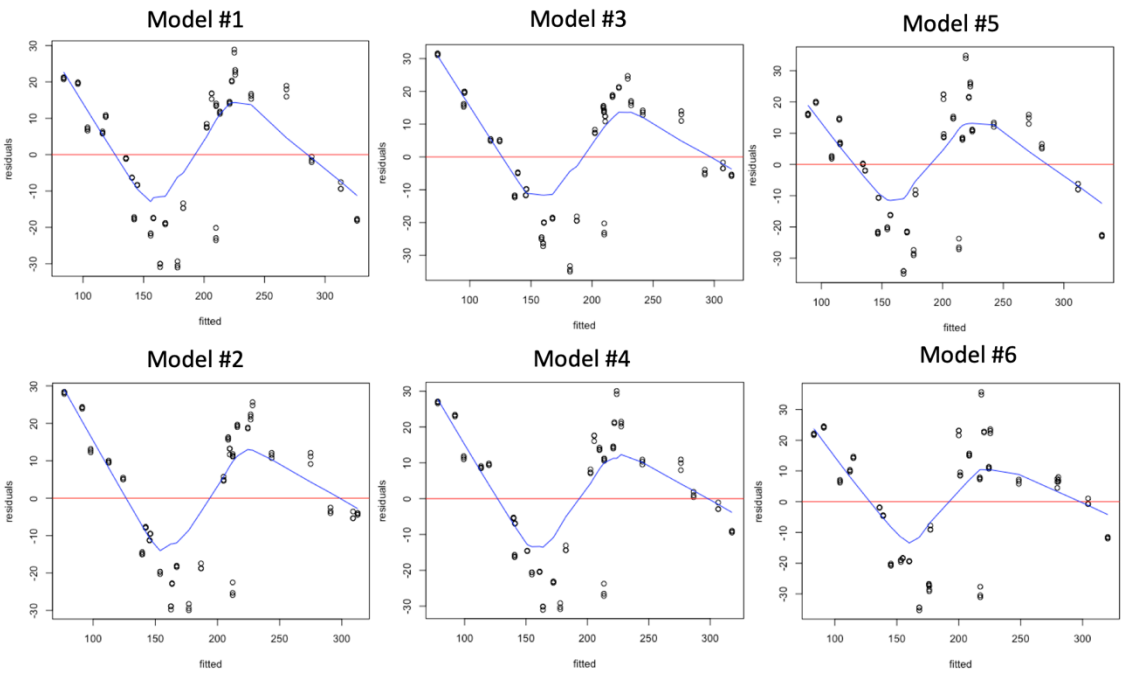

Figure 74 – Residual plots against fitted values for regression models of waiting time– scenario 2.

Residuals variance independence is also analyzed graphically in Figure 74. The models present a similar behavior for the residuals, and the lowess curve suggests a non-independence of the variance. The conditional variance of each group is calculated and the ratio to the largest to the smallest variance of groups is analyzed (Cohen et al., 2003), results are in the following table.

Table 37 – Ratio between maximum and minimum conditional variance of residual groups for average waiting time regression – scenario 2.

Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
3.04	1.72	1.85	1.38	1.3	1.72

Although all models have low ratio between the conditional residual variance calculated, as presented on Table 37, the graphical analysis suggests the presence heteroscedasticity. Independent of model form, the models presented have similar behavior in the residual's variance. One of the reasons could be because the majority of the scenarios have low rejection rates, with 75% of output data bellow 0.31 rejection rate, and rejections have an impact on the average waiting time, this imbalance in the data might cause the residuals variance behavior for the waiting time models.

For the case of heteroscedasticity, the standard errors calculated in the regressions presented are not correct, therefore, heteroscedasticity-robust standard errors are calculated and new statistical tests are performed to access the validity of the models (Stock & Watson, 2015). Table 38 presents the models and the respective lower significance level for p-values of the coefficient estimates with the updated standard errors.

Table 38 - Heteroscedasticity-robust standard errors for waiting time regression models

Model Number	F-statistic(p-values on joint Hypotheses)	Lower Coef. Signif.
1	< 2.2e-16	*
2	< 2.2e-16	***
3	< 2.2e-16	***
4	< 2.2e-16	***
5	< 2.2e-16	*
6	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The hypothesis testing is valid after calculating heteroscedasticity-robust standard errors. One implication of heteroskedasticity in the error terms is that OLS is no longer BLUE - Best Linear Unbiased Estimator, and under certain conditions, other regression estimators can be more efficient than OLS (Stock & Watson, 2015). For all the models presented, the degree of variance explained are higher than 91% for acceptable levels of significance, therefore, for the objectives of this study, this degree of fitting is satisfactory.

Given the results presented, model number 3 is chosen. Equation model parameters are shown in the following table. Although the interaction term for model 1 increases the fitting, there is no evidence for a strong interaction between the variables for the analyzed constrained operation within the dataset.

Table 39 - Equation model for waiting time – scenario 2

Parameter	Estimate	Heteroscedasticity-robust std. Error	t value	Pr(>|t|)
Intercept (β_0)	238.817	18.653	12.80	< 2e-16 ***
$\ln B$ (β_1)	-45.506	1.218	-37.35	< 2e-16 ***
$\ln Q$ (β_2)	31.514	2.292	13.75	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error 17.65 on 82 degrees of freedom
Multiple R-squared: 0.9273, Adjusted R-squared: 0.9255
F-statistic: 731.6 on 2 and 65 DF, p-value: < 2.2e-16

The model selected is able to take into account the non-linear relationship between the waiting time and fleet size and demand.

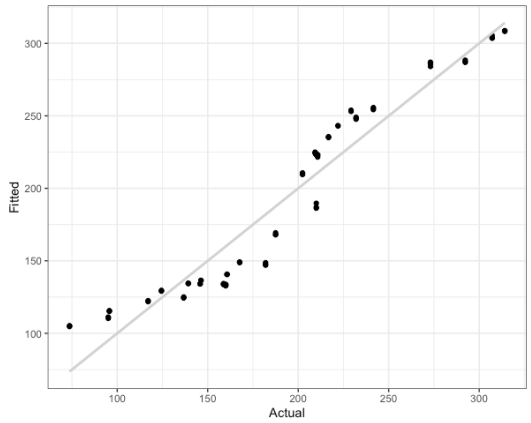

Figure 75 - Fitted vs Actual values for the model selected for average waiting time. Scenario 2.

4.2.4.Rejection Rate

The rejection rate is the ratio between rejected trips and potential demand.

4.2.4.1. Correlation and relationship between variables

Following is a correlation table for the input variables and in-vehicle time.

Table 40 – Correlation between input variables and rejection rate

Variable	**Correlation**
Potential Demand	0.26
Fleet size	-0.60
Vehicle capacity	0.02

According to Table 40, the variable that has a higher linear correlation with the rejection rate is the fleet size, with a correlation of -0.60.

The negative relationship with fleet size is expected, since a low level of supply is expected to increase rejections and vice-versa.

Vehicle capacity has almost null correlation with rejections, depicting the importance of demand and fleet size for this variable. Figure 76 contains the scatter plots of rejection rate against the independent variables.

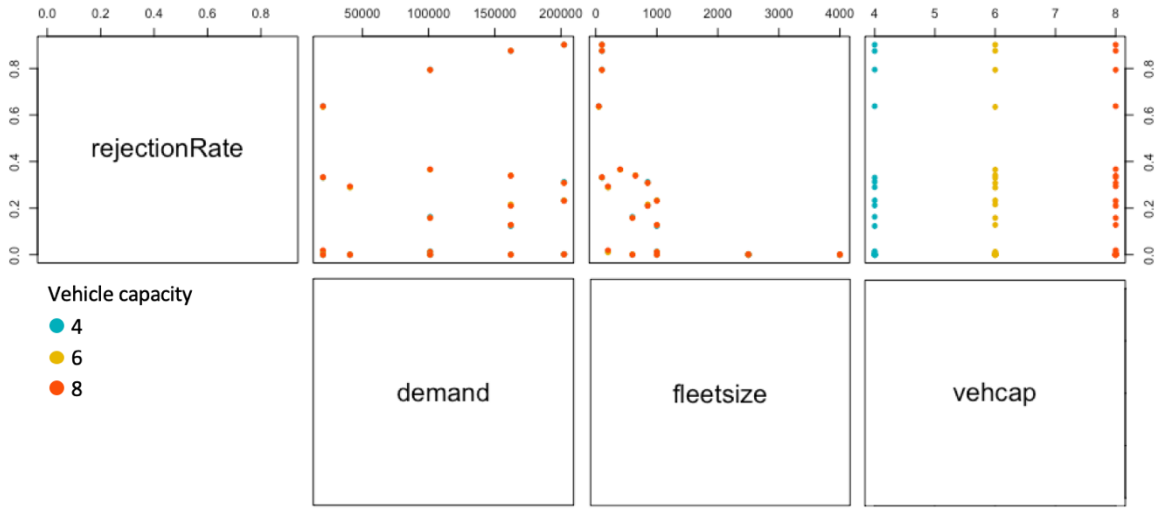

Figure 76 – Rejection rate x independent variables, scenario 2.

Figure 77 and Figure 78 show the relationship between rejection rate and vehicle capacity and fleet size. As for the other variables in scenario 2, vehicle capacity doesn't influence the rejection rate and fleet size and demand are the most important variables.

The fleet size relationship with rejection rate data suggests a non-linear relationship with an asymptotic behavior towards the limit values of 0 and 1. As the number of vehicles increase, the rate of decrease in the rejection rate due to the increase in fleet size diminishes.

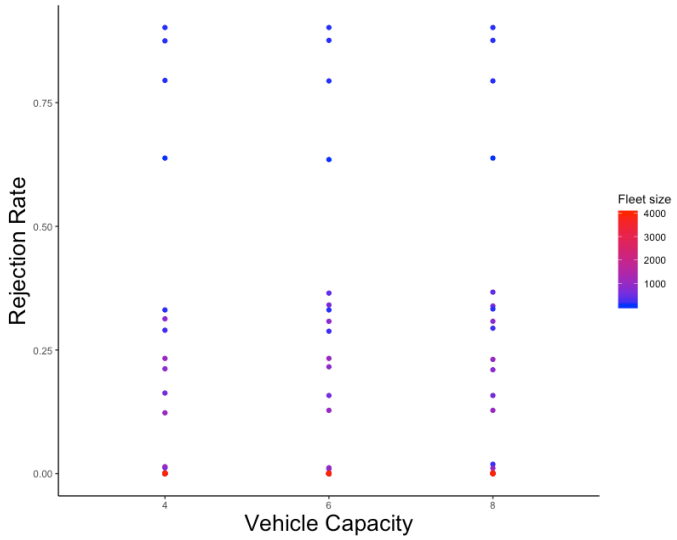

Figure 77 – Rejection Rate according vehicle capacity. Scenario 2.

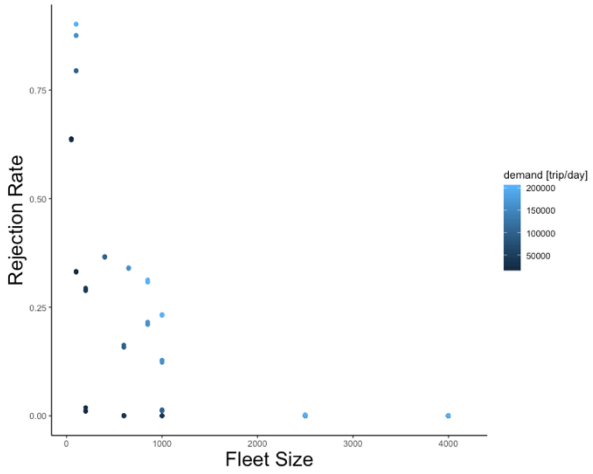

Figure 78 – Rejection Rate according to fleet size, scenario 2.

Potential demand has a positive correlation with rejection rate, however, the data suggests a non-linear relationship depending on the ratio between demand and supply, as observed in Figure 79. For lower ratios (higher fleet sizes), the influence of demand is almost null and increases as the ratio diminishes. For higher ratios (lower supply), the rejections are higher, and demand have a stronger influence on it.

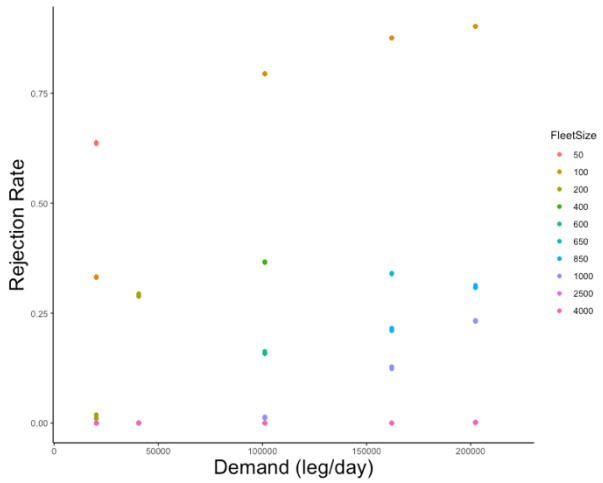

Figure 79 – Rejection Rate according number of trips colored by fleet size. Scenario 2.

