University of Amsterdam

Masters Thesis

Assessing Charging Infrastructure through Data-Driven Agent-Based Simulations of Electric Vehicles

Author:
Seyla Wachlin

Supervisors:
Dr. Michael H Lees
Drs. ir. Jurjen R Helmus

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Computational Science in the

Section Computational Science
Informatics Institute

August 2017

UNIVERSITY OF AMSTERDAM

Abstract

Faculty of Science
Informatics Institute

Master of Science in Computational Science

Assessing Charging Infrastructure through Data-Driven Agent-Based Simulations of Electric Vehicles

by Seyla Wachlin

Electric vehicles (EVs) are a promising solution for sustainability problems, but providing an optimal charging infrastructure is a challenging problem that arises with the increase of EVs. Infrastructure optimization methods, such as policy incentives and roll-out strategies, can help create this optimal charging infrastructure. Simulation models are needed to provide insight into the effects of these infrastructure optimization methods before they are carried out. To the best of our knowledge existing simulation models are not validated or only validated by small amounts of data, which greatly decreases the predictive certainty of these models.

In this thesis we present the Simulation of Electric Vehicle Activity (SEVA) model; a predictive, data-driven simulation model which is validated by a large real-world dataset. From this dataset we are able to extract the behavior of individual EV users in terms of when, where and how long they charge. By simulating the charging transactions of each individual EV user we can accurately predict the future behavior of these EV users, in terms of in which area.

The SEVA model is extended by using a discrete choice model to create an Improved Charge Pole Selection (ICPS) process, resulting in the SEVA-ICPS model. A literature review revealed that the factors which influence how an EV user selects a charge pole (CP) are distance, charging speed and cost (divisible into charging fees and parking fees). We performed a data analysis to further understand these factors. Next, these factors are used as explanatory variables within three logistic regression choice models, which were fitted using the large real-world dataset. All three choice models were implemented within the SEVA model and the resulting SEVA-ICPS model is validated. This validation shows that all three choice models perform significantly better than randomly selecting a CP in a specific area. Furthermore our ICPS process not only accurately

simulates the decision process of the EV users, but also accurately predicts if and how new CPs are used.

We show the capabilities of the SEVA-ICPS model with two case studies. The first case study aims to understand the effects of non-habitual EV users, such as car sharing users, on the habitual EV users. In this case study the SEVA-ICPS model is extended to include these non-habitual EV users and their behavior. This case study shows that while the charging infrastructures of Amsterdam and Utrecht are relatively well prepared for an increase in non-habitual users, in The Hague and Rotterdam an increase of one non-habitual agent for each habitual agent in the system caused the percentage of failed connection attempts of habitual agent to increase significantly.

To prevent this strong increase in the percentage of failed connection attempts and to prepare the cities for the increase of non-habitual EV users a number of different rollout strategies can be implemented in practice. The second case study then analyses how the effects of an increase of non-habitual EVs can be decreased by testing four different rollout strategies. While all four tested rollout strategies are effective in decreasing the percentages of failed connection attempts in all four cities, specifically adding CPs at locations where many failed connection attempts occur most strongly increases the robustness of all cities. A policy implication is that if municipalities are able to keep track of where EV users experience failed connection attempts, this could give them the information they need to determine where it is most efficient to place new CPs. This then helps them to optimize their charging infrastructure.

Acknowledgements

This thesis brings to an end an intense two year master programme in Computational Science at the University of Amsterdam. Having completed my BSc in Computer Science, I opted for this master programme because of its strong interdisciplinary and scientific nature. The commitment of staff members to this programme has been an important source of inspiration. In particular, I would like to acknowledge and thank the people whose support has been crucially important for me to be able to produce this thesis.

I was very fortunate to have Mike Lees and Jurjen Helmus as my supervisors, as they encouraged me to pursue this project from its inception and as they expressed such a strong commitment in guiding me through the various stages of writing this thesis. I would like to thank Mike for his support whenever I seemed to have reached a dead end, and for his invaluable and always challenging comments that pushed me to go further and increase the quality of this thesis. I thank Jurjen for his continuous support, his enthusiasm about the project in its various phases and for his inspiring ideas.

I greatly enjoyed working many, many days with Igna Vermeulen in the Computational Science common room - even when we couldn't agree on who got to sit in the best chair. It is thanks to Igna that writing this thesis really did feel like fun. And besides the unceasing 'gezelligheid', she was also always there to brainstorm with me and to support me whenever I got stuck.

Finally, I thank my partner, Jordy Perlee, who read the thesis at various stages of completing and always responded with positive criticism. For his patience and encouragement, I am very grateful. I also thank my parents, Annelies Moors and Steven Wachlin, for their loving support and unwavering trust in my various endeavours.

Contents

Abstract iii

Acknowledgements v

1 Introduction 1

2 The SEVA Model 5

2.1 Review of Relevant Literature . . . 5
2.1.1 System Evaluation Metrics . . . 6
2.1.2 Infrastructure Optimization: Data Analytics . . . 8
2.1.3 Infrastructure Optimization: Computational Models . . . 10
2.1.4 Conclusion . . . 15
2.2 Model Description . . . 17
2.2.1 Overview: Entities, State Variables and Scales . . . 17
2.2.2 Overview: Process Overview and Scheduling . . . 21
2.2.3 Design Concepts: Data-Driven . . . 22
2.2.4 Design Concepts: Observations . . . 23
2.2.5 Design Concepts: Behavior . . . 23
2.2.6 Details: Initialization . . . 25
2.2.7 Details: Submodels . . . 28
2.3 Model Metrics . . . 31
2.3.1 Clustering Metrics . . . 31
2.3.2 Competition Metrics . . . 31
2.3.3 Run Time Metrics . . . 31
2.3.4 Validation Metrics . . . 32
2.4 Sensitivity Analysis and Model Evaluation . . . 34
2.4.1 Simulation Setup . . . 34
2.4.2 Data Processing . . . 36
2.4.3 Clustering and Cluster Analysis . . . 37
2.4.4 Selection Process . . . 45
2.4.5 Model Validation . . . 47
2.5 Conclusion . . . 50

3 A Model of Charge Pole Selection 51

3.1 Review of Relevant Literature . . . 51
3.2 Data Analysis . . . 54
3.2.1 Distance . . . 55
3.2.2 Charging Speed . . . 56

3.2.3 Charging Fees . . . 58
3.2.4 Parking Fees . . . 61
3.2.5 Conclusion . . . 66
3.3 Choice Model . . . 68
3.3.1 Unweighted: Model per Cluster . . . 69
3.3.2 Weighted Model . . . 73
3.4 Model Description . . . 77
3.5 Model Evaluation . . . 80
3.5.1 Effect of Cluster Size . . . 84
3.5.2 Validation of Clusters Containing Newly Added CPs . . . 86
3.6 Conclusion . . . 89

4 Case Studies: Non-Habitual Users and Rollout Strategies 91
4.1 Non-Habitual Agents . . . 91
4.2 Non-Habitual Agents Experiments . . . 94
4.3 City Differences Concerning City Preparedness . . . 97
4.3.1 Mean Arrival Times of Failed Connection Attempts . . . 97
4.3.2 Overall Available Charging Infrastructure . . . 99
4.3.3 Available Charging Infrastructure in Range . . . 101
4.4 Effects of Rollout Strategies . . . 103
4.4.1 Setup of the Experiment . . . 103
4.4.2 Rollout Strategy Experiment . . . 106
4.5 Conclusion . . . 108

5 Conclusion and Future Work 111
5.1 Conclusion . . . 111
5.2 Possibilities for Future Work . . . 113

A SEVA Model Parameters 115

B Choice Model Parameters 117

C Number of Failed Connection Attempts 119

List of Figures 120

List of Tables 127

Bibliography 129

Chapter 1

Introduction

Cities struggle with sustainability problems such as air quality and CO_2 emissions. A promising solution is an increased use of electric vehicles (EVs), both through private use and through car-sharing schemes, rather than fossil fuel combustion cars. EVs have a small carbon footprint and less negative impact on air quality than fuel cars. The increase in EV usage would thus decrease local emissions, improve air quality and help facilitate sustainable cities.

When municipalities attempt to stimulate the use of EVs, a chicken and egg problem arises. Potential EV users are hesitant to buy electric vehicles if there is insufficient charging infrastructure, while municipalities are hesitant to provide new charge poles (CPs) if the number of EV users is limited. Nonetheless, cities are creating and expanding their charging infrastructure to facilitate demand and to increase EV adoption [1–3]. Many cities now have more mature and city-wide charging infrastructures. For these cities, the struggle remains how to further rollout charging infrastructure in the most efficient way, both in terms of cost and use. Both over-capacity and under-capacity of the CPs are situations which municipalities hope to avoid. Predicting the future usage of CPs for a given population of EV users would enable policy makers to create a more optimal charging infrastructure. They could avoid over-capacity or under-capacity of new and existing CPs by placing these poles at optimal locations [4].

Predictive simulation models provide insight into the effects of incentives and rollout strategies before they are implemented in practice and thus allow for scenario testing. While many simulation models on this topic exist today, these are, to the best of our knowledge, not validated or only validated using small amounts of data [5–12]. This greatly decreases the predictive certainty of the models, which is mentioned as a limitation in both [6] and [11].

* * *

The main objective of this thesis is to understand the effects of rollout strategies and thus enable the creation of an optimal charging infrastructure. This goal is achieved by developing a predictive, data-driven simulation model for the use of CPs by EV users, validated by the use of a real-world dataset. This dataset contains over 3 million charging transactions conducted by approximately 40 thousand electric vehicle users at around 5 thousand CPs located throughout the four major cities in the Netherlands (Amsterdam, The Hague, Rotterdam and Utrecht) [13]. The use of this large dataset is what sets our model apart from existing models. Data analysis of this dataset has already proven to help answer behavioral questions [14–16]. However, even with such a rich dataset, pure data analysis may not be enough to answer questions regarding what-if scenarios because pure data analysis can offer hindsight and insight but no foresight regarding events that have not previously occurred. More specifically, data analysis cannot help understand the effects of rollout strategies that have never been implemented. Our simulation model, in contrast, can give insight into policy related challenges.

Our model is designed and built in multiple stages. The first stage is the Simulation of Electric Vehicle Activity (SEVA) model, which covers the basic components of the system, namely the agents (EV users together with their EVs) and the environment (the collection of CPs and their metadata). The behavior of the agents is based on the actual behavior of the EV users as found in the data. Agents in the system can be in one of three possible states (connected, disconnected or selecting CP) and transactions between these states are controlled by the connection process, disconnection process and a naive implementation of the CP selection process. This SEVA model rather successfully predicts the behavior of habitual users (in which area, at which time and for how long they charge) by simulating each charging transaction of each electric vehicle user. This enables the simulation of each individual user in an agent-based model.

In this thesis the SEVA model is extended by using a discrete choice model to create an Improved Charge Pole Selection (ICPS) process. This is necessary as the SEVA model is not policy sensitive with regards to CP selection and thus insufficiently suitable to gain insight in the effects of rollout strategies. It is not able to fully predict the decision making process of agents and it cannot predict accurately if and how agents will use new CPs. In order to design the ICPS process, we first perform a comprehensive literature review to discover the factors which influence how an EV user selects a CP. These factors are then individually examined through a data analysis and used as the explanatory variables for a logistic regression choice model. This choice model is then fitted, implemented and validated within the SEVA model to create the SEVA-ICPS model.

Lastly, we show the capabilities of the SEVA-ICPS model through two applied case studies. The first case study allows us to gain insight in both the effects of non-habitual users on competition for the use of CPs and the preparedness of cities for non-habitual users. The second case study evaluates how four rollout strategies can improve this preparedness and thus decrease the effects of non-habitual users.

To summarize, the main question addressed in this thesis involves understanding rollout strategies in the Netherlands. Its contributions include:

1. An agent-based model framework (SEVA) which is able to utilize EV charging data and predict future charge behavior;
2. A discrete choice model to create an improved CP selection process which results in the SEVA-ICPS model;
3. An analysis of the robustness of the current charging infrastructures in the major cities of the Netherlands;
4. A policy evaluation of four rollout strategies in the major cities of the Netherlands.

This thesis is structured as follows. Chapter 2 focuses on the SEVA model and presents a review of the relevant literature, the model description, the model metrics and the model evaluation. Chapter 3 concentrates on the SEVA-ICPS model and contains a review of the relevant literature, the data analysis, the choice model, the implementation of this choice model and the model evaluation. Chapter 4 discusses the effects of non-habitual users and shows how these non-habitual agents behave, how they are implemented within the SEVA-ICPS model, how they affect the habitual users and how prepared and robust the different cities are for these non-habitual users. Furthermore, this chapter examines how different rollout strategies effect the robustness and preparedness of the charging infrastructures in the four cities against non-habitual users. We end with our overall conclusions and recommendations for future work in Chapter 5.

Chapter 2

The SEVA Model

The work in this chapter was performed in collaboration with Igna Vermeulen. I personally take credit for 50% of the work done in this chapter. This chapter is being prepared for submission to the journal *Transportation Research Part E: Logistics and Transportation Review*.

In this chapter we describe the SEVA model, a data-driven simulation framework. This framework is a fully functional and validated model that can simulate charging transactions. A charging transaction contains the location at which a specific user charges and the times at which the user starts and ends the charging transaction. The main purpose of the SEVA model is to provide insights into the effects of incentives and rollout strategies by enabling scenario testing. Furthermore the SEVA model is modular to allow for extensions for further research into specific areas, as is done in Chapters 3 and 4.

In this chapter we will first set forth an overview of relevant literature in Section 2.1, from which we can conclude what aspects of the system the model should capture. We continue with a detailed description of the model in Section 2.2. Then we look at possible output metrics of the model and which of those can be used to validate the model in Section 2.3. In Section 2.4 we show an extensive sensitivity analysis on the model parameters and the model validation regarding both the behavior of EV users and the use of CPs in the model.

2.1 Review of Relevant Literature

This section offers a review of the relevant EV literature. We discuss the literature which captures how to measure effects of infrastructure optimization on the charging infrastructure in Section 2.1.1. These metrics can then be used to study the effects of infrastructure optimization, such as rollout strategies and incentives, through data analytics or computational models. The existing studies for both of these methods are set out in Section 2.1.2 and 2.1.3, respectively.

2.1.1 System Evaluation Metrics

System evaluation metrics are needed to quantitatively evaluate the effects of rollout strategies and incentives on the overall charging infrastructure using either data analytics or computational models. In this section we start by reviewing existing work on system evaluation metrics.

A crucial question is how to measure the performance of charging infrastructures [17]. As Helmus and van den Hoed [18] point out, measuring the number of kWh charged, the number of users at a CP and the number of transactions at a CP are frequently used to indicate the performance of the system. However, we should first discuss what it actually means for the system to perform better or worse. The definition of a good performing system may be very different when employing different points of view. To determine indicators for system performance, we therefore need to define when the system is performing better or worse. Helmus and van den Hoed [18] suggest that in order to find a good performance measure, we first need to look at what different stakeholders want. They provide an overview of different stakeholders along with their concerns or objectives and the desired results, as can be seen in Table 2.1. The municipality is the first stakeholder and is mainly responsible for the rollout of charging infrastructure. Its concerns focus on air quality improvement in a cost-effective manner. The EV users (and potential new EV users) want optimal access to charging infrastructure, while the non EV using residents the city (non EV users) are more concerned about the utilization of the charging infrastructure such that the parking pressure does not increase. We also have commercial parties that want to facilitate a positive business case. And lastly, the grid operators want to optimally manage the energy grid.

The performance of the charging infrastructure from the point of view of the stakeholders is key to indicating if their objectives are being met. However, the actual performance of the system is often unknown to stakeholders. How to measure this performance is also unclear. van den Hoed and Helmus [14] argue that these objectives and concerns provide insight in the performance metrics that are relevant. They continue with assigning a performance indicator for each of the goals, which are measured in the last column of Table 2.1. Note that the indicators differ considerably between the different goals. Thus measuring the performance of the charging infrastructure is highly dependent on how performance is defined.

Stakeholder	Concern / objective	Result indicators	Performance indicators
Municipality	Achieve sustainability goals in a cost-effective way	Air quality improvements due to charging infrastructure Climate change improvements due to charging infrastructure Achieved cost effectiveness of charging infrastructure	Number of kWh charged
EV users / candidates	Stimulate electric mobility by enabling charging	Increased accessibility of charging infrastructure Growth in number of users of charging infrastructure	Growth in capacity utilization, number of users, percentage of long chargers and charging time ratio
Residents (non EV users)	Optimize utilization of charging infrastructure and manage parking pressure	Increased level of utilization of charging infrastructure	Percentage of low utilized stations
Commercial parties in the EV chain	Facilitate a positive business case	CI-costs reduced charging infrastructure-benefits increased Business case charging infrastructure improved	Costs / benefits-ratio, percentage of CPs with positive business case, kWh charged per potential kWh charged.
Grid operators	Safeguard grid quality	Risks of power outage / grid-congestion reduced Smart charging options facilitated	Peak power level, percentage CPs with smart charging capability

TABLE 2.1: *The stakeholder concerns regarding public charging infrastructure from [18].*

Wolbertus and van den Hoed [19] show that the analysis of charging infrastructure utilization has become important in recent years. However, analysis is usually based on individual CPs and their characteristics. A broader view of the system performance could be achieved by looking at a whole region, especially as the density of the CPs is increasing.

Once system performance metrics are selected, they can be used to evaluate the effects of rollout strategies and incentives by using either data analytics or computational models. In the next sections we review existing work in both areas.

2.1.2 Infrastructure Optimization: Data Analytics

The majority of studies which analyze charging infrastructure rollout optimization use small amounts of real-world data because few datasets are available containing substantial amounts of real-world EV data. However, studies which use insubstantial data or data from controlled environments might not capture all essentials of the behavior of real-world EV users. The capturing of EV behavior is a necessity to analyze rollout strategies [4, 14].

To the best of our knowledge there currently exists only one large-scale real-world dataset, namely the CHIEF dataset [14]. It contains over 3 million charging transactions conducted by approximately 40 thousand EV users at around 5 thousand CPs located throughout metropolitan areas in the Netherlands. The Netherlands is considered one of the frontrunners in electric mobility [20], which makes analysis on this dataset relevant for most other countries willing to invest in charging infrastructure. Not only its sheer size sets this dataset apart, but also the fact that it contains different types of real-world EV users in completely uncontrolled environments. Examples of those user types are home users, fleet users, taxis and car sharing users. A detailed description of the dataset creation, data preparation and data filtering is described in [14] and [16]. The CHIEF dataset only contains public CPs, but still manages to capture more than 80% of all charging transaction in the Netherlands. Furthermore, Table 2.2 contains the data variables, examples and descriptions of the dataset.

	Value	Description
Transaction ID	123456	A unique ID of the transactions.
Location ID	1234	A unique ID of the CP.
User ID	ABCD1234	A unique ID of the user.
Start time connection	2016-06-15 18:23:11	The time and date at which the transaction started.
End time connection	2016-06-16 08:32:35	The time and date at which the transaction stopped.
kWh	8.9	The number of kWh that were transferred in the transaction.
City	Amsterdam	The city in which the CP is located.
PostalCode	1000AB	The postalcode-6 in which the CP is located.
Longitude	5.0000000	The longitude of the CP.
Latitude	52.0000000	The latitude of the CP.

TABLE 2.2: *An sample entry of a charging transaction in the CHIEF dataset [2, 4].*

The IDO-LAAD project [13] uses this CHIEF dataset. In van den Hoed and Helmus [14] a data analysis is provided regarding the public charging infrastructure (e.g. the

the number of charging transactions, number of unique users, amount of charged energy) in the city of Amsterdam using the CHIEF dataset. One of the results shows that car2go's (a car sharing scheme with electric cars in Amsterdam [21]) have a major influence on the infrastructure. They also found a lack of effective CP usage, as the ratio between charge times and connection times is low. Thus there are still opportunities for municipalities to implement incentives and policies in order to increase the effectiveness of the CP infrastructure, such as incentives for users to move their car once it is fully charged [14]. Furthermore, this study showed that the demand facilitation by the municipalities of Amsterdam (i.e. the public charging infrastructure development) positively effects the use of the charging infrastructure. But even with this increased usage of EVs, conventional cars are still the norm. A possible explanation for this is an underdeveloped charging infrastructure and limited battery capacity, which cause range anxiety among EV users and deter others from adopting EVs. The deployment of more CPs could increase the EV adoption and thus next to incentives and policies, rollout strategies are also essential [15].

There are multiple studies that focus on rollout strategies [9, 22]. However, these are limited in the sense that they look at the early phase of the CP rollout (100-500 CPs) or because only a small dataset is used. Spoelstra and Helmus [16] describe and analyze two rollout strategies deployed in the Netherlands using the CHIEF dataset, namely demand-driven and strategic strategies. A demand-driven rollout means that CPs are placed near people's homes after a request is made while a strategic rollout refers to the placement of CPs near public facilities and at other strategic locations. They conclude that both strategies are needed as demand-driven rollout provides infrastructure in residential areas whereas strategic rollout ensures that public locations have CPs [15, 16].

Yet another study [15] using the CHIEF dataset argues that CP failure needs to be considered in rollout strategies. Failure occurs when EV users cannot charge at their usual or preferred CP and need to find an alternative. Glombek and Quax [15] analyze the vulnerability of CPs and find that in terms of redistribution of the failure of a CP, vulnerable CPs are located mostly on the outskirts of the city. In comparison, in terms of the maximum number of EV users impacted by a failure, vulnerable CPs are location in the city center. This holds true for all analyzed cities (Amsterdam, Utrecht and Rotterdam). Glombek and Quax [15] recommend that additional CPs should be implemented in both the city center and on the outskirts of the city, as these areas are most vulnerable to failures.

To conclude, studies show that current rollout strategies increase EV adoption, but that rollout strategies may be improved to avoid over-capacity and under-capacity. The studies also show that future strategies should focus on placing CPs in both residential areas and public locations, while at the same time decreasing the number of vulnerable CPs both on the outskirts of cities and in the city centers. Furthermore, the charging infrastructure can also be optimized without adding new CPs, namely through the implementation of incentives and policies. This may result in a more effective usage of existing CPs.

Given that the previously mentioned studies focus on data analysis, they are not able to predict which rollout strategies and policy incentives will work best in the future. To achieve this predictive models are needed, which is what the next section will focus on.

2.1.3 Infrastructure Optimization: Computational Models

Various papers in the field of simulation and modeling studies discuss EV charging infrastructure or EV behavior, each with its own focus. Sweda and Klabjan [6] and Hess et al. [5] are concerned with minimizing charging costs, Zhu et al. [7] focus on improving the adoption of EVs, and Uhrig et al. [8], Xi et al. [9], Yi and Bauer [10], Momtazpour et al. [11], Paffumi et al. [12], Daina et al. [23] all focus on optimizing the energy demand.

Table 2.3 shows an overview of the modelling papers that are available in our field. The objectives, modeling methods, data usage and factors influencing EV user behavior of these papers are presented. When we consider the column summarizing the data usage, it is remarkable that none of the models uses large amounts of real-world EV data. Mainly road data, population data, parking data, stated responce data or driving data of fossil-fuel vehicles are used. Several of the authors point out the lack of EV data as limitation [6, 11]. We typically see either models that use data about fossil-fuel vehicles and make assumptions about the behavior of EV users or models which use stated response data about EV use. The former contain biases which cannot be validated without available data of EV usage and the latter carry biases inherent to hypothetical choice situations [23]. To the best of our knowledge, currently no EV model exists that is validated using large amounts of real-world data about EV usage (i.e. revealed preference data).

Another interesting column contains the factors that influence the behavior of EV users. A wide range of factors are used to model this behavior. It is worth looking into the literature on EV user behavior in more depth, since the behavior of the EV user is at the core of modeling EV systems.

This importance of understanding charging behavior is pointed out by Azadfar et al. [24], Spoelstra [2], Banez-Chicharro et al. [25], Franke and Krems [3], Sweda and Klabjan [6] and Helmus [26]. They each state that understanding charging behavior is essential

Objective	**Modeling method**	**Data usage**	**Factors influencing charging behavior**	**Source**
Improving EV adoption	Agent-based model	None	State of charge	Hess et al. [5]
Improving EV adoption	Agent-based model	U.S road, population, workflow and zip code tabulation area data	Vehicle price, fuel cost, personal greenness, social influence, long distance penalty and infrastructure penalty	Sweda and Klabjan [6]
Minimizing charging costs	Genetic algorithm	Road data	Distance to destination and charging cost	Zhu et al. [7]
Optimizing energy demand	Monte Carlo simulations	Car park arrival and departure data	Arrival and departure times	Uhrig et al. [8]
Optimizing energy demand	Discrete event simulations	Using Ohio data containing information about demographics, socioeconomic data, vehicle ownership and usage (not EV specific)	EV arrivals, departures, and state of charge	Xi et al. [9]
Optimizing energy demand	Mathematical optimization model	Population and road data	Distance to destination, charging service quality and transportation energy consumption	Yi and Bauer [10]
Optimizing energy demand	Clustering techniques and network models	Synthetic data people and activities and data about electricity consumption	State of charge	Momtazpour et al. [11]
Optimizing energy demand	Agent-based model	Driving data (by fuel vehicles) of one month in parts of Italy	Time, preferences / history, state of charge, power of CP	Paffumi et al. [12]
Optimizing energy demand	Agent-based model	Stated response experiments	State of charge, prices of CPs, charging speeds of CPs	Daina et al. [23]

TABLE 2.3: *The objectives, modeling methods, data usage and factors influencing EV user behavior of various existing EV models.*

to optimize the charging infrastructure and to promote a more efficient utilization of the infrastructure. Sweda and Klabjan [6] and Helmus [26] furthermore state that this understanding will allow charging behavior to be simulated in an agent-based model, which can in turn help find optimal rollout strategies.

Many studies have been performed to analyze charging behavior of EV users [26]. The behavior of EV users is analyzed in pilot tests [27] and semi-controlled environments [4]. Most behavior studies focus on either the psychological side of charging behavior, making use of inquiries, interviews and stated response experiments [23, 28, 29], or the effect of EVs on the power grid [30]. The former look at behavior as a decision making process, while the latter capture behavior as the charging profiles of EV users [26]. We see approaches such as defining strategies for EV users. For example, De Gennaro et al. [31] define the charging behavior of an EV user as a portfolio of charging transactions. However this study focuses mainly on the traveling behavior and less on the charging behavior itself. Helmus [26] also states that the mean behavior of EV users might be a bad estimate of their behavior. An example of this is an EV user who mainly charges at one of two charge times (8am or 6pm), but never at the mean of those two times (2pm).

Helmus and van den Hoed [4] argue that to fully understand charging behavior, we need to acknowledge that there are different user types such as residents, commuters or city visitors and that these different user types exhibit different charging behavior. In their study they show the six different users types they found in the CHIEF dataset, namely commuters, car sharing scheme users, taxis, visitors and two types of residents. The typical activity patterns of the user types is seen in Figure 2.1. This could define charging behavior to some extent, but it is likely that more dimensions of charging behavior need to be considered to fully capture the complex behavior of EV users. The next paragraphs will set out the current research about the various charging behavior dimensions and their corresponding influencing factors.

Franke and Krems [3] use the user-battery interaction style as main variable for EV user behavior. This variable provides a balance between how much users are influenced by either their state of charge or the time between transactions. With a low user-battery interaction style the time between transactions is the key factor and state of charge is less important.

Spoelstra [2] performed a literature review to address the various factors which influence charging behavior. These can be grouped into three categories, namely driver related factors (range anxiety, planning, mobility pattern and EV experience), vehicle related factors (battery size, vehicle range and vehicle type) and environment related factors (CP density). Spoelstra [2] also states that charging behavior itself can be conceptualized by

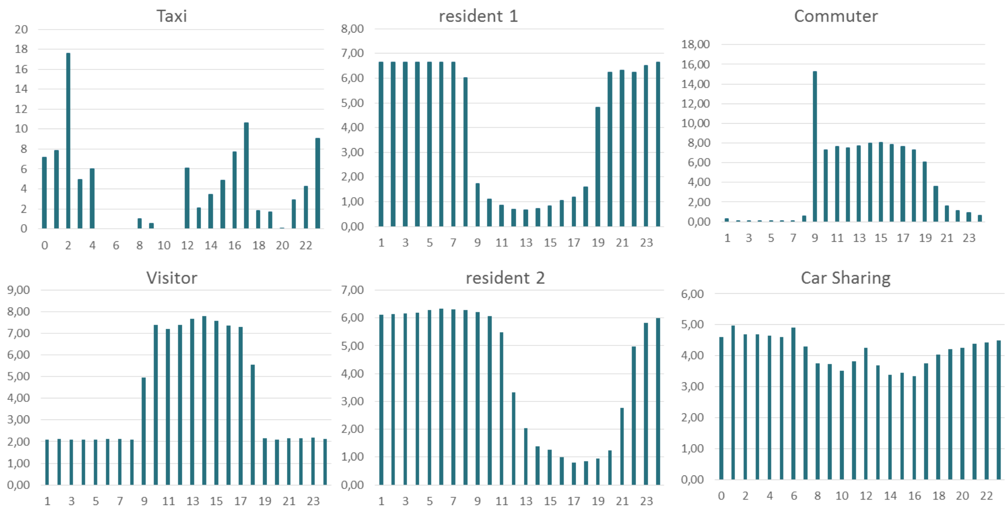


Figure 2.1: *Six illustrations of typical activity patterns found in the CHIEF dataset [4]. The horizontal axis present the hours of the day and the vertical axis show the intensity of activity.*

six dimensions, namely CP location, CP type, charging frequency, time of day, charging duration and energy transfer.

Azadfar et al. [24] also investigated the factors which influence charging behavior, and the dimensions of this charging behavior. They consider the time of day, the duration of the charging transaction, the frequency of charging and the energy required to charge the vehicle batteries. They state that factors which influence this behavior are the EV penetration rate, the charging infrastructure, the battery performance of EVs, the costs and incentive programs. They identify two key factors which most strongly influence charging behavior, namely the charging infrastructure (environment related) and the battery performance (vehicle related).

According to Helmus and van den Hoed [4] charging behavior can be defined as the successful result of the intentional behavior to charge a specific EV at a specific CP for a specific duration. The dimensions of charging are therefore time of charging, location of charging and duration of charging. This paper considers these three influencing factor all as relevant influencing factors.

Helmus [26] elaborates about these factors by naming the when (start connection time, end connection time, connection profile and time between transactions), the where (distance of subsequent transactions, CP volatility, neighborhood volatility and city volatility) and the what (initial state of charging, kWh charged) as relevant dimensions of charging behavior. Surprisingly, they only name driver related factors as factors which influence these dimensions, as their definitions implicitly focus only on psychological processes to make a decision rather than on physical factors such as the environment or the vehicle.

In summary, factors which influence charging behavior can be categorized into three groups, namely driver related factors, infrastructure related factors and vehicle related factors. A summary of the specific factors within these groups is presented in Table 2.4. Furthermore, Table 2.5 shows which influencing factors and dimensions of charging behavior are mentioned in literature. Here we see a trend that authors only mention six dimensions, namely when (time of day), where (location), what (charging duration, interarrival duration), energy transfer and CP type. We leave out the charging frequency factor, simply because this can be inferred from the charging duration and interarrival duration.

Influencing factors	**Examples**
Driver related	EV experience, degree of trip planning, degree of charging planning, social interaction, personality traits.
Environment related	CP area, CP density, parking pressure, ratio of types of CP, infrastructure policy, charging infrastructure, charging opportunities.
Vehicle related	Vehicle type, battery size, range, consumption.

Table 2.4: *The factors which influence charging behavior [2, 4, 24].*

Author	**Year**	**Influencing factors**	**Dimensions**
Franke and Krems [3]	2013	Driver, vehicle and environment related	Time and state of charge
Spoelstra [2]	2014	Driver, vehicle and environment related	Charging time of day, CP location, charging duration, CP type, charging frequency and energy transfer
Azadfar et al. [24]	2015	Vehicle and environment related	Time of day, duration, frequency and electricity required
Helmus and van den Hoed [4]	2016	Driver, vehicle and environment related	Time, location and duration
Helmus [26]	2016	Driver related	When, where and what

Table 2.5: *A comparison of the changing definition of charging behavior.*

The two other influencing factors, namely the amount of energy transferred and the state of charge are also left out in our model. This is due to research by Spoelstra [2], which shows that EV users have routine charging behavior, and that many EV users have similar charging routines. This is furthermore confirmed by the work of Smith et al. [32]. According to Spoelstra [2], routine behavior occurs because EV users perceive little range anxiety in their routine travels, and therefore are not constantly monitoring their state of charge. Contrary to Eggers and Eggers [33], Spoelstra [2] furthermore states that EV users only actively plan their charging behavior if ‘their mobility pattern was both unpredictable and common trip distances were long. In other situations, they relied on routine, trust and the predictability of their mobility’. Furthermore, EV users tend to use CPs which they know and have used before, and they usually do not deviate from their preferred CP(s). Lastly, Spoelstra [2] found that often the duration of charging transactions is longer than required. This corresponds to the ‘low user battery interaction’ concept set out in Franke and Krems [3]. All in all, we can conclude that EV users are creatures of habit, and that their state of charge does not play a large role in their charging behavior and their charging decisions.

Helmus and van den Hoed [4] hypothesize that combining the connection patterns in Figure 2.1 (when and what) with probabilities of which locations the users are likely to charge at (where), will enable us to predict how CPs will be used. This in turn can help analyze the effects of rollout strategies and incentives on the charging infrastructure. This hypothesis further confirms our notion that the dimensions where, when and what are sufficient to capture the complex charging behavior of EV users.

2.1.4 Conclusion

In this section we have shown that system evaluation metrics are needed to quantitatively evaluate the effects of rollout strategies and incentives on the overall charging infrastructure using either data analytics or computational models. We have furthermore summarized the existing literature regarding both data analytics and computational models.

Through this literature overview we set out to create a road map for designing simulations which model charging behavior in order to study the effects of rollout strategies and incentives on system performance metrics. This in turn can help policy makers optimize their charging infrastructure. We present our five main conclusions. Firstly, from the point of views of different stakeholders, the definition of a good performing charging infrastructure can be very different. Therefore, result indicators and performance indicators, matched with the appropriate stakeholder, can offer a good way of evaluating infrastructure optimization. Secondly, large amounts of real-world data are essential to analyze infrastructure optimization. Thirdly, to help policy makers optimize their charging infrastructure, predictions about the effects of infrastructure optimization are a necessity, which means simulations and computational models. Pure data analytics is not enough. Fourthly, these computational models need to be either based on or validated with large amounts of real-world data in order to test model assumptions. Fifthly, the three dimensions when, where and what are able to capture the complex behavior of EV users within such a computational model. This literature review enables the creation of behavioral computational models to study the effects of infrastructure optimization on system performance metrics in order to help policy makers to optimize their charging infrastructure.

2.2 Model Description

This section presents a comprehensive description of the SEVA model. In the following subsections we will gradually go into more detail about the workings of this model, following the ODD (Overview, Design concepts, Details) protocol [34].

2.2.1 Overview: Entities, State Variables and Scales

As with any agent-based model the environment and the agents are the two main entities in the model. The environment is defined as the collection of all CPs together with all the information about these CPs (i.e. their spatial location, whether they are occupied and their placement date). Agents are defined as EV users together with their EVs. Every agent is aware of the environment, meaning they know where CPs are located and whether they are occupied. The only form of communication between agents in the system is via the occupation of CPs. If several agents are connected at a CP such that all sockets[1] are taken, then no additional agent can connect to this CP.

	State variables	**Description**
Agent	ID	Unique identifier.
	Given charging transactions	List of charging transactions from the given data.
	Connected	Indicates if an agent is connected to a CP.
	Time next activity	Date and time of the next activity.
	Active center	Location of the active cluster. Note that this is the center the agent is connected to or the center the agents plans to connect to next.
	Active CP	ID of the active CP. Note that this is the CP the agent is connected to or 'none' if the agent is disconnected.
Environment	CP occupation	This variable stores whether each socket of each CP is occupied (and by which agent).
	CP meta-data	Meta-data about each CP, for example the location (longitude, latitude) where the CP is located.
Simulation handler	Agents	The agents contained in this simulation.
	Current time	The current time of the simulation.
	Sensors	Sensors to keep track of system metrics.

TABLE 2.6: *The entities of the model with their state variables.*

The overall simulation is managed by a simulation handler. This manager controls the actions of the agents and also the interactions between the agents and the environment.

[1]Note that each CP has two or more sockets and thus multiple agents can be connected to a single CP at the same time.

Furthermore, this simulation handler has a collection of observers which keep track of metrics for the experiments and system evaluation, such as the average simulation time and the agent validation scores. Thus the model contains three entities, (1) the agents, (2) the environment and (3) the simulation handler. These entities and descriptions of their state variables can be found in Table 2.6.

In the literature we find that EV users are extremely habit-based and frequent only a limited number of CPs routinely[2, 32]. In order to capture this part of charging behavior, we define agents to have several areas which they frequently charge in. In these areas agents regularly display the same type of activity. Each area contains one or more CPs, which the EV user had used in the data, forming a *cluster* of the agent. The details on how to determine the clusters of an agent are discussed in Section 2.2.6. Until then it is sufficient to know that each cluster is unique to a single agent and contains CPs which are close together and at which the agent exhibits similar charging behavior. Even if two agents have clusters with the exact same CPs, we still view them as different clusters.

The *center* (c) is the average of the locations of the CPs (p) in the cluster, weighted by the number of charging transactions of the user at each CP in the data. This center is an approximation of the location of the real destination of the user. The centers can, for instance, represent the EV user's home or work location, or any place where the user will charge their EV frequently. Implicitly this assumes that distance to the destination is the determining factor in choosing a CP. A schematic diagram of a cluster with its CPs can be seen in Figure 2.2. Note that the center is located closest to the CP (p_1) with most transactions by the user. We formalize the calculation of the longitude (c_{lon})

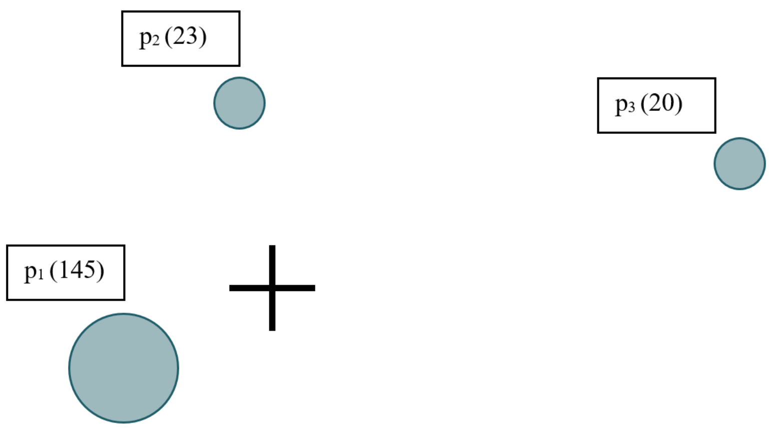


FIGURE 2.2: *An example of a cluster of an EV user. The center (indicated by the cross) shows the weighted average of the CP locations. The cluster contains the three CPs p_1, p_2 and p_3. Each CP has the number of transactions at that CP indicated between brackets.*

and latitude (c_{lat}) of the center with the following equations:

$$c_{\text{lon}} = \frac{\sum_{p \in c} p_{\text{lon}} \cdot p_{\text{tr}}}{c_{\text{tr}}} \tag{2.1}$$

$$c_{\text{lat}} = \frac{\sum_{p \in c} p_{\text{lat}} \cdot p_{\text{tr}}}{c_{\text{tr}}}. \tag{2.2}$$

Here p_{lon} and p_{lat} are the longitude and latitude of the CPs in the cluster. The number of transactions at a CP and within a cluster are indicated with p_{tr} and c_{tr}, respectively.

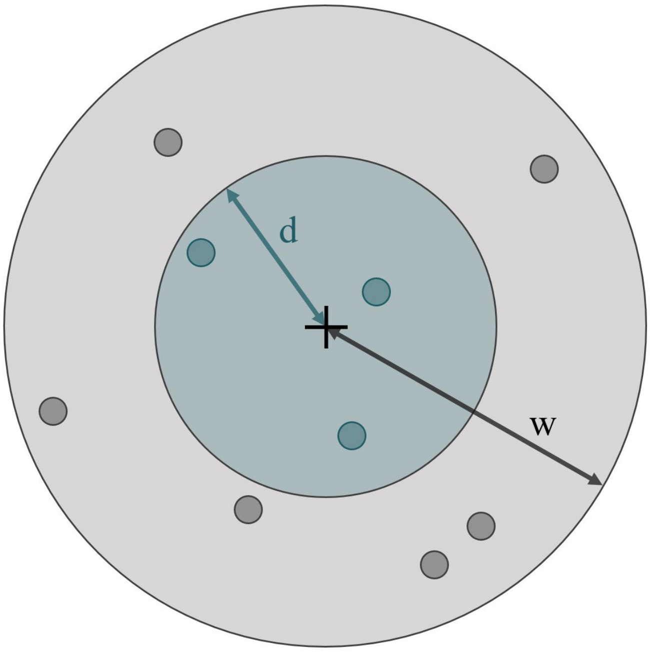
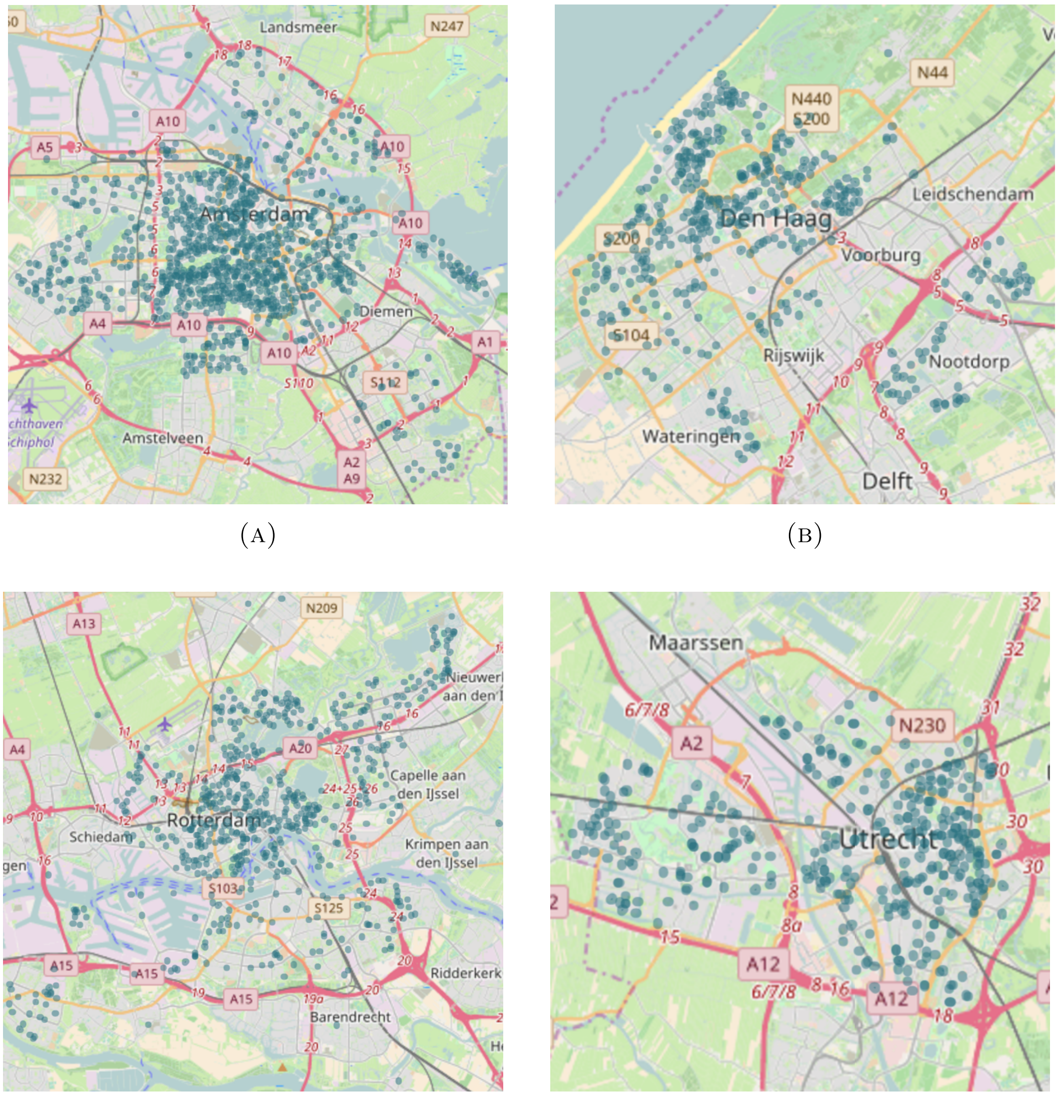


FIGURE 2.3: *A cluster of an agent, where the cross indicates the center, the small blue circles indicate previously visited CPs and the small gray circles indicate previously unvisited CPs within range of the walking preparedness.*

Figure 2.3 shows an example of a cluster of an agent. The *maximum distance* d of an agent is defined as the maximum of the distances between the center and any of the CPs in the cluster plus 10%. The *walking preparedness* w of an agent is defined as the maximum of d and the minimum radius (with a default value of 150 meters as can be seen in Table A.3). Thus if the maximum distance d is less than the minimum radius then the walking preparedness w is equal to this minimum distance. If d is greater than the minimum radius, then the w is equal to d. An agent only has one maximum distance and one walking preparedness, even if it has multiple clusters. The largest maximum distance (and walking preparedness) over all of the agents clusters is taken and applied to all of the agent's clusters. While the centers are fixed for an agent throughout a simulation, the clusters of an agent might expand to contain new CPs within the walking preparedness of an agent if the agent uses those during the simulation.

For each agent, its clusters and the behavior (captured in the dimensions of what, when and where) it exhibits in the cluster is purely extracted from the data. This means no assumptions need to be made about either the centers or the behavior exhibited at their clusters. We do not need to specify certain centers to be home or work locations and we do not need to define what behavior is exhibited at these specified location.



(A) (B) (C) (D)

Figure 2.4: *The CPs in the dataset plotted on a map for the four cities contained in the dataset, namely* (A) *Amsterdam,* (B) *The Hague,* (C) *Rotterdam and* (D) *Utrecht.*

All experiments are run with longitude and latitude combinations that fall within the boundaries of the Netherlands because the dataset is from this country. However, the spatial scale of the model is continuous and all valid combinations of longitude and latitude can be used. In Figure 2.4 the (majority of the) CPs in the dataset are plotted on a map, showing the scope of the CP locations.

The temporal scale is discretized using bins of T minutes. T is an input parameter of the model and the value for this variable can be found in Table A.2. The simulation can be run for a chosen number of days, months or years. The exact start date and stop condition that are used for the experiments and analysis are considered input parameters. They can be found in Appendix A (Table A.1).

With the increasing popularity of EVs the number of users and CPs in the dataset also increases over time. In Figure 2.5 we can see this growth. For each month we show the number of active EV users and used CPs in the dataset.

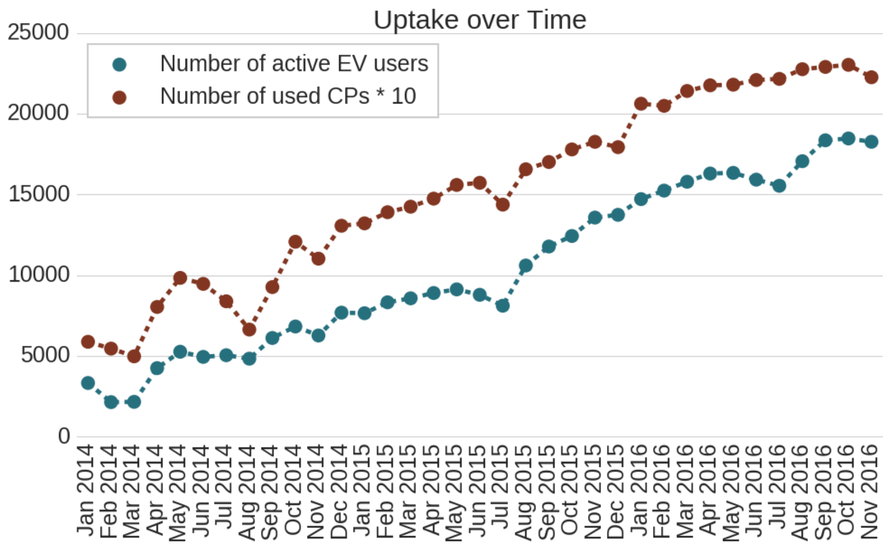


FIGURE 2.5: *The growth of the number of EV users and CPs in the dataset over the years.*

2.2.2 Overview: Process Overview and Scheduling

Agents in the system can be in one of three states, namely *connected*, *disconnected* or *selecting CP*. The transitions between these states are controlled by the processes *connection*, *disconnection* and *CP selection*. The execution loop connecting the states and processes can be seen in Figure 2.6. When a connected agent executes its next activity, it will disconnect using the disconnection process. When an agent is disconnected, this agent will choose a cluster using the connection process and then it will try to choose a CP to connect to by using the CP selection process. Table 2.7 provides an overview of the processes. A more extensive description on the processes is given in Section 2.2.7.

The simulation handler controls the actions of the agents by storing all agents in a time ordered queue, where the time of an agent is the time of its next activity. The simulation handler sequentially pops agents from this queue. Once an agent is popped, it executes its next activity (either a connection or a disconnection activity) and then recalculates

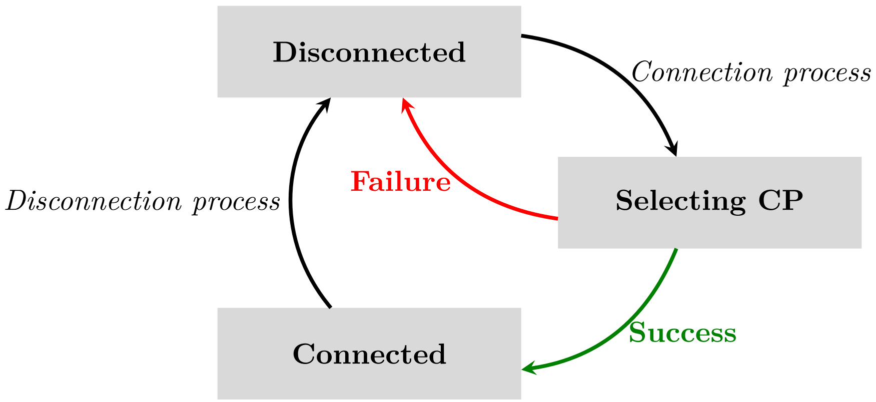


FIGURE 2.6: *The activity loop of an agent. Depending on the state of the agent (the grey boxes) the different processes are called. Note that the colored lines indicate the two possible outcomes of the CP selection process, namely success or failure.*

Process	**Input**	**Output**	**System interaction**
Disconnection process	Time at connection and cluster of connection.	Time at disconnection.	Removes agent from system at time of disconnection.
CP selection process	Cluster of connection.	CP of connection.	Adds agent to system at time of connection.
Connection process	Time at disconnection and cluster of previous connection.	Time at next connection and cluster of next connection.	-

TABLE 2.7: *An overview of the input, output and system interactions of each process.*

the time of its next activity. The simulation handler then pushes the agent back into the ordered queue.

2.2.3 Design Concepts: Data-Driven

The dataset used for this data-driven model is called the CHIEF dataset and has already been introduced in Section 2.1.2. An example of a charging transaction in this dataset can be seen in Table 2.2. Additional information about (the creation of) the dataset can be found in [14] and [16]. The size of the dataset as well as the geographical diversity provides a reliable base for understanding the factors influencing the charging behavior of EV users in the Netherlands [2]. Section 2.1 has furthermore shown why this dataset is not only unique but necessary for any study regarding EV behavior and modeling.

2.2.4 Design Concepts: Observations

Simulated charging transactions for each agent are the main output of the model. These transactions can then be analyzed and transformed into other required outputs. Activity patterns are created using the charging transactions summarizing the behavior of agents. An activity pattern captures the activity of a (group of) agents and/or CPs over the 24 hours of a day. It can be constructed in the following way. First a day is split up in bins of a fixed size. For the SEVA model we have a default bin size of 20 minutes[2], as can been seen in Table A.2. Each bin then holds the number of transactions that took place in this time interval. Thus if we want to create the activity pattern of a single agent, we would look at each of its transactions and add one to each bin that overlaps with the transaction period. The same can be done for the CPs. We can also define the activity pattern of a group of CPs, and simply take into consideration all transactions that occurred at a CP in this group. Comparing the simulated activity patterns of agents with their 'real' activity patterns is a method of model validation.

In addition to this validation metric, the model also contains clustering, competition and run time metrics (see Table 2.8). The clustering metrics contain information about the clusters of the agents. Competition metrics measure how much competition occurs in the system. To improve efficiency and understand the effects of certain parameters on simulation time, we implemented the possibility of measuring the run time of certain aspects of the simulation. We will go into more detail for each of the metrics in Section 2.3.

2.2.5 Design Concepts: Behavior

The most important internal design choice is how to model all dimensions of the charging behavior of the agents. To minimize the assumptions and to make optimal use of the dataset available, we constructed a data-driven method of modeling the behavior of the agents[3]. Given where and when an agent has previously charged, we extract probabilities from the data of where and when the agent will charge next. Having the luxury of a large dataset with EV charging transactions is a big advantage that is not seen in literature (as shown in Section 2.1). We assume that users do not change their behavior frequently and thus we determine when and where an agent will next charge based on its history. This yields the idea of a Markov model where clusters form the nodes in the system and connections between nodes indicate the probability of going to that node. Each cluster (or node) has arrival probabilities. This provides us with a way of

[2] An extensive explanation on the choice of this value is presented in Section 2.4.2.

[3] It would also be possible to do rule-based modeling. The reason this is not done is that it would mean trying to interpret *why* users do what they do while the data only tells us *what* they do.

	Metric	Description
Clustering	Number of CPs	The total number of CPs in the simulation at which at least one agent in the system has had one or more charging transactions.
	Number of clusters	The number of clusters for every agent in the simulation.
	Maximum distance (d)	For every agent in the simulation the maximum distance between a cluster and any of its CPs.
	Walking preparedness (w)	The walking preparedness for every agent in the simulation, defined as the maximum distance plus 10% with a minimum of the default value (see Table A.3).
	Number of CPs per cluster	The average number of CPs per cluster for every agent in the simulation.
Competition	Number of agents per CP	For every CP the number of agents that have been charging at that CP at least once during the simulation.
	Selection process attempts	For every agent in the simulation the consecutive failures and successes of the selection process, where a fail indicates that no CP within the desired cluster could be selected at the preferred time due to every CP being occupied.
Run time	Run time per agent initialization	The run time for initializing a single agent, for every agent in the simulation.
	Run time per simulation	The run time for a complete simulation.
Validation	Validation error per agent	For every agent in the system the validation error (see Section 2.3 for details).
	Validation error per CP	For every CP the validation error (see Section 2.3 for details).

TABLE 2.8: *The optional output metrics of the model.*

deciding *where* an agent will appear next. However, we also need to determine *when* the agent would appear. This too, can be extracted from the data when looking at the inter-arrival times and connection times of each user. Each agent has one or more clusters, specific to that agent, and each of these clusters has three types of distributions: (1) connection duration distributions, (2) disconnection duration distributions and (3) arrival distributions. Samples from these distributions approximate the behavior of the agents. More detailed explanations of these distributions will be discussed in the next section.

2.2.6 Details: Initialization

This section will describe the process of initializing the agents and the environment. The simulation handler receives the input parameters as specified in Table A.1, A.2, A.3 and A.4. While these tables contain default values[4] for each of the input parameters, the values can be changed for a simulation run. Whenever we deviate from these default values, it is explicitly stated. With the use of these parameters the raw data is loaded and pre-processed. In the preprocessing we remove invalid entries from the dataset, add information about the number of sockets per CP and the parking zone of a CP, merge CPs at the exact same location (i.e. charge hubs) to a single CP and split it into training data and test data based on the start and end dates for both training and test in the input parameters. Next, an instance of the environment is created, where all CPs present in the dataset are loaded and set to be unoccupied. The meta-data (e.g. longitude, latitude, number of sockets and placement date) of these CPs is also stored. For details on the meta-data refer to Section 2.2.3. Next, instances of the agents are created according to the agent selection method specified in the parameters (see agent selection method in Table A.1). When an agent is initialized, it first gets a unique ID. The next step is to capture, summarize and store the behavior of the agents. This is done by generating the clusters and distributions of the agent using the charging transactions of the agent. We will go into more details about this process in the next sections.

Clustering: the Creation of the Clusters

For each individual agent we cluster the CPs this agent uses in the dataset into one or more clusters. Clusters are formed with CPs that are close together (physical location) and where the agent exhibits similar behavior (activity patterns). The clustering requires several parameters, which are all listed in Table A.4.

We consider all CPs which the agent has visited more than s_{cs} (on default 10) times in the training data. For each of those CPs we determine the activity pattern of the agent at this CP as well as the longitude and latitude. This data is the input for the clustering algorithm. However, we need some pre-processing to ensure that sensible clusters are created. First we want the longitude and latitude to be in the same range of values. Therefore we shift the longitude with h (see Table A.4), which is the mean of the latitude values of all CPs in the data minus the mean of the longitude values.

Intuitively CPs will now be clustered when they have similar activity patterns, longitude and latitude. This means that CPs which are physically close to each other and where

[4] The default values for the input parameters are motivated in Section 2.4.

the user exhibits the similar temporal behavior will be clustered together. However, to increase the influence of distance (rather than the temporal behavior) on the clustering we also multiply the longitude and latitude with a scaling factor f. The resulting data is clustered using the Birch algorithm without a pre-specified number of clusters [35]. The input parameters for the algorithm were tuned to the values found in Table A.4. A sensitivity analysis and argumentation for these values is given in Section 2.4. Lastly we check all of the resulting groups of CPs for having at least s_c (on default 20) charging transactions and at least a fraction of f_c (on default 0.08) of the total number of transactions of the agent. Groups of CPs which satisfy these criteria form the clusters of the agent. Agents are only valid, and thus present in the simulation if they have at least one cluster. A center itself is located at the mean location (longitude and latitude) of all CPs within that cluster, weighted by the number of transactions at each of the CPs as explained in Section 2.2.1.

Summarizing Behavior: the Creation of the Distributions

The next step of the initialization is the creation of the various distributions of the agent. As mentioned in Section 2.2.5 every agent in the simulation has three types of distributions. An agent has several distributions of the same type, namely

- An arrival distribution per cluster;
- A connection duration distribution per cluster per bin (time interval with length of the bin size);
- A disconnection duration distribution per bin.

The precision, and thus the number of discrete intervals in each distribution, is determined by the bin size (see Table A.2). This bin size determines the length of each bin for every distribution in the model. The disconnection duration distributions take into account the disconnection durations appearing in the data for an agent. For each bin we have a disconnection duration distribution, i.e. if the agent starts disconnecting at 1:05pm it will call upon the disconnection distribution of the 1:00pm to 1:20pm bin. Each of those distributions has bins with durations (i.e. a bin containing disconnection durations of 0 to 20 minutes, 20 to 40 minutes, etc.). The value in such bin is the number of occurrences in the data for that disconnection duration seen at the disconnection time. To clarify, a disconnection duration of 30 minutes starting at 1:05pm would add to the value with a duration of 20 to 40 minutes (if the bin size is 20 minutes) of the disconnection distribution only if this specific disconnection distribution belongs to the bin containing the start time 1:05pm. In the end we normalize each distribution and thus get the disconnection duration distribution for the agent. When the agent disconnects from a CP, a sample is drawn from the disconnection duration distribution of the agent belonging to the time of disconnection to find the length of the disconnection and thus determine the time at which the agent will start its next connection. The connection duration distributions are similar, except that now the duration of transactions are considered instead of the duration of the disconnection and the connection distributions are cluster specific, meaning that for each cluster we have a set of distributions.

The clusters of every agent have one more distribution, namely the arrival distribution. The unit of this distribution is time and the bins indicate a time interval within a day (for example between 1:00pm and 1:20pm). The value of that bin indicates the number of occurrences with which an agent connected to any CP in that cluster in the data. This distribution is used to sample a start time of connection at the initialization of the simulation (see Section 2.2.6 on how this is used). Note that while the arrival distribution has a range of 24 hours, the connection and disconnection distributions have a range that depends on the maximum duration of (dis)connection in the data and can thus exceed 24 hours.

Initial State

Once the behavior of the agents is captured, we can move on to the last step of the initialization of an agent, which is to determine its initial state. This initial state sets the remaining state variables of the agent: is connected, time next activity, active center and active CP (see Table 2.6).

First we determine whether the agent is connected at the beginning of the simulation. For this we calculate the overall (normalized) activity pattern of the agent and take the value of the bin in which the start time of the simulation falls. Because this value is normalized, it is equal to the probability that the agent is connected at the start time. Thus the agent is connected if a randomly drawn number is smaller than this probability. What remains is to determine when the agent will (dis)connect and where the agent is connected (if it is connected) or will connect to (if it is disconnected).

When the agent starts out disconnected we need to set a time and place for it to first connect. To do this, we choose a cluster with probabilities according the number of transactions in the data for each cluster. Once we have this cluster, we sample an arrival time from that cluster's arrival distribution. This gives us the time of the next activity of the agent, namely connecting to a (yet to be determined) CP at said cluster.

When the agent starts out connected we want to decide where this agent is connected as well as when its next activity of disconnecting will take place. For each center we count the number of occurrences when the agent was connected at this time of day (in practice this means getting the value of the activity pattern of each cluster). These values are then used as the probabilities with which we select a cluster. For this cluster, we sample an arrival time from the cluster's arrival distribution. We now ensure this arrival time is in the 24 hours previous to the start of the simulation (as that was likely when the agent connected to the cluster where he is now connected). A CP is then selected using the selection process as described in Section 2.2.2 (and in more detail in Section 2.2.7). Then we sample from the cluster's connection duration distribution to determine when the agent will disconnect from this CP and set this as the time of the next activity. Note that this next activity time might be before the actual start time of the simulation, however this is not a problem as we simulate according to which activity is earliest. This ensures that the activity that is scheduled first, will be handled first irrespective of whether this is before or after the given start time of the simulation.

The last step of the initialization is to create the ordered queue containing the activity times of the agents. It is implemented using a heap data structure. This data structure contains all the agents in the simulation together with their next activity times. It can efficiently pop the activity that is up next and push the newly calculated time of the next activity onto the heap.

2.2.7 Details: Submodels

The modularity of the SEVA model is very strong as each process (connection, disconnection and CP selection) has its own internal working, independent of the rest of the simulation. Therefore each individual process can be adjusted, updated and improved without interference from the rest of the processes. Also new processes, like state of charge modeling, may be added if this is desirable.

We will go into more detail about each of the three processes: the connection process, the CP selection process and the disconnection process. An overview of the input, output and system interaction of each of the three processes has already been showed in Table 2.7. Refer back to Figure 2.6 for the relation between the different states of the agent and the processes.

The *connection* process determines when and at which cluster an agent will connect. Once the next cluster has been determined, the *CP selection* process determines at which CP in the cluster the agent will connect. When a CP has successfully been selected (indicated with the green arrow in Figure 2.6), the agent becomes connected and the next process to call upon is the disconnection process. When the CP selection process fails (the red arrow in Figure 2.6) the agent stays disconnected and the connection process is called upon again after t_r (on default 20 minutes) time (see Table A.3).

The *disconnection* process determines when an agent will disconnect from its current CP. To disconnect the agent, the socket of the CP to which the agent was connected will be freed. This means that socket is then available for (other) agents.

Connection Process

In the connection process we first decide *when* the agent will appear next before we decide on *where* this will be. This is because an EV user is tied to charging every so often to ensure its battery will not run out. When an agent calls upon this process, it is by definition in a connected state. First the state of the agent is changed to disconnected and the environment is updated such that the socket of the CP at which the agent was connected becomes free again. After this we need to decide where the agent will appear next and at what time this will happen. We also need to calculate how long the agent will stay disconnected.

We sample from the disconnection duration distribution belonging to the current time bin. This will yield the duration of the disconnection and combined with the current time we determine the time at which the agent will connect. However, we also need to check if this time is valid for that agent. For example, if an agent is known to connect during work hours at a certain location, it would not make sense for the agent to connect there at 2am. We consider a time to be valid if within the bin of that time, the agent is known to have disconnected from the cluster in the training data. If this is not the case, we check the bins preceding and succeeding the bin we originally chose. We have the extra restriction that the connection duration should be at least the bin size (see Table A.2). If the newly chosen bin is still invalid, we check the second bins preceding and succeeding the originally chosen bin. And so on, until we find a valid bin. Once a valid bin is selected, we set the time of the next activity (namely connecting to the selected cluster) of the agent to be the current time plus the found duration.

With the time at which the agent will be connecting determined, we next determine *where* the agent will be connecting. Given the time of connection, we give each of the clusters of the agent a probability according to the frequency with which it arrived at the cluster at that time. We then draw a cluster according to those probabilities and this will be the cluster we will connect to next. The active cluster of the agent is updated to be this new cluster.

CP Selection Process

The CP selection process determines which CP an agent should connect to when we already know the cluster it will connect to from the connection process. It is called upon at the time when the agent wants to connect again, as decided by the connection process. With a chance determined by the habit probability parameter p_h, with default value 0.4 (see Table A.3), the agent will select a CP based on habit. That is, a CP is selected with a chance directly proportional to the number of transactions the agent had at that CP in both the training data and in the simulation up to that point.

With a chance of $1-p_h$ the next CP to select is selected based on distance. That is, the CP closest to the center (while still within the range of the agent's walking preparedness) with an available socket is selected. It is possible that no CP could be selected using the habit method because none of the agent's CPs were available. In this case we also select the CP based on distance.

In the end there are two possible options, either the agent has selected a CP successfully or it failed to select a CP. If the agent was successful its state is changed to connected and a socket of the CP to which it connects becomes occupied[5]. If the agent was unsuccessful this means that all relevant CPs are at the moment occupied and the connection process will be called upon again after t_r (20 minutes on default) time (see Table A.3).

Disconnection Process

The last process is the disconnection process and determines when an agent will disconnect again once it is connected to a CP. This process is called upon right away after a successful CP selection.

We sample from the connection duration distribution that belongs to cluster the agent is connected to and the bin of the current time in order to get a duration for the current connection (and thus decide when to disconnect). The time for the next activity of the agent will then be the current time plus the sampled duration. We check the next activity time on validity in the same way as we have done in the connection process for the time of connection.

[5] The socket to connect to within a CP is simply the first one available, since the sockets of a single CP are interchangeable it does not matter which one is picked. We do, however, keep track of which agent is connected to which socket.

2.3 Model Metrics

In this section we go into detail about each of the metrics that capture the output of the model. An overview of all the metrics is already given in Table 2.8. The metrics are divided into clustering, competition, run time and validation metrics.

2.3.1 Clustering Metrics

All clustering metrics capture and output information regarding the clusters of the agents and thus the clustering of CPs. The method of clustering the CPs of the agents into clusters is fundamental as it captures where agents exhibit which behavior. We want to be sure to measure and verify the outcomes of this process. The first two metrics we look at are the number of clusters and the number of CPs in the clusters. Furthermore, the total number of CPs in the simulation is also of importance, as this allows us to compare this with the values per cluster and agent. Lastly we take into consideration the size of the clusters, that is we look at the radii of the clusters in the form of the maximum distance and the defined walking preparedness (see Table 2.8), which were explained in Section 2.2.1.

2.3.2 Competition Metrics

To capture the competition in the system we introduce competition metrics. In the literature the number of unique users per CP is often mentioned as a metric to measure competition (as shown in Section 2.1). Therefore the number of unique agents per CP is one of the competition metrics. We also look at how many times an agent fails to select a CP. This occurs when the agent selected a cluster and tries to pick a CP in the neighborhood of this cluster. If all CPs in this neighborhood are occupied, the agent fails to select a CP. Taking the percentage of failed selections can also give insight on the competition in the system.

2.3.3 Run Time Metrics

We gain insight about the computation time of the simulation by using the run time metrics. As the model is developed for the purpose of doing large-scale simulations, it is important that computation time stays within bounds even when the size of the experiments grows. For this purpose we introduce two metrics that time the most computationally intensive parts of the model, namely the initialization of agents and the simulation itself.

2.3.4 Validation Metrics

Lastly, we discuss the validation metrics used. In order to state whether the simulations are producing accurate charging transactions, we want to validate them. We do this by first dividing the data that we have into a training set and a test set. Ultimately we want to predict the use of charging transactions in the future by using the data up till the present. The dataset that was actively used in this thesis contained charge transaction up to (and including) January 2017. For this reason the data is split at a specific date. Data before this date will then be the training data and after this date will be the test data. In this way we can test if using data until a specific date (training data) allows us to predict charging transactions after that date (test data). Default values for these parameters can be found in Table A.2.

To determine if the simulated data matches the training and test data, we need a validation metric. We can view the simulated charging transactions in two different ways, namely from the perspective of agents and from the perspective of the CPs. Therefore we want to introduce metrics to measure how well individual agent behavior validates as well as how the whole charging infrastructure (the CPs) validates. Since we have defined agent behavior in terms of activity patterns, as seen in Section 2.1, we use these to summarize charging transactions. Thus we wish to compare the activity patterns of the training transactions with the activity patterns of the simulated transactions, as well as to compare the activity patterns of the test transactions with the activity patterns of the simulated transactions.

We compare every bin in one activity pattern with the corresponding bin in the second activity pattern. The absolute difference between the two i-th bins of the activity patterns is defined as d_i. Then d_i is a value between 0 (meaning both are the same) and 1 (meaning one bin is 0 and the other is 1). Note that when the bin of an activity pattern has a value of 0.5 the maximum value for d_i would be 0.5 as well. Because we always look at the (fixed) pattern from the training or test data and compare this with the simulated data, we want to adjust the error measure slightly. Given the value of the bin of the activity pattern belonging to the data, we still want the error of the bin to be between 0 and 1. Therefore we look at the maximum error a bin can have (given the data-based bin value) and normalize with this factor. The equation used to calculate the error value e_i between the training or test data p_i and the simulated data of the i-th bin is as follows:

$$e_i = 100 \cdot \frac{d_i}{\max(1 - p_i, p_i)}$$

The error value of the whole activity pattern can then be viewed as the average of the values of all its bins. The scaling factor of 100 makes sure the values are in a more readable region, as without it the error values tend to get very small. The error values can in theory obtain values between 0 and 100, though in practice most are below 5. With this we can define the agent validation as follows. For every cluster of the agent we determine the error value for the activity patterns based on the training or test data and the (mean) activity pattern for the (various) simulation run(s). The error value for that single agent is then defined as the mean of all these values.

For the CP validation every CP gets an error value based on the activity patterns at that CP. However we need to be careful here, since we only want to consider transactions from the data at the CP that are from agents that are in the training data. Therefore we only look at the CPs that are in a cluster of an agent at the start of the simulation, and only look at the transactions in the data at this CP made by an agent in the simulation.

2.4 Sensitivity Analysis and Model Evaluation

In this section we will discuss various experiments we have performed on sensitivity analysis and model evaluation. The main purpose is to validate the model and choose sensible default values for the parameters. For each of the experiments we keep all the parameters on the default values as stated in Table A.1, A.2, A.3 and A.4 and vary the single parameter which we look at in the experiment.

2.4.1 Simulation Setup

We first determine how many agents are needed and how many repeats of the simulation should be done in order to get sensible results. The effect of varying the number of agents in the system on the validation values can be seen in Figure 2.7. Note that the mean value for the agent validation remains constant when we increase the number of agents. Thus competition that occurs within the system when more agents are present, does not influence the agent validation. The CP validation however, is influenced by the number of agents in the system. We can see that with more agents the CP validation gradually becomes better. Remember that the CP validation is based on the activity patterns at the CPs. Those activity patterns match the training and test data better with more agents in the system, and thus more agents composing the activity pattern for a single CP. For this reason we prefer a higher number of agents for the experiments. Since the CP validation evens out at 2000 agents, we decided to do the experiments using 2000 agents in the system.

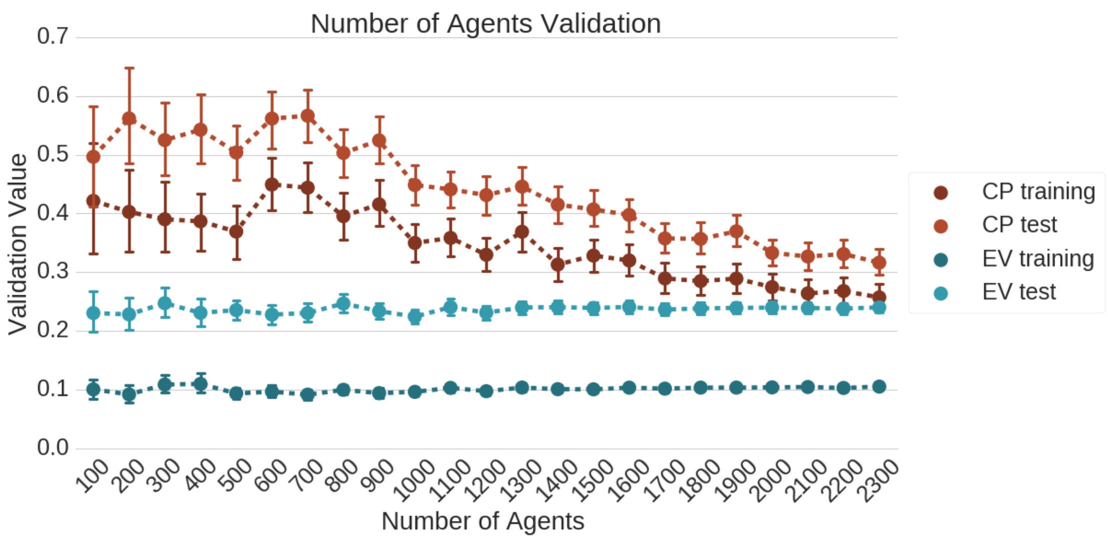


FIGURE 2.7: *The results of the experiment where the number of agents is varied. Per value of the parameter we plot the mean validation values with the 95% confidence interval.*

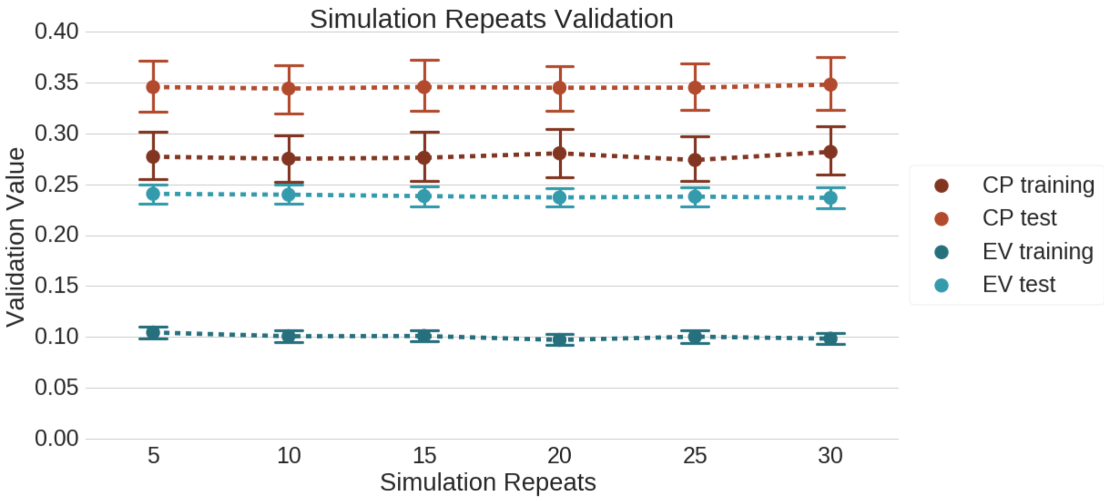


FIGURE 2.8: *The results of the experiment where the number of simulation repeats is varied. Per value of the parameter we plot the mean validation values with the 95% confidence interval.*

The results of the experiment of how many simulation repeats are required can be seen in Figure 2.8. We vary the simulation repeats from 5 up to 30 and find that for both the CP validation and the agent validation this has minimal influence. There are multiple activities for each agent in a single simulation (on average 200 activities) as well multiple agents in a single simulation. These repeating activities and a high number of agents causes the need for less repeats of the experiments. The initialization, as well as the simulation itself, do contain randomness though, which is why we lower the number of repeats no lower than five. The results in Figure 2.8 show that five simulation repeats is sufficient.

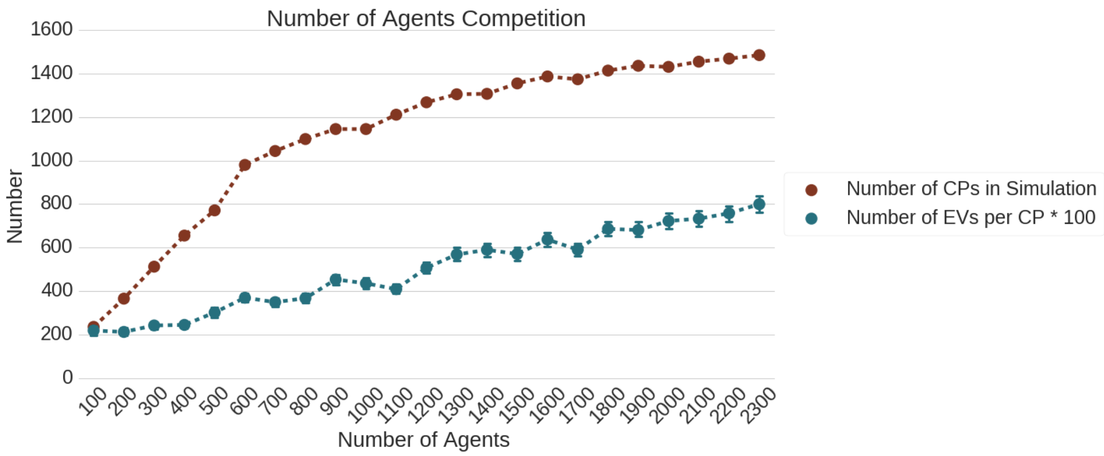
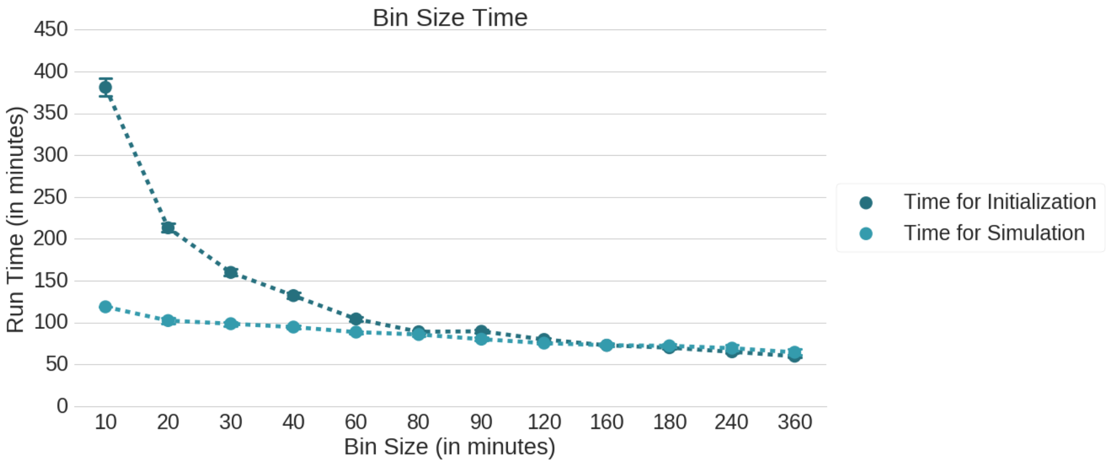


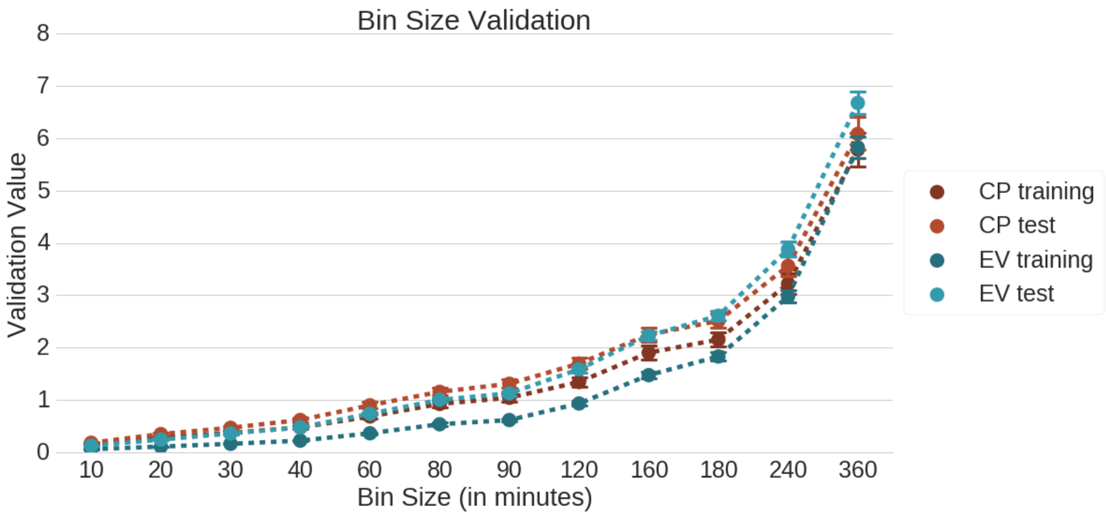
FIGURE 2.9: *The number of CPs in the simulation and the number of agents per CP in the simulation for various numbers of agents in the simulation. Per value of the parameter we plot the mean numbers with the 95% confidence interval.*

Varying the number of agents in the simulation also influences the number of CPs that are used by all agents in the simulation. We can look at the number of used CPs in the simulation and the number of agents per CP as the number of agents vary in Figure 2.9. These results strengthen the statement that a high number of agents is needed in the simulations. Higher values indicate that more competition is present in the system, which makes the system more interesting to study.