4.2.4.2. Multiple Regression

Different types of non-linear relationships are analyzed and used as regressors for the rejection rate function in order to account for the non-linear behavior observed.

The following Figure 80, Figure 81 and Figure 82 show the correlation of average rejection with transformed variables, this exploratory measure helps to identify better the type of non-linear relationship that accounts for the curvilinear existing relationship and to eliminate the nonlinearity.

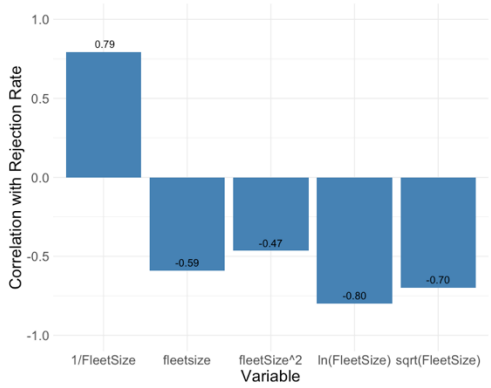

Figure 80 – Correlation between rejection rate and non-linear functions of fleet size – scenario 2.

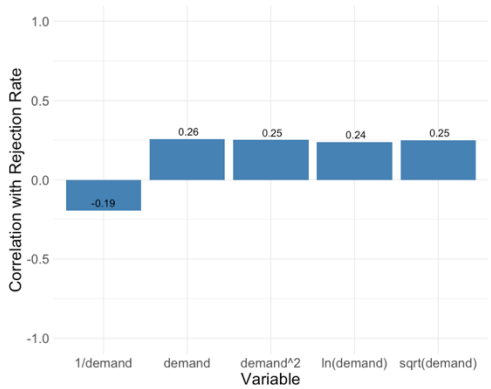

Figure 81 – Correlation between rejection rate and non-linear functions of demand – scenario 2.

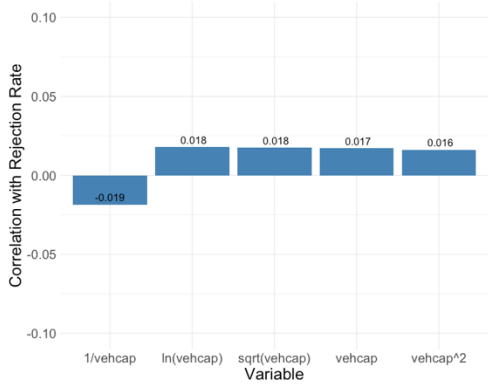

Figure 82 – Correlation between rejection rate and non-linear functions of vehicle capacity – scenario 2.

Vehicle capacity remains with almost null correlation and there is no significant change on the correlation value for demand. For fleet size there is significant increase in correlation for logarithmic and reciprocal functions.

Based on the previous exploratory data analysis and theoretical explanations, multivariate regression models were performed and are summarized on Table 41. For

modeling demand the potential demand was divided by the hours in the day, therefore potential demand used for the models are in rides/h.

Table 41 – Best fitted models average rejection rate regression – scenario 2.

Number	Model Form	R^2	Adjusted R^2	Residual SE	F-statistic(p-values on joint Hypotheses)	Lower Coef. Signif.
1	$-\beta_0 - \beta_1 * \frac{1}{B} + \beta_2 * \ln Q + \beta_3 * \frac{1}{B} * \ln B$	0.939	0.936	0.071	< 2.2e-16	***
2	$\beta_0 + \beta_1 * \frac{1}{B} - \beta_2 * \frac{1}{Q} - \beta_3 * \frac{1}{B} * \frac{1}{Q}$	0.934	0.932	0.073	< 2.2e-16	**
3	$-\beta_0 + \beta_1 * \frac{1}{B} + \beta_2 * \sqrt{Q} + \beta_3 * \frac{1}{B} * \sqrt{Q}$	0.931	0.929	0.075	< 2.2e-16	*
4	$-\beta_0 + \beta_1 * \frac{1}{B} + \beta_2 * Q - \beta_3 * Q^2 + \beta_4 * \frac{1}{B} * Q^2$	0.923	0.919	0.080	< 2.2e-16	*
5	$-\beta_0 + \beta_1 * \frac{1}{B} + \beta_2 * Q + \beta_3 * \frac{1}{B} * Q$	0.917	0.913	0.082	< 2.2e-16	*
6	$\beta_0 - \beta_1 * \ln B + \beta_2 * Q - \beta_3 * \ln B * Q$	0.859	0.854	0.107	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

All models provide a good fit with satisfactory R^2, with model 1 being able to explain 94% of data variance.

The normality of residuals assumption is analyzed graphically and shown in Figure 83, all models follows the normal distribution curve with some small deviations, higher for model 4, 5 and 6. From statistical tests of Table 42, all models can't reject the null hypothesis of normality at 0.05 significance level from Kolmogorov-Smirnov test, however the results from the Shapiro-Wilk test allows this null hypotheses to be rejected for all models but 6. The sample size and distribution of values is one of the reasons to this discrepancy between both test results and graphical analysis.

Table 42 – Tests for OLS regression residuals normality – rejection rate (scenario2)

Number	Kolmogorov-Smirnov test (p-value)	Shapiro-Wilk normality test (p-value)
1	0.5462	0.0008
2	0.0439	0.0001
3	0.4334	0.0102
4	0.3384	0.0005
5	0.1106	0.0054
6	0.8494	0.2668

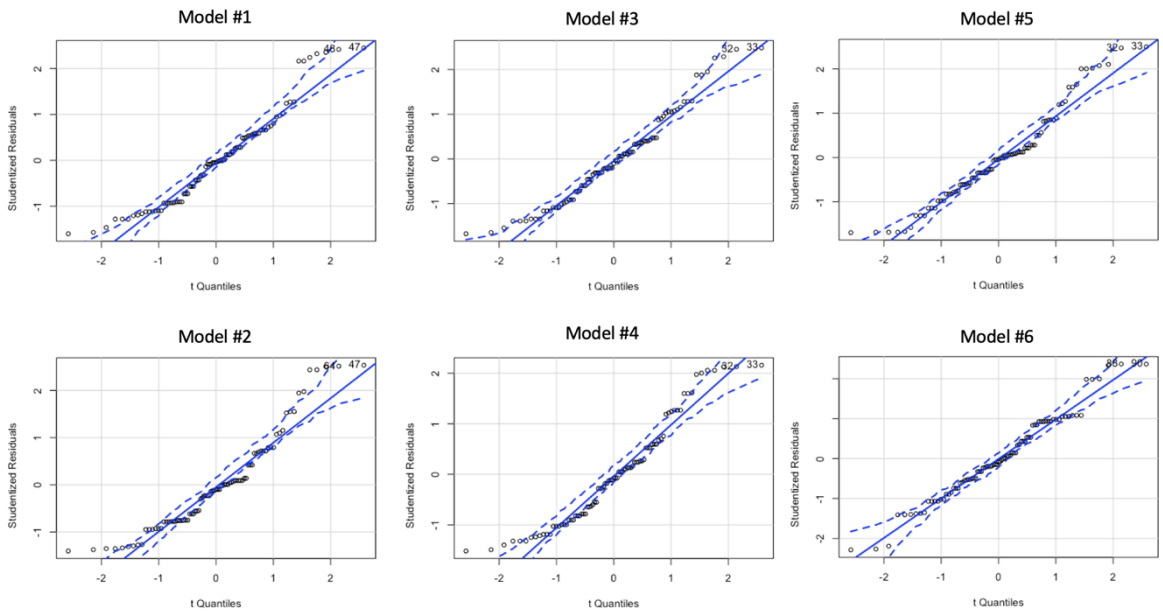

Figure 83 – Normal Q-Q plots for residuals of regression models for rejection rate including 0.95 confidence level– scenario 2.

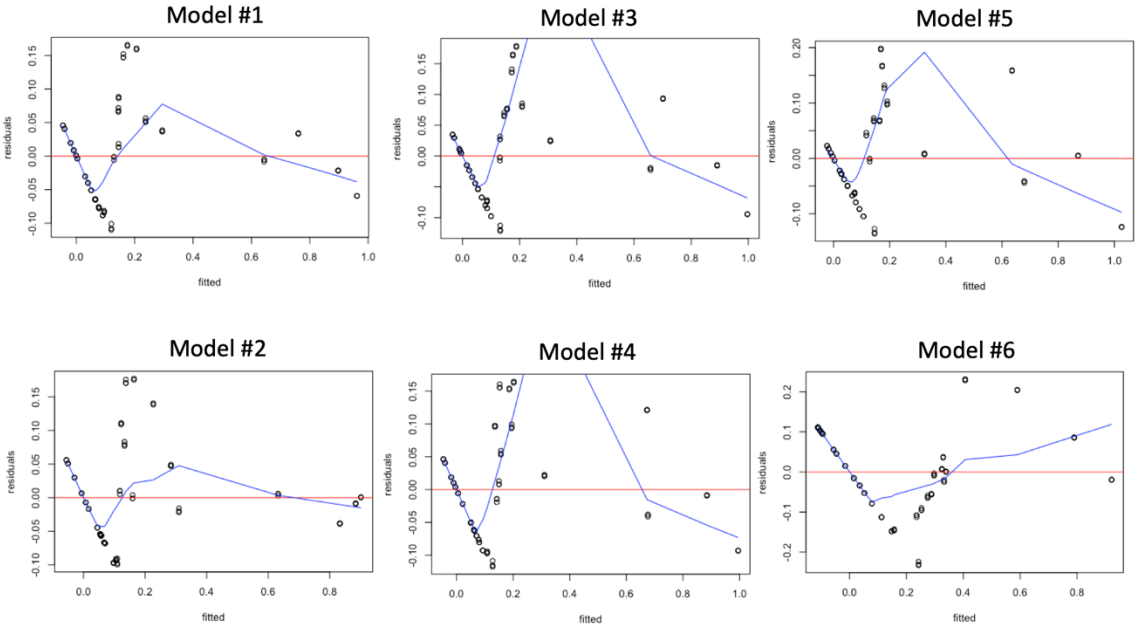

Figure 84 – Residual plots against fitted values for regression models of rejection rate – scenario 2.

Residuals variance independence is also analyzed graphically in Figure 84. The imbalance in the dataset is responsible for the concentration of residuals for low values of rejection rate. To analyze better the variance of residuals, the residuals of each model are ordered according to their values and divided in sets of 3 groups, with approximately equal number of elements, the conditional variance of each group is calculated and the ratio to the largest to the smallest variance of groups is analyzed (Cohen et al., 2003), results are in the following table.

Table 43 – Ratio between maximum and minimum conditional variance of residual groups for average waiting time regression – scenario 2.

Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
8.14	12.27	9.1	8.35	13.3	5.91

Therefore, the models present a low difference in the conditional variance even with the concentration of low values, with only model 2 and model 5 with ratio values above 10.0.

Given the results presented, model number 1 is chosen, equation model parameters are shown in the following table.

Table 44 - Equation model for rejection rate – scenario 2

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
Intercept (β_0)	-0.37208	0.08557	-4.348	3.96e-05 ***
$\frac{1}{B}$ (β_1)	-128.29865	13.34802	-9.612	4.92e-15 ***
$\ln Q$ (β_2)	0.04715	0.01061	4.445	2.77e-05 ***
$\frac{1}{B} * \ln Q$ (β_3)	24.22909	1.76686	13.713	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.07067 on 81 degrees of freedom
Multiple R-squared: 0.9386, Adjusted R-squared: 0.9363
F-statistic: 412.8 on 3 and 81 DF, p-value: < 2.2e-16

The model selected is able to take into account the non-linear relationship between the rejection rate and fleet size and potential demand. The interaction term included represents the system state regarding supply and demand, which influences directly the operation in system that allows request rejections. This interaction term moderates the curvilinearity of the relationships between rejection rate and independent variables. The asymptotic property of the reciprocal function used for the fleet size parameter and the logarithmic function for demand also matches the expected behavior observed in the simulation output. The coefficient signals are also in line with the expected results.

4.3. Total cost functions

The variables modelled on items 4.1 and 4.2 are used in the total cost functions for the scenarios analyzed. The equations for the variables and the total cost functions are presented below. The objective functions for the optimization problem are defined.

4.3.1. Scenario 1 – Operation with no request rejections

The distance detour factor equation for this scenario is defined as follows:

$$\rho(B, C, Q) = \beta_0 + \beta_1 * C - \beta_2 * \ln B - \beta_3 * \frac{1}{Q} \quad (15)$$

Average occupancy:

$$O_B(B, C, Q) = \beta_0 + \beta_1 * C - \beta_2 * \ln B + \beta_3 * Q - \beta_4 * Q^2 + \beta_5 * C * Q \quad (16)$$

By inserting (15) and (16) on (3):

$$D_B = \frac{Q * l * (\beta_0 + \beta_1 * C - \beta_2 * \ln B - \beta_3 * \frac{1}{Q})}{\beta_0 + \beta_1 * C - \beta_2 * \ln B + \beta_3 * Q - \beta_4 * Q^2 + \beta_5 * C * Q}$$

Average waiting time:

$$t_w(B, C, Q) = -\beta_0 + \beta_1 * \ln Q - \beta_2 * \ln B + \beta_3 * C - \beta_4 * \ln Q * C \quad (17)$$

By inserting (15) on (8):

$$t_v(Q, B, C) = \frac{Q * l * (\beta_0 + \beta_1 * C - \beta_2 * \ln B - \beta_3 * \frac{1}{Q})}{v_m}$$

Therefore, operator and user cost are defined as follows:

$$C_o = c_0^{rs} * B + c_1^{rs} * \left[Q * l \left(\frac{\beta_0 + \beta_1 * C - \beta_2 * \ln B - \beta_3 * \frac{1}{Q}}{\beta_0 + \beta_1 * C - \beta_2 * \ln B + \beta_3 * Q - \beta_4 * Q^2 + \beta_5 * C * Q} \right) \right]$$

$$C_u = P_w * Q * (-\beta_0 + \beta_1 * \ln Q - \beta_2 * \ln B + \beta_3 * C - \beta_4 * \ln Q * C) + P_v * \left(\frac{Q * l * (\beta_0 + \beta_1 * C - \beta_2 * \ln B - \beta_3 * \frac{1}{Q})}{v_m} \right)$$

In order to find the optimum fleet size for given demands and vehicle capacities that minimizes the total cost, the optimization problem is:

$$Min\ C_t = c_0^{rs} * B + c_1^{rs} * \left(\frac{Q * l * \left(\beta_0 + \beta_1 * C - \beta_2 * \ln B - \beta_3 * \frac{1}{Q}\right)}{-\beta_0 + \beta_1 * C - \beta_2 * \ln B + \beta_3 * \ln Q} \right) + P_w * (-\beta_0 + \beta_1 * \ln Q - \beta_2 * \ln B + \beta_3 * C - \beta_4 * \ln Q * C) + P_v * \left(\frac{Q * l * \left(\beta_0 + \beta_1 * C - \beta_2 * \ln B - \beta_3 * \frac{1}{Q}\right)}{v_m} \right)$$

-Subject to: $100 \leq B \leq 4000$

and $4 \leq C \leq 8$

Which must be solved numerically for the different levels of demand and vehicle capacity. The upper and low bounds for the fleet size value are set in order to not go beyond the range of the data used for the variable's multiple regression modeling.

4.3.2.Scenario 2 - Operation with rejections

The distance detour factor equation for this scenario is defined as follows:

$$\rho(B, Q) = \beta_0 + \beta_1 * \ln B + \beta_2 * \frac{1}{Q} - \beta_3 * \ln B * \frac{1}{Q} \quad (18)$$

Average occupancy:

$$O_B(B,Q) = \beta_0 + \beta_1 * \ln B + \beta_2 * \ln Q \quad (19)$$

Average waiting time:

$$t_w(B,Q) = \beta_0 - \beta_1 * \ln B + \beta_2 * \ln Q \quad (20)$$

Rejection rate:

$$r(B,Q) = -\beta_0 - \beta_1 * \frac{1}{B} + \beta_2 * \ln Q + \beta_3 * \frac{1}{B} * \ln B \quad (21)$$

By inserting (18), (19) and (21) on (9) and (3):

$$D_B = \left(1 - \beta_0 - \beta_1 * \frac{1}{B} + \beta_2 * \ln Q + \beta_3 * \frac{1}{B} * \ln B\right) * Q * l * \left(\frac{\beta_0 + \beta_1 * \ln B + \beta_2 * \frac{1}{Q} - \beta_3 * \ln B * \frac{1}{Q}}{\beta_0 + \beta_1 * \ln B + \beta_2 * \ln Q}\right)$$

By inserting (18) and (21) on (8):

$$t_v(Q,B,C) = \left(1 - \beta_0 - \beta_1 * \frac{1}{B} + \beta_2 * \ln Q + \beta_3 * \frac{1}{B} * \ln B\right) * Q * l \left(\frac{(\beta_0 + \beta_1 * \ln B + \beta_2 * \frac{1}{Q} - \beta_3 * \ln B * \frac{1}{Q})}{v_m}\right)$$

Therefore, operator and user cost are defined as follows:

$$C_o = c_0^{rs} * B + c_1^{rs} * \left(1 - \beta_0 - \beta_1 * \frac{1}{B} + \beta_2 * \ln Q + \beta_3 * \frac{1}{B} * \ln B\right) * Q * l * \left(\frac{\beta_0 + \beta_1 * \ln B + \beta_2 * \frac{1}{Q} - \beta_3 * \ln B * \frac{1}{Q}}{\beta_0 + \beta_1 * \ln B + \beta_2 * \ln Q}\right)$$

$$C_u = P_w * \left(1 - \beta_0 - \beta_1 * \frac{1}{B} + \beta_2 * \ln Q + \beta_3 * \frac{1}{B} * \ln B\right) * Q * (\beta_0 - \beta_1 * \ln B + \beta_2 * \ln Q) + P_v$$
$$* \left(1 - \beta_0 - \beta_1 * \frac{1}{B} + \beta_2 * \ln Q + \beta_3 * \frac{1}{B} * \ln B\right) * Q * l \left(\frac{(\beta_0 + \beta_1 * \ln B + \beta_2 * \frac{1}{Q} - \beta_3 * \ln B * \frac{1}{Q})}{v_m}\right)$$

Accordingly, for the constrained operational scenario, the minimization problem is:

$$Min\ C_t = c_0^{rs} * B + c_1^{rs} * \left(1 - \beta_0 - \beta_1 * \frac{1}{B} + \beta_2 * \ln Q + \beta_3 * \frac{1}{B} * \ln B\right) * Q * l * \left(\frac{\beta_0 + \beta_1 * \ln B + \beta_2 * \frac{1}{Q} - \beta_3 * \ln B * \frac{1}{Q}}{\beta_0 + \beta_1 * \ln B + \beta_2 * \ln Q}\right) + P_w$$
$$* \left(1 - \beta_0 - \beta_1 * \frac{1}{B} + \beta_2 * \ln Q + \beta_3 * \frac{1}{B} * \ln B\right) * Q * (\beta_0 - \beta_1 * \ln B + \beta_2 * \ln Q) + P_v$$
$$* \left(1 - \beta_0 - \beta_1 * \frac{1}{B} + \beta_2 * \ln Q + \beta_3 * \frac{1}{B} * \ln B\right) * Q * l \left(\frac{(\beta_0 + \beta_1 * \ln B + \beta_2 * \frac{1}{Q} - \beta_3 * \ln B * \frac{1}{Q})}{v_m}\right)$$

-Subject to: $100 \leq B \leq 4000$

and $4 \leq C \leq 8$

The total cost minimization problems are going to be solved numerically and results are shown on chapter 5. To solve the optimization, a similar approach to Tirachini et al. (2010a) is performed. The algorithm implemented in Matlab solves the problem assuming a real number for the fleet size, then the solution for the value is rounded to the upper and lower

closest integer. The problem is solved again using the two rounded values for the lowest total cost.

4.4. Modeling of Automation Effect

The total cost models are analyzed adapted with the introduction of vehicle automation. The assumptions of Tirachini and Antoniou (2020) on the operator cost are used:

- A reduction in operating cost due to not having to pay the drivers or at least a fraction of them;
- An increase in capital cost due to the inclusion of automation technology.

The operator cost defined on equation (2) has two types of costs, one associated to the fleet operation per hour and other associated to the number of kilometers operated, the former accounts for driver salaries and vehicle capital cost. It is possible to model the vehicle cost as a linear function of vehicle capacity (Tirachini & Hensher, 2011). Therefore:

$$c_0^{rs} = c_0^{rs-h} + c_1^{rs-h} * K$$

and

$$c_1^{rs} = c_0^{rs-km} + c_1^{rs-km} * K$$

Based on the assumptions stated, the effect of automation will rely on c_0^{rs}, while there will be no change for c_1^{rs}, therefore operator cost for automated vehicles:

$$c_0^{rs-aut} = c_0^{rs-h-aut} + c_1^{rs-h-aut} * K \ = \alpha * c_0^{rs-h} + \beta * c_1^{rs-h} * K$$

Where $\alpha = \frac{c_0^{rs-h-aut}}{c_0^{rs-h}}$ is the relative change in the fixed parameter of the hourly unit operator cost and $\beta = \frac{c_1^{rs-h-aut}}{c_1^{rs-h}}$ is the relative change in the marginal cost related to vehicle capacity, per vehicle-hour. Following the assumptions of Tirachini and Antoniou (2020), in which is plausible to assume $\beta > 1$ if the marginal cost due to vehicle capacity increases.

The effect of speed is also an important point, it is not clear yet if automated vehicles, both public and private will have increased speeds or decreased. However, the models developed in this study were based on simulations that used a speed of 30 km/h for the vehicles. Therefore, the effect of automation on speed is not going to be analyzed and the speed of 30 km/h is going to be used in the following items.

This study also uses the factor δ as the percentage of current human driving cost that is still required under automation (Tirachini & Antoniou, 2019), in which α is a linear function of:

$$\alpha = \alpha_0 + \alpha_1 * \delta \qquad (22)$$

The comparison between the modelled scenarios under the human driven and automated vehicles is analyzed in the numerical application and optimization solution performed in Chapter 5.

Chapter 5

Numerical Application and Results

5.1. Estimation of parameters

The model is applied using input data parameters from the city of Munich, which was the area of study used for the simulations performed. The data used is based on that used in Tirachini and Antoniou (2020).

For the fleet-based operator costs [€/veh-hour], vehicle capital cost, driver cost, and charging infrastructure cost were included. For the spatial based running costs (€/km-h), the costs of energy consumption and maintenance were included. The fleet of vehicles is considered to be electric for all scenarios.

The energy cost per kilometer is calculated by multiplying the energy consumption and cost of energy, and the maintenance cost per kilometer is calculated by using the original value in per hour from Tirachini and Antoniou (2020) and dividing it by the average speed. As previously explained, the average speed of 30 km/h is used for all cases. Table 45 presents the cost parameters used.

Table 45 – Operator cost parameters, Munich. Source: Tirachini and Antoniou (2020)

Parameter	Unit	Car	Van	Mini bus
Vehicle capacity	[pax/veh]	5	8	44
Vehicle capital cost	[€/veh-hour]	1.4	2.1	7.7
Driver cost	[€/veh-hour]	15.3	15.3	15.3
Charging infrastructure cost	[€/km-hour]	0.8	1.0	1.5
Energy consumption	[KWh/km]	0.14	0.15	0.64
Cost of energy	[€/kwh]	0.23	0.23	0.23
Energy cost	[€/km]	0.03	0.03	0.15
Vehicle maintenance cost	[€/km]	0.03	0.03	0.05
Average speed	[km/h]	30.0	30.0	30.0
Total cost human-driven vehicle (veh)	[€/veh-hour]	17.5	18.4	24.50
Total cost human-driven vehicle (km)	[€/km]	0.06	0.07	0.20
Total cost automated vehicle (veh), $\delta = 0$	[€/veh-hour]	3.0	4.3	12.0
Total cost automated vehicle (veh), $\delta = 0.5$	[€/veh-hour]	10.65	11.95	19.70

The values of the parameters c_0^{rs} and c_1^{rs} are estimated from linear regressions (Figure 85 and Figure 86) using the operator costs from Table 45.

Table 46 – Estimation of operator cost parameters

Parameter	Unit	Value	Automated ($\delta = 0$)	Automated ($\delta = 0.5$)
c_0^{rs-h}	[€/veh-hour]	16.80	2.17	9.82
c_1^{rs-h}	[€/veh-hour]	0.17	0.22	0.22
c_0^{rs-km}	[€/km-hour]	3.96E-02	-	-
c_1^{rs-km}	[€/km-hour]	3.66E-03	-	-
α_0	-	-	0.13	0.13
α_1	-	-	-	0.91
β	-	-	1.29	1.29

The values of α and β are estimated from the ratio between parameters c_0^{rs-h} and c_1^{rs-h}. The parameters α_0 and α_1 are estimated using (22).