2.4.2 Data Processing



(A)



(B)

FIGURE 2.10: *The results of the experiment that varies the bin size. Note that the horizontal axes do not have a linear scale. Per value of the parameter we plot* (A) *the mean run time and* (B) *mean validation value with the 95% confidence interval.*

Having established the optimal values for the number of agents (2000) and the number of repeats (5), a sensitivity analysis of the parameters of the model will be done. First we will look at the parameters that concern the data processing of the model.

Varying the bin size[6] (as introduced in Section 2.2.6) yields the results found in Figure 2.10. We note that the run time of the initialization of the agents decreases as the bin size increases. This is because the distributions of the agents are created in the initialization and those distributions become more fine-grained as the bin size gets smaller. When we look at the validation for the different bin sizes in Figure 2.10b, we can see that the model validates better for lower values of the bin size. We decided on a default bin size of 20 minutes as this reduces the run time by half compared to a bin size of 10 minutes, while the validation values only increase slightly.

Next, we look at the influence of the warmup period. The warmup period is the length of time at the beginning of the simulation which is not taken into consideration in the validation. The results of this experiment can be seen in Figure 2.11. Surprisingly, the warmup period has little effect on the validation metrics. This can be explained by the fact that the initialization is extensive and thus the agents are exhibiting proper behavior from the moment the simulation starts. We pick a warmup period of 7 days, since we would still want to offer the system some time for warmup to be sure inconsistencies at the initialization do not influence the validation.

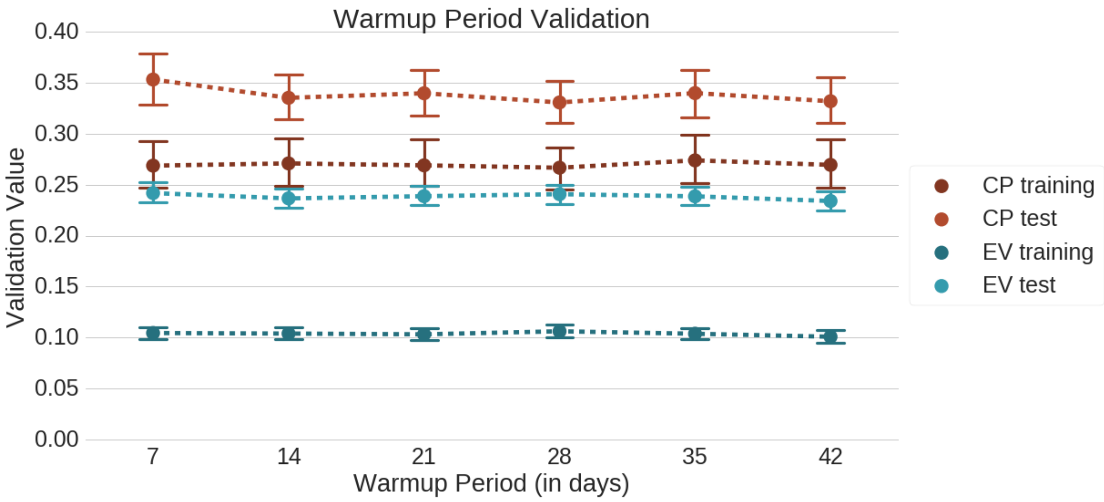


FIGURE 2.11: *The results of the experiment that varies the warmup period and measures the validation values for the agents and CPs. Per value of the parameter we plot the mean validation values with the 95% confidence interval.*

2.4.3 Clustering and Cluster Analysis

In this section we look at the sensitivity analysis of the various parameters concerning clustering as well as experiments concerning the analysis of the clusters of the agents.

[6]Note that only durations able to fill a day exactly, are valid as bin size.

One important choice for the location of the clusters (i.e. the centers) is whether to look at weighted or unweighted centers (see Section 2.2.1 for the definition of weighted centers). The agent and CP validation depending on weighting the centers can be seen in Figure 2.12. The weighted centers give slightly, but not significantly, better results than the unweighted centers. We also take into account that intuitively it makes more sense to weight the centers. Consider the example where an EV user uses one preferred CP almost all the time, but uses another CP close by on rare occasions. In that case we would want the center to be located closer to the preferred CP than to the occasional CP. With this reasoning and the fact that there is no significant difference in the validation values, we decided to use weighted centers.

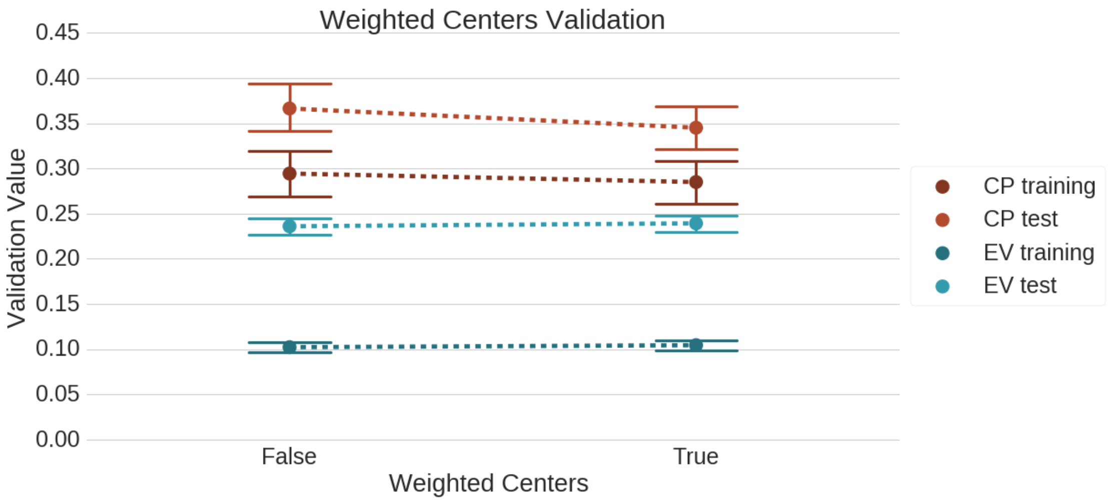
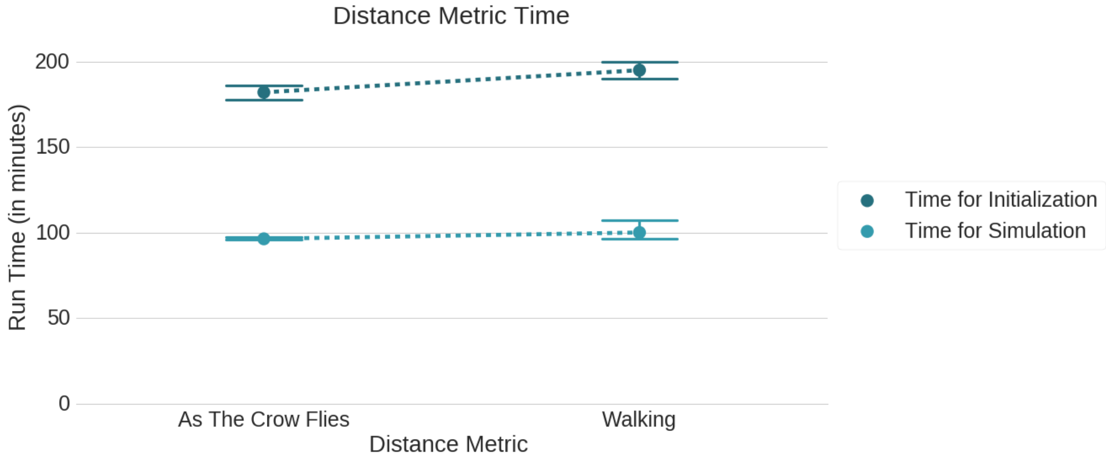
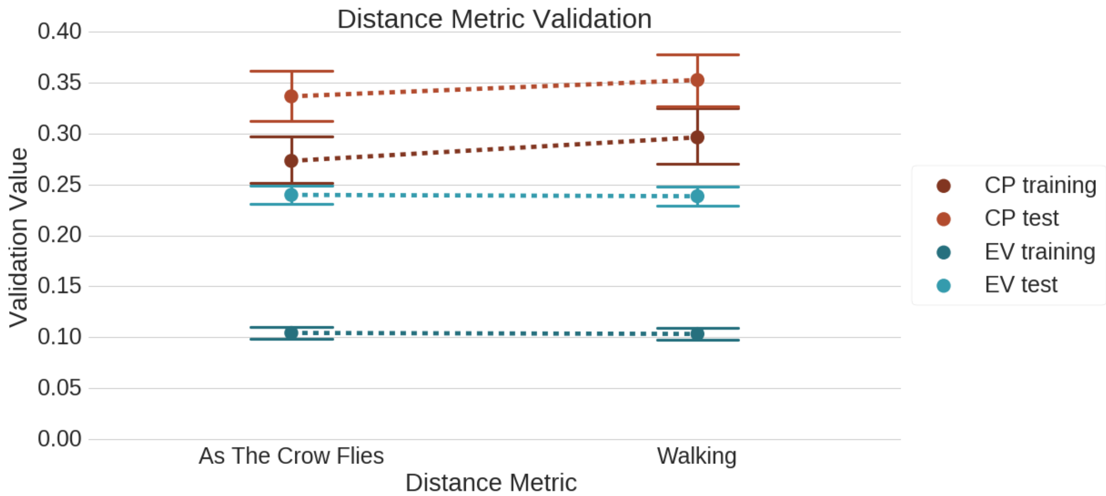


FIGURE 2.12: *The results of using weighted clusters on the validation metrics. Per value of the parameter we plot the mean validation values with the 95% confidence interval.*

We consider two options for distance metrics, namely walking distance and distance as the crow flies. We compare the validation for both measures as well as the run times of the simulation, since calculating walking distance is more computationally intensive than calculating distance as the crow flies. These experiment results can be seen in Figure 2.13. Note that, as expected, the run time is higher when we are using walking distance. However, this increase is minimal. For the validation values we can see that both distance metrics show similar results. While the CP validation got some higher values for walking distance, this is not a significant difference. To decide which metric to use, consider an example where two CPs are on opposite sides of a river. Those should not be considered close to one another if there is no bridge nearby. Thus for this reason we have a strong preference for walking distance. However, due to limitations in the software (OSRM [36]) used to compute the walking distance distances, distances in the Utrecht region could not be computed. As a result, agents in this region cannot be used in combination with walking distance and thus this yields a smaller agent database.



(A)



(B)

FIGURE 2.13: *The difference between the distance metrics walking and as the crow flies. Per value of the parameter we plot* (A) *the mean run time and* (B) *the mean validation values with the 95% confidence interval.*

While this is unfortunate, we still feel it is important to capture the distance between CPs accurately, thus we decided to use the walking distance metric.

To gain insight in the distance between a CP and the cluster it is a part of, we have plotted the histogram in Figure 2.14. Every agent in the simulation gets a maximum distance assigned that indicates the furthest distance the agent has between a cluster and one of its CPs (also see Section 2.2.1). Note that there are a lot of CPs at distance zero from their center, implying these are clusters containing just a single CP. For the other clusters we see that most CPs lie within 100 to 400 meters from their center.

Remember that we introduced the minimum radius within which to consider alternative CPs in Section 2.2.1. The influence of this minimum radius parameter can be seen in Figure 2.15. The difference in the validation values is minimal for the various values

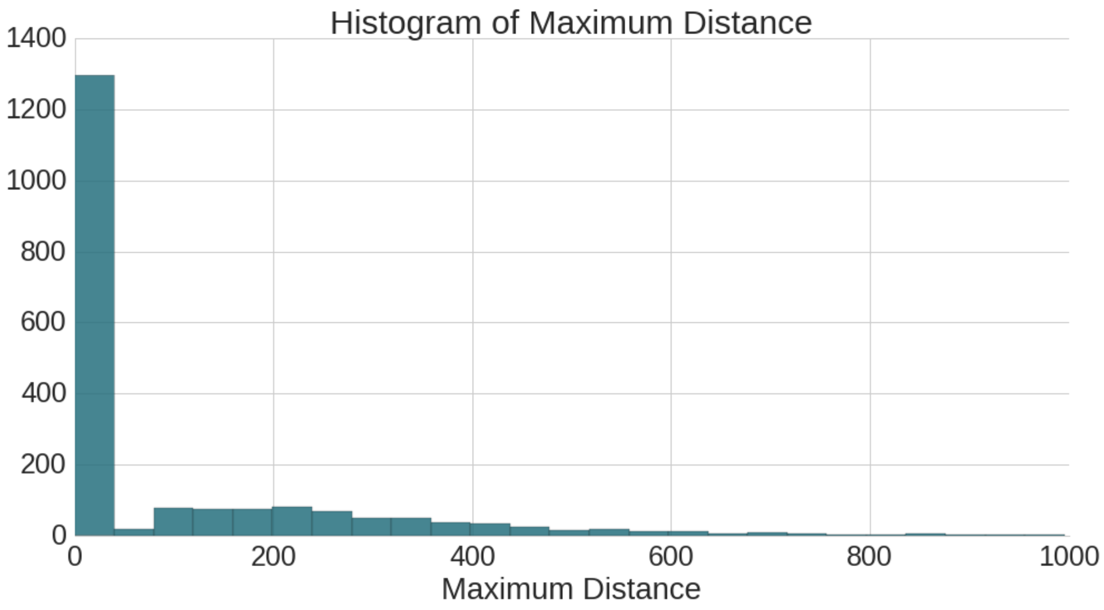


FIGURE 2.14: *The spread of the maximum distances between a center and the CP in the cluster furthest from the center.*

for the minimum radius. This combined with the results for the maximum distance, we decided to put the minimum radius on the value of 150 meters. This corresponds to findings from Kruger [37] as he found that EV users are willing to walk 200 to 400 meters and thus should have clusters with radii of 100 to 200 meters. So whenever an agent has a cluster that has a radius of less than 150 meters, it will still consider alternative CPs within a radius of 150 meters (the default minimum radius).

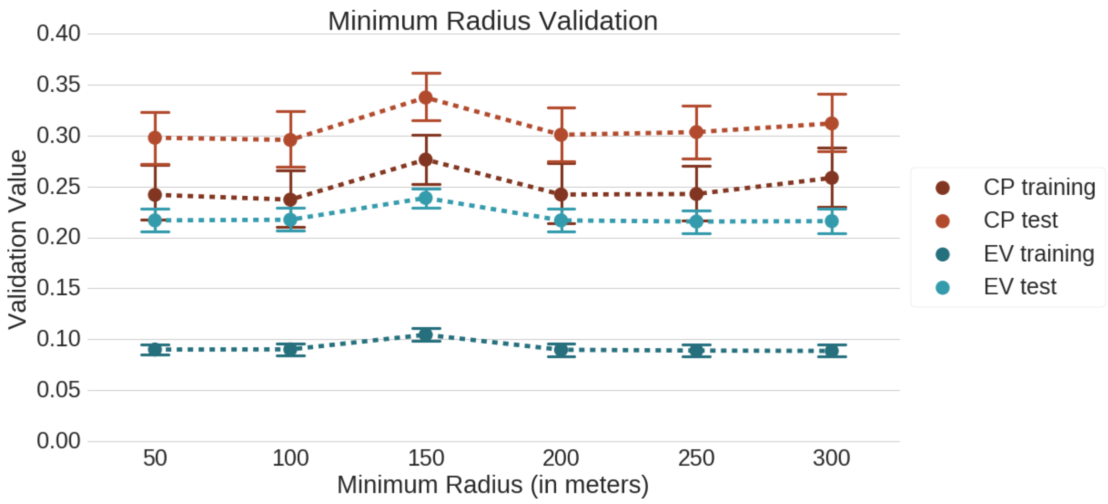


FIGURE 2.15: *The influence of the minimum radius parameter on the validation metrics. Per value of the parameter we plot the mean validation values with the 95% confidence interval.*

Closely related to the minimum radius and the maximum distance is the walking preparedness of the agents, also explained in Section 2.2.1. Remember that we defined the walking preparedness of an agent to be its maximum distance plus 10% (or the minimum radius if this maximum distance is below the minimum radius). The spread of the values for the walking preparedness can then be seen in Figure 2.16. Since, by definition, we put a minimum walking preparedness of 150 meters, we see a large peak at this value. However, we also see some agents that have a higher walking preparedness indicating that their clusters have a larger radius than the default minimum distance.

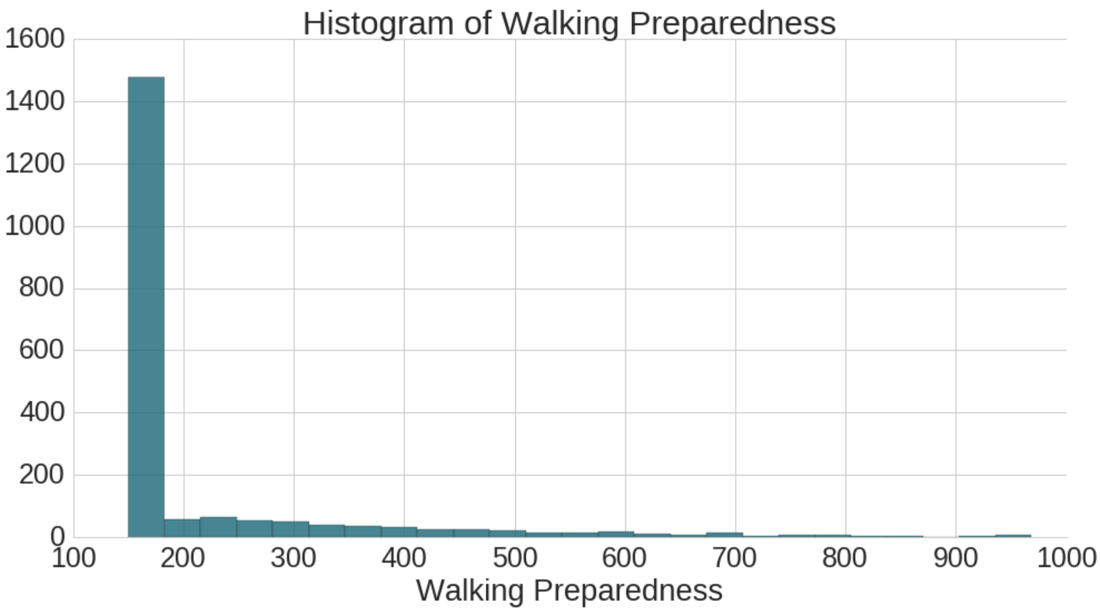


Figure 2.16: *The frequency of the walking preparedness over the agents in the simulation.*

The following results are direct input parameters for the clustering algorithm (as mentioned in Section 2.2.6) that determines which CPs are clustered together. For these experiments we consider the average number of CPs per cluster, the number of clusters per agent, the maximum distance and the walking preparedness.

The Birch threshold is an input parameter of the Birch clustering algorithm [35]. The influence of varying this parameter on the clustering is shown in Figure 2.17. We see that the average number of CPs varies most for low values of the Birch threshold. Around the value 1.5 this levels off, which is why we picked that value for this parameter.

The Birch algorithm has another input parameter, namely the Birch branching factor. We have done a similar experiment for this parameter, but concluded that this parameter had zero influence on the clustering. The resulting figure is therefor not included in this thesis.

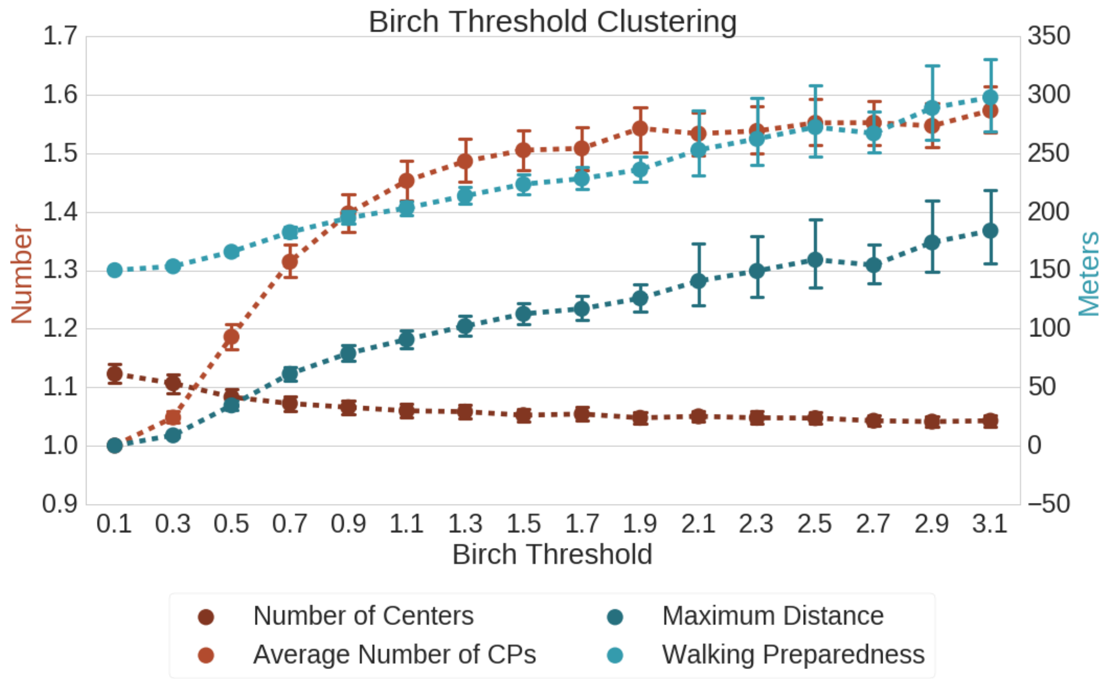


FIGURE 2.17: *The influence of the Birch clustering parameter on various metrics. Per value of the parameter we plot the mean values with the 95% confidence interval.*

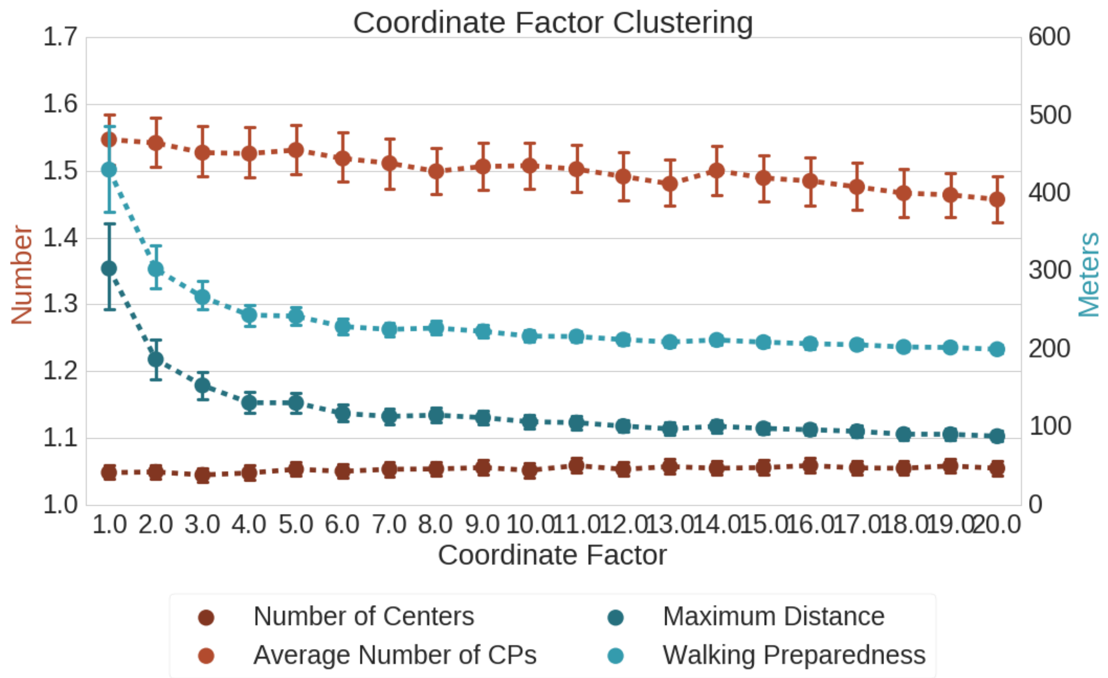


FIGURE 2.18: *The influence of the coordinate factor parameter. Per value of the parameter we plot the mean values with the 95% confidence interval.*

Giving the location more influence in the clustering is done by using the coordinate factor. The influence of this parameter is shown in Figure 2.18. We can see that the lower range of the values have a strong influence on the maximum distance and walking preparedness of the agents. We have picked the value for this parameter to be 8, since at this value the changes in y-values have leveled off.

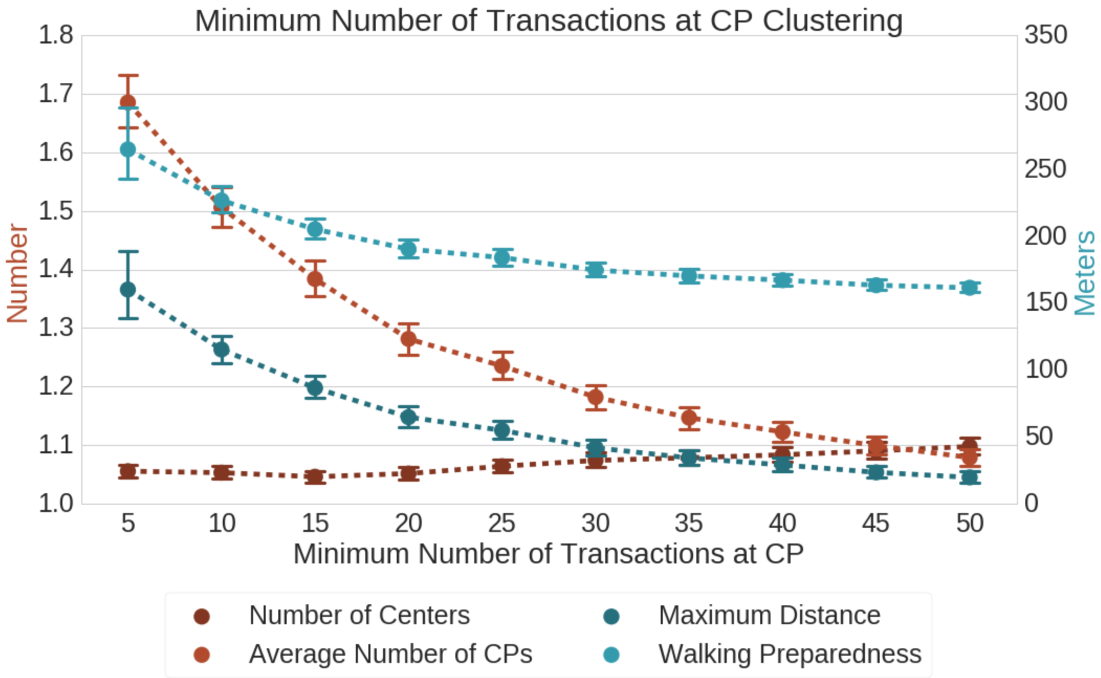


FIGURE 2.19: *The influence of the minimum number of transactions per CP. Per value of the parameter we plot the mean values with the 95% confidence interval.*

An agent needs to have a minimum number of transactions at a CP for that CP to be considered for clustering, as we do not want to cluster a CP if it has only been used, say, once. In Figure 2.19 we can see the influence of this parameter. We picked the value 10 based on these results, because we do not want the number of CPs per cluster and the maximum distance to drop too far.

Just like the minimum number of transactions per CP, we also set a minimum for the number of transactions within a cluster for this cluster to be considered valid for an agent. The results of the experiment varying this parameter can be seen in Figure 2.20. Based on these results we picked the minimum number of transactions to be 20 within a cluster.

The last clustering parameter which we look at is the threshold for the minimum fraction of transactions in a cluster. Each cluster of an agent should have at least this fraction of transactions within its cluster. We introduced this parameter to make sure agents with a lot of charge transactions do not get clusters at places where they have charged 20 times over a long period of time. The results of the experiment on this parameter are shown in Figure 2.21. We can see that the influence of this parameter is minimal. However, intuitively it makes sense that an agent should spend at least some fraction of

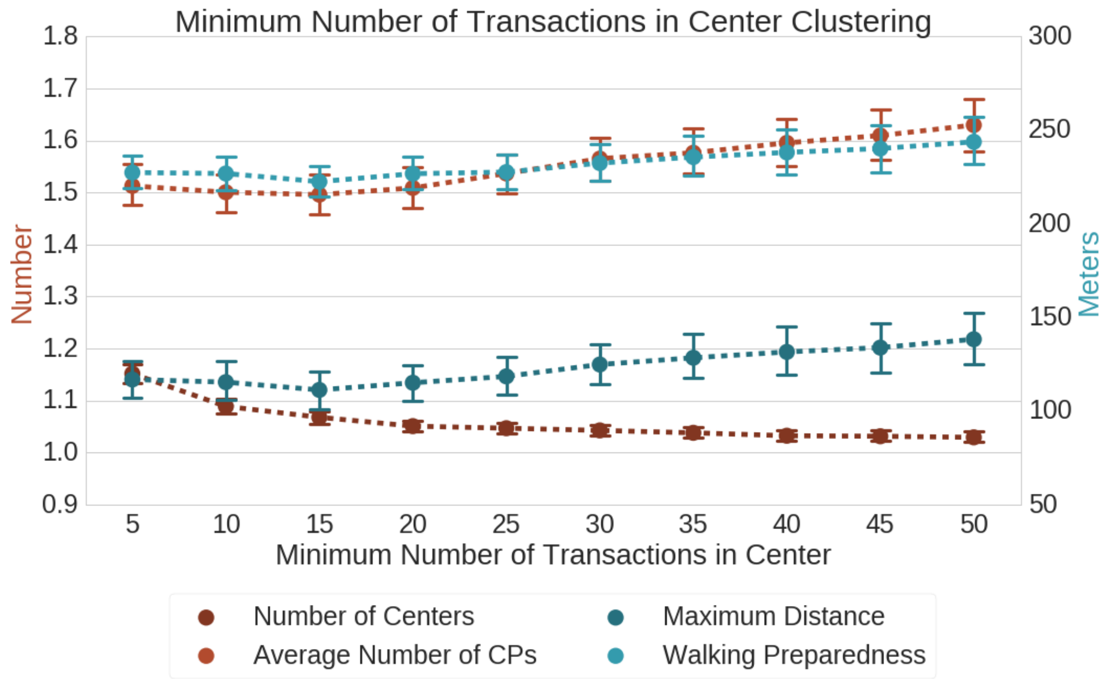


FIGURE 2.20: *The influence of the minimum number of transactions per cluster. Per value of the parameter we plot the mean values with the 95% confidence interval.*

its charge transactions within a cluster. Therefore we picked the value for this parameter to be 0.08.

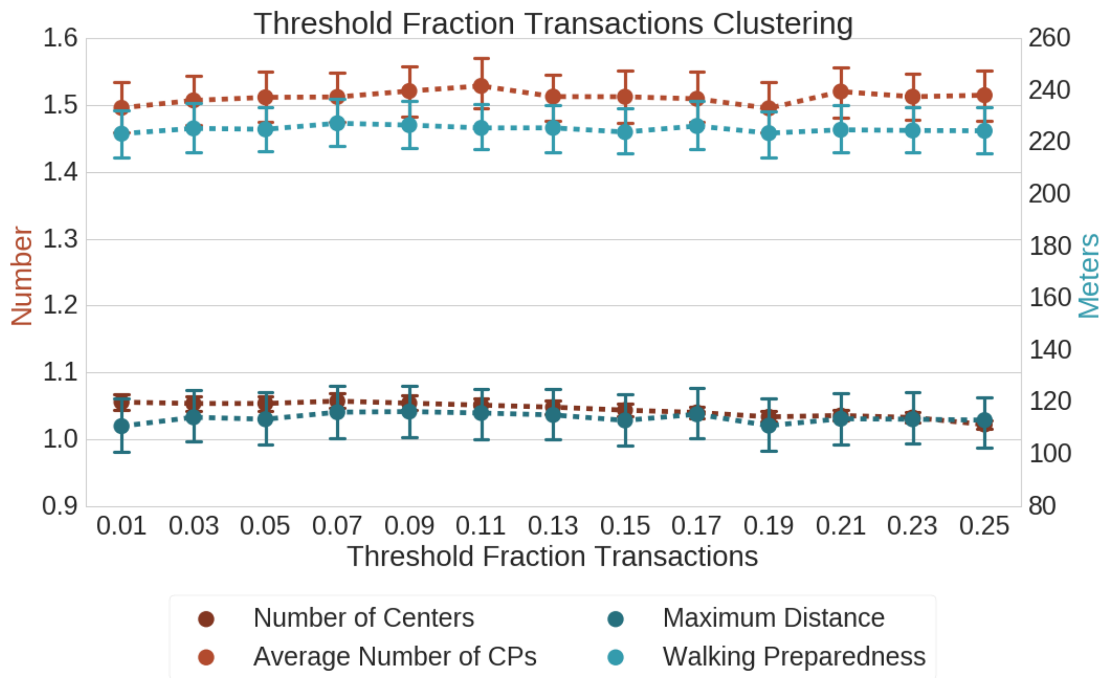


FIGURE 2.21: *The influence of the threshold for the minimum fraction of transactions in a cluster. Per value of the parameter we plot the mean values with the 95% confidence interval.*

2.4.4 Selection Process

In this last subsection of the sensitivity analysis and model evaluation we consider the parameters that influence the selection process of the agents.

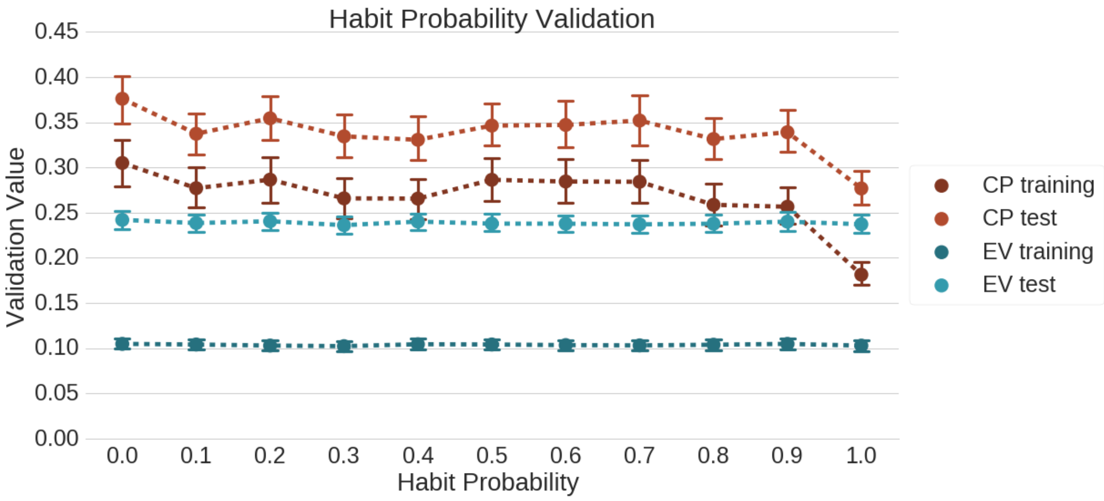


FIGURE 2.22: *The influence of the habit probability. Per value of the parameter we plot the mean values with the 95% confidence interval.*

The influence of the habit probability, as introduced in Section 2.2.7, can be seen in Figure 2.22. There is a clear dip in the CP validation at a habit probability of 1.0. This can be explained by reasoning that with a habit probability of one, we always do whatever is determined by the data and thus validation improves significantly. However, there is a problem with this, since new CPs will then never be used by the agents. In general, when the habit probability parameter is lower, new CPs within walking distance of an agent's cluster will be used more often. To allow for agents to learn to use new CPs, we thus want to lower the value for this parameter. In the end we decided on a value of 0.4 for this habit probability. With this value an agent with a cluster containing, say, two CPs would be able to shift its preference to a more optimally placed CP by picking that one based on distance roughly 60% of the time.

The last parameter for the sensitivity analysis is the retry time of a cluster. Whenever all CPs within the cluster of an agent are occupied, the agent will try to connect again after this retry time. The results of the experiment can be found in Figure 2.23. As we can see, the influence of this parameter is minimal, and thus we decided to keep the default value of 20 minutes, which is equal to the default bin size.

In the final part of this subsection we will look at another two experiments concerning the failures of the CP selection process.

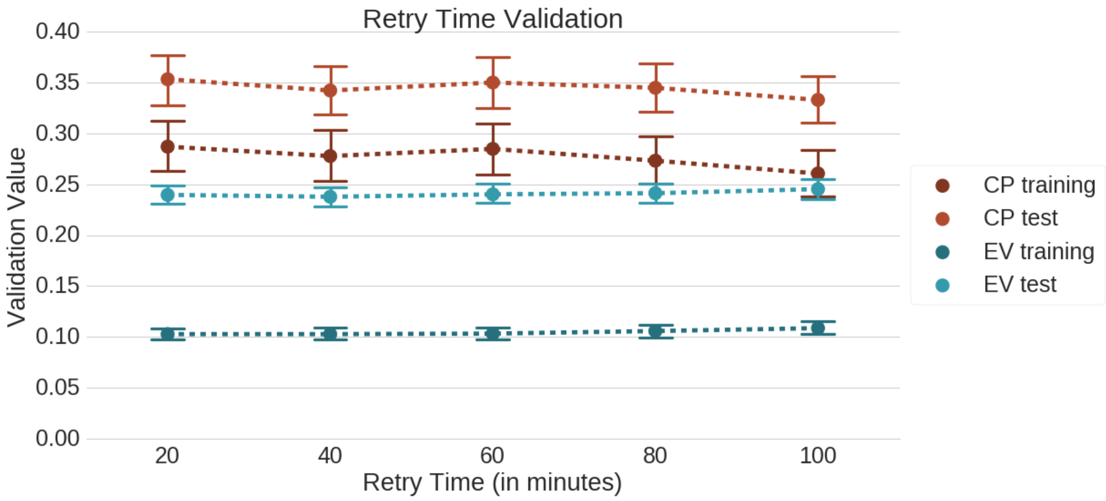


FIGURE 2.23: *The influence of the retry time parameter. Per value of the parameter we plot the mean values with the 95% confidence interval.*

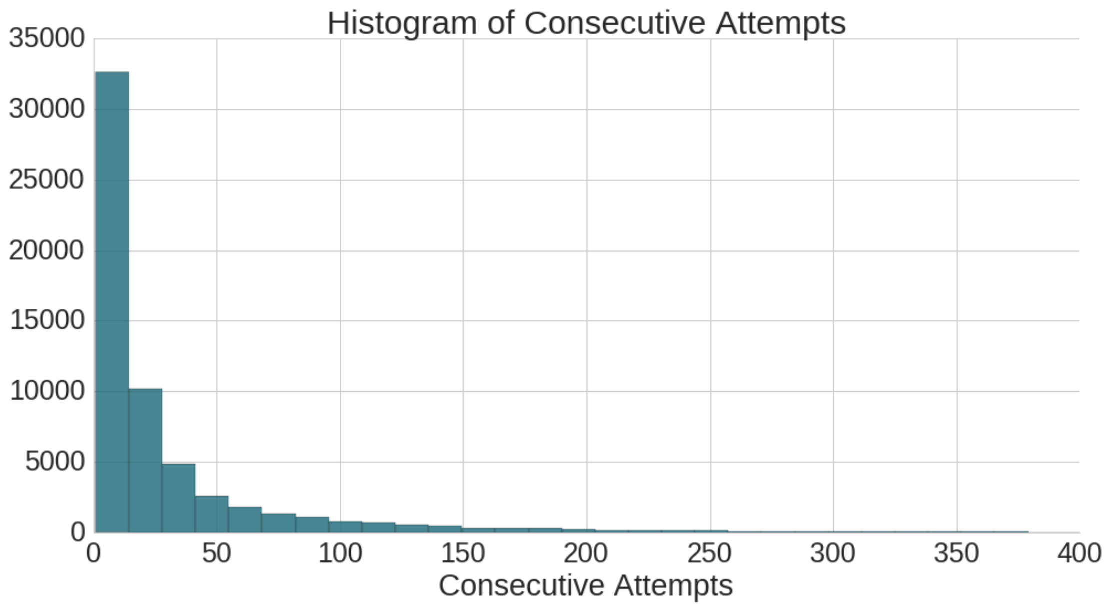


FIGURE 2.24: *The frequency of consecutive failed attempts to connect to a CP over a time period of 5 simulation years with 2000 agents.*

The consecutive failures to connect to a CP by an agent can be as seen in Figure 2.24. A failure to connect is caused by the CP selection process failing if all CPs within walking distance are occupied. Consecutive failures occur if at the next attempt (after the retry time) all CPs are again occupied. The frequencies imply the total number of occurrences at the number of consecutive failed attempts over the five simulation repeats (thus over five simulation years). While most EV users only fail an connection attempt once or twice in a row, we see that in some extreme cases in which up to 1400 attempts to connect to a CP were made in a row. Naturally this would not happen in reality. These extreme cases are rare and thus their influence on the simulation is considered minimal.

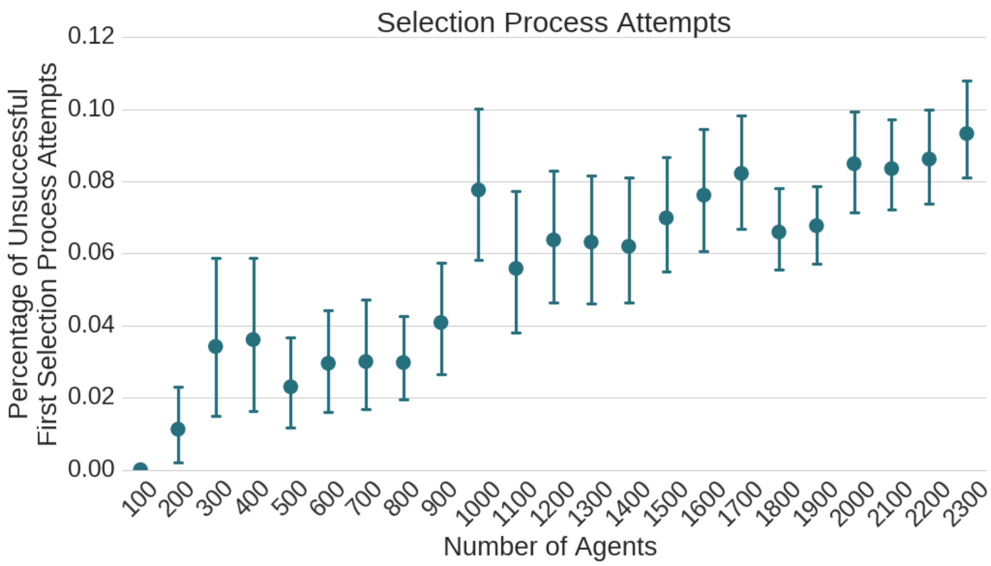
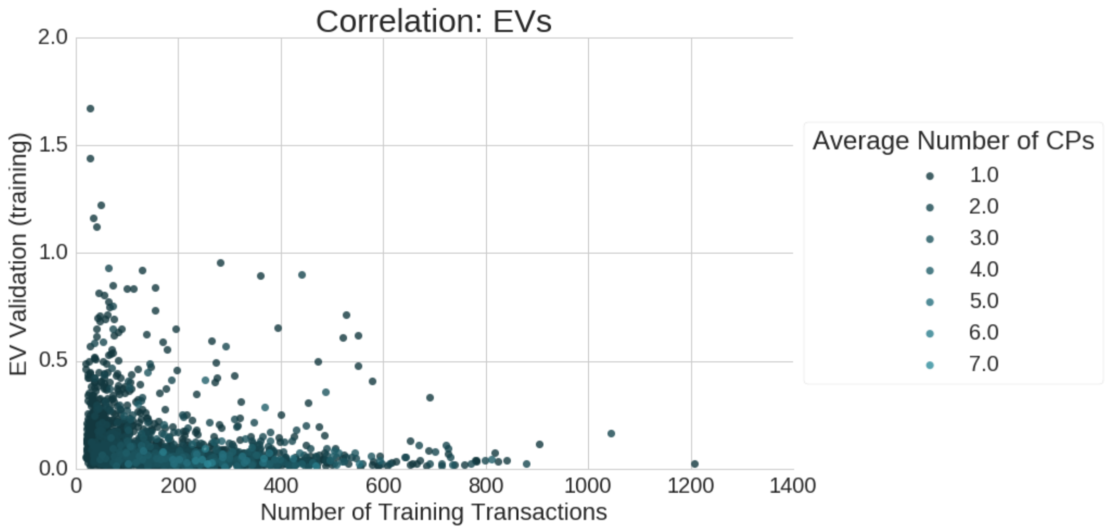


FIGURE 2.25: *The percentage of failed (first) attempts in the selection process. Per value of the parameter we plot the mean percentages with the 95% confidence interval.*

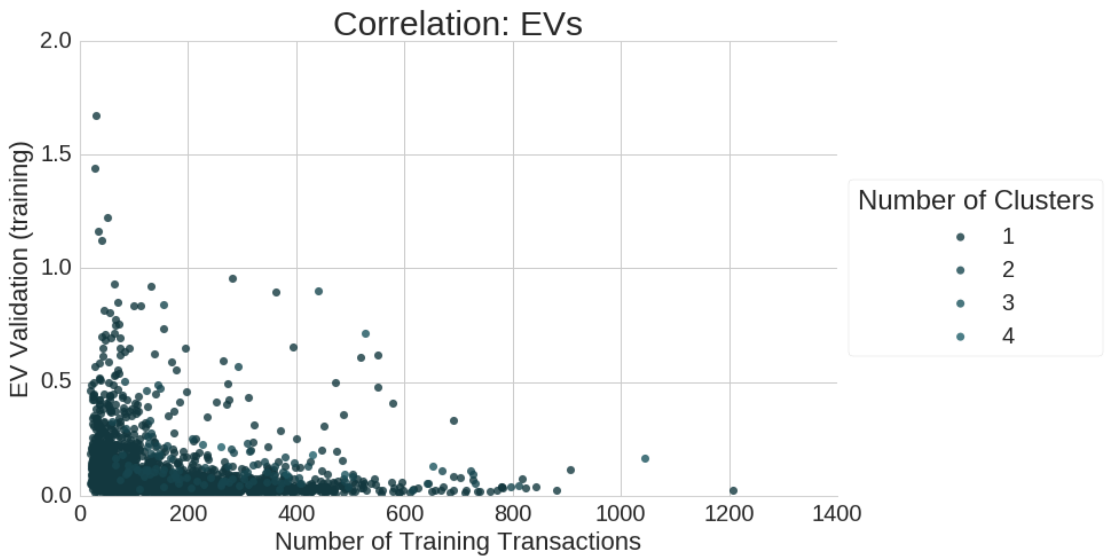
The final experiment concerns the percentage of failed connection attempts. The results of this experiment can be seen in Figure 2.25. Only the first failed selection attempts of an agent are considered. Thus any consecutive failed attempts are not taken into account, but a new failed attempt after a successful connection does count. We can see that the number of failed attempts increases as more agents are present in the system, indicating that more competition is present in the system.

2.4.5 Model Validation

We conclude this section with an indication of how well the model performs. First, we look into the correlation between the number of charging transaction and the agent validation as seen in Figure 2.26. Note that the higher (thus worse) validation values are mostly found at a smaller number of charging transactions. Thus naturally we want to include all the relevant charging transactions available.



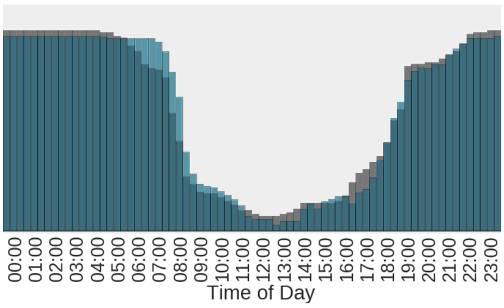
(A)



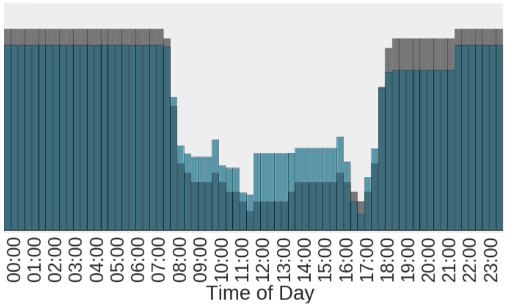
(B)

FIGURE 2.26: *The relation between the number of charging transactions and the EV validation. Each dot represents an agent in the simulation. Indicated are also* (A) *the number of clusters and* (B) *the average number of CPs per cluster each agent.*

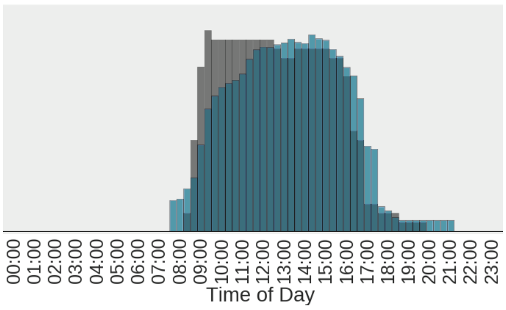
What remains is to gain insight in how well the model performs for certain validation metrics. For this we need to know what an activity pattern looks like for a certain validation value. For the chosen default values for the parameters we have on average an agent validation value of 0.10 for training and 0.23 for test and a CP validation of 0.29 for training and 0.33 for test. Examples of how well the activity patterns then match can be seen in Figure 2.27. Note that for the value around 0.1 we got a near perfect match of the activity pattern. For higher values up to 0.33 the pattern gradually fluctuates more. However, even for those higher values the patterns still roughly match. Therefore, we conclude that the agents are validating very well and the CPs are validating adequately.



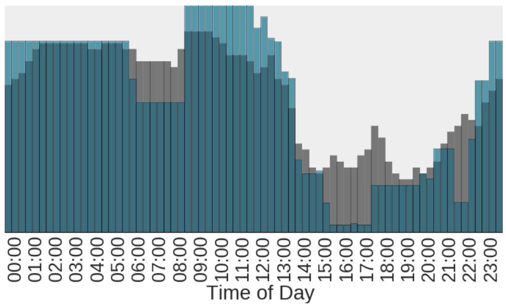
(A)



(B)



(C)



(D)

FIGURE 2.27: *Examples of activity patterns for the validation values* (A) 0.102*,* (B) 0.239*,* (C) 0.296 *and* (D) 0.333*. The real activity pattern is indicated in gray while the simulated pattern is blue.*

2.5 Conclusion

We have presented the core of the SEVA model which captures the charging transactions of EV users in the Netherlands. The model is data-driven, making use of a rich dataset to extract charging behavior. The internal workings of the model have been discussed in detail. We provided argumentation for the various choices of the parameters in the model using an extensive sensitivity analysis. Lastly, we validated the model from the agent's point of view as well as from the CP's point of view and found that the model validates well. As seen in the validation scores, the accuracy of the model is good. This further confirms that charging behavior is habit-based.

The strength of the SEVA model is that it is data-driven and it offers meaningful predictions. It provides a strong basis for further research on the assesment of charging infrastructure and the effects of infrastructure optimization methods.

Chapter 3

A Model of Charge Pole Selection

This chapter provides insight into how EV users select a CP given a set of choices through the creation of a discrete choice model. We start with an overview of relevant literature in Section 3.1 and a data analysis of factors which influence the decision process of the EV users in Section 3.2. We use the insights gained to develop our choice model, described in Section 3.3. We then show how this choice model is implemented within the SEVA model in Section 3.4 to create the SEVA-ICPS model and how well this SEVA-ICPS model validates in Section 3.5.

3.1 Review of Relevant Literature

This section presents existing choice models and related literature to understand which factors influence how EV users choose a CP.

Smart et al. [38] analyse data from a fleet of plug-in hybrid electric vehicle conversions. They discovered that few non-commercial EV users plugged into CPs away from home and they show that the majority of vehicles charged at one location (the residence).

A survey of analyzing the charging preferences of EV users in the Western Australian Electric Vehicle trial indicates that EVs were extremely sensitive to charging cost and showed a strong preference for low cost EV charging. This survey also showed that EV users are sensitive to the charging duration and prefer fast chargers. For this advanced discrete choice models were used, such as multinomial logit models and random parameters logit models [39].

Spoelstra [2] discovered that two out of three EV users charge in one or two location, which indicates that they use CPs that are already known to them. They are habit-based. He also found that the CP density did not influence charging behaviour.

Daina [40] also found that EV users are cost sensitive, using both a multinomial logit model and a mixed logit model.

A survey of panel data from a two-year field trial on EV use in Japan discovered that EV users prefer to charge at night, because the electricity cost is lower at this time. This once again shows that EV users are cost-sensitive. They used a multinomial logit model [41].

Helmus and van den Hoed [42] state that the CP volatility, a measure based on the distribution of user preferences regarding CPs in the vicinity of their destinations, is also an influencing factor. EV users with a high CP volatility will use more CPs in an area than users with a low CP volatility.

Kruger [37] analyzed the CHIEF dataset [14] to show that EV users are prepared to walk, on average, 600 meters and that the median is around the 200 to 400 meters. When compared to the Dutch average walking preparedness of non-EV users, these values seem high. Kruger [37] explains this by pointing out that EV users need to finding a parking spot with a charger while non-EV users are not bound by this.

Yu and MacKenzie [43] also showed that modeling the charging speed of a CP influences EV users, as they found that modeling charge energy (the amount of energy that can be taken on during a potential charging transaction) provided a better fit than dwell time and state of charge independently.

Wen et al. [44] found that people are cost sensitive and that they prefer high-power charging, even after the amount of energy they can get during the charging transaction is controlled for. According to Wen et al. [44], people are once again making choices based on heuristics instead of based on necessity or optimal 'utility' (to minimize costs and maximize energy). From this we can conclude that choices are frequently made based on habit.

Xu et al. [45] investigate the factors that influence an EV user's charging mode (normal or fast charger) and location (home/company or public CPs). A mixed logit model was successfully build to describe the EV behavior of nearly 500 EV users in Japan. They also show that EV users prefer to charge when the low-price electricity tariff begins. Furthermore, Xu et al. [45] found that private EV users are less likely to use fast chargers at night, as this could obstruct a person's natural sleep patterns. Lastly, Xu et al. [45] also found that the use of normal chargers is habit-driven while the use of fast chargers is demand-driven by state of charge. The more often you have used fast chargers in the past, the more likely you will be to use fast chargers in the future [45].

Daina et al. [23] developed a random utility model for joint EV drivers' activity-travel scheduling choices and charging choices. They also found that EV users are cost sensitive. They furthermore found that only 40% of their users prefer charging as fast as possible.

Sheppard et al. [46] find that explicitly modeling the factors which influence whether and where to charge increases modeling accuracy, and that this can therefore be used as a metric for deciding where to place new charging infrastructure (and it can be seen as a summary of all characteristics relevant to the need for charging infrastructure). They used a nested logit model. The 'charge attribute' factors which influence whether and where to charge are cost, distance to activity, availability and whether the charger is a home charger. Interestingly, pole capacity is found to not influence the decision.

State of charge is mentioned as a factor which influences charging behavior [23, 23, 41, 44, 47]. However, this factor only influences the likelihood of when an EV user charges but not at which CP. For this reason this factor is not included in our analysis. Habit and memory-based choices are also mentioned as factors which influence how EV users choose their CPs [2, 38, 42, 44–46]. As habit emerges from the choices based on the other factors, we assume this habit emerges from choices made due to other factors.

From this literature overview, we can conclude that there are three factors which influence the selection choice. These factors are distance, charging speed and cost. These factors and the corresponding literature is summarized in Table 3.1.

Influencing factors	**Related literature**
Cost	[23, 39–41, 44–46]
Speed	[39, 43, 44] but [23, 46] partially contradict this.
Distance	[37, 46]
Habit	[2, 38, 42, 44–46]
State of charge	[23, 40, 41, 44, 47]

TABLE 3.1: *A summary of literature related to the factors which influence how EV users choose a CP within a certain area.*

3.2 Data Analysis

Having found which factors may influence the selection process, this section focuses on capturing and understanding each factor individually using data analysis. In Section 3.3 we will also examine the relation between the factors. Note that the cost factor is divided into charging fees and parking fees, as these fees can influence EV users in different ways. Thus the influencing factors become distance, charging speed, charging fee and parking fee.

To analyze these factors, we use the habitual agents and their clusters as found in the SEVA model. There are 2551 valid[1] agents with 2688 clusters, 1112 in Amsterdam, 585 in The Hague, 617 in Rotterdam and 237 in Utrecht. We only look at agents with choices within range of their walking preparedness. This means only agents are taken into consideration which have CPs with different distances, charging speeds, charging fees, or parking fees to choose from within the range of their walking preparedness. Figure 3.1 shows a histogram of the number of choices within the range of the clusters for each agent. On average, clusters have 4.6 CPs in the range of their walking preparedness. The median and spread of this figure show that more than half of the clusters only have 1 or 2 CPs in range and therefore have few or no choices to make. This means most of our habitual users are using the same pole every day.

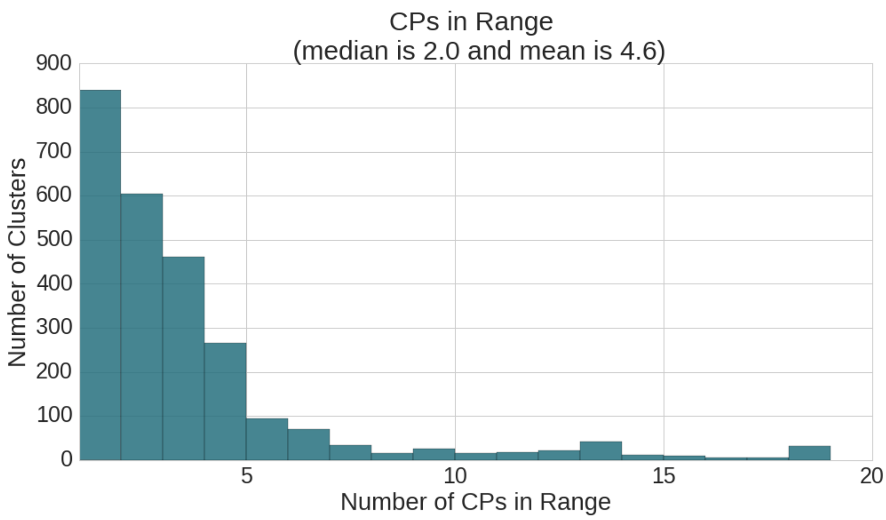


FIGURE 3.1: *The histogram of the number of CPs in the range of the clusters of each agent. The histogram is shown until the 95th percentile.*

[1]As mentioned in Section 2.2.6, agents are only valid if they have at least one cluster with 20 charging transactions and at least 8% of the total transactions of the agent in this cluster.

3.2.1 Distance

The first factor to analyze is distance. This is defined as the distance in meters from the center to a CP in that cluster (d_{cp} in Figure 3.2). Note that this is the distance 'as the crow flies'. In 1848 clusters (68.8% of all clusters) there are CPs with multiple distances and thus multiple choices regarding distance.

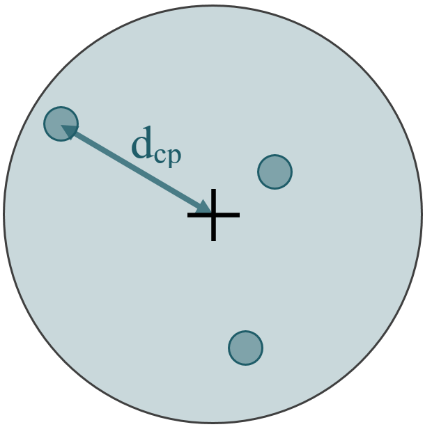
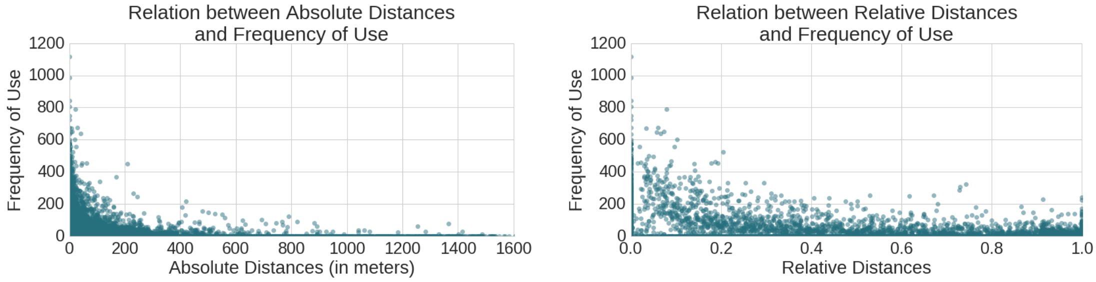


FIGURE 3.2: *A cluster of one agent with a cross denoting its center and circles denoting its CPs. The distance between a CP and its center is defined as d_{cp}.*

Figure 3.3 shows a significant, negative correlation between the both the absolute and relative distance[2] and the frequency of use of the CPs. This is to be expected given that the centers are determined by the weighted average locations of the used CPs in that cluster.



(A) (B)

FIGURE 3.3: *The relation between (A) the absolute distance or (B) the relative distance and frequency of use. There is a negative Pearson correlation of -0.22 ($p = 0.00$) and -0.27 ($p = 0.00$), respectively. Only clusters in which there is a choice regarding distances are taking into account.*

Figure 3.4 can help understand if the centers are in a logical location by showing a histogram of the relative distances of unused CPs. From this figure we can conclude that the further away from the center the more unused CPs there are, which corresponds

[2]Relative distances are calculated by dividing the absolute distance by the maximal absolute distance of the CPs in the same cluster.

with our expectations. This means that our centers are likely in the correct location[3]. Note that the peaks at the relative distances 0.0 and 1.0 are due to charging hubs. Only one of the CPs in the hub can be used at once, meaning all the other CPs in the hub remain unused.

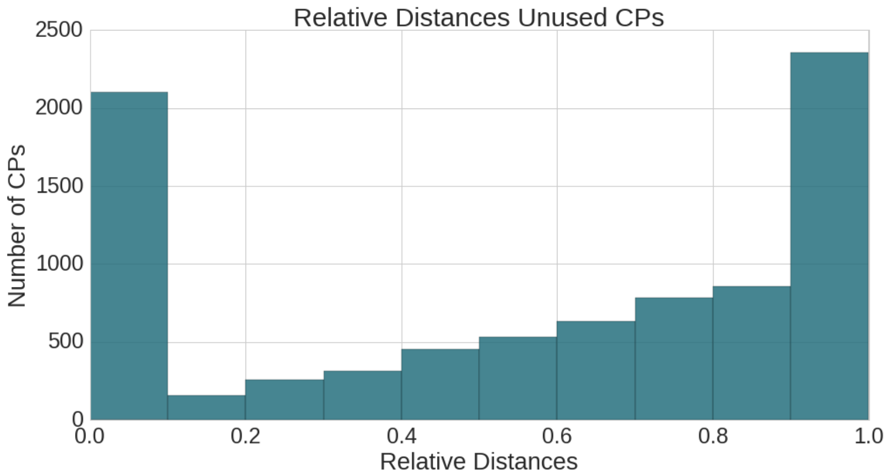


FIGURE 3.4: *The relative distances within clusters of unused CPs.*

3.2.2 Charging Speed

This section examines the effects of charging speed on the selection process. Charging speed is defined as the number of kW offered by each socket of a CP. Note that there are two types of CPs, namely slow chargers (less than 50kW) and fast chargers (50kW or higher). EVs can have different types of connectors, namely levels I, II and III. Only EVs with level III connectors can charge at a fast charger [48, 49]. The vast majority of CPs in our dataset are slow chargers.

In 755 clusters (28.1% of all clusters) there are CPs with multiple charging speeds in range and only these clusters will be used in the analysis in this section. Figure 3.5 shows the number of CPs in range for clusters in which there is a choice regarding the charging speeds of the CPs. Comparing Figure 3.5 to Figure 3.1 we see that both the mean and median have shifted to the right. The mean number of CPs in range increased from 4.6 to 10.8, and the median number of CPs in range increased from 2.0 to 5.0. Thus an average of 10.8 CPs in range is needed for a cluster to have a choice with respect to charging speeds. Note that there could be a geographic bias. The places where there are more operators are likely highly populated and these places are also more likely to have a higher density of CPs.

[3]This could be confirmed by validating the centers of the home clusters with the actual home locations of the agents. However, due to privacy restrictions these home locations were unavailable to us.

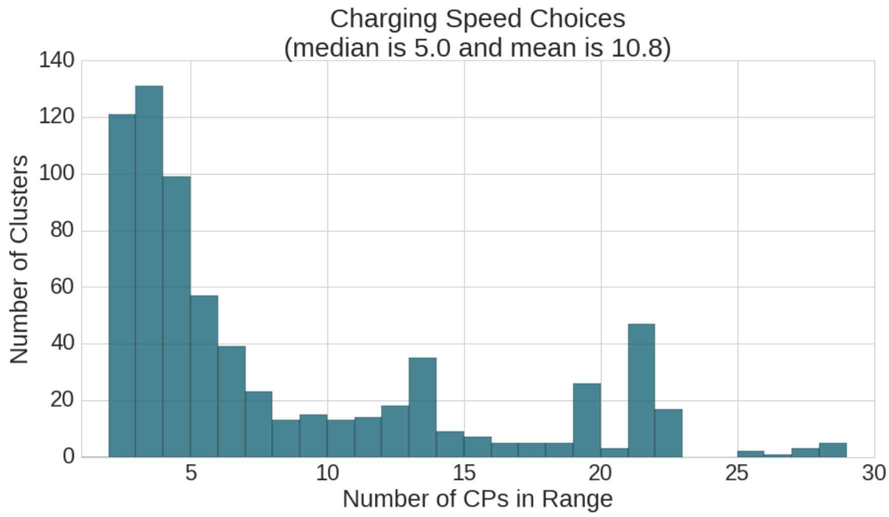


FIGURE 3.5: *A histogram of the number of CPs in range when there is a choice regarding charging speed within the cluster.*

Figure 3.6 shows the relative spread of the charging speeds, looking only at the clusters where there is a choice. This indicates that the majority of the CPs have a charging speed of either 16A or 20A, which corresponds to the relative charging speeds of 1.0 and 1.2.

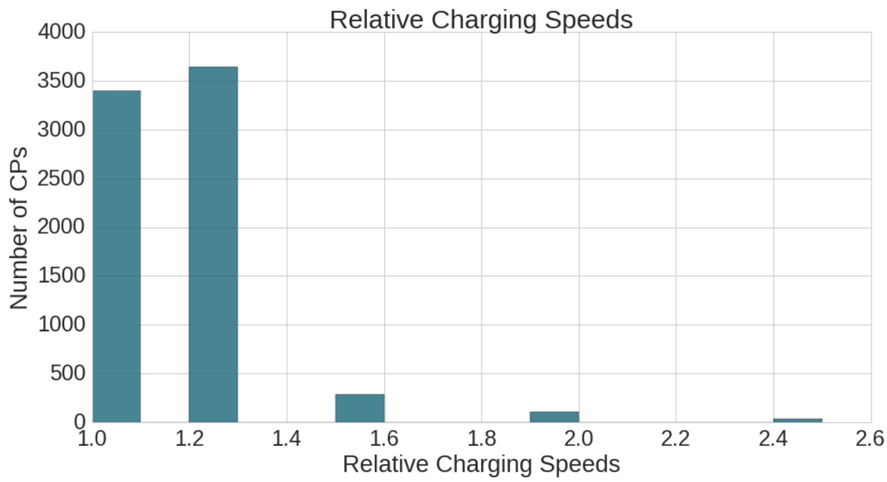


FIGURE 3.6: *The histogram of the relative charging speeds. The relative speeds are calculated by dividing the absolute speeds by the minimum speed in the corresponding cluster.*

Figure 3.7 shows the correlations between the relative charging speeds and the frequency of use of a CP. This figure shows no significant correlation, which was confirmed using a Pearson correlation test. The apparent decrease of the frequency of use for higher charging speeds is solely because there are fewer CPs with these higher charging speeds. From this analysis we can conclude that within the category of normal chargers, there is no correlation between the charging speed and the usage of CPs.

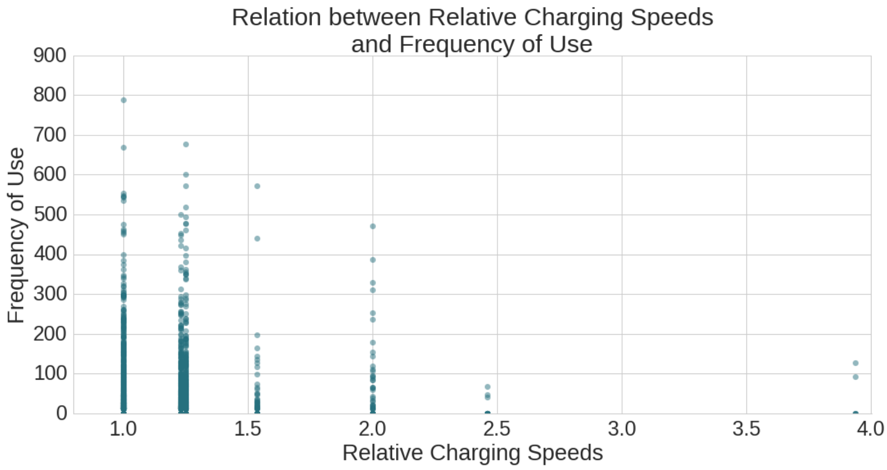


FIGURE 3.7: *The correlation between the relative charging speeds and the frequency of use.*

An explanation for this lack of correlation is that there is a lack of charging speed differences within clusters. The distribution of charging speeds can be seen in Figure 3.8. This figure shows that the maximal charging speed difference in a cluster is below 10kW in 86.5% of the clusters.

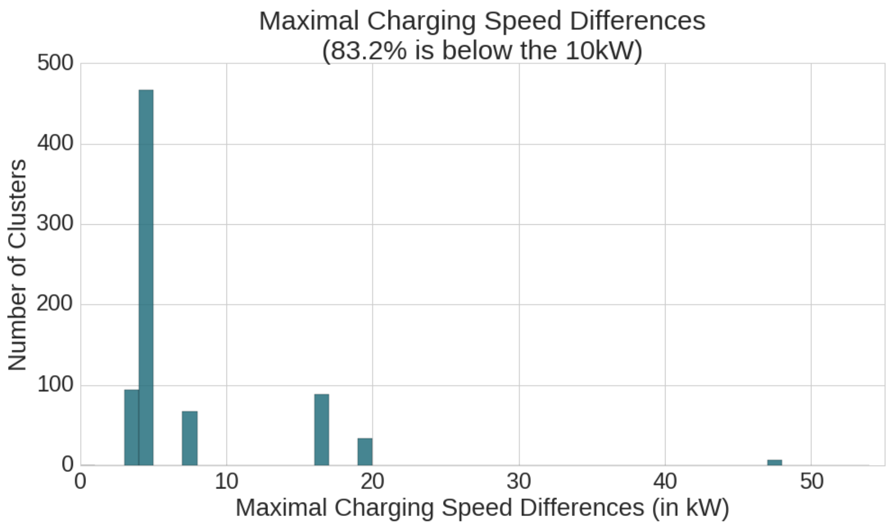


FIGURE 3.8: *The maximal difference in charging speeds within the clusters.*

3.2.3 Charging Fees

Next we analyze the effects of different charging fees. In the Netherlands, this fee consists of a start tariff and an additional kWh tariff. These tariffs can differ greatly depending on which CP operator installs and maintains the CP and which charge card an EV user has [50].

To analyze the effects of charging fees on the decision process of the agents, it is important to first examine which agents have a choice regarding these charging fees. 340 out of 2677 clusters (12.7%) have more than one CP operator and could therefore have a difference in charging fee within their cluster (depending on the charge cards of the EV users). As the remaining 2348 clusters are not interesting for our charging fee analysis, only these 340 clusters with possible choices regarding charging fees are taken into account in the analysis in this section.

Figure 3.9 shows the histogram of the number of CPs in range for the clusters in which there is a choice regarding the charging fee. When comparing this figure to Figure 3.1, we can see that the mean and median number of CPs in range has increased from 4.6 to 14.8 and from 2.0 to 7.0, respectively. This means that there need to be more CPs in range in order to get multiple choices regarding the parking fees. Note that this increase is greater than the increase needed to get choices in terms of charging speeds (see Figure 3.5).

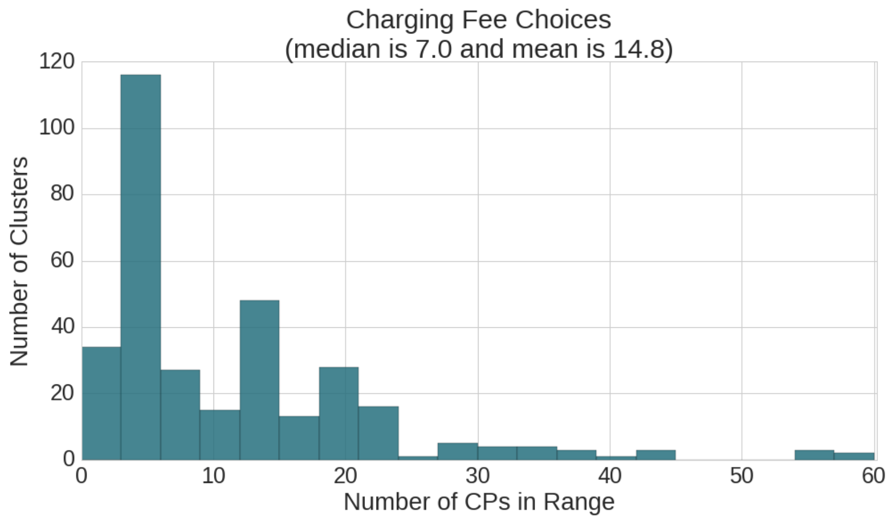
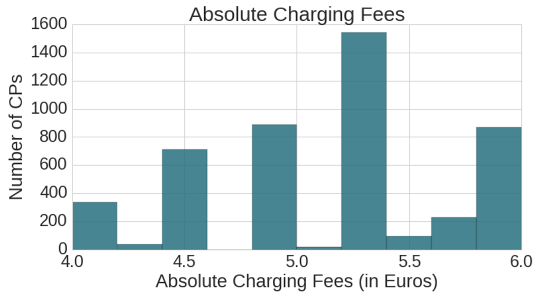
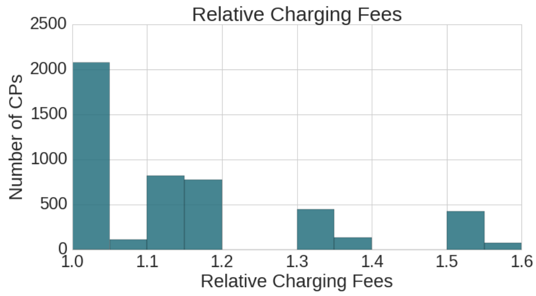


FIGURE 3.9: *The histogram of the number of CPs in the range of the walking preparedness of the clusters where there are multiple options regarding charging fees.*

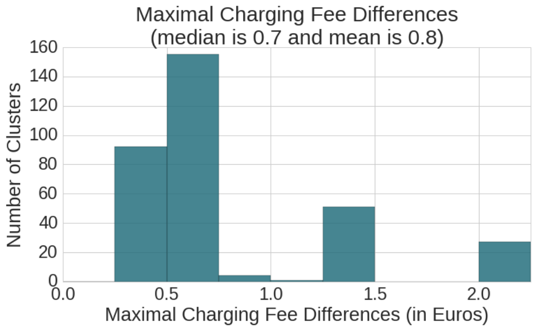
Figure 3.10 shows the absolute charging fees per CP and the maximal charging fee difference per cluster. These figures increase our understanding of the spread of prices in the cluster where there are choices regarding charging fees.



(A)



(B)



(C)

FIGURE 3.10: *The histograms of the* (A) *absolute and* (B) *relative charging fees are shown together with the* (C) *spread of the maximal difference in charging fees within the clusters.*

To understand the effect of the charging fees on the use of CPs, we now look at the relationship between the relative charging fees and the frequency of use for each CP[4], which can be seen in Table 3.2. Note that a small negative correlation was found between the relative charging fees and the frequency of use when all cities are taken together. However, when analyzing this relation within each city, Table 3.2 shows us that only in Utrecht there is a significant negative correlation between charging fee and frequency of use.

An explanation for the lack of correlation between the relative charging fees and the frequency of use can be seen in Figure 3.10c, which shows that the median of prices differences is 70 euro cents per charge transaction. We may hypothesize that for some EV users this is not a large enough difference to influence their choice of CP. We may furthermore hypothesize that EV users in Utrecht are more sensitive to charging fees, as they do not have to pay parking fees (which in other cities are often higher than the charging fees).

[4] To calculate the charging fee per CP, we assume the average kWh charged is 14.4, as done by Wolbertus [50].

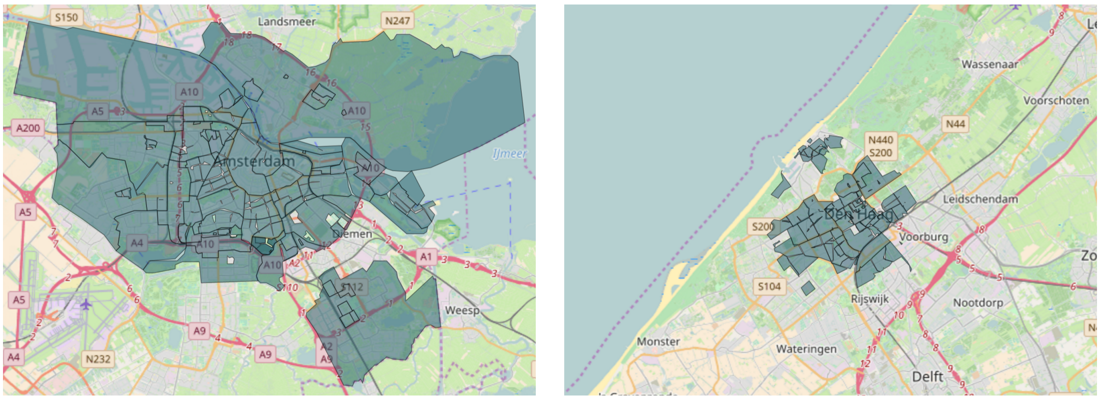
City	Pearson Correlation	P-value
All	-0.06	0.00
Amsterdam	-0.03	0.21
The Hague	0.03	0.74
Rotterdam	-0.01	0.85
Utrecht	-0.15	0.00

Table 3.2: *The relationship between the relative charging fees and the frequency of use for each CP are displayed by showing the Pearson correlation values together with the corresponding p-values for each city.*

3.2.4 Parking Fees

In three of the four cities analyzed in the Netherlands (Amsterdam, The Hague and Rotterdam), EV users pay parking fees when charging their EVs outside their single designated residential parking permit area[5]. The fees depend on the tariff areas[6], on the time of day and on the day of the week. For example, the parking fees for parking on a Sunday from 9am to 11am are different from those in the same period on a Monday.

The municipalities of The Hague and Amsterdam have provided us with their parking zone information. In this section we are thus able to analyze the effect of parking fees for the agents located in these municipalities. The permit areas for Amsterdam and The Hague can be seen in Figure 3.11.