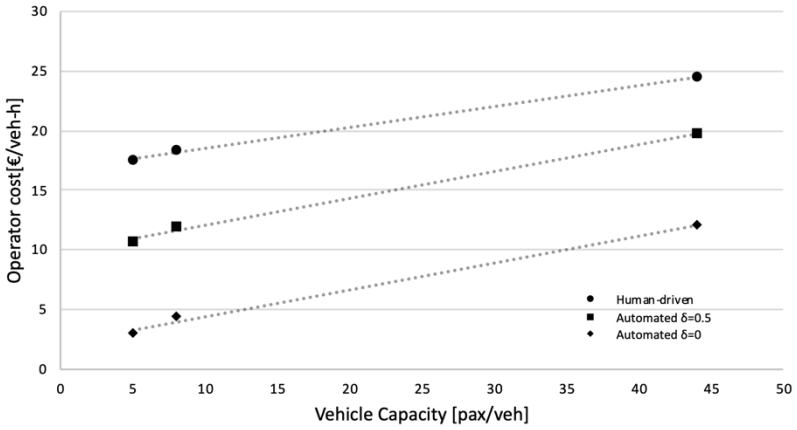

Figure 85 – Hourly operator cost of human driven and automated vehicles

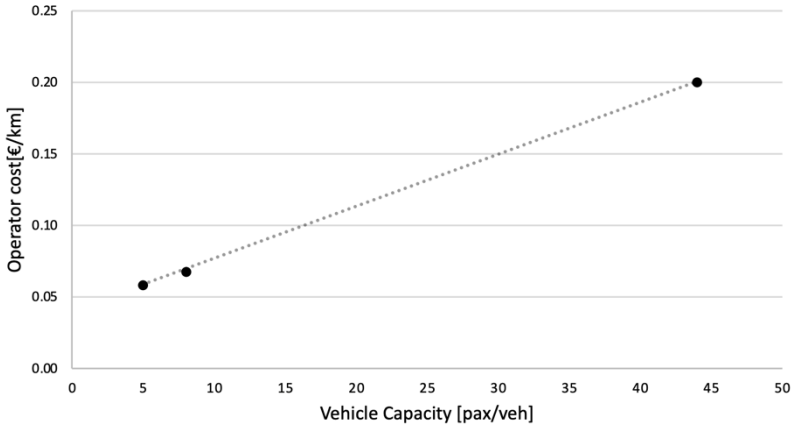

Figure 86 – Spatial basis operator cost

For all cases shown in Table 46, the value of R^2 is above 0.996.

5.2. Numerical application

The models are solved and analyzed using values to for the city of Munich. The two operational scenarios developed in this thesis are used for three different vehicles capacities, 4, 6 and 8. Two alternative scenarios of automated vehicles are also analyzed for each operational scheme.

The first alternative scenario for vehicle automation assumes full cost savings on all human driving cost due to vehicle automation ($\delta = 0$). The second alternative scenario assumes automation driving cost savings of 50% of human-driven vehicles ($\delta = 0.5$).

The values of waiting time and in-vehicle time savings are from (Tirachini & Antoniou, 2020).

User cost, trip and speed parameters used are shown in Table 47. Values for savings in waiting time and in-vehicle time are obtained from Tirachini and Antoniou (2020).

Table 47 – Parameters for model application

Parameter	Unit	Value
Value of waiting time savings P_w	[€/hour]	11.4
Value of in-vehicle time savings P_v	[€/hour]	5.2
Average speed	[km/hour]	30.0
Average trip distance	[km]	7.0

The average direct trip distance used for the city of Munich is 7.0 km (Buehler et al., 2016) and the average speed used is consistent with the speed used for the simulations performed. The evaluated demand varies from 800 pax/h to 8,500 pax/h, reflecting the low and high levels of demand used for the simulations performed.

5.2.1. Scenario 1 – Operation with no request rejections

The optimal fleet size and resulting average costs for users, operator and total (€/trip) for the scenarios without rejections are shown in Figure 87.

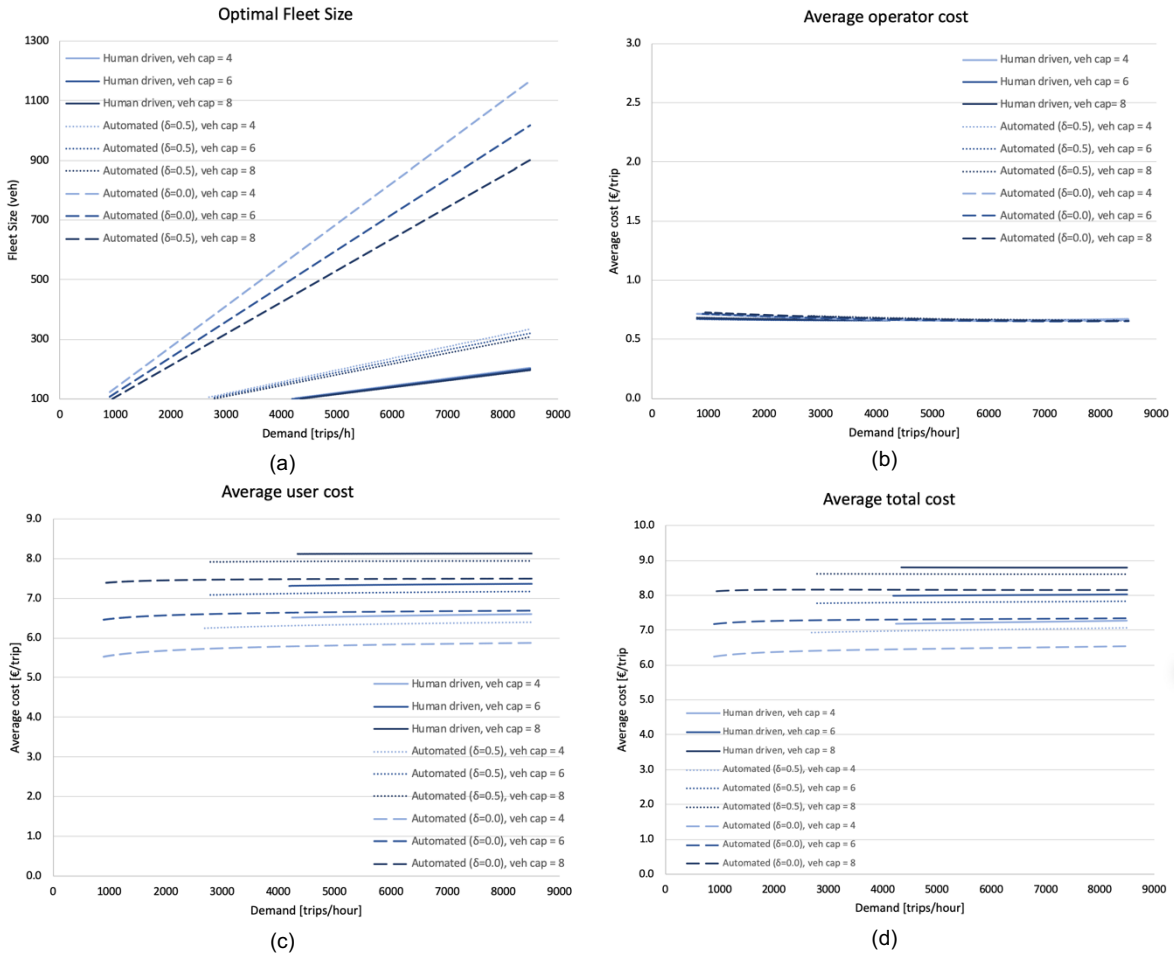

Figure 87 – Optimal fleet size and average costs – scenario 1. (a) Optimal fleet size. (b) Average operator cost. (c) Average user cost. (d) Average total cost.

The higher values for the optimal fleet sizes in the automated scenario are expected due to the reduction in driver's salary, being cheaper to provide a higher fleet with increasing demand. It is possible to observe that although the optimal fleet size increases with the demand, all the average costs don't decrease with patronage and remain almost constant, with a low rate of change. Regarding user costs, it depends on average waiting time and average in-vehicle time, the latter depending on the detour factor. As shown before, the fleet size has a negative relationship with the detour factor, while demand influences it positively, as shown before on 4.1.1. However, the main factor that influences detour for a system with no rejections is the vehicle capacity. This behavior is shown in Figure 88 (a), where the resulting detour factor remains almost constant for the increase in the demand and in the optimal fleet size. It is also possible to see that for all scenarios, lower vehicle capacities have lower average user costs due to the influence of vehicle capacity. For the no-rejection scheme, the demand has a higher influence than fleet size on waiting time; therefore, the effect of waiting time reduction increase in the optimal fleet size is hindered by the increase in the demand, Figure 88(b) shows the resulting average waiting time for the optimal fleet sizes and respective demand, leading to very small increases on the user cost. The high values of user

cost reflect a property of the no-rejection system, in which the higher degree of shareability increases the detour factor.

The average operator cost has similar values for all scenarios, with an insignificant difference. In contrast with user costs, it presents low values bellow 1.0 €/trip, which is mostly due to the high occupancy rates for this system operations scheme. Operator cost depends basically on the fleet cost and on the ratio between detour factor and average occupancy. While the marginal cost fraction related to fleet costs reduces, the marginal cost fraction related to the fleet mileage increases at the same rate.

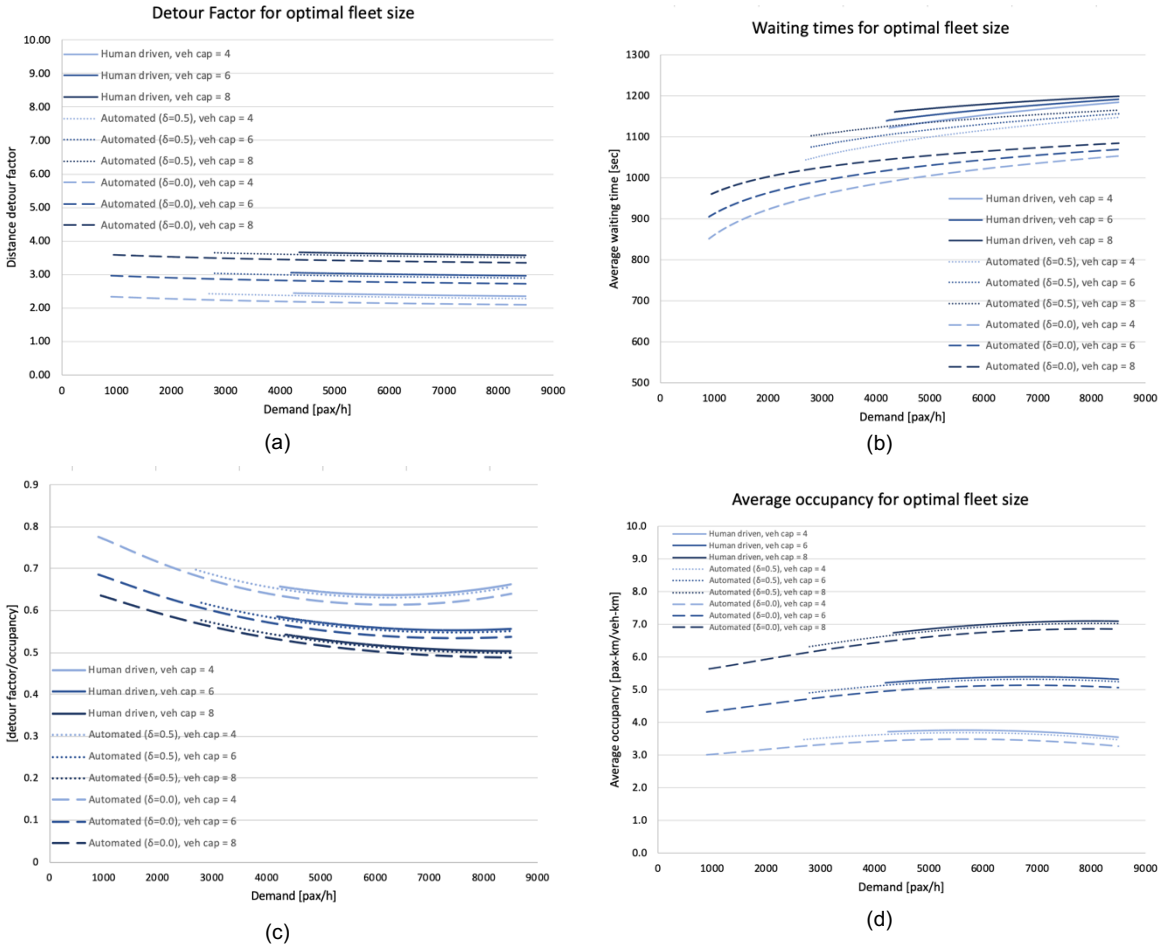

Figure 88 – System variables for the optimal fleet size. (a) Detour factor for optimal fleet size. (b) Average waiting time for optimal fleet size. (c) Ratio between detour factor and average occupancy for optimal fleet size. (d) Average occupancy for optimal fleet size.