(A) (B)

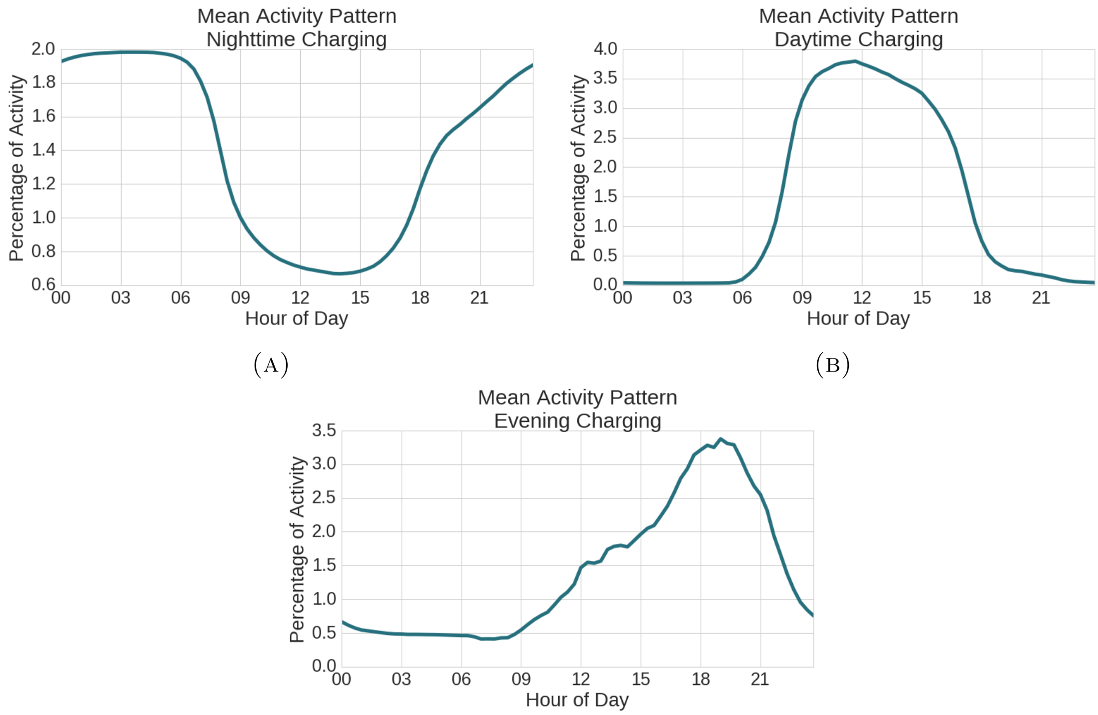
Figure 3.11: *The permit areas of* (A) *Amsterdam and* (B) *The Hague.*

The residential permit areas of the EV users in our model are not known. However, we can approximate them given the fact that the locations of these permit areas correspond with the home locations of the EV users. These home locations are also not known, but can, in turn, be approximated by using the activity patterns of each cluster. Figure 3.12

[5]Note that we only consider residential permit areas in this thesis. In future work non-residential permit areas (i.e. permits for parking at work) could also be considered.

[6]In The Hague tariff areas correspond to the permit areas but in Amsterdam the tariff areas are disconnected from the permit areas.

shows the three most prevailing activity patterns, and Figures 3.12a and 3.12c clearly show clusters which are mostly used in the evening and at night, strongly indicating that this is where the agent lives. A cluster with one of these two activity patterns is considered the 'home' cluster of an agent. The permit area in which the center of this 'home' cluster can be found is most likely the permit area of this agent. For 82.8% of the agents (922 agents) in Amsterdam and for 83.6% of the agents (489 agents) in The Hague we were able to identify their residential permit area in this way. The remaining agents in these two cities only have work clusters with less than 0.05% of their activity at 3am, meaning we cannot approximate their home address nor their residential permit area.



(A) (B) (C)

FIGURE 3.12: *The prevailing activity patterns. The activity patterns of all clusters were clustered using the Birch algorithm [35] with a Birch threshold of* 0.01 *and a branching factor of* 100. (A) *Nighttime charging* - 1983 *clusters.* (B) *Daytime charging* - 549 *clusters.* (C) *Evening charging* - 160 *clusters.*

There are four possible ways the clusters of the agents can interact with the permit areas, as is illustrated in Figure 3.13. Clusters can be located completely within the permit area, clusters can exist without a permit area when agents never charge at a home cluster and thus have no residential permit area, clusters can be partially inside and partially outside the permit area and clusters can be located completely outside the permit area.

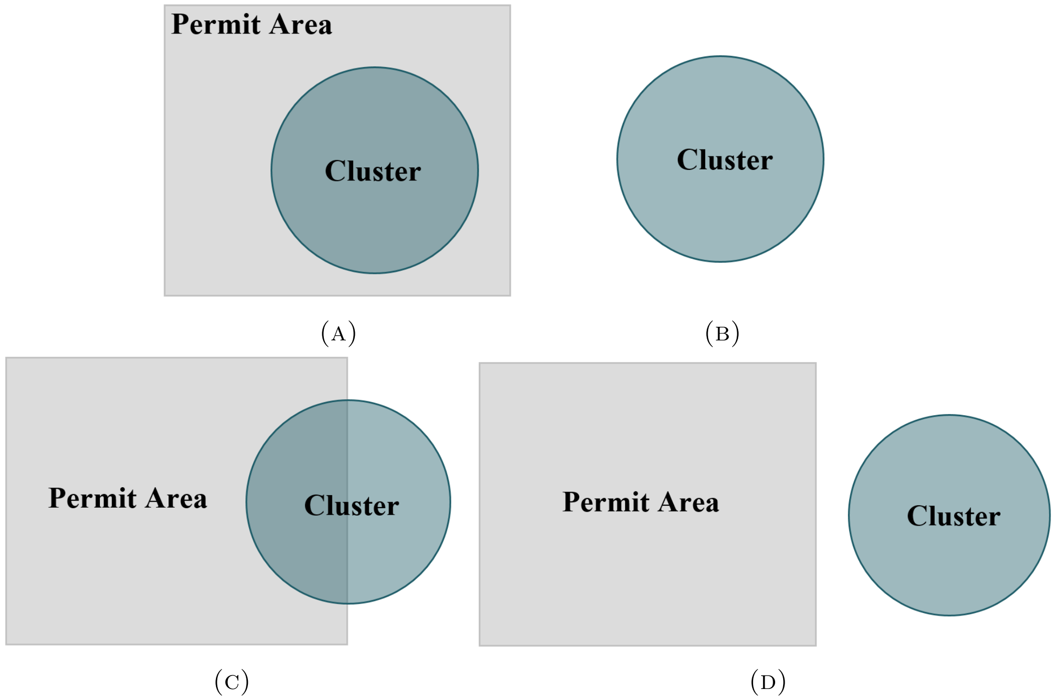


FIGURE 3.13: *Different types of interactions between the cluster and the permit area.* (A) *Cluster is completely inside the permit area of the agent.* (B) *Cluster of an agent without a permit area.* (C) *Cluster is partially inside and partially outside the permit area of the agent.* (D) *Cluster completely outside the permit area of the agent.*

Interaction type	Clusters in Amsterdam	Clusters in The Hague
Inside permit area	862	491
No permit area	192	99
Outside permit area	40	13
Inside and outside permit area	89	12

TABLE 3.3: *The cluster counts for each interaction type.*

Table 3.3 shows how the clusters interact with their permit areas for both Amsterdam and The Hague. As the majority of clusters fall within the permit area, we will first look at these clusters in more detail. Of the EV users with residential permit areas, 95% and 97% of the clusters fall completely within the permit areas for the cities of Amsterdam and The Hague, respectively. EV users clearly show a strong preference for charging and parking within their permit area, likely to avoid the additional parking fees. These clusters then fall completely within the permit areas because centers are weighted by the frequency of use of the CPs in the clusters. By not charging outside the permit areas, EV users are essentially pulling their centers away from the borders of the permit areas[7].

[7]This could be further analyzed if the actual home locations of the agents were known.

We continue with a price analysis for the remaining interaction types shown in Table 3.3[8] for the city of Amsterdam. Due to pricing differences the cities are studied separately, and because the parking fee distributions for the city of The Hague shown similar patterns, they will not be discussed in length in this thesis. Figures 3.14a and 3.15a show the distributions of parking fees (per hour) per transaction for the Amsterdam clusters for whose agents we could not identify a residential permit area. These figures show an average charging fee of 17.00 euro per transaction and an average charging fee of 2.72 euro per hour. As the agents probably do not pay for their parking fees in work clusters themselves (this is likely arranged via their job), they could show how EV users act when they are not conscious of the parking fees.

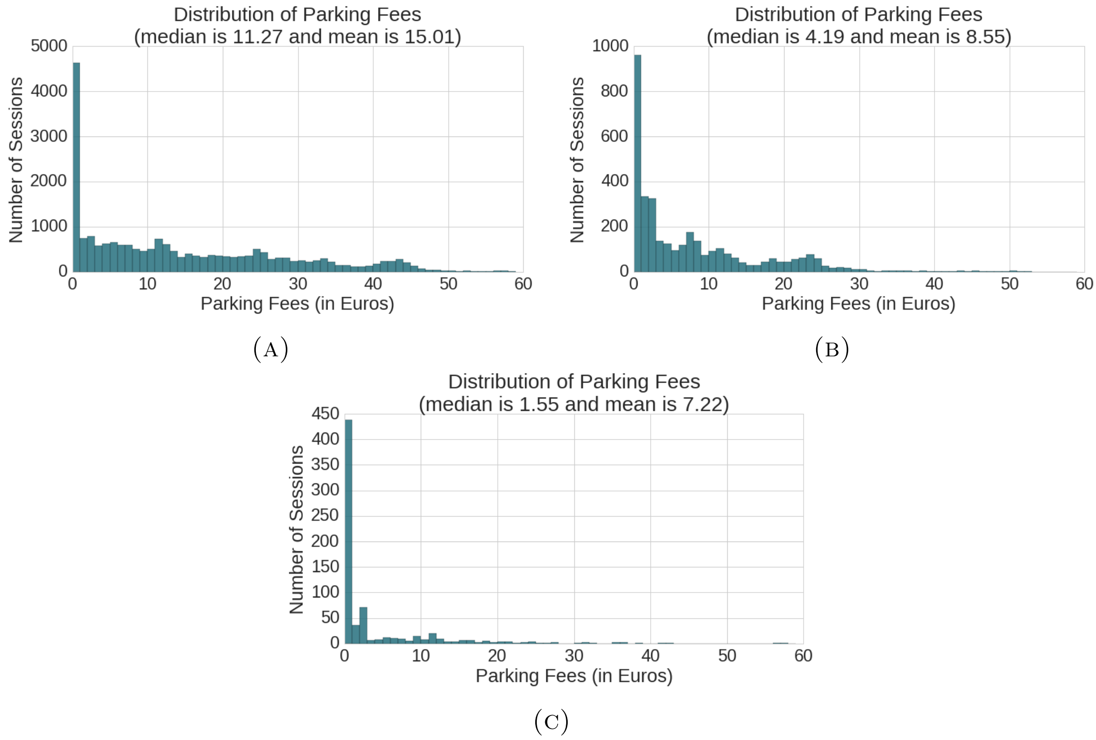


Figure 3.14: *Parking fees per transaction in Amsterdam for* (A) clusters belonging to agents without a permit area, (B) clusters completely outside the permit area and (C) clusters both inside and outside the permit area.

Figures 3.14b and 3.15b show similar distributions for the clusters which fall completely outside the permit area. Within these clusters there are no CPs within the permit area, and thus there is no possibility to choose a CP within the permit area. However, there is still the choice of when to charge and park, which does influence the parking fee per transaction. A clear difference can be seen between these figures and Figures 3.14a and 3.15a, as both the median and mean parking fees (per hour) per transaction are lower. The contextual difference between these clusters is that the clusters in the first figures

[8]Note that a dozen agents are removed from this analysis because we have manually discovered these agents' clusters are incorrectly determined.

do not have a residential permit area (they only charge at work) while the clusters in the second figure have a residential permit area. An explanation for the difference in average parking fees (per hour) could be that the EV users with cluster(s) outside the permit area are more price conscious because they have a permit area. It is likely that EV users who only charge at work do not have to pay for their parking fee and therefore do not take these costs into consideration.

Figures 3.14c and 3.15c show the parking fees for clusters which partially overlap with the permit areas. Here we see an even greater difference in parking fees, compared to Figures 3.14a and 3.15a. This indicates that EV users are even more price conscious when there is a choice not only in the time, but also in the location (as there is now a choice between CPs inside and outside the permit area).

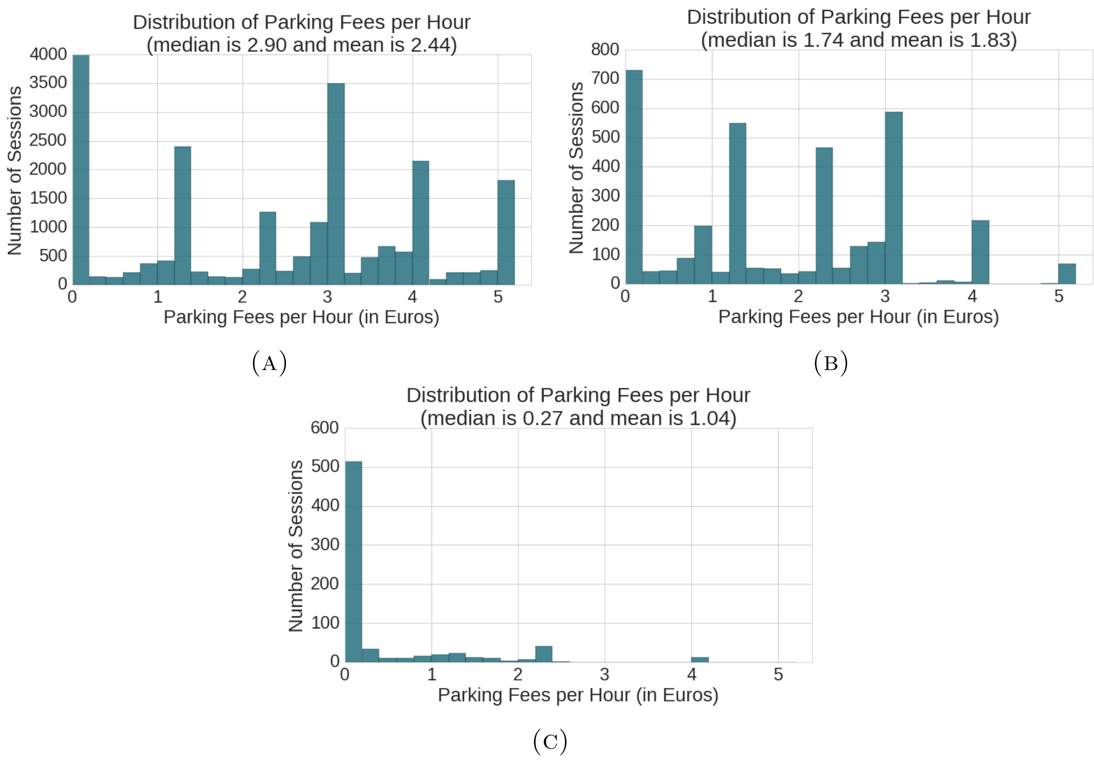


FIGURE 3.15: *Parking fees per hour per transaction in Amsterdam for* (A) clusters belonging to agents without a permit area, (B) clusters completely outside the permit area and (C) clusters both inside and outside the permit area.

The clusters which are partially inside the permit area show the behavior of EV users when they have a choice (in both the when and where dimensions) regarding parking fees. For this reason these clusters will be studied in more detail. Figure 3.16 shows the histograms of the percentage of transactions physically or economically outside the permit area. ‘Physically outside’ means the transaction was at a CP located outside the permit area while ‘economically outside’ means the transaction occurred at a CP located outside the permit area at a time when the parking fee at that CP was not zero.

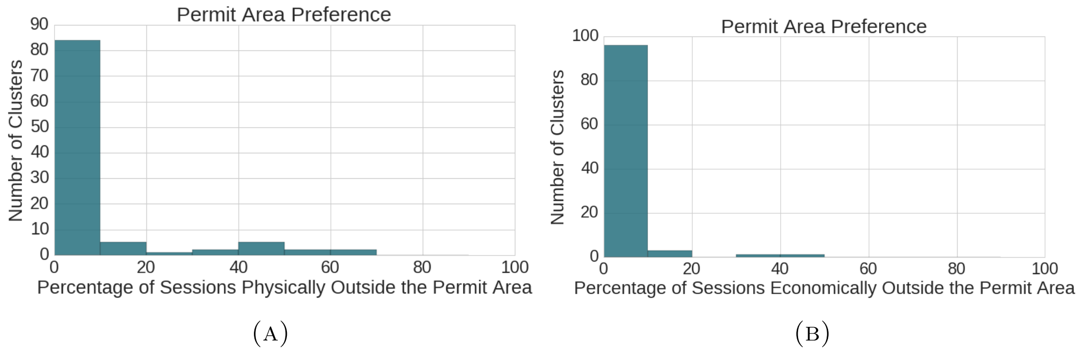


FIGURE 3.16: *Histograms of the percentage of transactions* (A) *physically and* (B) *economically outside the permit area.*

When comparing the two subfigures, it becomes clear that EV users are very conscious and aware about when they can charge for free outside their permit areas. Furthermore, Figure 3.16b shows that the percentage of transactions economically outside the permit area is between 0% and 10% for the majority of the clusters. This indicates a strong preferences for charging within the permit area when there is a choice.

3.2.5 Conclusion

In this section we have analyzed the four factors which could influence the decision process of EV users. For the first factor, distance, we have found that in 68.8% of all clusters there is a choice regarding this factor. We furthermore found a significant negative correlation between the distances of CPs and the frequency of use of CPs, indicating that this factor strongly influences how EV users select their CP.

For the second factor, charging speed, 28.1% of all clusters contain choices. Within these clusters no correlation was found between the charging speeds of CPs and their frequency of use. An explanation is that most of the CPs in our dataset are slow chargers, with a limited difference in charging speeds between them. Within the group of slow chargers, the charging speed does not significantly effect the frequency of use.

For the third factor, charging fees, 12.7% of the clusters contain choices. For this factor a small but significant correlation was found between the fees and the frequency of use. When examining this relation for each city individually, only Utrecht shows a significant correlation. An explanation for the lack of correlation for the other cities is that the difference in price between the options available is low, with a mean difference of less than one euro.

For the last factor, parking fees, 7.5% of clusters contain choices. We found that in the vast majority of these clusters, the percentage of transactions outside the permit area is between 0% and 10%. This indicates a strong preference for charging within the residential permit area and thus avoiding parking fees.

3.3 Choice Model

The previous section focused on the individual effects of the influencing factors. This section examines the effects of these variables in relation to each other by creating a discrete choice model. This model will then form the basis for the selection process implementation. Therefore, this choice model should be able to predict which CP the agent will choose, given the CPs in a cluster and their parameters (such as charging speed and distance to the center).

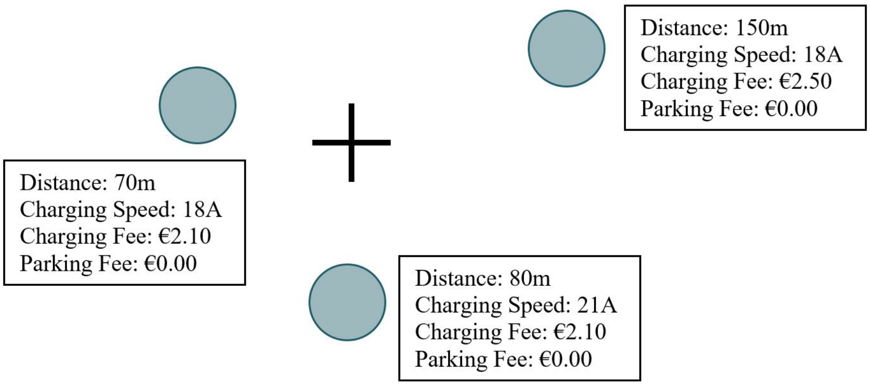


FIGURE 3.17: *An example of the choice EV users must make when selecting the CP to connect to. The circles indicate the CPs, while the cross indicates the center of the cluster to which the agent will connect.*

Multinomial logit models [40, 41], mixed logit models [39, 40, 43–45], random parameter logit models [39], nested logit models [46] and latent class logit models [43, 44] are popular techniques used for modeling charging choices. Researchers tend to prefer mixed logit models and latent class models as they offer a way to model the heterogeneity across individuals. As we will create a separate regression model for each cluster of each individual agent, there is no need to use such models. Therefore we will use a (binary) logistic regression model to model the charging choices.

A binary logistic regression model captures the relationship between the explanatory variables $X_1, ..., X_n$ and the outcome variable Y using coefficients $\beta_0, ..., \beta_n$ as can be seen in the following set of equations:

$$P(Y = 1|X) = \frac{e^a}{1 + e^a}, \text{ with} \tag{3.1}$$

$$a = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_n X_n. \tag{3.2}$$

The coefficients $(\beta_0, ..., \beta_n)$ are computed by fitting the model to training data. More information about the model used can be found in [51].

In our model the outcome variable Y indicates what the probability is that a CP would be chosen. The explanatory variables are the relative distance ($X_{1,relative}$), the relative charging speed ($X_{2,relative}$), the relative charging fee ($X_{3,relative}$) and the relative parking fee ($X_{4,relative}$) multiplied by the presence of a residential permit area (X_5). This results in the Equation 3.3.

$$a = \beta_0 + \beta_1 X_{1,relative} + \beta_2 X_{2,relative} + \beta_3 X_{3,relative} + \beta_{4,5} X_{4,relative} X_5 \tag{3.3}$$

For $i \in \{1, 2, 3, 4\}$, the $X_{i,relative}$ of a CP is calculated by dividing the absolute value of $X_{i,absolute}$ of that CP by the maximum $X_{i,absolute}$ found at CPs in the corresponding cluster. When there are no differences between the $X_{i,absolute}$ values within a cluster, the $X_{i,relative}$ values of all CPs in that cluster are set to zero. For each parameter $X_{i,absolute}$ is calculated as described in Section 3.2.

We have seen in the previous section that the presence of a permit area makes EV users more conscious about the parking fees. The interaction variable relative parking fee multiplied by the presence of a residential permit area is calculated by multiplying the relative parking fee by zero if no residential permit area was found. As a result the relative parking fees of agents who do not have a permit area (or live in cities for which we do not have the permit areas) are zero. This means that this variable is not taken into account for these agents. For agents who have a residential permit area, the relative parking fees remain the same as they are multiplied by one. Thus X_5 is zero if the agent does not have a permit area and one if the agent has a permit area.

3.3.1 Unweighted: Model per Cluster

One logistic regression model was created for each cluster of those agents that have ever chosen more than one CP within a single cluster. Creating one model for each cluster allows us to investigate differences and heterogeneity between agents. In a third of the clusters multiple choices have been made. Note that there is no apparent need for agents to choose multiple CPs when they can always charge at their preferred CP.

Table 3.4 shows the coefficient values at the CPs for each explanatory variable. We can see that in this cluster the EV user is sensitive to distance, charging speed and charging fee. More specifically, both the coefficients of the relative distance and relative charging fee variables are negative, meaning that CPs with lower distances and lower charging fees are more likely to be chosen. The coefficient for the charging speed variable is positive, meaning that CPs with higher charging speeds are preferred. The coefficient for the charging fee is zero as the agent has no choices regarding charging fees.

Coefficient	Value
β_0	-2.752
β_1	-2.071
β_2	6.887
β_3	-2.838
$\beta_{4,5}$	0.000
Accuracy	0.96

TABLE 3.4: *An example of the resulting logistic model for one cluster of one agent in Amsterdam.*

Figure 3.18 shows a summary of the coefficient values for all models. Here we can see the distributions of each coefficient, together with the relations between these coefficients, for each city. The coefficient for distance (β_1) has a larger range of values compared to the other coefficients. The coefficient for the parking fee multiplied by the permit area variable ($\beta_{4,5}$) is either zero or negative. Only agents in the cities of Amsterdam and The Hague have non-zero values for this $\beta_{4,5}$ coefficient. This corresponds to the fact that permit areas are only available in these two cities (the permit areas of Rotterdam are not available and Utrecht does not have permit areas). No correlations were found between the different explanatory variables. From this figure we can distinguish no clear user groups or clear differences between users within a single city. This seems to indicate a lack of heterogeneity between habitual EV users. This confirms that no mixed logit model or latent class logit model is needed to capture the choices made by the habitual agents.

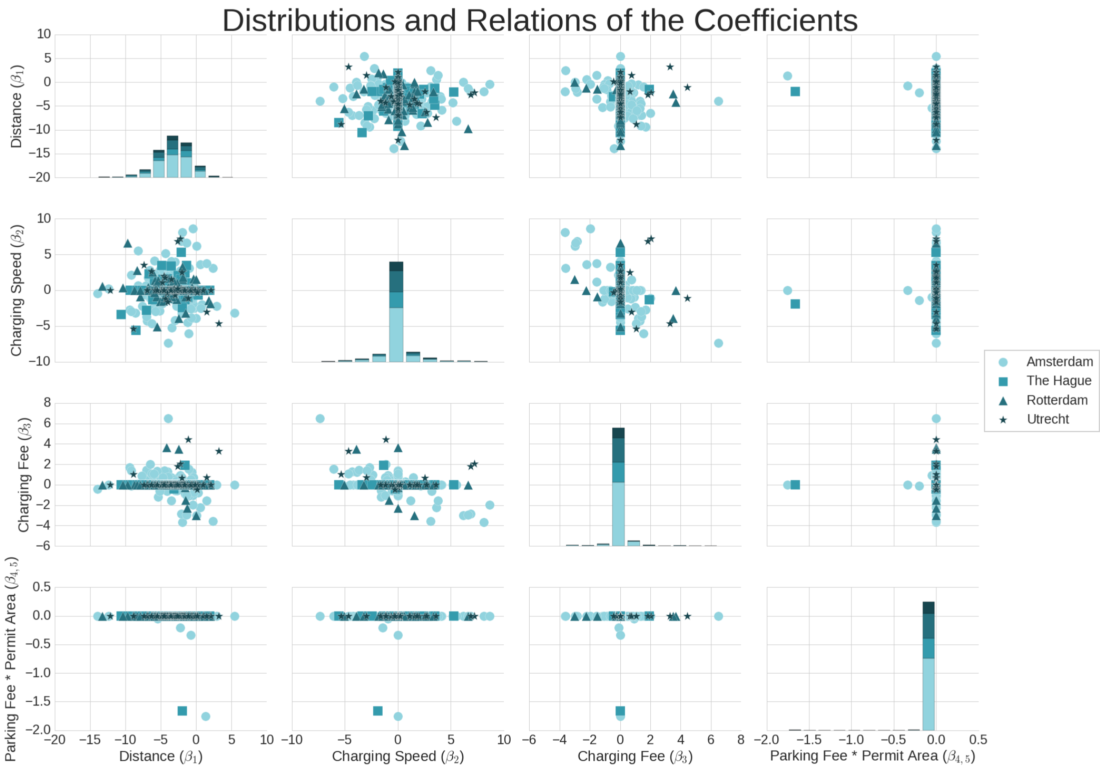


FIGURE 3.18: *The distributions of the coefficient values and the relations between them. Each dot represents the logistic regression model for one cluster of one agent.*

Figure 3.19 shows the confidence interval for each coefficient and city. This figure indicates that the relative distance ($X_{1,relative}$) has a strong significant negative effect on whether a CP is chosen (odds-ratio of 0.04). In contrast, the relative charging speed ($X_{2,relative}$), relative charging fee ($X_{3,relative}$) and the relative parking fee multiplied with the permit area indicator ($X_{4,relative}X_5$) have little to no (significant) effects on whether a CP is chosen. This figure also shows the difference between the city of Utrecht and the other cities. The reason that the confidence bars are larger for Utrecht is that there is less data available for this city. 58% of the agents are located Amsterdam, 18% are located in Rotterdam, 15% in The Hague and 9% in Utrecht.

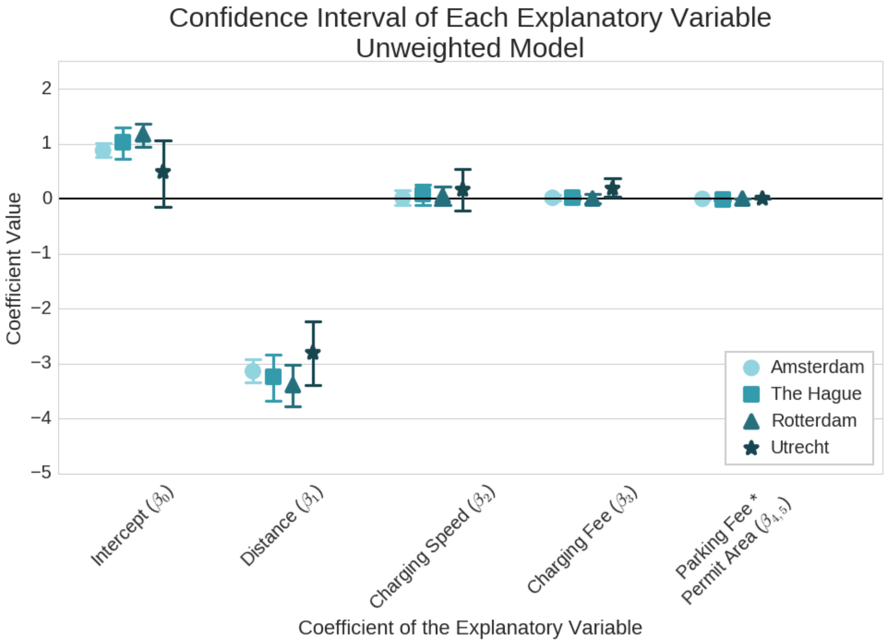


FIGURE 3.19: *A summary of the resulting logistic models by visualizing the 95% confidence intervals for the values of each of the coefficients of the explanatory variables for each city. The mean (overall) values are* 0.85 *for* β_0, -3.12 *for* β_1, 0.01 *for* β_2, 0.03 *for* β_3 *and* -0.01 *for* $\beta_{4,5}$. *The mean accuracy of the individual models is* 0.81.

In Section 3.2 we have seen that charging fees and parking fees have an effect on the selection of a CP. However in Figure 3.19 the coefficients of these variables are near zero. A possible explanation for this is that there are many clusters in which there is no choice regarding these variables (which was also shown in Section 3.2). If, and only if, there is no choice in a cluster, the coefficient for said cluster becomes (exactly) zero. This means that if many clusters have no choice, this could result in a mean coefficient of near zero. Figure 3.20 confirms this, by showing the percentages of coefficients equal to zero for each explanatory variable. This figure shows that few clusters have choices regarding charging fees and parking fees, confirming our hypothesis that clusters with few choices 'pull' the coefficient values to zero.

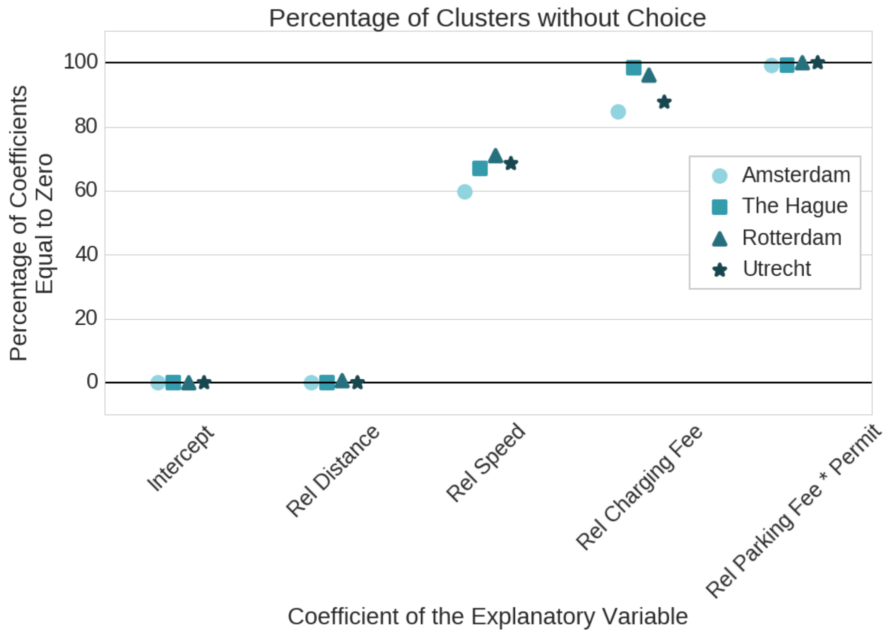


FIGURE 3.20: *The percentage of clusters without choice for each explanatory variable.*

3.3.2 Weighted Model

To get coefficients which are not biased due to lack of choice, we need to weight the agents with more choices more strongly. We can do this by creating one logistic regression model per city. This means adding all training data within one city together, which we can easily do as the explanatory variables are all relative. By doing this we ensure that agents with small amounts of training data do not skew the coefficient values and that agents with large amounts of training data influence the model more strongly.

Figure 3.21 shows that there is a strong correlation between the amount of training data in a cluster and the number of CPs in range. As Section 3.2 showed that more CPs in range means more choices, we can conclude that more training data in a cluster means more choices in a cluster. Thus clusters in which there are more choices influence the model more strongly.

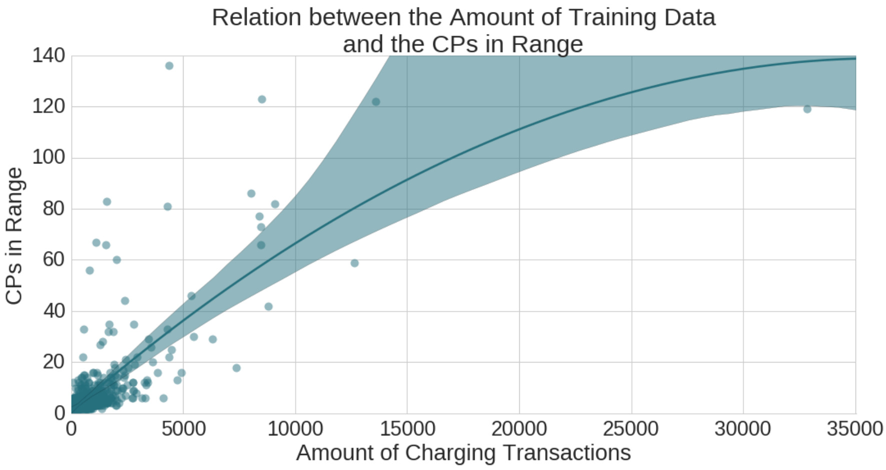
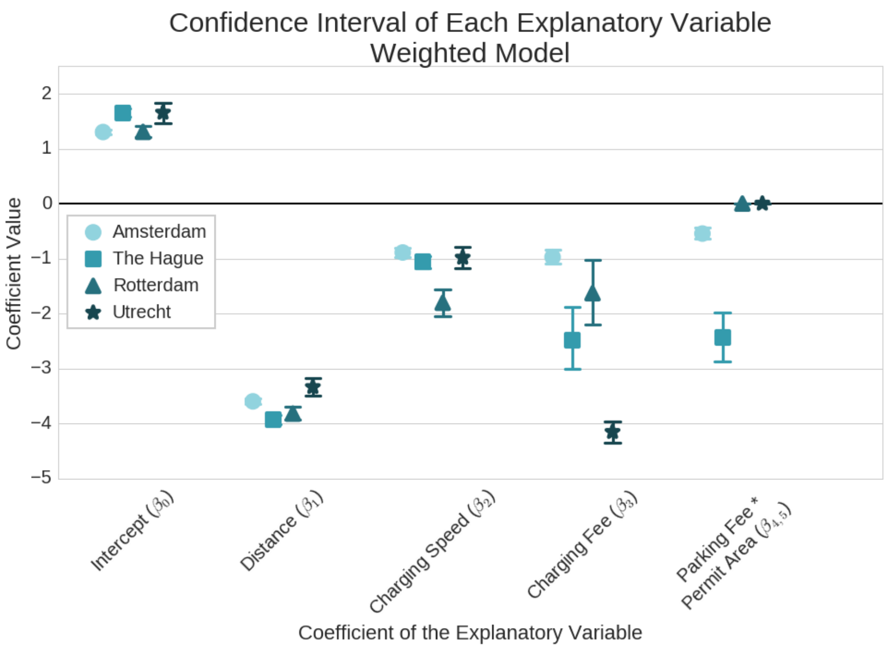


FIGURE 3.21: *The correlation between the amount of training data of a cluster and the number of CPs in range in this cluster. One dot represents one cluster. A Pearson correlation of* 0.73 *was found* ($p = 0.00$).

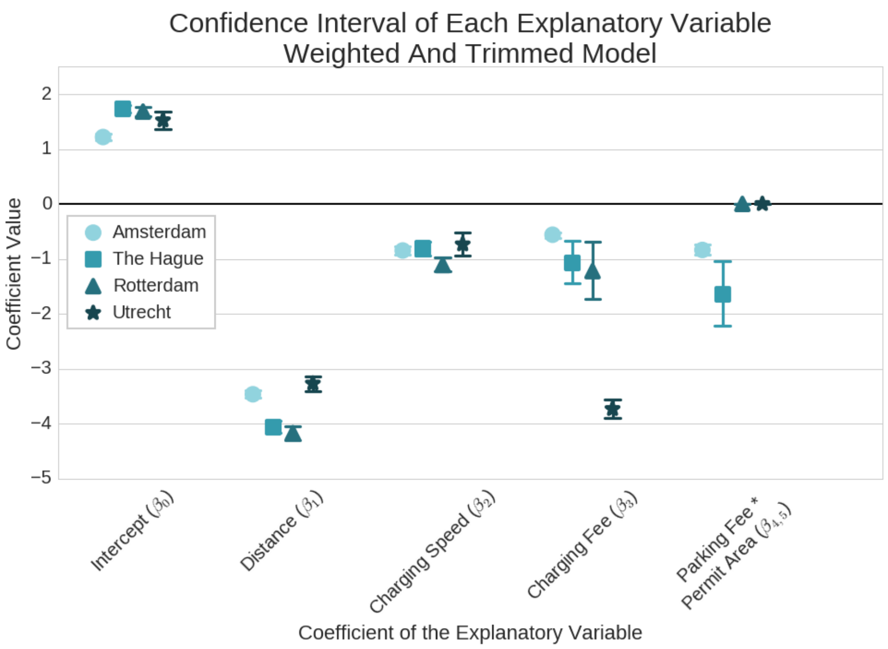
Figure 3.22a shows the 95% confidence intervals for the coefficient values for the logit model per city (the ‘weighted’ approach). Figure 3.22b shows the confidence intervals when the outliers are removed (the ‘weighted and trimmed’ approach) by taking the 99^{th} percentile. The result of removing these outliers is that the coefficient values for different cities group closer together for the charging speed and charging fee variable. For the intercept, distance and charging speed variables there are no significant differences between the four cities. For the charging fee variable we see the negative effect of the charging fee is stronger for the city of Utrecht than for the cities of Amsterdam, The Hague and Rotterdam. Furthermore we see the variable parking fee multiplied by the permit area indicator is zero for both Rotterdam and Utrecht, as these cities do not have (known) permit areas. In both Amsterdam and The Hague there is a negative coefficient for this variable, meaning that there is a negative effect of parking fees on the selection of a CP.

A summary of the coefficient values for each modeling approach can be found in Appendix B. Note that these coefficients are learned by using the existing habitual agents, given the training data of the years 2014, 2015 and 2016.

As new data becomes available, this model could be rerun to recalculate the coefficients. In the future we expect there to be more choices regarding the different influencing factors, as the number of CPs is expected to keep increasing. More choices might mean that the influence of the factors becomes more pronounced and that this would be reflected in the coefficients. However, our ‘weighted’ and ‘weighted and trimmed’ models already weights agents with more choices more strongly, meaning the coefficients might stay the same. Furthermore, new and unforeseen policies could influence the existing factors or create new factors which influence the decision process of the agents.



(A)



(B)

FIGURE 3.22: *A summary of the resulting logistic models by visualizing the 95% confidence intervals for the values of each of the coefficients of the explanatory variables for each city.* (A) *shows the 'weighted' model while* (A) *shows the 'weighted and trimmed' model.*

3.4 Model Description

The logistic regression function is implemented into the SEVA model by replacing the existing selection process, and thus creating the SEVA-ICPS model. The selection process takes as input the cluster in which the agent wants to connect and returns as output the selected CP within this cluster at which the agent will connect.

All CPs in range of the walking preparedness of an agent will be taken into consideration. Each of these CPs is assigned a probability of being chosen, given the relative distance of each CP to the center ($X_{1,relative}$), the relative charging speed at each CP ($X_{2,relative}$), the relative charging fee for the transaction at each CP ($X_{3,relative}$) and the relative parking fee for the transaction at each CP multiplied by the permit area indicator ($X_{4,relative}X_5$). The equation to calculate the probability of being chosen ($Y = 1$), given the explanatory variables (X) of the CP can be seen in Equation 3.4.

$$P(Y = 1|X) = \frac{e^a}{1 + e^a}, \text{ with} \tag{3.4}$$

$$a = \beta_0 + \beta_1 X_{1,relative} + \beta_2 X_{2,relative} + \beta_3 X_{3,relative} + \beta_{4,5} X_{4,relative} X_5$$

The coefficient values (β_i) are dependent on the city in which the cluster is located and on the logistic regression approach. These β_i values can be found in Table B.1.

The SEVA model does not contain the state of charge and kWh charged per transaction. However the kWh charged per transaction is needed to calculate the relative charging fee for each transaction at each CP this must be added to the SEVA model. The number of kWh charged per transaction is not normally distributed (as can be seen in Figure 3.23), meaning this number cannot be approximated by the mean number of kWh charged in the training data.

Figure 3.24a shows that there is a relation between the connection duration, which is known in the model, and the kWh charged. We see that the number of kWh charged levels off. This is probably because when EV users charge at home, they often leave their EVs plugged in all night, and do not unplug them even when their batteries are full. Some transactions do appear to have a small number of kWh charged while the connection duration is long. By taking the mean for each connection duration bin (as can be seen in Figure 3.24b), we can approximate the number of kWh charged in a data-driven way.

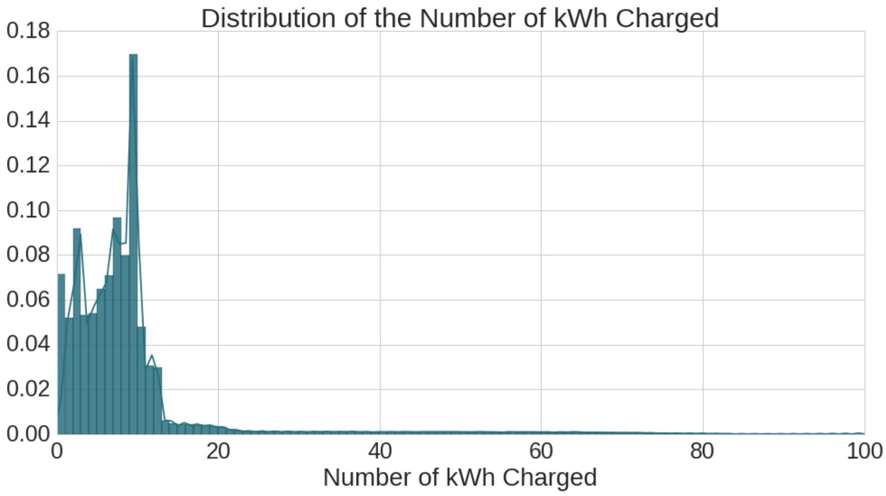


FIGURE 3.23: *The distribution of the number of kWh charged. The normaltest confirmed that this distribution significantly differs from a normal distribution.*

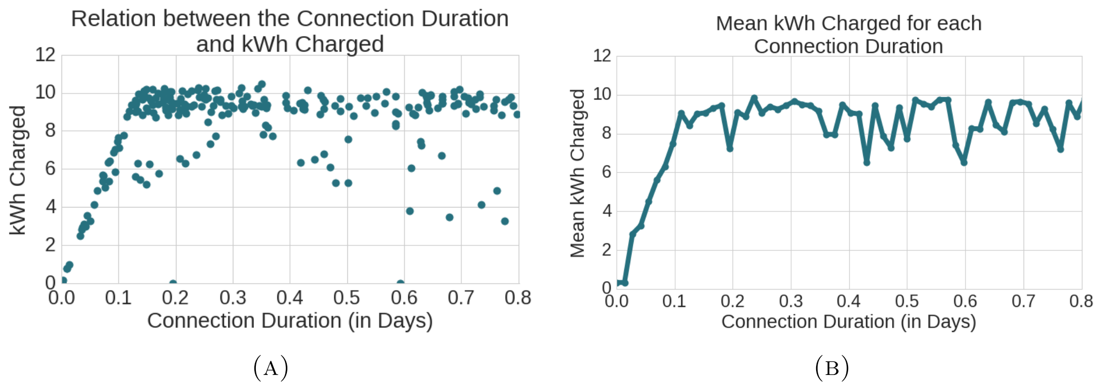


FIGURE 3.24: *An example of the relation between the connection duration and the number of kWh charged for one agent.* (A) *Scatterplot and* (B) *mean value per 20 minutes to show the relation between the number of kWh charged and the connection duration. Each dot represents one transaction of the one agent's training data.*

Listing 3.1 shows precise details for the updated selection process. To summarize, an agent chooses a CP in range based on probabilities given by choice model function. If this CP is occupied, the agent chooses again excluding this CP as it is now known to be occupied. When a CP is chosen which is occupied, this is defined as a failed connection attempt. This process continues until either the agent chooses an unoccupied CP or until all CPs in range are found to be occupied. If the latter occurs, the agent tries to connect again at a later time (the value of this reconnection time t_r can be found Table A.3).

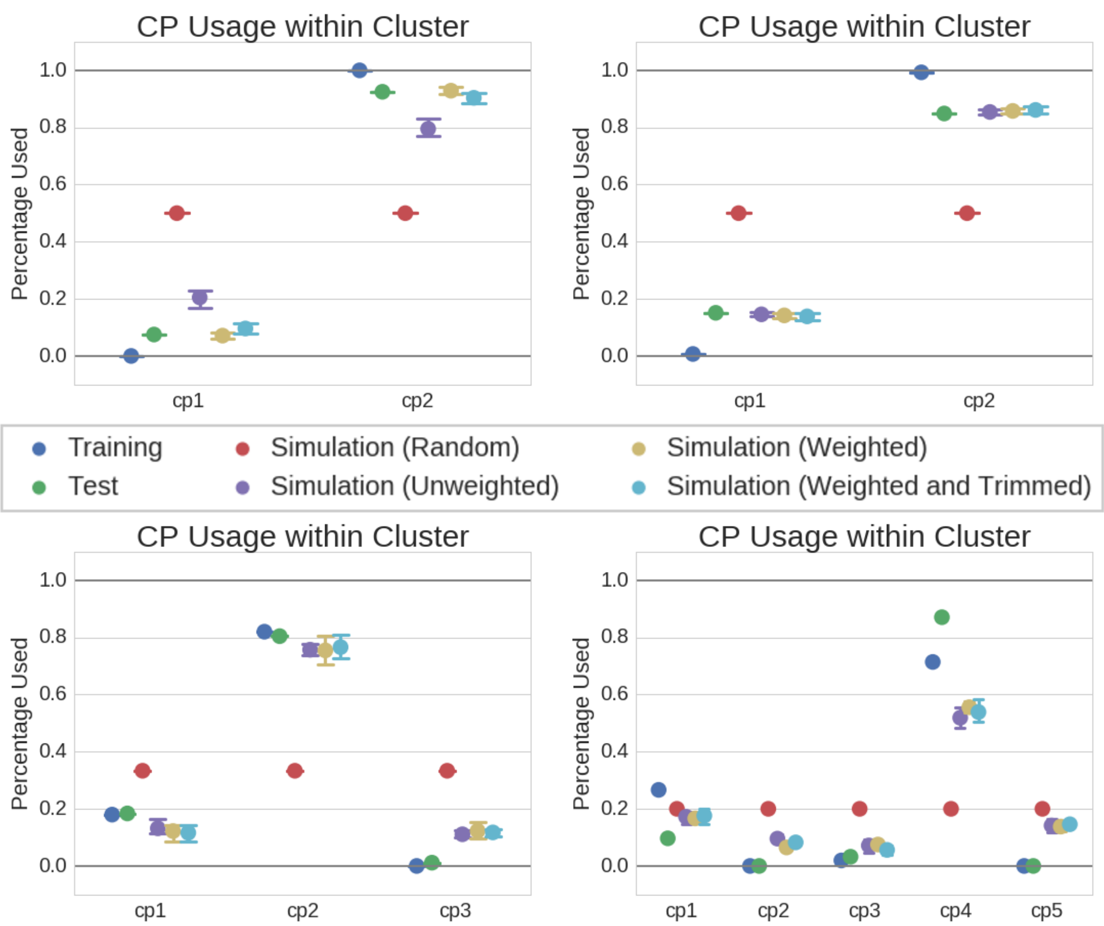
```
def select_cp_using_choice_model(cps_in_range):
    chosen = False
    cps_tried = {}

    while not chosen or cps_tried != cps_in_range:
        # Determine probability of choosing each CP, using the choice model.
        # Exclude any CPs that are already tried.
        probs = []
        for cp in cps_in_range:
            if cp not in cps_tried:
                probs.append(get_prob_using_choice_model(cp))
        chosen_cp = choose_cp_based_on_probs(probs)

        if is_occupied(chosen_cp):
            # Failed connection attempt.
            cps_tried.append(chosen_cp)
            if cps_tried == cps_in_range:
                # All CPs are occupied.
                # The agent must try to connect at a later time.
                active_cp = None
        else:
            # Successful connection attempt, a CP has been chosen.
            chosen = True
            active_cp = chosen_cp
    return active_cp

def get_prob_using_choice_model(cp):
    city_of_cp = get_city_of_cp(cp)
    beta0, beta1, beta2, beta3, beta45 = get_betas()

    a = beta0[city_of_cp] +
        beta1[city_of_cp] * get_distance(cp) +
        beta2[city_of_cp] * get_charging_speed(cp) +
        beta3[city_of_cp] * get_charging_fee(cp) +
        beta45[city_of_cp] * get_parking_fee(cp) * permit_indicator(agent)
    return numpy.exp(a) / (1 + numpy.exp(a))
```

LISTING 3.1: *Pseudo code showing the improved CP selection process.*

3.5 Model Evaluation

Figure 3.25 shows examples of the usage of the CPs in four clusters when the new selection process is used. The dots represent the percentage of use for each CP, for either the training data, the test data or the simulation data. Four simulation approaches are shown. With the random approach, each CP is chosen with equal probability, This approach is used as a benchmark case. The 'unweighted', 'weighted' and 'weighted and trimmed' approaches use the logistic regression model described in Section 3.3 with coefficients as shown in Table B.1.

(A) (B)



(C) (D)

FIGURE 3.25: *Examples of the usage of CPs within clusters.*

To quantitatively analyze the performance of the selection process, we propose a performance metric based on the percentage of usage for each CP within a cluster for the training data ($u_{tr}(cp)$), the test data ($u_{te}(cp)$) and the simulation data ($u_{si}(cp)$). Equations 3.5, 3.6 and 3.7 show that these usages are calculated by dividing the number of transactions in the training, test or simulation data ($|cp|_{tr}$, $|cp|_{te}$ and $|cp|_{si}$ respectively) by the total number of transactions in the cluster. $|\mathrm{cp}|_{\mathrm{si}}$ is dependent on the used modelling approach, as each approach can cause the number of transactions at each CP to differ.

$$\mathrm{u_{tr}(cp)} = \frac{|\mathrm{cp}|_{\mathrm{tr}}}{\sum_{\mathrm{cp}' \in \mathrm{cluster}} |\mathrm{cp}'|_{\mathrm{tr}}} \tag{3.5}$$

$$\mathrm{u_{te}(cp)} = \frac{|\mathrm{cp}|_{\mathrm{te}}}{\sum_{\mathrm{cp}' \in \mathrm{cluster}} |\mathrm{cp}'|_{\mathrm{te}}} \tag{3.6}$$

$$\mathrm{u_{si}(cp)} = \frac{|\mathrm{cp}|_{\mathrm{si}}}{\sum_{\mathrm{cp}' \in \mathrm{cluster}} |\mathrm{cp}'|_{\mathrm{si}}} \tag{3.7}$$

The training error of a cluster ($\mathrm{e_{tr,si}}$), test error of a cluster ($\mathrm{e_{te,si}}$) and inherent behavior change within a cluster ($\mathrm{e_{tr,te}}$) is calculated by taking the absolute difference in usage of each CP in that cluster, weighted by the percentage of usage of each CP (see Equations 3.8, 3.9 and 3.10). This ensures deviations of CP usage for CPs with more usage are weighted more strongly. Note that this error measure is scaled between 0 and 1, where 0 is the best possible score as this indicates there is no deviation for any of the CPs in the cluster and 1 is the worst possible score. The error scores of each of the clusters in Figure 3.25 can be seen in Table 3.5.

$$\mathrm{e_{tr,si}(cluster)} = \sum_{\mathrm{cp} \in \mathrm{cluster}} |\mathrm{u_{tr}(cp)} - \mathrm{u_{si}(cp)}| \cdot \mathrm{u_{tr}(cp)} \tag{3.8}$$

$$\mathrm{e_{te,si}(cluster)} = \sum_{\mathrm{cp} \in \mathrm{cluster}} |\mathrm{u_{te}(cp)} - \mathrm{u_{si}(cp)}| \cdot \mathrm{u_{te}(cp)} \tag{3.9}$$

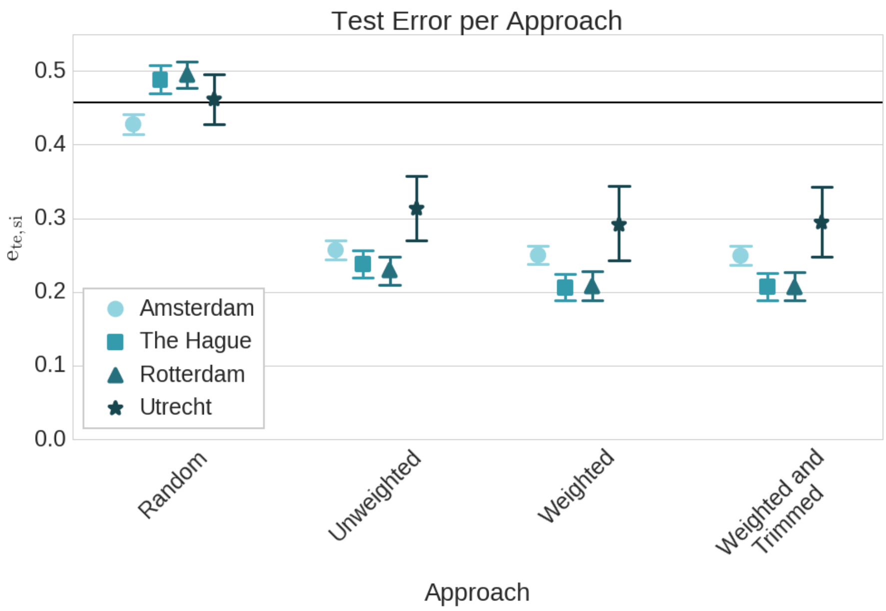
$$\mathrm{e_{tr,te}(cluster)} = \sum_{\mathrm{cp} \in \mathrm{cluster}} |\mathrm{u_{tr}(cp)} - \mathrm{u_{te}(cp)}| \cdot \mathrm{u_{tr}(cp)} \tag{3.10}$$

Error Type	(a)	(b)	(c)	(d)
$\mathrm{e_{tr,si}}$ (Random)	0.50	0.49	0.43	0.39
$\mathrm{e_{tr,si}}$ (Unweighted)	0.20	0.14	0.06	0.17
$\mathrm{e_{tr,si}}$ (Weighted)	0.07	0.14	0.06	0.14
$\mathrm{e_{tr,si}}$ (Weighted and Trimmed)	0.10	0.13	0.04	0.15
$\mathrm{e_{te,si}}$ (Random)	0.43	0.35	0.41	0.60
$\mathrm{e_{te,si}}$ (Unweighted)	0.13	0.01	0.05	0.32
$\mathrm{e_{te,si}}$ (Weighted)	0.07	0.01	0.05	0.28
$\mathrm{e_{te,si}}$ (Weighted and Trimmed)	0.10	0.01	0.04	0.30
$\mathrm{e_{tr,te}}$	0.07	0.14	0.01	0.16

TABLE 3.5: *The error scores for each of the clusters in Figure 3.25.*

Next we examine how each approach performs for all clusters. Figure 3.26 shows the mean training and test errors for the different approaches and different cities. All three modelling approaches perform significantly better than the benchmark case (random) in the training, as the error scores decrease by at least 50%. Compared to the test data all approaches still perform significantly better, with error scores decreasing by approximately 30%. Utrecht performs best on the training data but worst on the test data, which indicates that the model is overfitting for Utrecht, perhaps because of the limited amount of data for this city.

(A)



(B)

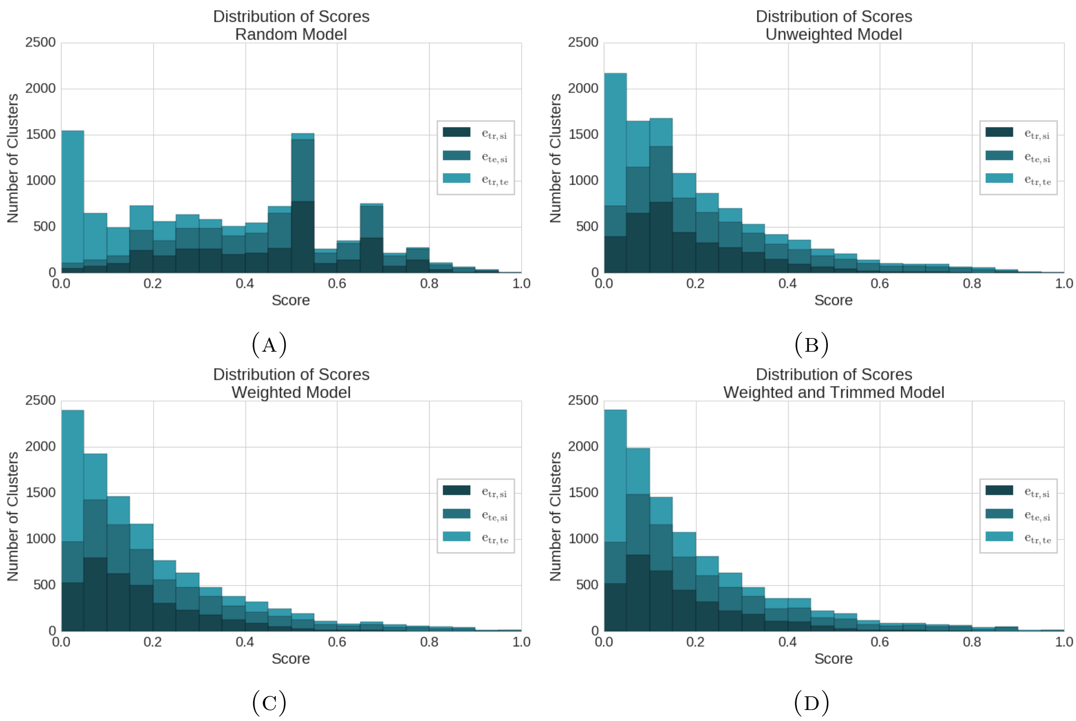
FIGURE 3.26: *The mean* (A) *training and* (B) *test errors with a 95% confidence interval for each modelling approach and each city.*

Table 3.6 shows the mean overall errors for each approach. A t-test was used to show that the differences between the approaches are significant between all approaches except for the approaches ‘weighted’ and ‘weighted and trimmed’. Furthermore this table shows that the average, weighted deviation in usage percentage per CP is 0.18 and 0.24 for the training and test data, respectively.

Approach	$e_{tr,si}$	$e_{te,si}$
Random	0.44	0.46
Unweighted	0.19	0.25
Weighted	0.17	0.23
Weighted and Trimmed	0.17	0.23

TABLE 3.6: *The mean overall error for both the training data and the test data. The differences are significant between all approaches except between ‘Weighted’ and ‘Weighted and Trimmed’.*

Figure 3.27 shows the distributions of error scores for each of the modelling approaches. The difference between the unweighted approach and the two other non-random approaches can be seen in the third bin, which is higher for the unweighted approach. For the ‘weighted’ and ‘weighted and trimmed’ approaches, we see that the distributions of training errors, test errors and inherent behavior change follow similar trends. This indicates that the training and test error scores can be, in part, explained by the inherent behavioral change of agents within each cluster.

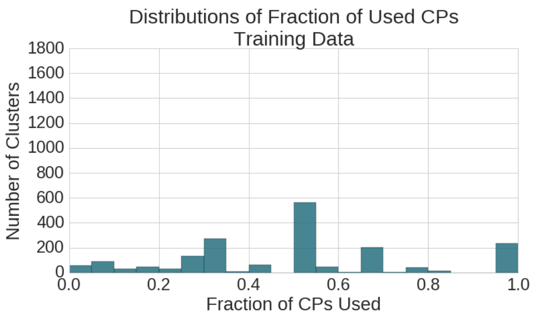


(A) (B) (C) (D)

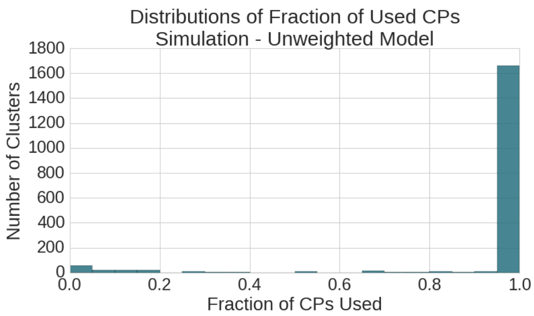
FIGURE 3.27: *The distribution of error scores for each of the error metrics in Equations 3.8, 3.9 and 3.10.*

3.5.1 Effect of Cluster Size

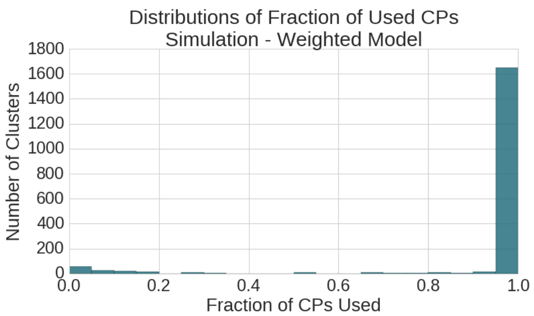
Next we examine the effect of the size of the clusters on the validation of the clusters. Cluster size is defined as the number of CPs used in either the training data, the test data or the simulation data within a cluster. This does not have to be equal to the number of CPs in range, although Figure 3.28 shows that for all our simulation approaches the cluster size is equal to the number of CPs in range in 93% of the clusters. Note that the benchmark approach (random) is not shown, as all clusters always use all CPs in this modelling approach.



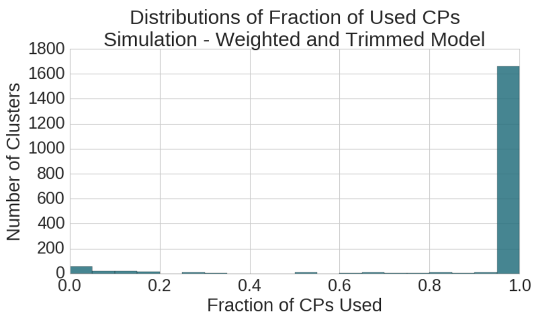
(A)



(B)



(C)



(D)

FIGURE 3.28: *The distributions of the fractions of CPs used within a cluster for each approach compared to the actual fraction in the training data. In the training data 13% of the clusters use all CPs in range. In each the three approaches 89% of the clusters use all CPs in range.*

Figure 3.29 shows the error scores for all cluster sizes for the 'weighted and trimmed' approach[9]. We see that the higher the cluster size, the higher the training and test error scores. This corresponds to what we expect as the more choices there are, the harder it is to choose the correct one. However we do see that this increase appears to be logistic, rather than linear or exponential. From this we can conclude that the modelling approach can work for clusters with large cluster sizes.

[9] Both the 'weighted' and 'weighted and trimmed' approaches are significantly better than the other proposed model and the benchmark case. Between these approaches there is no significant difference and we chose 'weighted and trimmed' as it better matches with the data analysis in section 3.2.

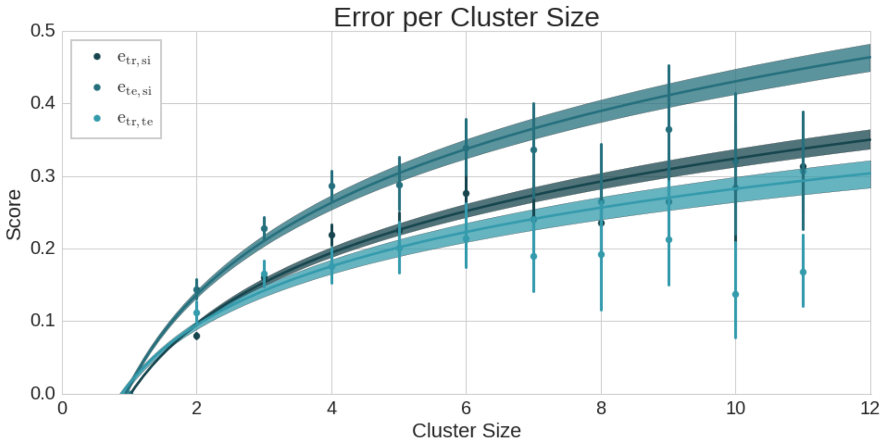


FIGURE 3.29: *The effect of the cluster size on the error scores with 95% confidence interval. A regression fit is plotted through these points, also with a 95% confidence interval for the regression estimate. The 'weighted and trimmed' approach is used.*

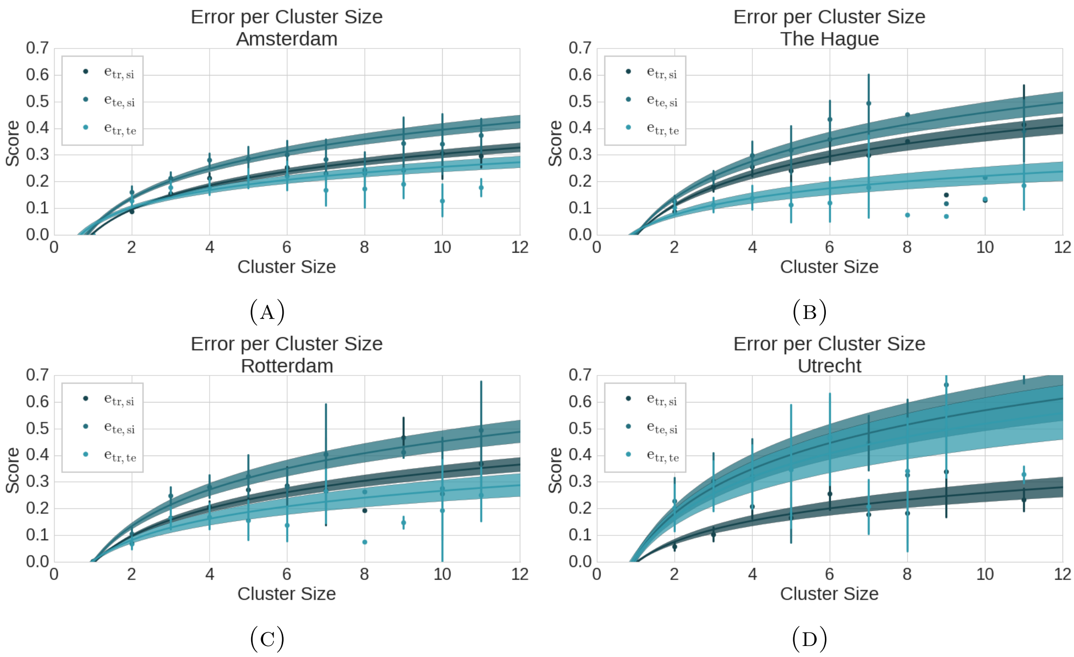


FIGURE 3.30: *The mean error score per cluster sizes for each city. The 'weighted and trimmed' modelling approach is used.*

Figure 3.30 shows the error per cluster size for each individual city. Firstly, this figure shows each city also follows the logistic increase seen for all cities together in Figure 3.29. Secondly, Utrecht has a much higher inherent behavioral change (the light blue line) compared to the other cities. This behavioral change between 2014-2015 and 2016 indicates that external factors influenced the behavior of the EV users in Utrecht, making it harder to correctly predict this city for 2016. This furthermore offers an explanation for the higher $e_{te,si}$ for Utrecht in Figure 3.26b. A possible external factor which caused this behavioral change is the ending of the concession period in Utrecht on January 1st, 2016.

3.5.2 Validation of Clusters Containing Newly Added CPs

In this section we examine how clusters containing newly added CPs validate. Newly added CPs are defined as CPs which have been added after the end date of the training data (January 2016). Clusters with newly added CPs in range thus contain CPs which have been added after January 1st, 2016.

Figure 3.31 shows the differences in this distribution of test scores between clusters with newly added CPs and clusters without newly added CPs. While these types of clusters both follow a similar distribution per approach, the clusters with newly added CPs appear to contain more high test error values. This is confirmed in Figure 3.32, which shows that, on average, the mean test error of the clusters with new CPs is significantly higher compared to clusters without new CPs. This is not what we would want, as we aim to develop an improved selection process which can accurately predict if and how new CPs are used.

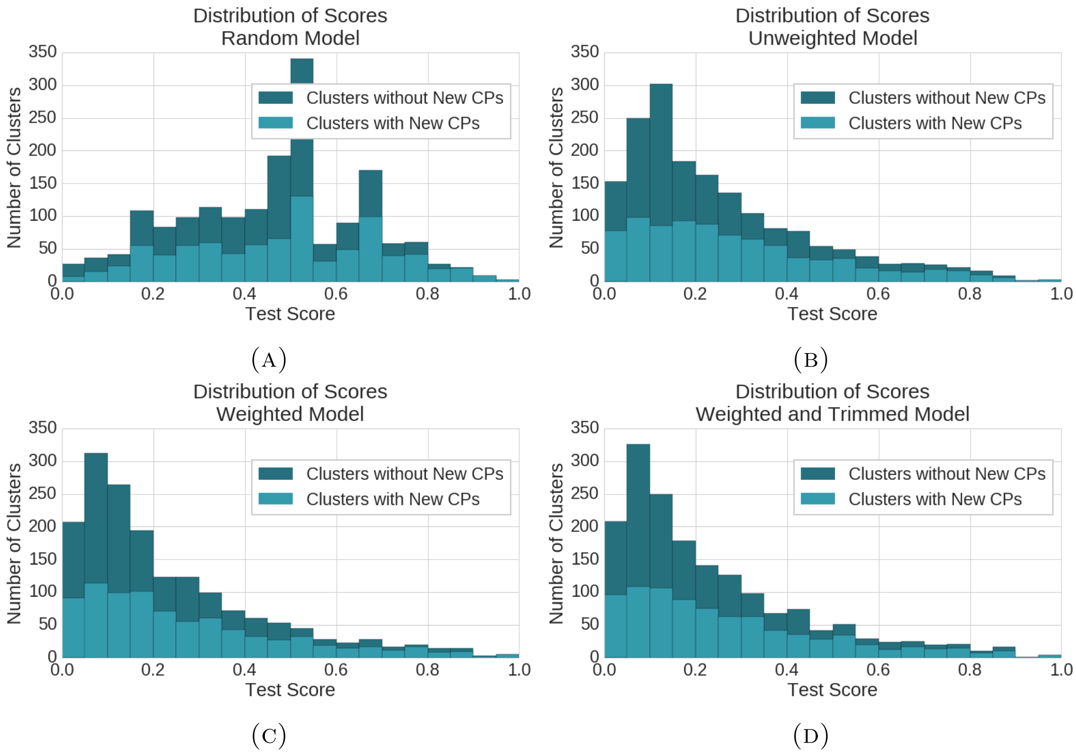


FIGURE 3.31: *The distribution of test scores ($e_{te,si}$) for each of the approaches. The stacked bars show the difference between clusters with newly added CPs and clusters without newly added CPs.*

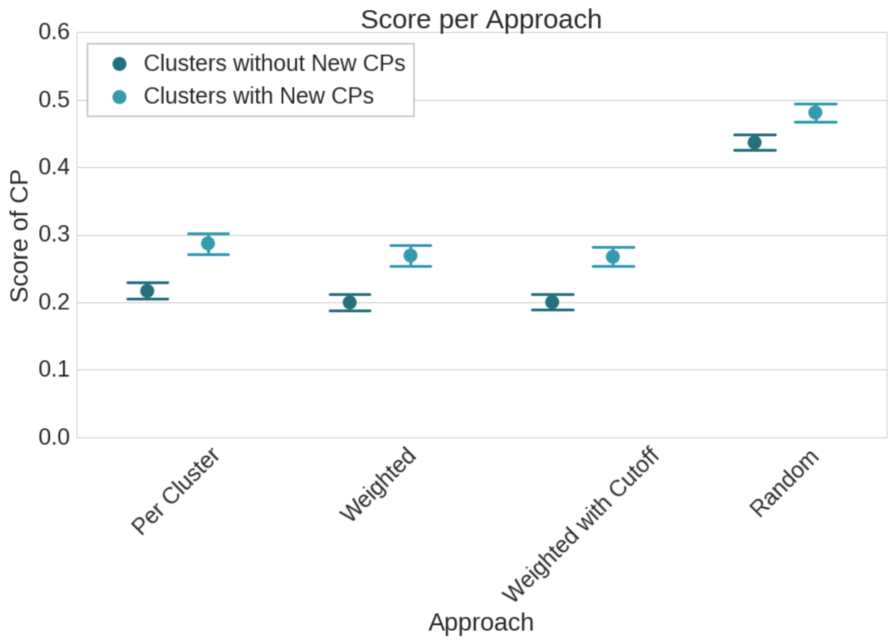


FIGURE 3.32: *The mean test score of each approach, for both clusters which have newly added CPs and clusters which do not have newly added CPs. A 95% confidence interval is shown.*

A possible explanation why clusters containing newly added CPs have higher test error errors is that clusters with new CPs are more likely to have larger cluster sizes. As we saw in Figure 3.29, larger clusters sizes result in higher training and test errors. Figure 3.33 indeed shows that clusters which have CPs which have been placed after 2015 have a larger cluster size. Figure 3.34 furthermore shows that when we compared test errors per cluster size between clusters with and without newly added CPs, there is no significant difference between these cluster types. Thus the difference in test score between these types was solely due to the difference in cluster sizes.

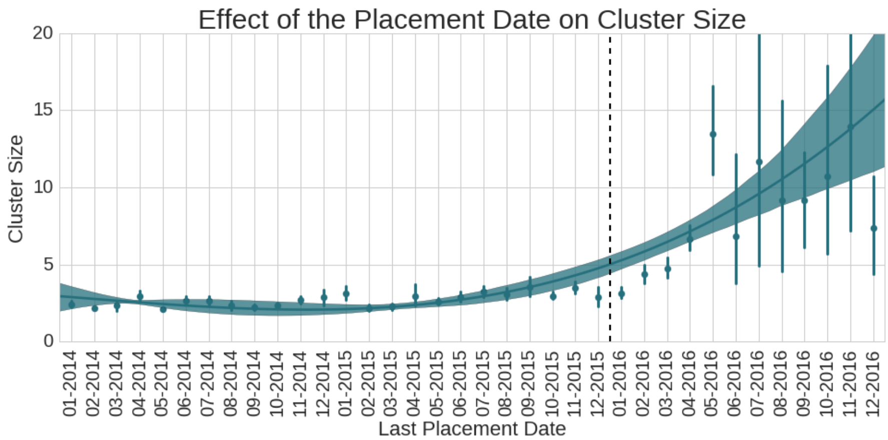


Figure 3.33: *The effect of the last placement date of a cluster on its cluster size. The last placement date of a cluster is defined as the latest placement date of all of the CPs in the cluster. A fit with 95% confidence interval is shown.*

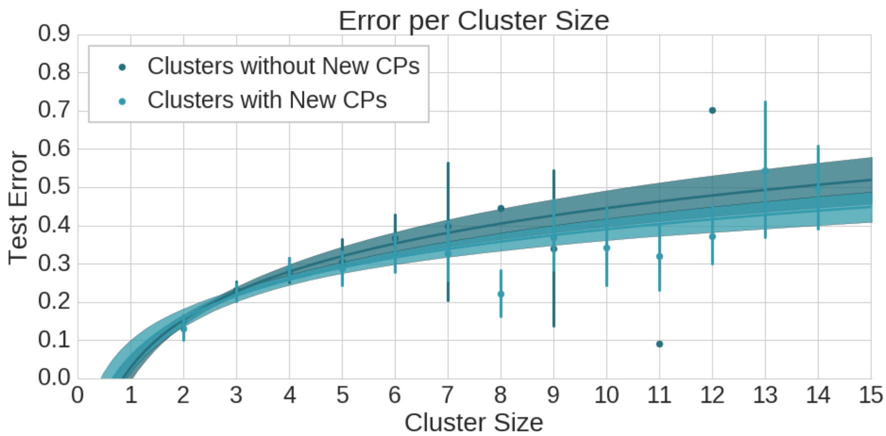


Figure 3.34: *The test error of clusters with and without newly added CPs for each cluster size. Two logistic fits with 95% confidence interval are shown. The 'weighted and trimmed' approach is used.*

3.6 Conclusion

To conclude, each of our three logistic regression models has both training and test errors which are significantly lower than those of our random model (the benchmark case). This is true for all cities. The two best logit models are 'weighted' and 'weighted and trimmed', and between these models there is no significant difference. The following sections will continue with the 'weighted and trimmed' model, as this matches best with the data analysis in Section 3.2.

When cluster sizes increase and there are more CPs to choose from, the training and test errors of the 'weighted and trimmed' approach increase. However, this increase is not linear, but rather logistic.

A limitation of this modelling approach is that all CPs in range have a (small) probability of being selected. This does not match with reality as in the training data only 13% of the clusters use all CPs in range, while with the logistic choice models 89% of the clusters use all CPs. However, while all CPs are used at least once, many CPs in range are used very infrequently (which matches reality) as they have very small probabilities of being selected.

Another limitation of this modelling approach is that there is no learning or adaption in the choice model. This means that agents do not have memory and will continue to choose CPs even if they have previously experienced that these CPs are often occupied. A possible extension for this model would be to include this memory in the choice models.

Chapter 4

Case Studies: Non-Habitual Users and Rollout Strategies

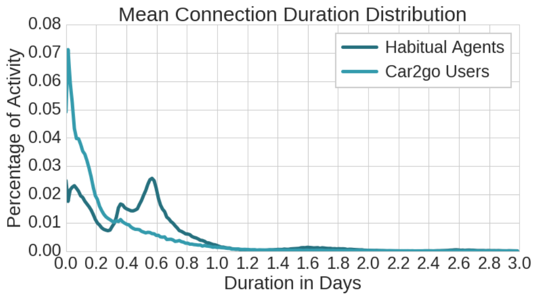
This chapter provides an analysis of the preparedness of cities for an increase of non-habitual EV users by studying the effects of these EV users on the habitual users in the system. Furthermore, this chapter allows us to gain insight into how these effects can be reduced using different rollout strategies. Non-habitual users are EV users with frequent short charging transactions spread uniformly through a specific city, similar to a random walker.

The goal of this chapter is to provide insight into the robustness and preparedness of different cities for these non-habitual users. This goal is achieved by adding non-habitual agents to the SEVA-ICPS model. Section 4.1 describes how we capture, implement and validate the non-habitual agents, while Section 4.2 shows how the non-habitual agents affect the habitual agents. Section 4.3 then analyses why non-habitual agents affect the habitual agents differently in different cities. Section 4.4 analyzes how cities can combat the effect of non-habitual agents through the use of different rollout strategies.

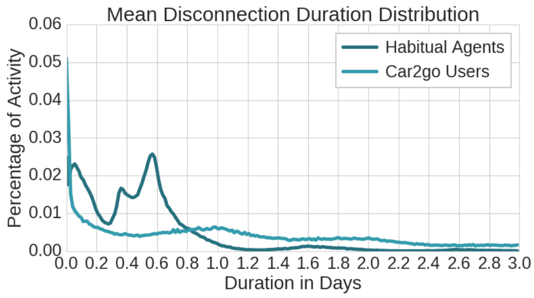
4.1 Non-Habitual Agents

To study the effect of non-habitual agents and their occupation of existing CPs on habitual agents, we first need to capture the behavior of these non-habitual users. Capturing this behavior enables us to add EV users with similar behavior to the simulation model. As we have shown in Chapter 2, the connection duration distribution, disconnection duration distribution and arrival distribution of an EV user is enough to fully capture its behavior with regards to when and for how long they charge. To realistically capture these distributions, it is preferable to base them on real-world data as this increases their predictive value. In terms of when and for how long the non-habitual agents charge, they are similar to free-floating electric car sharing schemes as these cars also have short and frequent charging transaction which are relatively uniformly distributed throughout the day. Thus we will capture these distributions using data from a successful car sharing scheme in Amsterdam, namely the car2go users [21]. As mentioned in Section 2.1.2, car2go users have a major influence on the charging infrastructure. These car2go users are a subset of non-habitual users that we are using to approximate all non-habitual EV users.