Moreover, the average total costs follow the trend of operator and user costs and are almost constant. The values in Figure 87(d) range from 6.2 to 8.8 €/trip. When comparing these values to values obtained for other modes in other studies, it is possible to see that this system is not competitive in terms of pricing. The total costs for the optimal frequency of automated public transport estimated on Tirachini and Antoniou (2020) falls below 4.0 €/trip to all scenarios analyzed for demands higher than 1,000 trips/hour, reaching even lower values. If translated to cost in per passenger-km, the best values for full automated vehicles of capacity 4 in the analyzed system are about seven times higher than the costs of travel for

pooled automated electric vehicle estimated on Sperling (2018) and about two times the cost for autonomous taxis without pooling estimated on Bösch et al. (2018)

The numerical application and scenarios analyzed show that the provision of such system presents no economies of scale. The main reason relies mostly on the high values for the user cost and the constant marginal costs for the operator.

Such high values are a characteristic of the no-rejection framework, in which every request has to be served as if it was acting as public transport. In this case, for a door-to-door service, the detour factor is too high. Therefore this system turns not being competitive because of too long waiting and in-vehicle times. Another aspect of long detours to take into account in the acceptance of this systems is that with higher occupancies and longer detours, there are more strangers in the same intimate environment of the car, and this might also be an obstacle for such systems success (Bösch et al., 2018).

5.2.2.Scenario 2 – Operation with rejections

The optimal fleet size and resulting average costs for users, operator and total (€/trip) for the scenarios with rejections are shown in Figure 89.

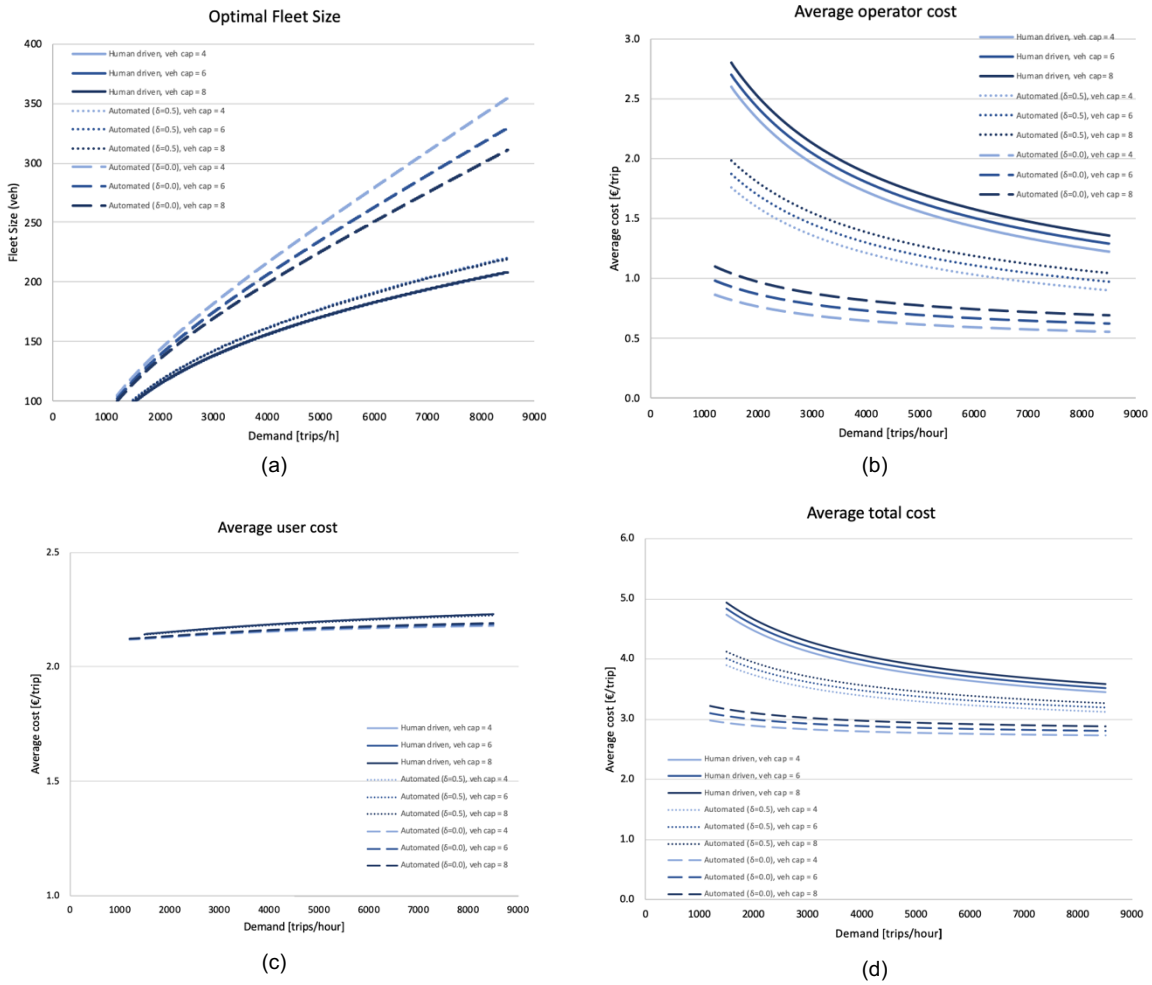

Figure 89 – Optimal fleet size and average costs – scenario 2. (a) Optimal fleet size. (b) Average operator cost. (c) Average user cost. (d) Average total cost.

The user cost remains almost constant while there is a decrease in the average operator cost, leading to a decrease in average total cost with patronage. It shows that, under this framework, the system presents economies of scale.

There is a major overall reduction in the average user costs when comparing to scenario 1. This depicts the importance of the operational strategies, given that for a number of scenarios simulated, the rejection rate was zero, even though constraints were used. The increase in demand increases the waiting time, therefore, there is a small increase in user cost, and the increased fleet size is not enough to offset this effect. Regarding the user cost related to in-vehicle time, the detour factor remains constant for the optimal fleet size values, similar to scenario 1 but with lower values.

There is a reduction in operator costs as demand increases. This happens because in this case both the marginal costs related to the fleet capital costs and the marginal costs related to the fleet mileage reduces with patronage.

It can also be seen also that the resulting average occupancy increase, this relationship was described on item 4.2.2, and leads to a decrease in the ratio between detour factor and occupancy, Figure 90 depicts this behavior. It is also possible to note that the average operator costs are lower for a lower vehicle capacity, being the fleets of seat capacity 4 the ones with the lowest costs. For this scenario, the vehicle capacity doesn't influence user costs, as described on the regression models on Chapter 4., contrasting with the opposite relationship shown for scenario 1.

There is almost no impact of automation on user costs. Average waiting times present a small reduction with automation for full cost savings ($\delta = 0.0$) , the effect for half cost savings ($\delta = 0.5$) are not substantial. For a demand of 5,000 trips/hour, for example, with four-seat capacity vehicles, the average waiting time is reduced from 273 sec to 256 sec, a reduction of 6.3%. This lower change in average waiting times compared to scenario 1 is mostly due to the constrained operations depending on it.

Regarding operator costs, there is a major reduction due to vehicle automation, for 5,000 trips/hour and four-seat vehicle capacity, the average operator cost is reduced from 1.6 €/trip to 1.1 €/trip for $\delta = 0.5$ and to 0.6 €/trip for $\delta = 0.0$, a reduction of 29% and 61%, respectively. This is happening not only because of the lower fleet size cost, but also due to the decrease in the rejection rate. The reductions aforementioned are respective to 3.72% and 23.0% reductions in the rejection rate. Figure 90 depicts this behavior.

The decrease in operator costs due to automation is responsible for the decrease in the average total cost. From Figure 89 it is possible to see that the full automation effect is able to bring this service provision to competitive levels in terms of pricing, for a human-driven vehicle with six-seat vehicle in a scenario of 5,000 trips/hour, the average total cost is 3.8 €/trip, for the same six-seat vehicle and demand in the fully automated scenario the cost is 2.8

€/trip, a 26% reduction. By comparing these values with the values for automated public transport estimated on Tirachini and Antoniou (2020), they are competitive if both systems are automated and have full cost savings ($\delta = 0.0$). For values with no change in running time of public transport with automation, automated public transport appears to be more competitive, if the running time is doubled for automated transport than the shared DRT service seems to be more competitive. By comparing the most competitive price for four-seat full automated vehicles with values from Sperling (2018) and Bösch et al. (2018), it is still higher but a lower degree when compared to values of scenario 1. The total cost for a demand of 2.000 pax/h for example, is equivalent to a price of 0.6 €/pax-km.

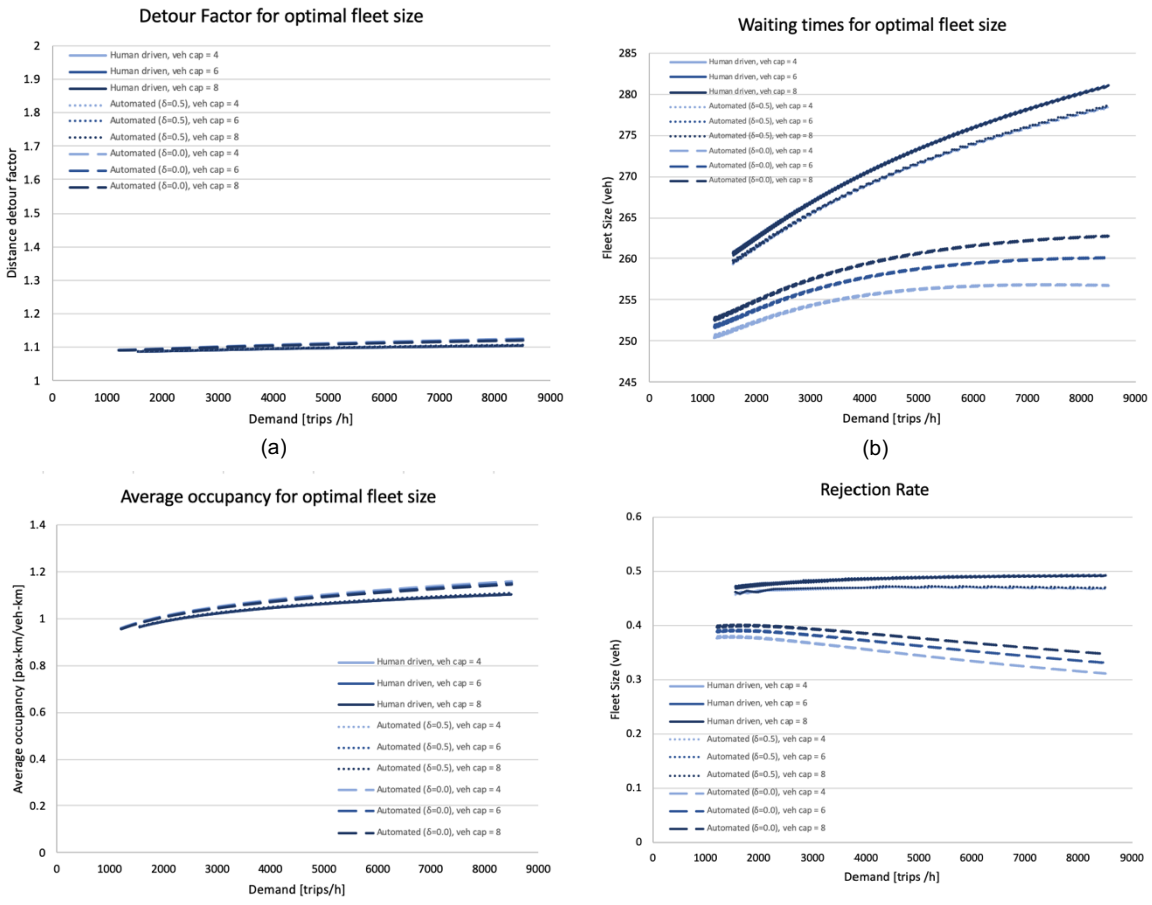

Figure 90 - System variables for the optimal fleet size – scenario 2. (a) Detour factor for optimal fleet size. (b) Average waiting time for optimal fleet size. (c) Ratio between detour factor and average occupancy for optimal fleet size. (d) Average occupancy for optimal fleet size.

Chapter 6

Conclusion

This thesis analyzes the performance of shared DRT systems in two perspectives. An operational scenario without rejections: In the face of the upcoming automation and the raising questionings about substituting fixed-route public transport for on-demand shared service, this type of operation is what a true public transport should offer. And an operational scenario with rejections: Because of the shared DRT system dynamics and inherent costs associated with fulfilling all the trips in a door-to-door service, recent research has been studying this system mostly in frameworks that allow for rejection. These systems are important in order to shed light on the impact of different strategies and rejection levels on the overall system performance.

For that, a total cost minimization model was developed based on using a hybrid approach, in which variables are analytically modeled and derived from multiple regression analyses using output data from agent-based simulations. The model is extended to take into account the effects of vehicle automation.

Total costs are optimized in a numerical application using data from Munich, and the performance of the system is analyzed for the different scenarios. The causal relationship between the system design variables and the economic aspects of the system are explored. The conclusions are summarized next.

6.1. Conclusions

Three research questions were proposed in Chapter 1. The study performed in the thesis proposes the following conclusions.

Question 1*: Can total cost models be formulated with the utilization of analytical and simulation data output for shared DRT services?*

A total cost model for shared DRT services is developed in Chapter 2, and system variables are derived analytically. The variables that depend on the high level of dynamics of a shared DRT system are modeled as functions of the input design variables that have the theoretical causal relationship. Linear regression models were developed using output simulation data using as independent variables fleet size, capacity, and demand. The system output variables modeled were: Distance detour factor, average occupancy, average waiting time, and trip rejection rate. The models have shown a good degree of accuracy and are in line with expected theoretical behavior. After plugging in the functions, the total cost model presented is satisfactory and allows for system performance analysis.

Question 2: *How do the operational scenarios analyzed for shared DRT perform under the minimization of total costs?*

The average total costs for a system without rejections are very high when compared to a system with rejections. This happens because in a no-rejection scenario, all trips are fulfilled with the cost of higher detours, average waiting times and in-vehicle travel time. This leads to higher values of user cost. Although user costs are high, the high levels of occupancy equalize marginal costs to average operator cost at low levels. In a system with rejections, it has been shown that economies of scale are present. The operational strategy that takes into account waiting times and in-vehicle travel times is responsible for maintaining the user costs at low levels while the average operator cost decreases. However, with exception to full automation cost savings scenarios, even with the increase of optimal fleet with demand, the resulting rejection rates don't decrease and are accountable for more than 40% of the trips. Such behavior has to be addressed by researches and planners if such systems are proposed as a substitution for traditional fixed-route public transport. Furthermore, the results also depict the importance of the operational strategy on this type of service; there is a clear trade-off between the shareability of the trips and the user cost. Dispatching and routing strategies play an important role in the system performance.

Question 3: *What is the expected impact of automation on average costs of shared DRT systems for optimal provision?*

It is shown that automation enables the provision of a bigger optimal fleet. It is estimated in Chapter 5 that the average total costs of shared DRT systems reduce. However, the costs become only competitive for the system with rejections. The system without rejections still have higher costs related to the user, and the operator costs are still equalized with marginal costs. Moreover, the results suggest that if automation enables full cost savings on human driving, the average total cost of the shared DRT system becomes competitive. In this scenario, the increase in the optimal fleet size is enough to enable for the marginal reduction of the rejection rate. However, for the best scenarios, its value is still higher than 30%. Although the savings in automation enable a bigger fleet, the impact on waiting times is rather low on the system with rejections, mainly because of the constrained operation. The higher impact of automation on the waiting times of the system without request rejection is not enough to reduce the user costs substantially.

6.2. Limitations and future work

This thesis contributes to the literature in different methodological and practical aspects; however, the limitations of the work performed have to be considered to guide future research better.

Regarding the modeling of the system variables using output data from agent-based simulations, the values obtained for the regression coefficients here are specific to the network, synthetic population trip characteristics, and dispatching and routing strategies simulated.

In terms of modeling, the assumption for the total in-vehicle travel time doesn't take into account time spent onboarding and deboarding of passengers, therefore further research to incorporate such times is recommended.

The performance of the systems analyzed does not take into account the externalities involved, such as congestion effects or emissions, and the demand is inelastic. Therefore, the extension for models that take into account these topics is an interesting topic of research.

The comparison performed relied on the same transportation system, with differences in the operational scheme (with and without rejections). To better compare the performance of shared DRT systems against fixed-route public transport, the development of frameworks at network levels for automated public transport is needed.

The different dispatching and routing algorithms are key system inputs on determining its performance, therefore studies that test the robustness of the total cost minimization models under different strategies, network, and trip characteristics are of valuable contribution.

References

Alonso-Mora, J., Samaranayake, S., Wallar, A., Frazzoli, E., & Rus, D. (2018). Correction for Alonso-Mora et al., On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proc Natl Acad Sci U S A, 115*(3), E555. doi:10.1073/pnas.1721622115

Badia, H., & Jenelius, B. (2019). *Feeder transit services in different development stages of automated buses: comparing fixed routes versus door-to-door trips*. Paper presented at the 22nd EURO Working Group on Transportation Meeting, EWGT 2019, Barcelona, Spain.

Balac, M., Becker, H., Ciari, F., & Axhausen, K. W. (2019). Modeling competing free-floating carsharing operators – A case study for Zurich, Switzerland. *Transportation Research Part C: Emerging Technologies, 98*, 101-117. doi:10.1016/j.trc.2018.11.011

Becker, H., Balac, M., Ciari, F., & Axhausen, K. W. (2019). Assessing the welfare impacts of Shared Mobility and Mobility as a Service (MaaS). *Transportation Research Part A: Policy and Practice*. doi:10.1016/j.tra.2019.09.027

Bischoff, J., Fuhrer, K., & Maciejewski, M. (2018). *Impact assessment of autonomous DRT systems*. Paper presented at the mobil.TUM 2018 ”Urban Mobility - Shaping the Future Together” - International Scientific

Conference on Mobility and Transport.

Bischoff, J., & Maciejewski, M. (2016). Simulation of City-wide Replacement of Private Cars with Autonomous Taxis in Berlin. *Procedia Computer Science, 83*, 237-244. doi:10.1016/j.procs.2016.04.121

Bischoff, J., Maciejewski, M., & Nagel, K. (2017). *City-wide Shared Taxis:*

A Simulation Study in Berlin. Paper presented at the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC).

Bösch, P. M., Becker, F., Becker, H., & Axhausen, K. W. (2018). Cost-based analysis of autonomous mobility services. *Transport Policy, 64*, 76-91. doi:10.1016/j.tranpol.2017.09.005

Boyd, J. H., Asher, N. J., & Wetzler, E. S. (1978). Nontechnological innovation in urban transit: A comparison of some alternatives. *Journal of Urban Economics, 5*(1), 1-20. doi:https://doi.org/10.1016/0094-1190(78)90033-5

Bruun, E. (2005). Bus Rapid Transit and Light Rail:Comparing Operating Costs with a Parametric Cost Model. *Transportation Research Record, 1927*(1), 11-21. doi:10.1177/0361198105192700102

Buehler, R., Pucher, J., Gerike, R., & Götschi, T. (2016). Reducing car dependence in the heart of Europe: lessons from Germany, Austria, and Switzerland. *Transport Reviews, 37*(1), 4-28. doi:10.1080/01441647.2016.1177799

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed. ed.): Lawrence Erlbaum Associates Publishers.

Delle Site, P., & Filippi, F. (1998). Service optimization for bus corridors with short-turn strategies and variable vehicle size. *Transportation Research Part A: Policy and Practice, 32*(1), 19-38. doi:https://doi.org/10.1016/S0965-8564(97)00016-5

Dia, H., & Javanshour, F. (2017). Autonomous Shared Mobility-On-Demand: Melbourne Pilot Simulation Study. *Transportation Research Procedia, 22*, 285-296. doi:https://doi.org/10.1016/j.trpro.2017.03.035

Enoch, M., Potter, S., Parkhurst, G., & Smith, M. (2006). *Why do demand responsive transport systems fail?* Paper presented at the Transportation Research Board 85th Annual Meeting, Washington DC.

Fu, M., Rothfeld, R., & Antoniou, C. (2019). Exploring Preferences for Transportation Modes in an Urban Air Mobility Environment: Munich Case Study. *Transportation Research Record, 2673*(10), 427-442. doi:10.1177/0361198119843858

Hörl, S. (2017). Agent-based simulation of autonomous taxi services with dynamic demand responses. *Procedia Computer Science, 109*, 899-904. doi:https://doi.org/10.1016/j.procs.2017.05.418.

Hörl, S., Balac, M., & Axhausen, K. W. (2019). *Dynamic demand estimation for an AMoD system in Paris*. Paper presented at the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France.

Hörl, S., Ruch, C., Becker, F., Frazzoli, E., & Axhausen, K. W. (2019). Fleet operational policies for automated mobility: A simulation assessment for Zurich. *Transportation Research Part C: Emerging Technologies, 102*, 20-31. doi:10.1016/j.trc.2019.02.020

Horni, A., Nagel, K., & Axhausen, K. W. (2016). *The Multi-Agent Transport Simulation MATSim.* London: Ubiquity Press.

Infrastruktur, B. f. V. u. d. (2019). *Mobilität in Deutschland Kurzreport - Stadt München, Münchner Umland und MVV-Verbundraum*. Retrieved from https://www.muenchen-transparent.de/dokumente/5499206/datei

Inturri, G., Le Pira, M., Giuffrida, N., Ignaccolo, M., Pluchino, A., Rapisarda, A., & D'Angelo, R. (2019). Multi-agent simulation for planning and designing new shared mobility services. *Research in Transportation Economics, 73*, 34-44. doi:https://doi.org/10.1016/j.retrec.2018.11.009

Jansson, J. O. (1980). A Simple Bus Line Model for Optimisation of Service Frequency and Bus Size. *Journal of Transport Economics and Policy, 14*(1), 53-80. Retrieved from www.jstor.org/stable/20052563

Jara-Diaz, S. R. (1990). Consumer's surplus and the value of travel time savings. *Transportation Research Part B: Methodological, 24*(1), 73-77. doi:https://doi.org/10.1016/0191-2615(90)90033-U

Jara-Díaz, S. R., & Gschwender, A. (2003). From the Single Line Model to the Spatial Structure of Transit Services: Corridors or Direct? *Journal of Transport Economics and Policy, 37*(2), 261-277. Retrieved from www.jstor.org/stable/20053933

Javanshour, F., Dia, H., & Duncan, G. (2019). Exploring System Characteristics of Autonomous Mobility On-Demand Systems Under Varying Travel Demand Patterns. In *Intelligent Transport Systems for Everyone's Mobility* (pp. 299-315).

Jokinen, J. (2016). *Economic Perspectives on Automated Demand Responsive Transportation and Shared Taxi Services - Analytical models and simulations for policy analysis.* (Doctor of Science). Aalto University Helsinki Retrieved from http://urn.fi/URN:ISBN:978-952-60-6870-1

Jokinen, J., Sihvola, T., Hyytiä, E., & Sulonen, R. (2011). *Why urban mass demand responsive transport?* Paper presented at the 2011 IEEE Forum on Integrated and Sustainable Transportation Systems, Vienna.

Kim, M. E., & Schonfeld, P. (2014). Integration of conventional and flexible bus services with timed transfers. *Transportation Research Part B: Methodological, 68*, 76-97. doi:https://doi.org/10.1016/j.trb.2014.05.017

Kuah, G. K., & Perl, J. (1988). Optimization of Feeder Bus Routes and Bus‐Stop Spacing. *Journal of Transportation Engineering, 114*(3), 341-354. doi:doi:10.1061/(ASCE)0733-947X(1988)114:3(341)

Kühnel, N., & Zilske, M. (2019). JOSM MATSim plug-in (Version 0.8.7). Retrieved from https://github.com/matsim-org/josm-matsim-plugin

Leich, G., & Bischoff, J. (2019). Should autonomous shared taxis replace buses? A simulation study. *Transportation Research Procedia, 41*, 450-460. doi:https://doi.org/10.1016/j.trpro.2019.09.076

Li, X., & Quadrifoglio, L. (2010). Feeder transit services: Choosing between fixed and demand responsive policy. *Transportation Research Part C: Emerging Technologies, 18*(5), 770-780. doi:10.1016/j.trc.2009.05.015

Loeb, B., & Kockelman, K. M. (2019). Fleet performance and cost evaluation of a shared autonomous electric vehicle (SAEV) fleet: A case study for Austin, Texas. *Transportation Research Part A: Policy and Practice, 121*, 374-385. doi:https://doi.org/10.1016/j.tra.2019.01.025

Maciejewski, M. (2016). Dynamic Transport Services. In A. Horni, K. Nagel, & K. W. Axhausen (Eds.), *The Multi-Agent Transport Simulation MATSim* (pp. 145-152). London: Ubiquity Press.

Maciejewski, M., Bischoff, J., Hörl, S., & Nagel, K. (2017). *Towards a Testbed for Dynamic Vehicle Routing Algorithms*, Cham.