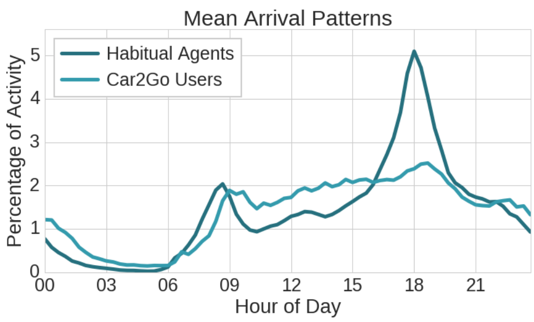
Figure 4.1 shows the behavior of the non-habitual EV users captured within the connection duration distribution, the disconnection duration distribution and the arrival distribution. In the when and what dimension, car2go users appear to have short and frequent charge transactions. Car2go users arrive throughout the day, with peak moments at 9am and 6pm, corresponding with people going to work and coming home. Throughout the day the arrivals increase.



(A)



(B)



(C)

FIGURE 4.1: *The mean* (A) *connection duration distributions,* (B) *disconnection duration distributions and* (C) *arrival patterns for both habitual agents and car2go users.*

Car sharing users have a bias towards CPs in certain areas, such as train stations and city centers. Our non-habitual agents are randomly spread through the city, as we are not trying to simulate car sharing schemes, but rather the effect of random occupations of CPs on the habitual agents. Thus to capture where our non-habitual agent (will) charge, each non-habitual agent would need to have a defined cluster from which to choose CPs. Given that these agents are confined to one city, we can simply say that they have one cluster which contains all CPs of their city. From this cluster they then randomly select a CP.

As with the habitual agents, we can validate how well the activity patterns of our non-habitual agents match with that of the car2go users from which we extracted the behavior. Figure 4.2 compares the mean simulated activity patterns of 100 non-habitual agents with their training data, namely the mean activity pattern of the car2go users in Amsterdam. These non-habitual agents have a validation score of 0.04, which is even better than the habitual agents' mean validation score of 0.1 (seen in Figure 2.7).

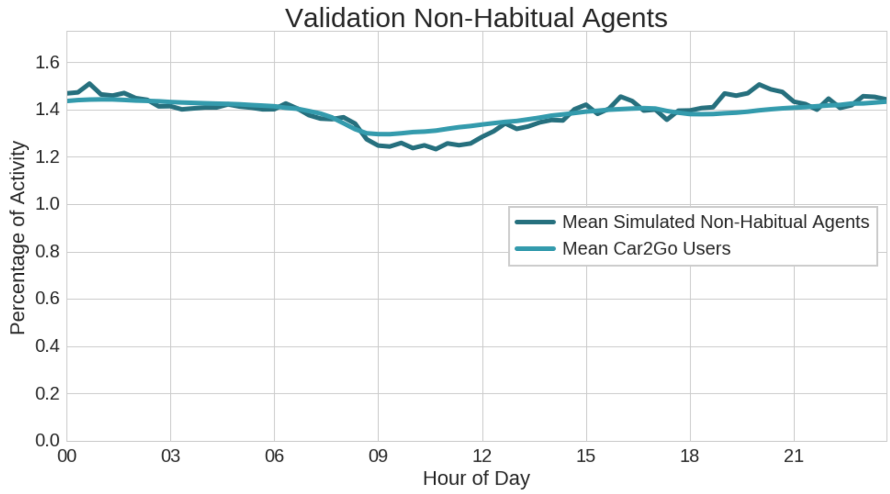


FIGURE 4.2: *The mean simulated activity pattern of 100 non-habitual agents is shown together with the mean activity pattern of all car2go data from Amsterdam. The simulation ran for a year. The agent validation score is* 0.04.

4.2 Non-Habitual Agents Experiments

As shown in Section 2.1.1, different stakeholders have different points of view and goals. A system evaluation metrics needs to be chosen based on one or more stakeholders. As we want to examine the effect of non-habitual agents on the habitual agents, our point of view is that of the (potential) EV users. As seen in Table 2.1, one of the corresponding result indicators for this stakeholder is 'increased accessibility of charging infrastructure'.

Therefore, to quantify the influence of non-habitual agents on the habitual agents, we measure the percentage and number of times a habitual agent chooses an occupied CP. This is equal to the percentage and number of failed connection attempts. Note that in the SEVA-ICPS model, the complete selection process fails if all CPs in range are occupied. When this occurs, the agent retries to connect (a 'retry attempt') after a certain amount of time (t_r, see Table A.3). The exact value of this variable has not been validated and therefore is not known to be correct. To ensure this variable does not skew our results, we exclude all failed connection attempts which occurred during these retry attempts. In our experiments we vary the numbers of non-habitual agents per habitual agent between 0.0 and 2.0. A value of 1.0 means that one non-habitual agent is added for every habitual agent in the system. In absolute terms this means that for a value of 1.0, 1112 non-habitual agents would be added in Amsterdam, 585 in The Hague, 617 in Rotterdam and 237 in Utrecht.

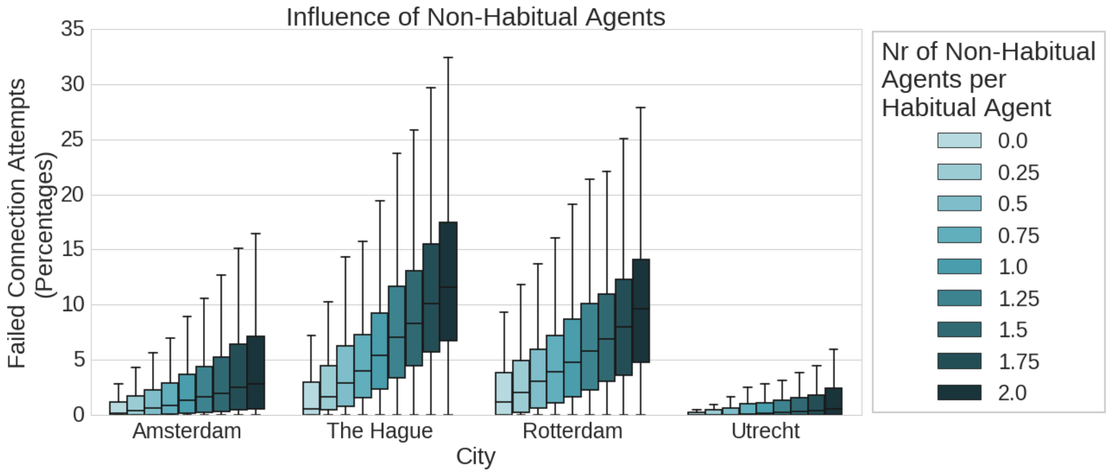


FIGURE 4.3: *The percentage of failed connection attempts for each city and for the different numbers of non-habitual agents.*

Figure 4.3 shows the boxplot containing the percentage of failed connection attempts for each agent within the different cities and for the different numbers of non-habitual agents. Note that The Hague and Rotterdam have the highest percentages of failed connection attempts, with medians of 12% and 9% respectively when two non-habitual agents are added for each habitual agent. Utrecht has the lowest percentage of failed connection attempts, with a median of 2% when two non-habitual agents are added for each habitual agent. A nearly identical trend can be observed when we look at the number of failed connection attempts, as can be seen in Appendix C (Figure C.1).

Next we examine how the percentages of failed connection attempts increase when the number of non-habitual agents increases. Figure 4.4 shows the mean for each city and each percentage, with a linear fit. In the brackets of the legends, the slope of each line is shown. The Hague has the steepest slope and when the number of non-habitual agents increases by 100%, the mean failed connection attempts in The Hague increase by 5.3%. The Hague is followed by Rotterdam and Amsterdam with increases of 3.8% and 1.8% respectively. Lastly, Utrecht has the most gradual increase with an increase of 0.6%. The numbers of failed connection attempts also shows that The Hague has the steepest slope, followed by Rotterdam, Amsterdam and lastly Utrecht as can be seen in Appendix C (Figure C.2).

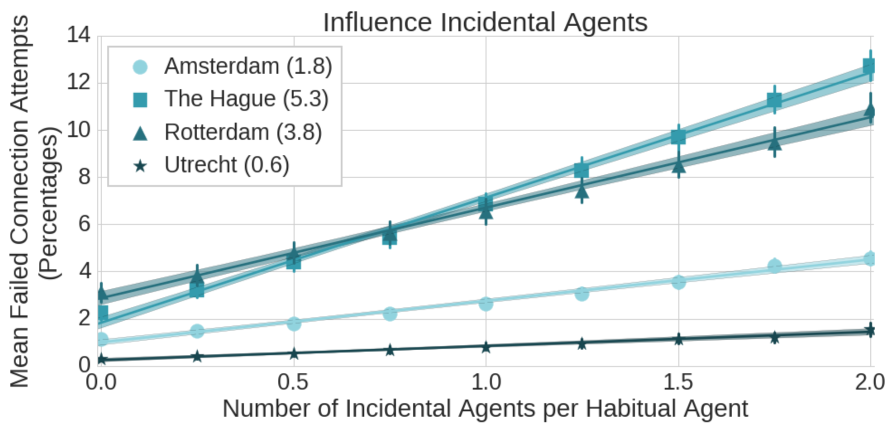


FIGURE 4.4: *The mean percentage of failed connection attempts for each city and for the different numbers of non-habitual agents, with a linear fit. The slope of this fit is shown in the brackets.*

Figure 4.5 shows the distribution of the percentages of failed connection attempts of all agent in each city, when two non-habitual agents are added for each habitual agent. This figure shows that the effects of non-habitual agents on habitual agents are less severe in the cities of Amsterdam and Utrecht, as in these cities a significant percentage of the agents has a very low percentage of failed connection attempts and is therefore not strongly effected by the non-habitual agents. This cannot be said for the cities of The Hague and Rotterdam, as there are much less agents with low percentages of failed connection attempts.

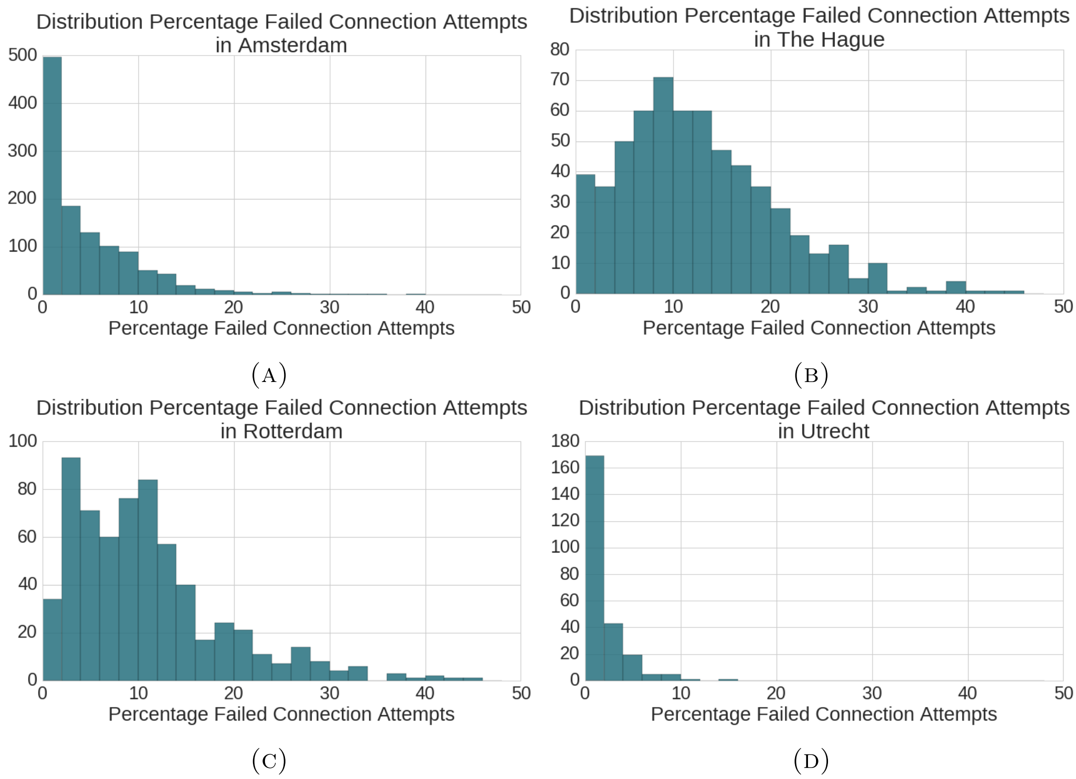


FIGURE 4.5: *The distributions of failed connection attempts for the agents in the different cities when two non-habitual agents are added for each habitual agent.*

The results in this section show that the effect of non-habitual agents on the habitual agents is greater in the cities of The Hague and Rotterdam. These cities seem to be less robust against an increase of non-habitual EV users and less prepared for this type of EV user, compared to the cities of Amsterdam and Utrecht.

4.3 City Differences Concerning City Preparedness

This section will attempt to discover why the cities of The Hague and Rotterdam perform significantly worse when compared to Amsterdam and Utrecht by looking for differences between the cities in terms of charging infrastructure and agent behavior.

4.3.1 Mean Arrival Times of Failed Connection Attempts

To understand why the percentages and numbers of failed connection attempts are higher in these cities, we first look at when these failed connection attempts occur.

Figure 4.6 shows the mean arrival time for failed connection attempts in the different cities. During two moments in the day (8am and 6pm) there is more congestion at the CPs within all four cities. While nearly twice as many EV users arrive at 6pm compared to 8am (see Figure 4.7a), we see that when there are no non-habitual agents, the majority of failed connection occur at 8am, indicating that during the morning rush hour the CPs are most congested and the charging infrastructure is less robust, while in the evening rush hour there are nearly no failed connection attempts.

The infrastructure in all cities is less able to handle the demand during the morning rush hour. When more non-habitual agents are added into the system, we see that in the cities of The Hague and Rotterdam more failed transactions occur at both 8am and 6pm. In Amsterdam and Utrecht only the congestion at 8am increases while the failed connection attempts at 6pm barely increase. It is likely that there is a longer peak for The Hague at 6pm because the mean activity pattern for The Hague shows that there are mostly evening and night chargers in this city (see Figure 4.7b).

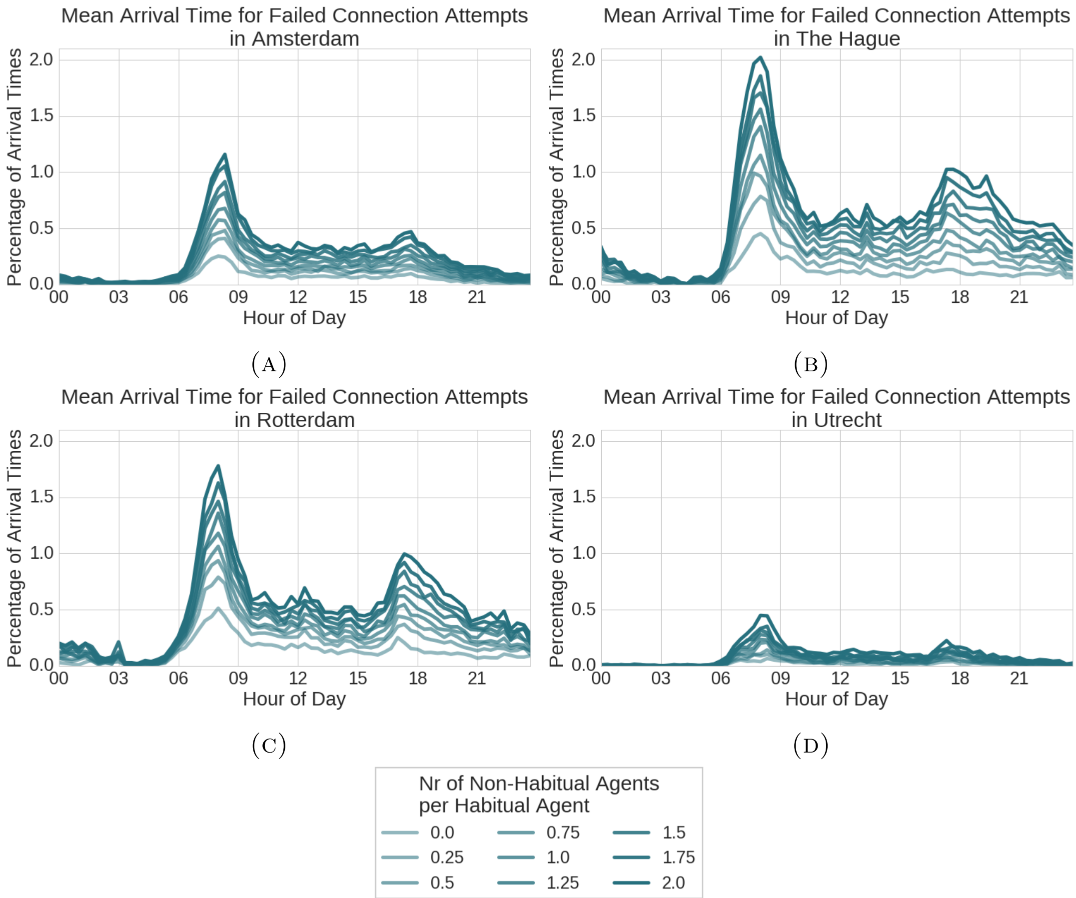


FIGURE 4.6: *The normalized mean arrival times for the the failed connection attempts in the four cities.*

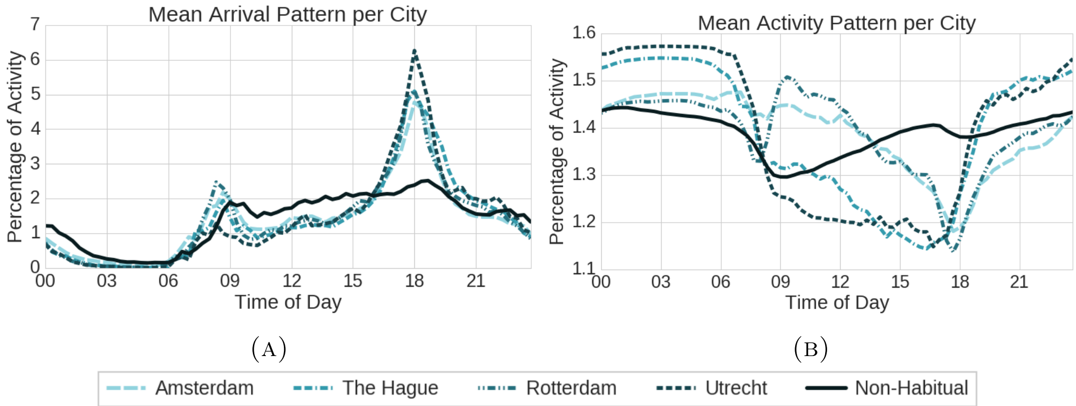


FIGURE 4.7: *Mean (A) arrival times and (B) activity patterns for EV users in the different cities. The light blue line indicates the mean arrival times for non-habitual users.*

4.3.2 Overall Available Charging Infrastructure

The (lack of) robustness of the different cities may be due to the differences in the number of CPs per habitual agents in each city. Figure 4.8 shows this number of CPs per habitual agent, together with the absolute numbers of habitual agents and CPs for each city. This figure shows that Amsterdam has the most CPs (approximately 1060), followed by Rotterdam (702), The Hague (503) and Utrecht (427). The number of CPs per habitual agent for the cities of Amsterdam, the Hague and Rotterdam are similar, indicating that this does not explain the differences in robustness between Amsterdam, The Hague and Rotterdam. As Utrecht has by far the highest number of CPs per habitual agent this may explain why Utrecht performs best.

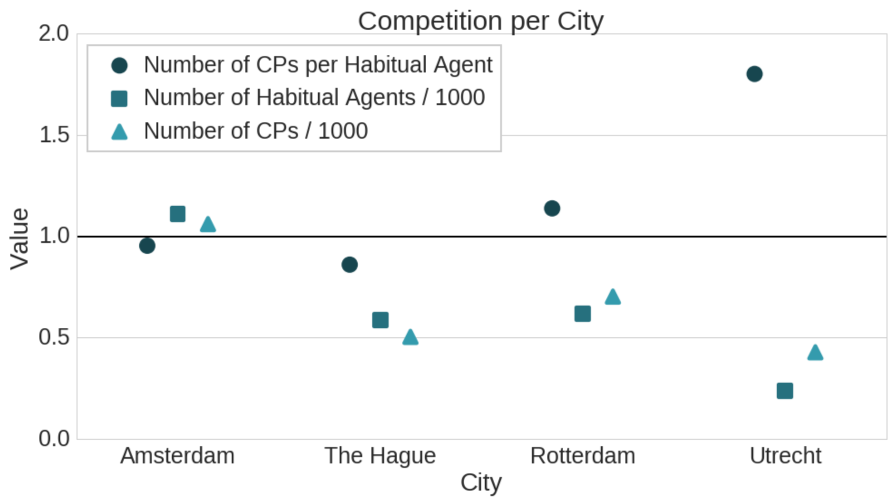


FIGURE 4.8: *The number of agents, the number of CPs and the ratio between agents and CPs in each city.*

However, note that each CP has at least two sockets per CP, so while the number of CPs does not explain the differences in robustness, the number of sockets could. Figure 4.9 shows the distribution of sockets per CP for each city. A clear difference can be seen between Amsterdam and Utrecht on the one hand and The Hague and Rotterdam on the other hand, as Amsterdam and Utrecht have more CPs with more than two sockets.

Figure 4.10 shows the number of sockets per habitual agent, and we can see that this value is lowest for the cities of The Hague and Rotterdam. Thus the number of sockets per habitual agent may explain the difference between the cities, as cities with more sockets per agent are more robust against non-habitual agents.

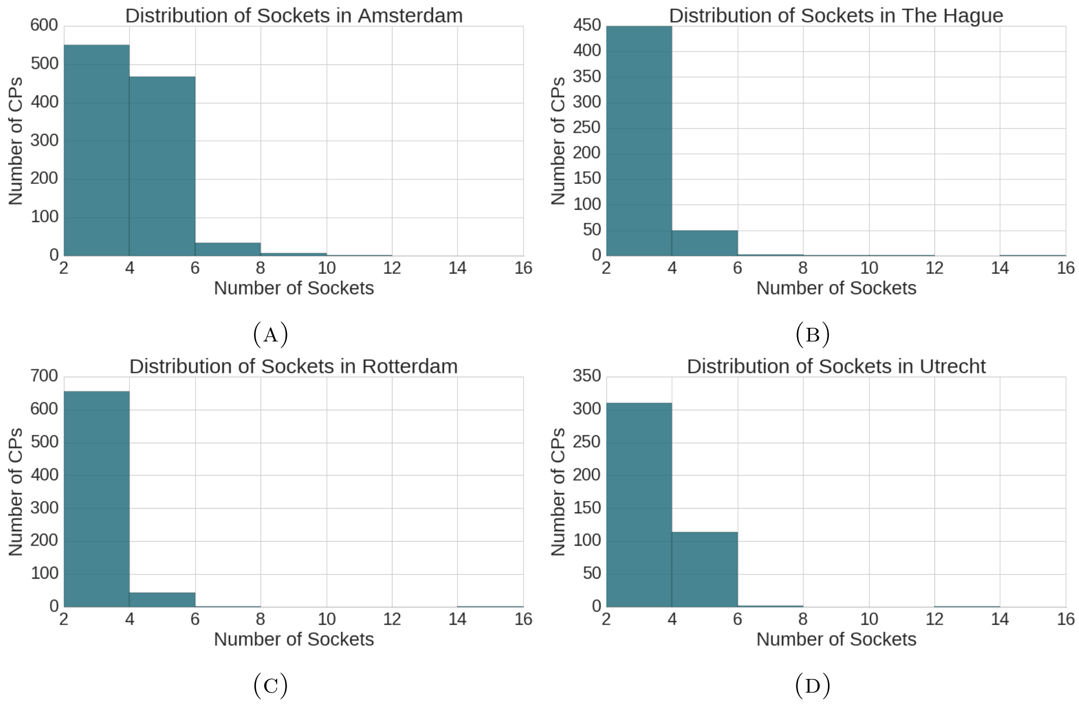


FIGURE 4.9: *The distribution of number of sockets per CP for the different cities.*

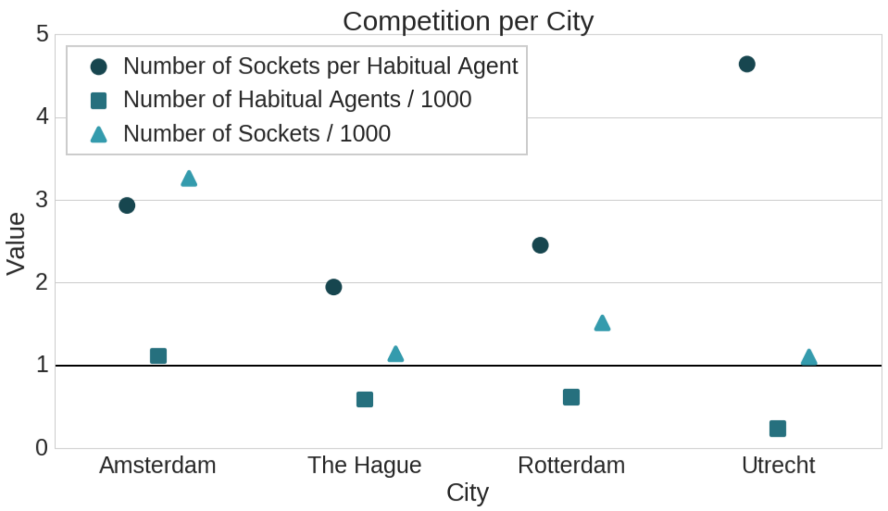


FIGURE 4.10: *The number of agents, the number of sockets and the ratio between agents and sockets in each city.*

4.3.3 Available Charging Infrastructure in Range

As every agent can only use sockets in range of its walking preparedness, we now look at the available charging infrastructure (i.e. CPs) in range, rather than the overall available charging infrastructure per city. Another possible difference between the cities is the distance in meters the EV users are willing to walk to find a CP.

In the SEVA model the maximum distance and walking preparedness are estimated for each EV user through clustering as was discussed in Section 2.2.6. Remember that the maximum distance is the maximum distance between the CPs in a cluster and the center of that cluster, while the walking preparedness for each EV user is calculated by taking the maximum of the maximum distance and the minimum radius as described in Section 2.2.1 and Figure 2.3. The minimum radius is an input parameter (see Table A.3) and it is currently set to 150 meters.

Both the maximum distance and walking preparedness are significantly lower for the cities of The Hague and Rotterdam as can be seen in Figure 4.11a and 4.11b. This means EV users in these cities are less willing to walk far to find CPs. The result of this is that they have less sockets in the range of their walking preparedness, as can be seen in Figure 4.11c.

The density of sockets per squared km (shown in Figure 4.11d) is also lower in the cities of The Hague and Rotterdam, furthermore showing that agents in these cities have less available charging infrastructure in range of their walking preparedness. Perhaps these lower densities result in EV users which are less willing to walk far to find their CPs, as they would have to walk a lot further to find one. Adding more CPs (and sockets) could therefore result in an increase in walking preparedness.

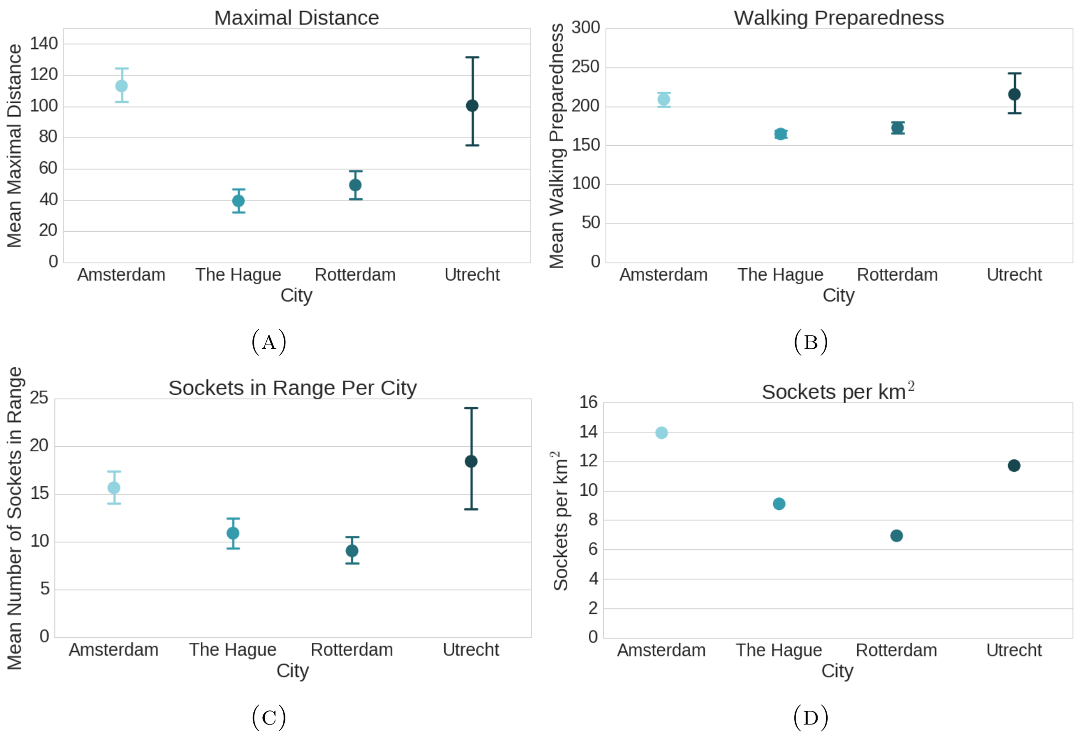


FIGURE 4.11: *The mean* (A) *maximum distance and* (B) *walking preparedness of the agents in each city with a 95% confidence interval.* (C) *The number of sockets in range of the walking preparedness for each city with a 95% confidence interval.* (D) *The density per city.*

4.4 Effects of Rollout Strategies

There are multiple methods of optimizing the charging infrastructure and decreasing the percentage of failed connection attempts, such as effective rollout strategies and incentives. In the previous section we hypothesized that by adding CPs in specific locations, we could reduce the percentages and numbers of failed connection attempts caused by the increase in non-habitual agents. In this section we will test this hypothesis by implementing and analyzing the effects of four different rollout strategies. Section 4.4.1 describes the setup of the experiment and Section 4.4.2 shows and analyzes the effects of the four different rollout strategies.

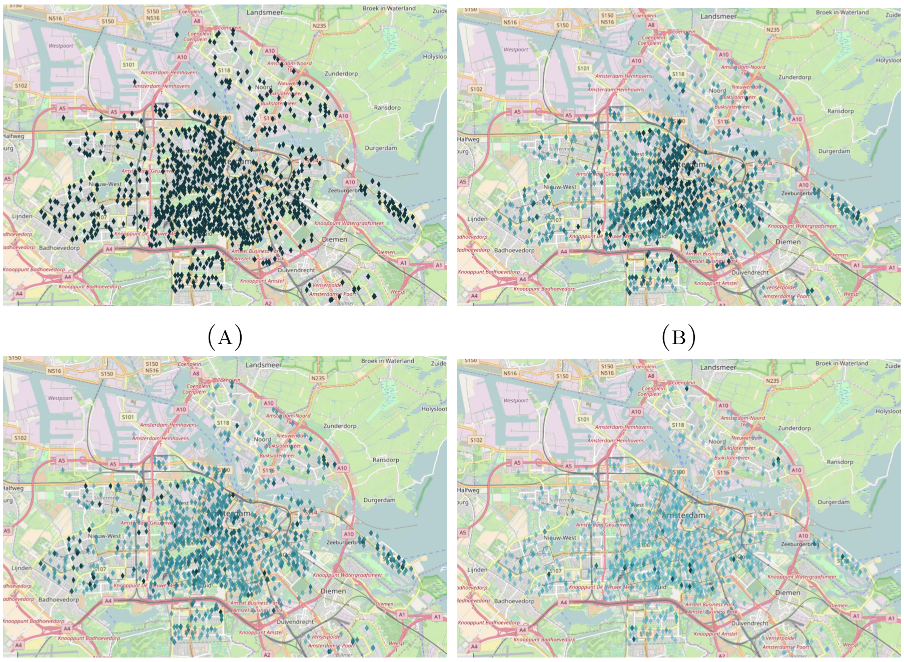
4.4.1 Setup of the Experiment

For this experiment we add one non-habitual agent for every habitual agent in the system (i.e., 1.0 in Figure 4.3), in order to ensure that there is enough competition in the system to see differences in percentages of failed connection attempts between the four rollout strategies[1].For each rollout strategy CPs are selected in a particular way. At the location (longitude, latitude) of these selected CPs new CPs with two sockets are added.

Four ways of selecting these CPs have been defined, namely (1) randomly selecting CPs; (2) selecting CPs with the highest number of unique users per week; (3) selecting CPs with the highest number of kWh charged per week; (4) selecting CPs with the highest number of failed connection attempts.

Figures 4.12, 4.13, 4.14 and 4.15 show the CPs of the cities of Amsterdam, The Hague, Rotterdam and Utrecht respectively. Each dot represents a CP, colored by the different ways of selecting a CP where to place a new CP with two sockets. Thus the darker the dots, the more likely this location will be selected to place a new CP.

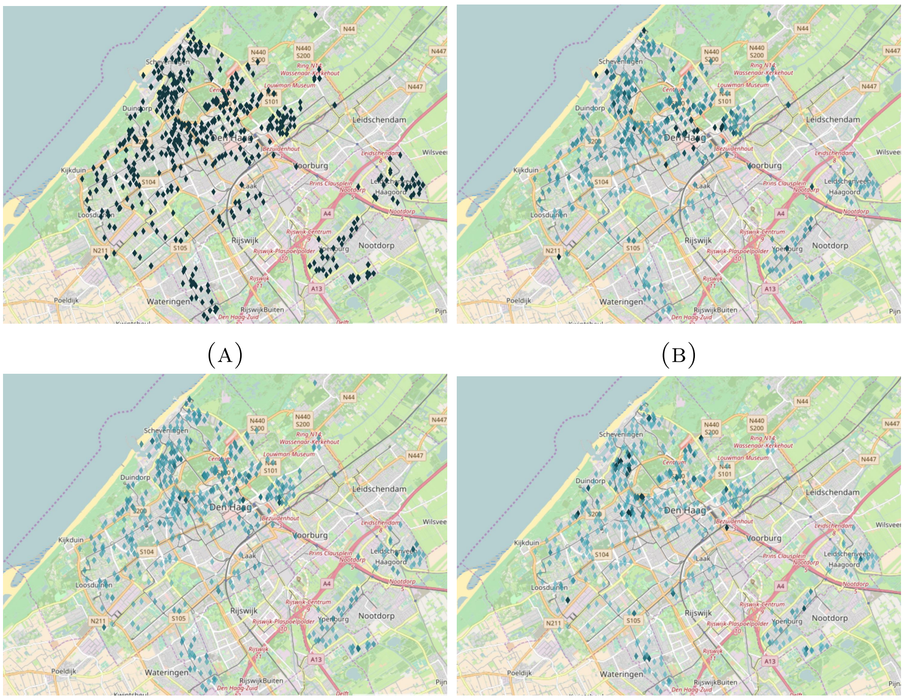
[1]Note that in the SEVA model we filter out EV users with infrequent behavior. This means that our simulation does not contain the total population of EV users. By adding the non-habitual users, we create a more realistic outcome and make it possible to study different rollout strategies.



(A) (B)

(C) (D)

FIGURE 4.12: *The CPs of Amsterdam, colored* (A) *uniformly,* (B) *by the number of unique users per week,* (C) *by the number of kWh charged per week and* (D) *by the failed connection attempts.*



(A) (B)

(C) (D)

FIGURE 4.13: *The CPs of The Hague, colored* (A) *uniformly,* (B) *by the number of unique users per week,* (C) *by the number of kWh charged per week and* (D) *by the failed connection attempts.*

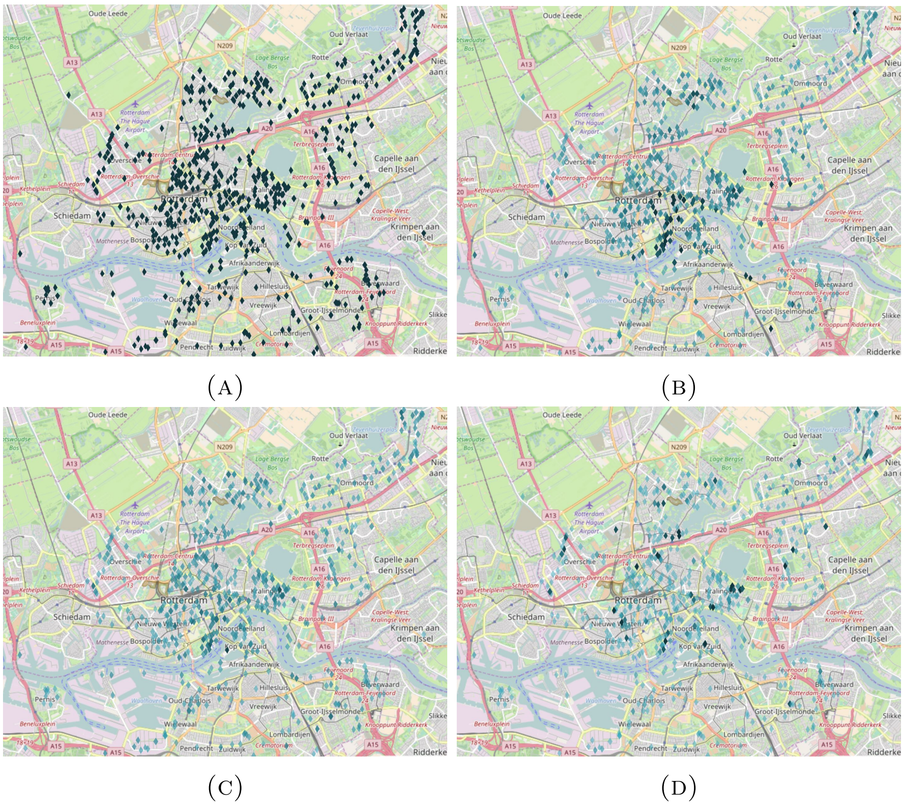


FIGURE 4.14: *The CPs of Rotterdam, colored* (A) *uniformly,* (B) *by the number of unique users per week,* (C) *by the number of kWh charged per week and* (D) *by the failed connection attempts.*

FIGURE 4.15: *The CPs of Utrecht, colored* (A) *uniformly,* (B) *by the number of unique users per week,* (C) *by the number of kWh charged per week and* (D) *by the failed connection attempts.*

4.4.2 Rollout Strategy Experiment

In this section we examine the effects of the different rollout strategies. Figure 4.16 shows how the percentage of failed connection attempts decrease as the number of CPs added increases, for each rollout strategy and each city. We see that for each city all four strategies significantly help decrease the percentage of failed connection attempts. Especially the robustness of the cities which started with higher percentages of failed connection attempts (The Hague and Rotterdam) greatly increases.

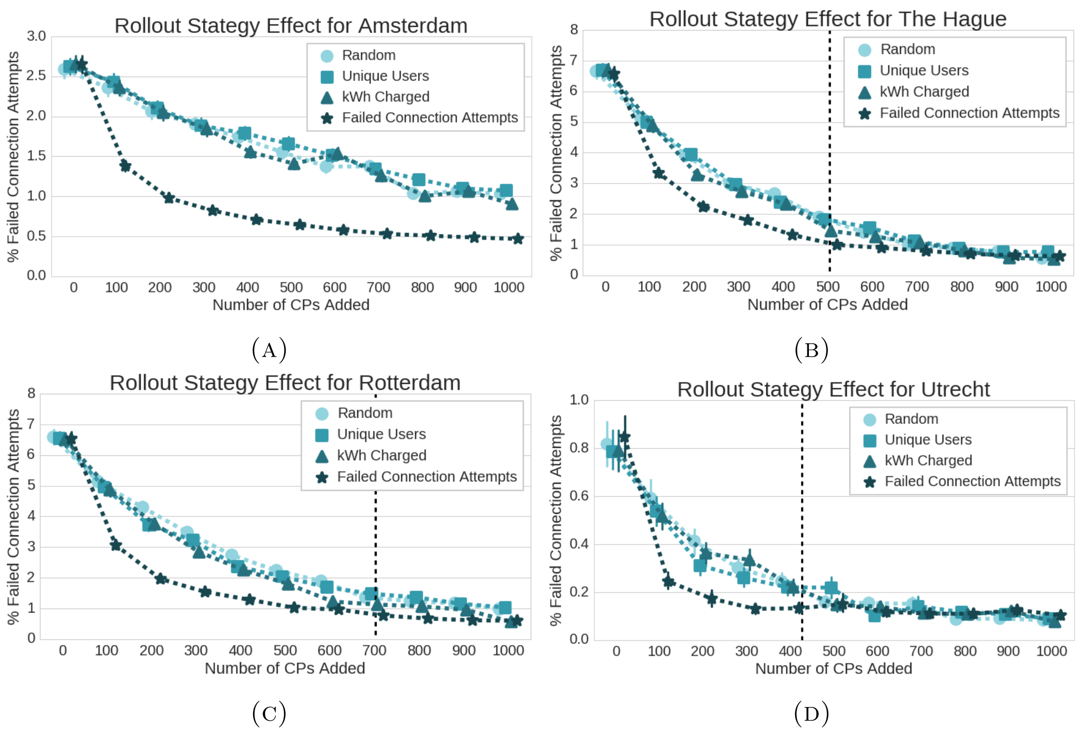


FIGURE 4.16: *The effects of the rollout strategies on the percentage of failed connection attempts for different numbers of CPs added and for different cities. The dotted line indicates how many CPs each city currently has.*

These figures show that while intuitively the strategies of selecting CPs with many unique users per week and CPs with large number of kWh charged per week make sense, our simulation shows that these strategies both do not perform better than randomly selecting CPs. Specifically selecting the CPs with the most failed connection attempts works best to decrease the percentage of failed connection attempts, and this is true for each city.