Maciejewski, M., & Nagel, K. (2012). *Towards Multi-Agent Simulation of the Dynamic Vehicle Routing Problem in MATSim*, Berlin, Heidelberg.

Maciejewski, M., Salanova, J. M., Bischoff, J., & Estrada, M. (2016). Large-scale microscopic simulation of taxi services. Berlin and Barcelona case studies. *Journal of Ambient Intelligence and Humanized Computing, 7*(3), 385-393. doi:10.1007/s12652-016-0366-3

Microsoft. (2019). Microsoft Excel for Mac (Version 16.32).

Mohring, H. (1972). Optimization and Scale Economies in Urban Bus Transportation. *American Economic Review, 62*, 591-604.

Moreno, A. T., Michalski, A., Llorca, C., & Moeckel, R. (2018). Shared Autonomous Vehicles Effect on Vehicle-Km Traveled and Average Trip Duration. *Journal of Advanced Transportation, 2018*, 10. doi:10.1155/2018/8969353

QGIS Development Team. (2019). QGIS Geographic Information System (Version 3.8.1). Open Source Geospatial Foundation Project. Retrieved from http://qgis.osgeo.org

R Development Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria.

Rothfeld, R., Balac, M., Ploetner, K. O., & Antoniou, C. (2018a). Agent-based Simulation of Urban Air Mobility. In *2018 Modeling and Simulation Technologies Conference*.

Rothfeld, R., Balac, M., Ploetner, K. O., & Antoniou, C. (2018b). Initial Analysis of Urban Air Mobility's Transport Performance in Sioux Falls. In *Aviation Technology, Integration, and Operations Conference*.

Shen, Y., Zhang, H., & Zhao, J. (2018). Integrating shared autonomous vehicle in public transportation system: A supply-side simulation of the first-mile service in Singapore. *Transportation Research Part A: Policy and Practice, 113*, 125-136. doi:https://doi.org/10.1016/j.tra.2018.04.004

Shyue Koong, C., & Schonfeld, P. M. (1991). Multiple period optimization of bus transit systems. *Transportation Research Part B: Methodological, 25*(6), 453-478. doi:https://doi.org/10.1016/0191-2615(91)90038-K

Sperling, D. (2018). *Three Revolutions - Steering Automated, Shared, and Electric Vehicles to a Better Future*. Washington, DC: Island Pres.

Stock, J. H., & Watson, M. W. (2015). *Introduction to Econometrics (Updated third edition)*: Pearson Education Limited.

Tirachini, A. (2014). The economics and engineering of bus stops: Spacing, design and congestion. *Transportation Research Part A: Policy and Practice, 59*, 37-57. doi:https://doi.org/10.1016/j.tra.2013.10.010

Tirachini, A., & Antoniou, C. (2019). *The economics of automated public transport: effects on operator cost, travel time, fare and subsidy*. Manuscript submitted for publication.

Tirachini, A., & Antoniou, C. (2020). The economics of automated public transport: Effects on operator cost, travel time, fare and subsidy. *Economics of Transportation, 21*. doi:10.1016/j.ecotra.2019.100151

Tirachini, A., & Hensher, D. A. (2011). Bus congestion, optimal infrastructure investment and the choice of a fare collection system in dedicated bus corridors. *Transportation Research Part B: Methodological, 45*(5), 828-844. doi:https://doi.org/10.1016/j.trb.2011.02.006

Tirachini, A., Hensher, D. A., & Jara-Díaz, S. R. (2010a). Comparing operator and users costs of light rail, heavy rail and bus rapid transit over a radial public transport network. *Research in Transportation Economics, 29*(1), 231-242. doi:10.1016/j.retrec.2010.07.029

Tirachini, A., Hensher, D. A., & Jara-Díaz, S. R. (2010b). Restating modal investment priority with an improved model for public transport analysis. *Transportation Research Part E: Logistics and Transportation Review, 46*(6), 1148-1168. doi:10.1016/j.tre.2010.01.008

Tsao, M., Milojevic, D., Ruch, C., Salazar, M., Frazzoli, E., & Pavone, M. (2019, 20-24 May 2019). *Model Predictive Control of Ride-sharing Autonomous Mobility-on-Demand Systems.* Paper presented at the 2019 International Conference on Robotics and Automation (ICRA).

Viergutz, K., & Schmidt, C. (2019). Demand responsive - vs. conventional public transportation: A MATSim study about the rural town of Colditz, Germany. *Procedia Computer Science, 151*, 69-76. doi:https://doi.org/10.1016/j.procs.2019.04.013

Wang, B., Medina, S. A. O., & Fourie, P. (2018). Simulation of Autonomous Transit On Demand for Fleet Size and Deployment Strategy Optimization. *Procedia Computer Science, 130*, 797-802. doi:https://doi.org/10.1016/j.procs.2018.04.138

Wang, L., Wirasinghe, S. C., Kattan, L., & Saidi, S. (2018). Optimization of demand-responsive transit systems using zonal strategy. *International Journal of Urban Sciences, 22*(3), 366-381. doi:10.1080/12265934.2018.1431144

Wang, S., Correia, G. H. d. A., & Lin, H. X. (2019). Exploring the Performance of Different On-Demand Transit Services Provided by a Fleet of Shared Automated Vehicles: An Agent-Based Model. *Journal of Advanced Transportation, 2019*, 16. doi:10.1155/2019/7878042

Zilske, M. (2016). How to Write Your Own Extensions and Possibly Contribute Them to MATSim. In A. Horni, K. Nagel, & K. W. Axhausen (Eds.), *The Multi-Agent Transport Simulation MATSim* (pp. 297-304). London: Ubiquity Press.

Appendix A.1 Simulation results – no rejection (Scenario 1)

ID	demand	fleetsize	vehcap	tripdist	distdetour	wt	vt	tott	vehdist	avgvehdist	occ
1	20198	200	4	6014.79	2.15887196	805.13	808.57	1613.7	43079573.4	215397.87	2.82005412
2	20198	200	6	7926.69	2.844768	835.96	1142.82	1978.79	38722273.7	193611.37	4.13465609
3	20198	200	8	9612.98	3.44961747	852.07	1447.06	2299.13	36154969.3	180774.85	5.3702983
4	20198	600	4	5682.11	2.03947869	692.73	753.3	1446.02	42743423	71239.04	2.6850273
5	20197	600	6	7390.78	2.65240001	760.89	1063.46	1824.35	38475155.6	64125.26	3.8794955
6	20198	600	8	9071.65	3.255186	788.46	1361.65	2150.1	36191473.6	60319.12	5.06277221
7	20198	1000	4	5527.52	1.98409855	644.98	727.62	1372.59	41803576.7	41803.58	2.67056865
8	20198	1000	6	7243.05	2.5996253	719.47	1034.1	1753.57	37870186.8	37870.19	3.86306845
9	20198	1000	8	8764.31	3.14544476	750.58	1310.1	2060.68	35525471.7	35525.47	4.98294674
10	20198	2500	4	5353.88	1.92162577	585.81	696.41	1282.21	41343601.6	16537.44	2.61558413
11	20198	2500	6	6942.83	2.49167567	664.23	980.02	1644.25	37561804.6	15024.72	3.7333478
12	20197	2500	8	8447.46	3.03131627	711.24	1256.29	1967.53	35261385.4	14104.55	4.83853222
13	20198	4000	4	5273.72	1.89290214	561.77	681.81	1243.58	41232454.3	10308.11	2.58336785
14	20197	4000	6	6792.61	2.43811401	645.15	958.99	1604.13	37321759.4	9330.44	3.67569894
15	20198	4000	8	8230.12	2.9537916	677.73	1215.98	1893.71	35260563.5	8815.14	4.71438762
16	40578	200	4	6252.07	2.25303067	957.68	855.76	1813.44	81507707.4	407538.54	3.11246939
17	40582	200	6	8250.25	2.97306306	972.84	1221.6	2194.44	71657934.1	358289.67	4.67235973
18	40582	200	8	10114.57	3.64469308	967.51	1566.6	2534.11	66186103.7	330930.52	6.20174715
19	40582	600	4	5785.85	2.08502856	820	790.53	1610.53	80363797.2	133939.66	2.92173059
20	40582	600	6	7604.29	2.74022544	869.84	1126.88	1996.72	71260765.1	118767.94	4.33053583
21	40582	600	8	9334.4	3.36345685	891.61	1447.22	2338.83	65923804.5	109873.01	5.74615837
22	40582	1000	4	5682.24	2.04763192	782.58	774.03	1556.61	79630162.5	79630.16	2.8958457
23	40582	1000	6	7437.82	2.6802472	842.06	1101.42	1943.48	70574481.9	70574.48	4.27692281

24	40582	1000	8	9222.1	3.3230757	871.93	1427.03	2298.96	66137841.6	66137.84	5.65865552
25	40582	2500	4	5498.37	1.98143744	708.77	739.06	1447.84	77305647.6	30922.26	2.8863978
26	40582	2500	6	7158.58	2.5796034	784.82	1052.47	1837.29	68454587.3	27381.83	4.24382799
27	40582	2500	8	8796.32	3.16966229	818.3	1359.82	2178.11	63888769.7	25555.51	5.58740229
28	40582	4000	4	5428.85	1.95629299	680.5	726.86	1407.36	77085482.1	19271.37	2.85804259
29	40582	4000	6	7080.45	2.5515319	763.12	1036.12	1799.24	68738517.5	17184.63	4.18017194
30	40582	4000	8	8629.88	3.10968737	798.52	1330.82	2129.34	63790968.4	15947.74	5.49008424
31	101089	400	4	6302.59	2.26712686	1511.71	876.56	2388.27	192416826	481042.06	3.30984785
32	101120	400	6	8141.95	2.92800066	1084.8	1229.75	2314.55	165527373	413818.43	4.97344149
33	101127	400	8	9938.84	3.57410664	1045.45	1572.95	2618.41	151264704	378161.76	6.64434749
34	101131	600	4	5956.65	2.14180156	1034.01	836.41	1870.42	183323270	305538.78	3.28600931
35	101131	600	6	7730.75	2.77990536	1018.26	1179.66	2197.92	158536365	264227.28	4.93147724
36	101131	600	8	9480.77	3.40914714	1013.51	1517.48	2530.99	145559077	242598.46	6.58701452
37	101131	1000	4	5712.41	2.05407748	933.23	798.53	1731.76	180927276	180927.28	3.19297364
38	101131	1000	6	7475.6	2.68810747	962.5	1139.33	2101.83	157363115	157363.12	4.80422256
39	101131	1000	8	9118.29	3.27886355	969.05	1461.87	2430.93	143865210	143865.21	6.4097622
40	101131	2500	4	5573	2.00393379	864.05	773.62	1637.67	180921922	72368.77	3.11517287
41	101128	2500	6	7259.94	2.61040936	916.47	1103.29	2019.76	157148682	62859.47	4.67190183
42	101129	2500	8	8878.19	3.19238777	932.74	1420.22	2352.97	143895415	57558.17	6.23954891
43	101129	4000	4	5492.56	1.97509448	823.32	759.18	1582.5	177249651	44312.41	3.1337247
44	101130	4000	6	7132.28	2.5647657	886.35	1081.92	1968.27	153959130	38489.78	4.68492825
45	101129	4000	8	8732.23	3.14008465	911.8	1396	2307.8	141424847	35356.21	6.24417636
46	162052	650	4	5926.85	2.13927861	1082.69	835.18	1917.87	287451283	442232.74	3.34081507
47	162069	650	6	7761.47	2.80141849	1062.18	1192.07	2254.24	247719631	381107.12	5.07779861
48	162070	650	8	9505.22	3.43068233	1055.44	1532.84	2588.28	225673928	347190.66	6.82618506
49	162072	850	4	5813.5	2.09818351	1068.58	824.58	1893.16	280890520	330459.43	3.35435163
50	162072	850	6	7572.37	2.73300706	1058.68	1169.41	2228.09	242182095	284920.11	5.06754701
51	162072	850	8	9260.8	3.34257334	1050.49	1502.32	2552.81	221144494	260169.99	6.78703932