When selecting the CPs with the most failed connection attempts, less CPs are needed to reduce the percentage of failed connection attempts. In Rotterdam for example, a value of 3% of failed connection attempts can be reached by adding 300 CPs when using the 'kWh charged' rollout strategy while only 100 CPs are needed when the 'failed connection attempts' strategy is used. This means 200 less CPs are needed to achieve the same decrease in the percentage of failed connection attempts. Under the assumption that the price to maintain a CP is 739 euros per year (excluding purchase price) [52], this means nearly 150.000 euros could be saved by using more efficient rollout strategies.

4.5 Conclusion

This chapter has analyzed the effects of non-habitual agents on the percentages and numbers of failed connection attempts among habitual agents in the cities of Amsterdam, The Hague, Rotterdam and Utrecht.

The behavior of non-habitual agents is captured by basing the connection duration distributions, the disconnection duration distributions and the arrival distributions on real-world data, namely that of a free floating car sharing scheme. These non-habitual agents are spread randomly throughout the city, rather than centered around hot-spots (train stations, shopping malls, etc.) as true car sharing cars would be. We have shown that these non-habitual agents validate well, when comparing the mean activity patterns.

Under the assumption that non-habitual agents select their CP uniformly, we see clear differences between the robustness and preparedness of cities for these non-habitual agents. Amsterdam and Utrecht are more robust and decidedly better prepared for non-habitual agents when compared to The Hague and Rotterdam. In The Hague and Rotterdam the mean and median of the percentages of failed connection attempts are significantly higher. When the number of non-habitual agents increases, the increase of these percentages is linear for all cities. However, the slope of these linear lines are significantly higher for the cities of The Hague and Rotterdam.

Multiple differences between the cities can explain the differences in robustness and preparedness. Firstly, we saw that when more non-habitual agents are added into the system, more failed transactions occur at both 8am and 6pm in The Hague and Rotterdam, while in Amsterdam and Utrecht most failed transactions occur at 8am. From this we can conclude that the cities of The Hague and Rotterdam are less capable of handling the increase in demand during the evening rush hour. Secondly, the number of sockets available per habitual agent differ per city, with The Hague having the smallest number of sockets per habitual agent, followed by Rotterdam, Amsterdam and then Utrecht. Less sockets per habitual agent could be a direct cause of a lack of robustness per city. However every agent cannot use all socket in its city, but rather only the sockets in the range of its walking preparedness. We found that the maximum distance and walking preparedness of agents in The Hague and Rotterdam are significantly lower compared to the other cities. This together with the fact that the density of sockets in these cities is also lower, results in less charging option within range of the walking preparedness of agents in these cities. This could furthermore explain the differences between cities.

Policy implications of this chapter are that all cities can increase the robustness of charging infrastructure by increasing the number of CPs at locations which are used during the morning rush hour. The municipalities of The Hague and Rotterdam can furthermore improve the robustness of their charging infrastructure by increasing the numbers of CPs at location which are used during the evening rush hour (residential areas). Increasing the number of available charging options in range of the walking preparedness could also improve the robustness of cities. This can either be done by using incentives to get EV users to increase their walking preparedness, or by increasing the density of CPs (and thus sockets) within cities.

Our rollout strategy analysis shows how different placement strategies help decrease the percentage of failed connection attempts. We can conclude that placing new CPs at locations where EV users have (many) failed connection attempts works best to decrease the percentage of failed connection attempts and therefore make the charging infrastructure more robust. A policy implication is that if municipalities are able to keep track of where EV users experience failed connection attempts (for example by asking people to inform them, using an app or website, when one of their chosen CP was occupied), this could give them the information they need to determine where to place new CPs.

Chapter 5

Conclusion and Future Work

This thesis has focused on understanding one method of charging infrastructure optimization, namely rollout strategies. A simulation model was needed to predict future usage of CPs and to predict the effectiveness of rollout strategies which have never been implemented in practice.

5.1 Conclusion

In this thesis we have presented the Simulation of Electric Vehicle Activity (SEVA) model, a predictive, data-driven model which is validated using a large real-world dataset. From this dataset we were able to extract the behavior of individual EV users in the terms of when, where and how long these users charge. By simulating the charging transactions of each individual EV user we were able to accurately predict the future behavior of these EV users. A limitation of the first phase of this model is that it cannot accurately simulate how EV users choose a CP within a specific area. This means that it is neither able to simulate the decision making process of the EV users, nor can it predict if and how new CPs are used. Therefore the SEVA model is insufficiently suitable to gain insight in the effects of rollout strategies and incentives.

For this reason we created an Improved Charge Pole Selection (ICPS) process within the SEVA model, which results in the SEVA-ICPS model. A literature review revealed that the factors which influence how an EV user selects a CP are distance, cost and charging speed. The factor cost has been split into charging fees and parking fees, as these fees may influence EV users in different ways. We then performed a data analysis to further understand these factors individually. These influencing factors are then used as explanatory variables within three logistic regression choice models, which were fitted using the large real-world dataset. All three choice models were implemented within the selection process of the SEVA model to create the SEVA-ICPS model. Lastly, the SEVA-ICPS model was validated, which shows that all three choice models perform significantly better than randomly selecting a CP in a specific area. This improved selection process shows no significant difference in performance between clusters with and clusters without new CPs. Thus our improved selection process not only accurately simulates the decision process of the EV users, it can also predict if and how new CPs are used.

The capabilities of the SEVA-ICPS model have been shown with through two case studies. The first case study aimed to analyze the robustness of the current charging infrastructures in the major cities of the Netherlands. This goal is achieved by studying the effects of non-habitual EV users, such as car sharing schemes, on habitual EV users. In this case study the SEVA-ICPS model was extended to include these non-habitual EV users and their behavior. There we assume that these non-habitual agents are uniform in where they charge, as we are not trying to simulate car sharing schemes, but rather the effect of random occupations of CPs on the habitual agents. This case study shows that while the charging infrastructures of Amsterdam and Utrecht are relatively well prepared for an increase in non-habitual users, the cities of The Hague and Rotterdam are not. In the case of The Hague and Rotterdam, an increase of one non-habitual agent for each habitual agent in the system causes the percentage of failed connection attempts of habitual agent to increase by 5.3% and 3.8%, respectively while for Amsterdam and Utrecht these increases were 1.8% and 0.6% respectively. There are multiple explanations for this difference between cities. Firstly, the cities of The Hague and Rotterdam are less capable of handling the increase in demand during the evening rush hour, indicating that the charging infrastructures in these cities is not as good in residential areas. Secondly, the cities of Amsterdam and Utrecht have significantly more sockets available per habitual agent. Thirdly, the walking preparedness of agents and the CP density in Amsterdam and Utrecht are significantly higher meaning these EV users have more CPs in range and are therefore more likely to find an unoccupied CP.

A possible way of improving the robustness and preparedness of the infrastructures in The Hague and Rotterdam would be to increase their number of CPs. This hypothesis is tested in the second case study, which presents a policy evaluation of four rollout strategies in the major cities of the Netherlands. It shows how the impact of an increase of non-habitual EVs can be diminished by different rollout strategies. These four rollout strategies are defined as (1) randomly selecting CPs to duplicate; (2) selecting CPs with the highest number of unique users per week to duplicate; (3) selecting CPs with the highest number of kWh charged per week to duplicate; (4) selecting CPs with the highest number of failed connection attempts to duplicate. All four tested rollout strategies are effective in decreasing the percentages of failed connection attempts in all four cities. We found that by specifically adding CPs at locations where many failed connection attempts occurred, the overall percentage of failed connection attempts decreases significantly faster in all cities compared to the other strategies. Currently the measure of failed connection attempts is not present in the dataset. A policy implication of our results is that municipalities should try to measure these failures and take this into account when designing their rollout strategies.

The rollout strategy which selects CPs with the highest number of unique users per week to duplicate and the rollout strategy which selects CPs with the highest number of kWh charged per week to duplicate perform no better than the rollout strategy which randomly duplicates CPs. This shows the necessity of a simulation over data analysis for this type of what-if questions. Intuitively the 'unique users' and 'kWh charged' strategies are logical but through our simulation we have discovered that they perform no better than the strategy of randomly placing CPs.

5.2 Possibilities for Future Work

The work presented in this thesis can be improved and extended by focusing on (1) possibilities within the SEVA model, (2) possibilities within the SEVA-ICPS model, (3) possibilities regarding the non-habitual agents and the analysis thereof and (4) possibilities regarding the rollout strategies and the analysis thereof. Each of these options will now be discussed in more detail.

The SEVA model was built to be modular and to allow for extensions. The possibilities for extending this model are near endless and the possibilities presented are by no means an exhaustive list. The SEVA model could be extended to include the state of charge of each EV and the kWh charged for each transaction. This could be realized by modelling the movement of each EV when it is not charging. This extension could be used to study the effectiveness of CPs, as it would then be possible to compare the charge times with the connection times and show which fraction of time EV users are connected to a CP without charging. Another possibility would be to compare the cluster locations of the agents to the actual home locations of the EV users to validate that these locations are correctly calculated. This could offer even more insight into how EV users select their CPs. Other extensions include the addition of more habitual agents to the simulation by duplicating existing agents, the addition of smart charging options, such as vehicle-to-grid [23], the addition of direct communication between agents to allow for social charging and the creation of a graphical user interface enabling policy makers to test case studies and incentives.

The discrete choice model in SEVA-ICPS could be extended by incorporating memory and adaption into the CP selection process. This would ensure that agents do not continuously try to charge at a CP that is always occupied. Another possibility for future work would be to try different choice models and compare their accuracy. We used a binary logistic regression model, but mixed logit models [39, 40, 43–45], random parameter logit models [39], nested logit models [46] and latent class logit models [43, 44] are also popular techniques used for modeling charging choices.

The analysis of non-habitual agents could be extended by increasing the complexity of the non-habitual agents. They could be given a bias towards CPs in specific areas such as train stations and city centers. This bias could furthermore be made dependent on the time of day, as for instance car sharing cars are more likely to start charging at a train station during the morning rush hour. The result of adding this bias could be that regions which seem robust in our current analysis will break down (more realistically) as non-habitual agents will use them more frequently during particular times of the day.

The analysis of infrastructure optimization through rollout strategies could be expanded in two ways. On the one hand, more (complex) rollout strategies could be implemented and analyzed, for instance the staggering of the release of new CPs through time. On the other hand, different infrastructure optimization methods could be implemented, such as policies and incentives to steer EV behavior. This in turn could increase the efficiency of the existing charging infrastructure without the need to place new CPs.

Appendix A

SEVA Model Parameters

Parameter	Value	Description
Agent selection method	{given, random}	Method that is used to select agents for the simulation.
Number of agents	2000	The number of agents in the simulation if the agent selection method is 'random'.
Simulation repeats	5	Number of times to repeat the simulation.
Start time	01-01-2016	The date at which the model should start simulating.
Stop condition	{time, number of activities executed per agent}	Method used to determine how long the simulation show run. If time is selected, the simulation will run until a certain maximum time. If number of activities executed per agent is specified, the simulation will stop running when all agents have executed at least a certain number of activities.
Warmup period	7 days	The period of time to take as warmup of the system. This period of time at the start of the simulation will not be taken into account for validation.

Table A.1: *The parameters present in the model concerning the simulation.*

Parameter	Value	Description
Bin size (T)	20 minutes	The length of a time interval in the simulation.
Start date training	01-01-2014	The date (inclusive) at which to start the training set.
End date training	01-01-2016	The date (exclusive) at which to end the training set.
Start date test	01-01-2016	The date (inclusive) at which to start the test set.
End date test	01-01-2017	The date (exclusive) at which to end the test set.

Table A.2: *The parameters present in the model concerning data preprocessing.*

Parameter	Value	Description
Habit probability (p_h)	0.4	Probability with which a CP in a center is selected based on habit. With a probability of $1 - p_h$ a CP is selected based on distance.
Retry time clusters (t_r)	20 minutes	Time to wait whenever no free CP at a center can be found.
Minimum radius	150 meters	Minimum distance within which an agent considers new CPs based on distance.

Table A.3: *The parameters present in the model concerning the agent's behavior.*

Parameter	Value	Description
Minimum transactions center (s_c)	20	The minimum number of transactions that clustered CP(s) need to have to be considered a center.
Minimum transactions CP (s_{cs})	10	The minimum number of transactions that a CP should have in order to be considered for clustering.
Minimum fraction transactions center (f_c)	0.08	The minimum fraction of all the transactions of an agent a center should have.
Longitude shift (c)	47.4	The value with which the longitude is shifted for clustering.
Coordinate factor (f)	8.0	The value with which the coordinates are scaled for clustering.
Birch threshold	1.5	The threshold parameter used in the Birch clustering.
Birch branching factor	50	The branching factor used in the Birch clustering.

Table A.4: *The parameters present in the model concerning the clustering agents' CPs.*

Appendix B

Choice Model Parameters

Coefficient	City	Unweighted	Approach Weighted	Weighted and Trimmed
Intercept (β_0)	Overall	0.85 ± 0.10	1.41 ± 0.07	1.45 ± 0.07
	Amsterdam	0.77 ± 0.12	1.18 ± 0.06	1.13 ± 0.04
	The Hague	1.03 ± 0.28	1.66 ± 0.07	1.64 ± 0.08
	Rotterdam	1.13 ± 0.21	1.31 ± 0.12	1.53 ± 0.09
	Utrecht	0.44 ± 0.57	1.48 ± 0.15	1.48 ± 0.16
Distance (β_1)	Overall	-3.17 ± 0.14	-3.69 ± 0.08	-3.68 ± 0.10
	Amsterdam	-3.13 ± 0.29	-3.50 ± 0.06	-3.52 ± 0.07
	The Hague	-3.23 ± 0.40	-4.03 ± 0.15	-3.98 ± 0.13
	Rotterdam	-3.39 ± 0.36	-3.85 ± 0.10	-4.04 ± 0.12
	Utrecht	-2.84 ± 0.57	-3.37 ± 0.14	-3.19 ± 0.19
Charging Speed (β_2)	Overall	0.01 ± 0.09	-1.13 ± 0.11	-0.99 ± 0.08
	Amsterdam	-0.01 ± 0.09	-1.01 ± 0.06	-0.89 ± 0.06
	The Hague	-0.01 ± 0.21	-1.05 ± 0.10	-1.11 ± 0.14
	Rotterdam	0.01 ± 0.16	-1.63 ± 0.23	-1.08 ± 0.10
	Utrecht	0.16 ± 0.37	-0.81 ± 0.26	-0.87 ± 0.22
Charging Fee (β_3)	Overall	0.03 ± 0.03	-2.44 ± 0.30	-1.62 ± 0.29
	Amsterdam	0.02 ± 0.05	-0.88 ± 0.08	-0.50 ± 0.06
	The Hague	0.02 ± 0.03	-3.05 ± 0.13	-0.88 ± 0.29
	Rotterdam	-0.01 ± 0.08	-1.76 ± 0.42	-1.55 ± 0.36
	Utrecht	0.17 ± 0.15	-4.09 ± 0.25	-3.55 ± 0.20
Parking Fee * Permit Area ($\beta_{4,5}$)	Overall	-0.01 ± 0.01	-0.54 ± 0.22	-0.62 ± 0.21
	Amsterdam	-0.01 ± 0.01	-0.39 ± 0.11	-0.77 ± 0.08
	The Hague	-0.01 ± 0.02	-1.79 ± 0.60	-1.72 ± 0.56
	Rotterdam	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
	Utrecht	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

TABLE B.1: *The mean coefficient values with a 95% confidence interval for each approach, each city and each coefficient.*

Appendix C

Number of Failed Connection Attempts

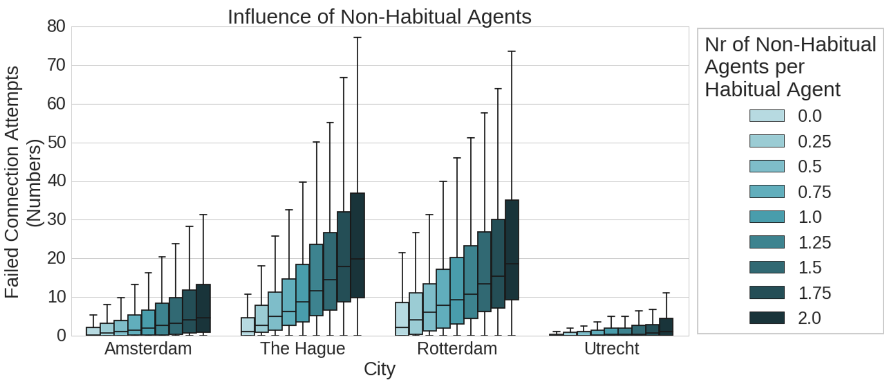


FIGURE C.1: *The number of failed connection attempts for each city and for the different numbers of non-habitual agents.*

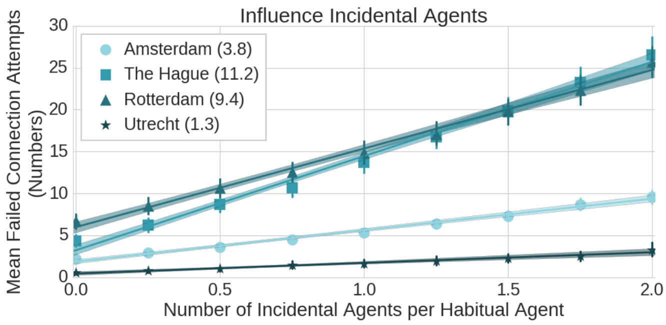


FIGURE C.2: *The mean percentage of failed connection attempts for each city and for the different numbers of non-habitual agents, with a linear fit. The slope of this fit is shown in the brackets.*

List of Figures

2.1 Six illustrations of typical activity patterns found in the CHIEF dataset [4]. The horizontal axis present the hours of the day and the vertical axis show the intensity of activity. 13
2.2 An example of a cluster of an EV user. The center (indicated by the cross) shows the weighted average of the CP locations. The cluster contains the three CPs p_1, p_2 and p_3. Each CP has the number of transactions at that CP indicated between brackets. 18
2.3 A cluster of an agent, where the cross indicates the center, the small blue circles indicate previously visited CPs and the small gray circles indicate previously unvisited CPs within range of the walking preparedness. 19
2.4 The CPs in the dataset plotted on a map for the four cities contained in the dataset, namely (A) Amsterdam, (B) The Hague, (C) Rotterdam and (D) Utrecht. 20
2.5 The growth of the number of EV users and CPs in the dataset over the years. 21
2.6 The activity loop of an agent. Depending on the state of the agent (the grey boxes) the different processes are called. Note that the colored lines indicate the two possible outcomes of the *CP selection* process, namely success or failure. 22
2.7 The results of the experiment where the number of agents is varied. Per value of the parameter we plot the mean validation values with the 95% confidence interval. 34
2.8 The results of the experiment where the number of simulation repeats is varied. Per value of the parameter we plot the mean validation values with the 95% confidence interval. 35
2.9 The number of CPs in the simulation and the number of agents per CP in the simulation for various numbers of agents in the simulation. Per value of the parameter we plot the mean numbers with the 95% confidence interval. 35
2.10 The results of the experiment that varies the bin size. Note that the horizontal axes do not have a linear scale. Per value of the parameter we plot (A) the mean run time and (B) mean validation value with the 95% confidence interval. 36
2.11 The results of the experiment that varies the warmup period and measures the validation values for the agents and CPs. Per value of the parameter we plot the mean validation values with the 95% confidence interval. . . . 37
2.12 The results of using weighted clusters on the validation metrics. Per value of the parameter we plot the mean validation values with the 95% confidence interval. 38

2.13 The difference between the distance metrics walking and as the crow flies. Per value of the parameter we plot (A) the mean run time and (B) the mean validation values with the 95% confidence interval. 39
2.14 The spread of the maximum distances between a center and the CP in the cluster furthest from the center. 40
2.15 The influence of the minimum radius parameter on the validation metrics. Per value of the parameter we plot the mean validation values with the 95% confidence interval. 40
2.16 The frequency of the walking preparedness over the agents in the simulation. 41
2.17 The influence of the Birch clustering parameter on various metrics. Per value of the parameter we plot the mean values with the 95% confidence interval. 42
2.18 The influence of the coordinate factor parameter. Per value of the parameter we plot the mean values with the 95% confidence interval. 42
2.19 The influence of the minimum number of transactions per CP. Per value of the parameter we plot the mean values with the 95% confidence interval. 43
2.20 The influence of the minimum number of transactions per cluster. Per value of the parameter we plot the mean values with the 95% confidence interval. 44
2.21 The influence of the threshold for the minimum fraction of transactions in a cluster. Per value of the parameter we plot the mean values with the 95% confidence interval. 44
2.22 The influence of the habit probability. Per value of the parameter we plot the mean values with the 95% confidence interval. 45
2.23 The influence of the retry time parameter. Per value of the parameter we plot the mean values with the 95% confidence interval. 46
2.24 The frequency of consecutive failed attempts to connect to a CP over a time period of 5 simulation years with 2000 agents. 46
2.25 The percentage of failed (first) attempts in the selection process. Per value of the parameter we plot the mean percentages with the 95% confidence interval. 47
2.26 The relation between the number of charging transactions and the EV validation. Each dot represents an agent in the simulation. Indicated are also (A) the number of clusters and (B) the average number of CPs per cluster each agent. 48
2.27 Examples of activity patterns for the validation values (A) 0.102, (B) 0.239, (C) 0.296 and (D) 0.333. The real activity pattern is indicated in gray while the simulated pattern is blue. 49

3.1 The histogram of the number of CPs in the range of the clusters of each agent. The histogram is shown until the 95$^{\text{th}}$ percentile. 54
3.2 A cluster of one agent with a cross denoting its center and circles denoting its CPs. The distance between a CP and its center is defined as d_{cp}. . . . 55
3.3 The relation between (A) the absolute distance or (B) the relative distance and frequency of use. There is a negative Pearson correlation of -0.22 ($p = 0.00$) and -0.27 ($p = 0.00$), respectively. Only clusters in which there is a choice regarding distances are taking into account. 55
3.4 The relative distances within clusters of unused CPs. 56

3.5 A histogram of the number of CPs in range when there is a choice regarding charging speed within the cluster. 57
3.6 The histogram of the relative charging speeds. The relative speeds are calculated by dividing the absolute speeds by the minimum speed in the corresponding cluster. 57
3.7 The correlation between the relative charging speeds and the frequency of use. 58
3.8 The maximal difference in charging speeds within the clusters. 58
3.9 The histogram of the number of CPs in the range of the walking preparedness of the clusters where there are multiple options regarding charging fees. 59
3.10 The histograms of the (A) absolute and (B) relative charging fees are shown together with the (C) spread of the maximal difference in charging fees within the clusters. 60
3.11 The permit areas of (A) Amsterdam and (B) The Hague. 61
3.12 The prevailing activity patterns. The activity patterns of all clusters were clustered using the Birch algorithm [35] with a Birch threshold of 0.01 and a branching factor of 100. (A) Nighttime charging - 1983 clusters. (B) Daytime charging - 549 clusters. (C) Evening charging - 160 clusters. . . 62
3.13 Different types of interactions between the cluster and the permit area. (A) Cluster is completely inside the permit area of the agent. (B) Cluster of an agent without a permit area. (C) Cluster is partially inside and partially outside the permit area of the agent. (D) Cluster completely outside the permit area of the agent. 63
3.14 Parking fees per transaction in Amsterdam for (A) 64
3.15 Parking fees per hour per transaction in Amsterdam for (A) 65
3.16 Histograms of the percentage of transactions (A) physically and (B) economically outside the permit area. 66
3.17 An example of the choice EV users must make when selecting the CP to connect to. The circles indicate the CPs, while the cross indicates the center of the cluster to which the agent will connect. 68
3.18 The distributions of the coefficient values and the relations between them. Each dot represents the logistic regression model for one cluster of one agent. 71
3.19 A summary of the resulting logistic models by visualizing the 95% confidence intervals for the values of each of the coefficients of the explanatory variables for each city. The mean (overall) values are 0.85 for β_0, -3.12 for β_1, 0.01 for β_2, 0.03 for β_3 and -0.01 for $\beta_{4,5}$. The mean accuracy of the individual models is 0.81. 72
3.20 The percentage of clusters without choice for each explanatory variable. . 73
3.21 The correlation between the amount of training data of a cluster and the number of CPs in range in this cluster. One dot represents one cluster. A Pearson correlation of 0.73 was found ($p = 0.00$). 74
3.22 A summary of the resulting logistic models by visualizing the 95% confidence intervals for the values of each of the coefficients of the explanatory variables for each city. (A) shows the ‘weighted’ model while (A) shows the ‘weighted and trimmed’ model. 76
3.23 The distribution of the number of kWh charged. The *normaltest* confirmed that this distribution significantly differs from a normal distribution. 78

3.24 An example of the relation between the connection duration and the number of kWh charged for one agent. (A) Scatterplot and (B) mean value per 20 minutes to show the relation between the number of kWh charged and the connection duration. Each dot represents one transaction of the one agent's training data. 78
3.25 Examples of the usage of CPs within clusters. 80
3.26 The mean (A) training and (B) test errors with a 95% confidence interval for each modelling approach and each city. 82
3.27 The distribution of error scores for each of the error metrics in Equations 3.8, 3.9 and 3.10. 83
3.28 The distributions of the fractions of CPs used within a cluster for each approach compared to the actual fraction in the training data. In the training data 13% of the clusters use all CPs in range. In each the three approaches 89% of the clusters use all CPs in range. 84
3.29 The effect of the cluster size on the error scores with 95% confidence interval. A regression fit is plotted through these points, also with a 95% confidence interval for the regression estimate. The 'weighted and trimmed' approach is used. 85
3.30 The mean error score per cluster sizes for each city. The 'weighted and trimmed' modelling approach is used. 85
3.31 The distribution of test scores ($e_{te,si}$) for each of the approaches. The stacked bars show the difference between clusters with newly added CPs and clusters without newly added CPs. 86
3.32 The mean test score of each approach, for both clusters which have newly added CPs and clusters which do not have newly added CPs. A 95% confidence interval is shown. 87
3.33 The effect of the last placement date of a cluster on its cluster size. The last placement date of a cluster is defined as the latest placement date of all of the CPs in the cluster. A fit with 95% confidence interval is shown. 88
3.34 The test error of clusters with and without newly added CPs for each cluster size. Two logistic fits with 95% confidence interval are shown. The 'weighted and trimmed' approach is used. 88

4.1 The mean (A) connection duration distributions, (B) disconnection duration distributions and (C) arrival patterns for both habitual agents and car2go users. 92
4.2 The mean simulated activity pattern of 100 non-habitual agents is shown together with the mean activity pattern of all car2go data from Amsterdam. The simulation ran for a year. The agent validation score is 0.04. 93
4.3 The percentage of failed connection attempts for each city and for the different numbers of non-habitual agents. 94
4.4 The mean percentage of failed connection attempts for each city and for the different numbers of non-habitual agents, with a linear fit. The slope of this fit is shown in the brackets. 95
4.5 The distributions of failed connection attempts for the agents in the different cities when two non-habitual agents are added for each habitual agent. 96

4.6 The normalized mean arrival times for the the failed connection attempts in the four cities. 98
4.7 Mean (A) arrival times and (B) activity patterns for EV users in the different cities. The light blue line indicates the mean arrival times for non-habitual users. 98
4.8 The number of agents, the number of CPs and the ratio between agents and CPs in each city. 99
4.9 The distribution of number of sockets per CP for the different cities. . . . 100
4.10 The number of agents, the number of sockets and the ratio between agents and sockets in each city. 100
4.11 The mean (A) maximum distance and (B) walking preparedness of the agents in each city with a 95% confidence interval. (C) The number of sockets in range of the walking preparedness for each city with a 95% confidence interval. (D) The density per city. 102
4.12 The CPs of Amsterdam, colored (A) uniformly, (B) by the number of unique users per week, (C) by the number of kWh charged per week and (D) by the failed connection attempts. 104
4.13 The CPs of The Hague, colored (A) uniformly, (B) by the number of unique users per week, (C) by the number of kWh charged per week and (D) by the failed connection attempts. 104
4.14 The CPs of Rotterdam, colored (A) uniformly, (B) by the number of unique users per week, (C) by the number of kWh charged per week and (D) by the failed connection attempts. 105
4.15 The CPs of Utrecht, colored (A) uniformly, (B) by the number of unique users per week, (C) by the number of kWh charged per week and (D) by the failed connection attempts. 105
4.16 The effects of the rollout strategies on the percentage of failed connection attempts for different numbers of CPs added and for different cities. The dotted line indicates how many CPs each city currently has. 106

C.1 The number of failed connection attempts for each city and for the different numbers of non-habitual agents. 119
C.2 The mean percentage of failed connection attempts for each city and for the different numbers of non-habitual agents, with a linear fit. The slope of this fit is shown in the brackets. 119

List of Tables

2.1 The stakeholder concerns regarding public charging infrastructure from [18]. 7
2.2 An sample entry of a charging transaction in the CHIEF dataset [2, 4]. . . 8
2.3 The objectives, modeling methods, data usage and factors influencing EV user behavior of various existing EV models. 11
2.4 The factors which influence charging behavior [2, 4, 24]. 14
2.5 A comparison of the changing definition of charging behavior. 14
2.6 The entities of the model with their state variables. 17
2.7 An overview of the input, output and system interactions of each process. 22
2.8 The optional output metrics of the model. 24

3.1 A summary of literature related to the factors which influence how EV users choose a CP within a certain area. 53
3.2 The relationship between the relative charging fees and the frequency of use for each CP are displayed by showing the Pearson correlation values together with the corresponding p-values for each city. 61
3.3 The cluster counts for each interaction type. 63
3.4 An example of the resulting logistic model for one cluster of one agent in Amsterdam. 70
3.5 The error scores for each of the clusters in Figure 3.25. 81
3.6 The mean overall error for both the training data and the test data. The differences are significant between all approaches except between ‘Weighted’ and ‘Weighted and Trimmed’. 83

A.1 The parameters present in the model concerning the simulation. 115
A.2 The parameters present in the model concerning data preprocessing. . . . 115
A.3 The parameters present in the model concerning the agent’s behavior. . . 116
A.4 The parameters present in the model concerning the clustering agents’ CPs. 116

B.1 The mean coefficient values with a 95% confidence interval for each approach, each city and each coefficient. 117

Bibliography

[1] F Leurent and E Windisch. Triggering the development of electric mobility: a review of public policies. *European Transport Research Review*, 3(4):221–235, 2011.

[2] J Spoelstra. Charging behaviour of dutch ev drivers. Master's thesis, Utrecht University, 2014.

[3] T Franke and J Krems. Understanding charging behaviour of electric vehicle users. *Transportation Research Part F: Traffic Psychology and Behaviour*, 21:75–89, 2013.

[4] J Helmus and R van den Hoed. Unraveling user type characteristics: towards a taxonomy for charging infrastructure. *IDO-laad*, 2015.

[5] A Hess, F Malandrino, M Reinhardt, C Casetti, K Hummel, and J Barceló-Ordinas. Optimal deployment of charging stations for electric vehicular networks. In *Proceedings of the first workshop on Urban networking*, pages 1–6. ACM, 2012.

[6] T Sweda and D Klabjan. An agent-based decision support system for electric vehicle charging infrastructure deployment. In *2011 IEEE Vehicle Power and Propulsion Conference*, pages 1–5. IEEE, 2011.

[7] Z Zhu, Z Gao, J Zheng, and H Du. Charging station location problem of plug-in electric vehicles. *Journal of Transport Geography*, 52:11–22, 2016.

[8] M Uhrig, L Weiß, M Suriyah, and T Leibfried. E-mobility in car parks-guidelines for charging infrastructure expansion planning and operation based on stochastic simulations. *EVS28*, 2015.

[9] X Xi, R Sioshansi, and V Marano. Simulation-optimization model for location of a public electric vehicle charging infrastructure. *Transportation Research Part D: Transport and Environment*, 22:60–69, 2013.

[10] Z Yi and P Bauer. Optimization models for placement of an energy-aware electric vehicle charging infrastructure. *Transportation Research Part E: Logistics and Transportation Review*, 2016.

[11] M Momtazpour, P Butler, M Hossain, M Bozchalui, N Ramakrishnan, and R Sharma. Coordinated clustering algorithms to support charging infrastructure design for electric vehicles. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, pages 126–133. ACM, 2012.

[12] E Paffumi, M De Gennaro, G Martini, and H Scholz. Assessment of the potential of electric vehicles and charging strategies to meet urban mobility requirements. *Transportmetrica A: Transport Science*, 11(1):22–60, 2015.

[13] Ido-laad website. http://www.idolaad.com/, 2017.

[14] R van den Hoed and J Helmus. Data analysis on the public charge infrastructure in the city of amsterdam. In *Electric Vehicle Symposium and Exhibition (EVS27), 2013 World*, pages 1–10. IEEE, 2013.

[15] M Glombek and R Quax. Simulating cascading failure in the charging infrastructure of electric vehicles a novel approach for improving charging station deployment. 2016.

[16] J Spoelstra and J Helmus. Public charging infrastructure use in the netherlands: A rollout - strategy assessment. *Proceedings of the European Electronic Vehicle congress*, 2015.

[17] J Wardle, Y Hübner, P Blythe, and J Gibbon. The provision of public recharging infrastructure for electric vehicles in the uk–is there a business case? 2015.

[18] J Helmus and R van den Hoed. Key performance indicators of charging infrastructure. *Proceedings of the 29th Electronic Vehicle Symposium*, pages 1–9, 2016.

[19] R Wolbertus and R van den Hoed. Benchmarking charging infrastructure utilization. pages 1–15, 2016.

[20] M Weeda, P Kroon, and D Appels. Elektrisch vervoer in nederland in internationaal perspectief. 2012.

[21] Car2go amsterdam. https://www.car2go.com/NL/nl/amsterdam, 2016.

[22] I Bunzeck, M Paukovic, and C Feenstra. User preferences for charging locations and charging schemes - a survey in eight eu countries. In *Presented at: D-Incert seminar*, volume 20. by: Publication date: ECN Policy Studies 20-5-2011, 2011.

[23] N Daina, A Sivakumar, and J Polak. Electric vehicle charging choices: Modelling and implications for smart charging services. *Transportation Research Part C: Emerging Technologies*, 81:36–56, 2017.

[24] E Azadfar, V Sreeram, and D Harries. The investigation of the major factors influencing plug-in electric vehicle driving patterns and charging behaviour. *Renewable and Sustainable Energy Reviews*, 42:1065–1076, 2015.

[25] F Banez-Chicharro, J Latorre, and A Ramos. Smart charging profiles for electric vehicles. *Computational Management Science*, 11(1-2):87–110, 2014.

[26] J Helmus. Typology of charging behavior of ev users. 2016.

[27] H Cai, X Jia, A Chiu, X Hu, and M Xu. Siting public electric vehicle charging stations in beijing using big-data informed travel patterns of the taxi fleet. *Transportation Research Part D: Transport and Environment*, 33:39–46, 2014.

[28] T Franke and J Krems. What drives range preferences in electric vehicle users? *Transport Policy*, 30:56–62, 2013.

[29] S Carroll and C Walsh. The smart move trial: Description and initial results. *Centre of Excellence for Low Carbon and Fuel Cell Technologies Report*, 2010.

[30] M Kintner-Meyer, K Schneider, and R Pratt. Impacts assessment of plug-in hybrid vehicles on electric utilities and regional us power grids, part 1: Technical analysis. *Pacific Northwest National Laboratory (a)*, pages 1–20, 2007.

[31] M De Gennaro, E Paffumi, and G Martini. Customer-driven design of the recharge infrastructure and vehicle-to-grid in urban areas: a large-scale application for electric vehicles deployment. *Energy*, 82:294–311, 2015.

[32] R Smith, S Shahidinejad, D Blair, and E Bibeau. Characterization of urban commuter driving profiles to optimize battery size in light-duty plug-in electric vehicles. *Transportation Research Part D: Transport and Environment*, 16(3):218–224, 2011.

[33] F Eggers and F Eggers. Where have all the flowers gone? forecasting green trends in the automobile industry with a choice-based conjoint adoption model. *Technological Forecasting and Social Change*, 78(1):51–62, 2011.

[34] V Grimm, U Berger, D DeAngelis, J Polhill, J Giske, and S Railsback. The odd protocol: a review and first update. *Ecological modelling*, 221(23):2760–2768, 2010.

[35] T Zhang, R Ramakrishnan, and M Livny. Birch: an efficient data clustering method for very large databases. In *ACM Sigmod Record*, volume 25, pages 103–114. ACM, 1996.

[36] Osrm. http://project-osrm.org, 2017.

[37] J Kruger. Loopbereidheid van voertuig gebruikers, 2016.

[38] J Smart, J Davies, M Shirk, C Quinn, and K Kurani. Electricity demand of phevs operated by private households and commercial fleets: effects of driving and charging behavior. *EVS25, Shenzhen, China*, 2010.

[39] F Jabeen, D Olaru, B Smith, T Braunl, and S Speidel. Electric vehicle battery charging behaviour: findings from a driver survey. In *Proceedings of the Australasian Transport Research Forum*, 2013.

[40] N Daina. *Modelling electric vehicle use and charging behaviour*. PhD thesis, Imperial College London, 2014.

[41] X Sun, T Yamamoto, and T Morikawa. Charge timing choice behavior of battery electric vehicle users. *Transportation Research Part D: Transport and Environment*, 37:97–107, 2015.

[42] J Helmus and R van den Hoed. Unraveling user type characteristics: Towards a taxonomy for charging infrastructure. In *EVS28 International Electric Vehicle Symposium and Exhibition*, pages 1–16, 2015.

[43] H Yu and D MacKenzie. Modeling charging choices of small-battery plug-in hybrid electric vehicle drivers by using instrumented vehicle data. *Transportation Research Record: Journal of the Transportation Research Board*, pages 56–65, 2016.

[44] Y Wen, D MacKenzie, and D Keith. Modeling the charging choices of battery electric vehicle drivers by using stated preference data. *Transportation Research Record: Journal of the Transportation Research Board*, pages 47–55, 2016.

[45] M Xu, Q Meng, K Liu, and T Yamamoto. Joint charging mode and location choice model for battery electric vehicle users. *Transportation Research Part B: Methodological*, 2017.

[46] C Sheppard, R Waraich, A Gopal, A Campbell, and A Pozdnukov. Modeling plug-in electric vehicle charging demand with beam, the framework for behavior energy autonomy mobility. 05/2017 2017.

[47] S Zoepf, D MacKenzie, D Keith, and W Chernicoff. Charging choices and fuel displacement in a large-scale demonstration of plug-in hybrid electric vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, pages 1–10, 2013.

[48] C Botsford and A Szczepanek. Fast charging vs. slow charging: Pros and cons for the new age of electric vehicles. In *International Battery Hybrid Fuel Cell Electric Vehicle Symposium*, 2009.

[49] L Jerram and J Gartner. Dc charging networks in the united states. *Navigant Research*, 2017.

[50] R Wolbertus. The jungle of charge tariffs in the netherlands. http://www.idolaad.com/shared-content/blog/rick-wolbertus/2016/charge-tarrifs.html, 2017.

[51] Sklearn. Machine learning in python. http://scikit-learn.org, 2017.

[52] J Verlaan. Laadpalen elektrische auto's binnen vijf jaar achterhaald. *NRC*, 2016.