52	162072	1000	4	5751.4	2.07564329	1043.76	814.87	1858.63	279472826	279472.83	3.33535434
53	162068	1000	6	7438.95	2.6848921	1035.33	1150.41	2185.74	239742610	239742.61	5.02879212
54	162070	1000	8	9079.14	3.27706451	1028.59	1474.82	2503.42	218880101	218880.1	6.72261724
55	162071	2500	4	5509.69	1.98841171	932.36	775.52	1707.88	277701764	111080.71	3.21553942
56	162072	2500	6	7169.07	2.58754209	972.33	1105.9	2078.23	239149939	95659.98	4.85848134
57	162072	2500	8	8743.06	3.1554622	980.38	1420.71	2401.08	217801331	87120.53	6.505953
58	162071	4000	4	5461.58	1.97110603	899.08	765.33	1664.41	275791706	68947.93	3.20953717
59	162071	4000	6	7060.85	2.54842692	944.33	1086.87	2031.2	237510837	59377.71	4.81807447
60	162071	4000	8	8614.29	3.1090663	959.18	1399.64	2358.82	216055758	54013.94	6.46180124
61	202226	750	4	6027.02	2.17258806	1670.83	851.96	2522.79	358436890	477915.85	3.39891298
62	202276	750	6	7737.25	2.78850971	1100.67	1191.92	2292.59	305206017	406941.36	5.12747488
63	202287	750	8	9432.64	3.39941905	1071.96	1525.7	2597.66	277888257	370517.68	6.8661591
64	202290	850	4	5839.79	2.10445916	1078.33	827.45	1905.78	349881018	411624.73	3.37624604
65	202293	850	6	7657.02	2.75916717	1084.21	1183.6	2267.81	301102442	354238.17	5.14427542
66	202293	850	8	9340.44	3.36585155	1074.17	1517.76	2591.92	273738178	322044.92	6.90260175
67	202294	1000	4	5789.4	2.08618006	1081.8	822.38	1904.18	345688628	345688.63	3.38790688
68	202293	1000	6	7528.95	2.71300804	1078.46	1167.59	2246.05	296892349	296892.35	5.12998698
69	202294	1000	8	9165.4	3.30273973	1068.72	1495.94	2564.66	269791155	269791.16	6.87237291
70	202292	2500	4	5512.13	1.98600247	969.76	779.43	1749.19	339858350	135943.34	3.28095456
71	202294	2500	6	7138.85	2.5722434	992.67	1105.5	2098.17	291619160	116647.66	4.95214164
72	202294	2500	8	8650.96	3.11718224	996.23	1414.45	2410.68	263740583	105496.23	6.63541665
73	202293	4000	4	5438.53	1.95944227	932.53	767.66	1700.19	339166215	84791.55	3.24375207
74	202293	4000	6	7029.74	2.53317574	968.13	1088.77	2056.91	289926145	72481.54	4.90492913
75	202294	4000	8	8540.16	3.07725791	977.89	1395.94	2373.83	263144043	65786.01	6.56531346

Appendix A.2 Simulation results – constrained operation (Scenario 2)

ID	demand	fleetsize	vehcap	tripdist	distdetour	wt	vt	tott	vehdist	avgvehdist	occ	realizedDemand	rejections	rejectionRate
1	20149	200	4	3185.41	1.1358453	186.91	317.6	504.5	61692902.8	308464.51	1.0300333	19949	200	0.01
2	20156	200	6	3188.16	1.13751949	186.3	318.84	505.14	61331936	306659.68	1.03906927	19989	167	0.01
3	20123	200	8	3192.16	1.13361175	189.71	316.76	506.47	61679495.2	308397.48	1.02545356	19814	309	0.02
4	20197	600	4	3256.57	1.16922	133.64	333.08	466.72	56748759.5	94581.27	1.15896256	20196	1	0
5	20197	600	6	3262.39	1.17131799	133.71	333.45	467.16	56775820.8	94626.37	1.16048042	20196	1	0
6	20198	600	8	3264.28	1.17185352	132.76	334.5	467.26	56770026.5	94616.71	1.16138624	20198	0	0
7	20197	1000	4	3262.09	1.17108834	124.38	332.76	457.13	54249304.8	54249.3	1.214415	20196	1	0
8	20197	1000	6	3266.66	1.17290161	125.06	333.13	458.2	54327006	54327.01	1.21443711	20197	0	0
9	20194	1000	8	3255.73	1.16870441	124.68	332.19	456.87	54398288.2	54398.29	1.20854825	20193	1	0
10	20197	2500	4	3238.37	1.16270227	110.8	326.71	437.52	51971846.3	20788.74	1.25841442	20196	1	0
11	20198	2500	6	3236.46	1.16201234	110.31	326.97	437.28	51770092.7	20708.04	1.26269851	20198	0	0
12	20198	2500	8	3249.01	1.16658109	111.22	328.35	439.57	51892835.5	20757.13	1.264534	20197	1	0
13	20198	4000	4	3232.55	1.1605835	105.24	324.2	429.44	51132229.7	12783.06	1.27690588	20198	0	0
14	20198	4000	6	3231.56	1.16023222	104.66	324.36	429.02	51047553.3	12761.89	1.27856896	20197	1	0
15	20195	4000	8	3231.25	1.1601001	105	324.82	429.82	50885689.3	12721.42	1.28232246	20194	1	0
16	37370	200	4	3487.21	1.09887724	247.68	333.67	581.35	100598872	502994.36	0.99847638	28804	8566	0.23
17	37413	200	6	3472.8	1.09673487	248.46	331.75	580.21	100777616	503888.08	0.99524022	28881	8532	0.23
18	37294	200	8	3487.21	1.09805373	249.03	333.25	582.27	100315595	501577.97	0.99653348	28667	8627	0.23
19	40580	600	4	3298.33	1.18838183	147.58	345.86	493.44	111172707	185287.84	1.20365197	40570	10	0
20	40575	600	6	3295.19	1.18688273	148.72	345.04	493.76	111455343	185758.91	1.19889496	40551	24	0
21	40578	600	8	3295.53	1.18758694	147.01	345.75	492.76	110964279	184940.47	1.20497821	40573	5	0
22	40582	1000	4	3312.45	1.19363266	134.24	348.31	482.54	107986465	107986.46	1.24477842	40580	2	0
23	40580	1000	6	3315.32	1.19473574	134.19	349.13	483.32	108057589	108057.59	1.24488349	40575	5	0
24	40582	1000	8	3311.63	1.19335868	133.61	348.55	482.16	108000801	108000.8	1.2443664	40582	0	0
25	40581	2500	4	3310.97	1.19327135	122.12	345.98	468.1	101915225	40766.09	1.31821246	40576	5	0

26	40578	2500	6	3305.88	1.19147127	122.51	345.82	468.33	102010979	40804.39	1.31469123	40568	10	0
27	40579	2500	8	3308.17	1.19220637	121.88	345.95	467.83	101684548	40673.82	1.31982531	40568	11	0
28	40580	4000	4	3299.69	1.18924033	115.61	343.17	458.78	99442019.2	24860.5	1.34626211	40572	8	0
29	40579	4000	6	3300.12	1.18948105	115.11	343.33	458.44	99165009.8	24791.25	1.35009888	40569	10	0
30	40581	4000	8	3297.07	1.18828321	115.43	343.19	458.62	99042293.1	24760.57	1.35078869	40577	4	0
31	90511	400	4	3620.58	1.10648689	253.32	349.56	602.88	214503904	536259.76	1.08498959	64281	26230	0.29
32	90561	400	6	3626.06	1.10782307	253.03	350.41	603.44	213719890	534299.73	1.08890441	64180	26381	0.29
33	90498	400	8	3628.51	1.10671193	253.9	350.42	604.32	214394703	535986.76	1.08412877	64057	26441	0.29
34	96938	600	4	3440.01	1.13519694	223.1	345.36	568.46	255057826	425096.38	1.14199972	84673	12265	0.13
35	97033	600	6	3426.26	1.13732129	221.58	344.86	566.44	254337865	423896.44	1.14714799	85155	11878	0.12
36	97059	600	8	3431.08	1.13523958	223.17	344.79	567.96	255345844	425576.41	1.14372986	85118	11941	0.12
37	100864	1000	4	3328.45	1.18654128	169.37	353.82	523.19	262702791	262702.79	1.26297308	99682	1182	0.01
38	100926	1000	6	3327.85	1.18833536	168.01	354.9	522.91	262168968	262168.97	1.26790607	99886	1040	0.01
39	100958	1000	8	3323.52	1.18753551	168.07	353.94	522.01	262280474	262280.47	1.26759341	100034	924	0.01
40	101125	2500	4	3376.5	1.21402684	134.05	366.98	501.03	250290680	100116.27	1.36400571	101110	15	0
41	101128	2500	6	3374.77	1.21330448	134.14	367.02	501.16	250130785	100052.31	1.3642188	101113	15	0
42	101126	2500	8	3375.24	1.21356508	134.17	366.87	501.04	250003164	100001.27	1.36509179	101112	14	0
43	101125	4000	4	3373.24	1.2128765	129.22	365.7	494.92	242870778	60717.69	1.40425058	101105	20	0
44	101128	4000	6	3373.69	1.21306447	129.64	365.69	495.33	243510303	60877.58	1.40076333	101106	22	0
45	101126	4000	8	3374.81	1.21334067	129.12	365.91	495.03	243015916	60753.98	1.40403732	101103	23	0
46	146628	650	4	3633.7	1.1231902	242.65	358.39	601.04	334994643	515376.37	1.16325817	107242	39386	0.27
47	146582	650	6	3641.42	1.12320866	242.99	358.9	601.89	334685198	514900.3	1.16268382	106863	39719	0.27
48	146570	650	8	3633.37	1.12196455	243.3	357.9	601.2	335278227	515812.66	1.16064071	107101	39469	0.27
49	152946	850	4	3522.89	1.14267504	223.38	356.72	580.1	374068702	440080.83	1.20276122	127712	25234	0.16
50	152667	850	6	3526.47	1.14107148	224.09	356.51	580.59	373966205	439960.24	1.19851407	127097	25570	0.17
51	153110	850	8	3521.77	1.14213024	223.59	356.31	579.9	375199067	441410.67	1.20188209	128045	25065	0.16
52	157106	1000	4	3419.56	1.15585812	209.84	352.7	562.54	395897117	395897.12	1.22794107	142164	14942	0.1
53	156978	1000	6	3425.25	1.15409108	210.67	352.87	563.55	395387109	395387.11	1.22364419	141249	15729	0.1

54	156982	1000	8	3423.21	1.15540472	209.57	353.14	562.71	394424622	394424.62	1.22602111	141263	15719	0.1
55	162058	2500	4	3376.64	1.21826034	140.7	372.83	513.53	393695080	157478.03	1.38949143	162006	52	0
56	162057	2500	6	3375.99	1.21809614	140.56	372.82	513.38	394023571	157609.43	1.38809149	162009	48	0
57	162061	2500	8	3377.21	1.21844401	140.76	372.69	513.45	394441931	157776.77	1.38711175	162008	53	0
58	162063	4000	4	3387.39	1.22203302	134.29	373.96	508.25	385780742	96445.19	1.42254653	162010	53	0
59	162061	4000	6	3386.15	1.22169146	134.6	374.07	508.67	386329799	96582.45	1.41998725	162008	53	0
60	162066	4000	8	3385.3	1.22135394	134.4	373.96	508.36	385532938	96383.23	1.42263529	162016	50	0
61	181227	750	4	3678.21	1.12553359	243.56	363.99	607.55	396640514	528854.02	1.18950027	128270	52957	0.29
62	181221	750	6	3688.22	1.12606746	243.35	364.83	608.17	397034569	529379.43	1.19062468	128170	53051	0.29
63	181117	750	8	3689.16	1.12639572	243.26	365.08	608.33	396571386	528761.85	1.19201211	128137	52980	0.29
64	184494	850	4	3656.97	1.13425017	235.37	365.58	600.95	420545619	494759.55	1.2076692	138880	45614	0.25
65	185079	850	6	3642.79	1.13343788	235.65	364.3	599.94	422636111	497218.95	1.20618107	139941	45138	0.24
66	184953	850	8	3642.81	1.13354286	235	364.59	599.58	421754469	496181.73	1.208355	139900	45053	0.24
67	189659	1000	4	3573.62	1.14390615	224.95	362.84	587.79	451803434	451803.43	1.22661715	155078	34581	0.18
68	189601	1000	6	3572.41	1.14459599	224.24	362.77	587.01	450415041	450415.04	1.23063987	155161	34440	0.18
69	190001	1000	8	3574.36	1.1447513	224.62	362.77	587.39	452245208	452245.21	1.22920559	155525	34476	0.18
70	202220	2500	4	3377.72	1.21530884	148.77	373.75	522.52	489413469	195765.39	1.39337123	201892	328	0
71	202220	2500	6	3374.89	1.21429497	149.18	373.53	522.71	488715147	195486.06	1.3941793	201890	330	0
72	202220	2500	8	3377.62	1.21495374	149.06	373.65	522.71	488973861	195589.54	1.3940853	201820	400	0
73	202266	4000	4	3400.13	1.22444254	136.53	378.8	515.33	479510406	119877.6	1.43341265	202150	116	0
74	202263	4000	6	3399.92	1.22430078	136.43	378.95	515.38	479611658	119902.91	1.43290101	202133	130	0
75	202266	4000	8	3399.33	1.22415004	136.42	378.84	515.27	479692496	119923.12	1.43257391	202156	110	0