CHALMERS UNIVERSITY OF TECHNOLOGY | UNIVERSITY OF GOTHENBURG

Modelling Electrical Vehicle Charging Demand with Historical Vehicle Movement Data

Master's thesis in Computer science and engineering

Kristina Markan
Victor Nilsson

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2022

Master's thesis 2022

Modelling Electrical Vehicle Charging Demand with Historical Vehicle Movement Data

Kristina Markan
Victor Nilsson

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2022

Modelling Electric Vehicle Charging Demand with Historical Vehicle Movement Data

Kristina Markan, Victor Nilsson

Supervisor: Divya Grover, Department of Computer Science and Engineering
Advisor: Jens Andersson, WirelessCar Sweden AB
Examiner: Nir Piterman, Department of Computer Science and Engineering

Master's Thesis 2022
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2022

Modelling Electric Vehicle Charging Demand with Historical Vehicle Movement Data

Kristina Markan, Victor Nilsson
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

In order to meet the environmental goals of 2030 set by the European Union, Sweden needs to build between 90 000 and 260 000 public charging poles in the upcoming eight years. In this thesis we have developed a mathematical model that given a area and time outputs the demand for charging in said area, along with introducing a redundancy value intended to show the tendency of vehicles in the area to stay plugged in to a charging station longer than necessary. Areas that show a large demand value and a low redundancy value can be good candidates for where to extend the charging infrastructure and build additional public charging poles. A prototype was built to test the model with real-world historical vehicle movement data provided by WirelessCar Sweden AB. The prototype was tested in three different areas: a supermarket parking area, a residential area, and an office area. Arranging the prototype output on a timeline for a regular weekday shows demand that corresponds with expected traffic in each area, and relative peak demand/redundancy between areas also follows expected patterns. However, assessing the precision of the output values would require more data and refinement of auxiliary functions used in the model, primarily by attaining a true State of Charge value provided by the vehicles.

Keywords: Electric vehicles, plug-in hybrid electric vehicles, charging infrastructure, vehicle data, mathematical modelling.

Acknowledgements

We would like to thank our supervisor Divya Grover for always quickly providing us with thoughtful feedback and ideas, and helping us with all of our questions. We want to thank our examiner Nir Piterman for taking an interest in our project and accepting a drastic pivot of the thesis subject. We would also like to thank our company advisor Jens Andersson at WirelessCar for assisting us with understanding and handling the data and trusting us to determine the direction of the project. Finally, we would like to thank WirelessCar and Volvo Cars for giving us permission to use the provided dataset for this thesis.

Victor Nilsson & Kristina Markan, Gothenburg, March 2022

Contents

List of Acronyms

BEV Battery Electric Vehicle.

CPEV Charging Point per Electric Vehicle.

EV Electric Vehicle.

HEV Hybrid Electric Vehicle.

OSM Open Street Map.

PHEV Plug-In Electric Vehicle.

SoC State of Charge.

List of Figures

List of Tables

1 Introduction

The size of the electrical vehicle fleet is ever increasing. In Sweden, the amount of electrical vehicles grew by over 80% in 2020 to surpass 200,000 in the beginning of 2021 [1]. As the numbers continue to grow, so does the demand for charging poles – the number of which also increase steadily, in Sweden seeing an increase of over 25% public charging poles in the same time period [2].

Previous studies has been done to determine the allocation and cost optimization of electrical vehicle (EV) charging facilities [3][4] as the need of the extension of infrastructure for EV battery charging has steadily increased. In this project, we will propose a model for determining the demand of charging poles by area based on real-life vehicle movement data. The goal of this model is to be useful as an aid when assessing where to spend resources on extending charging infrastructure, for which the aforementioned data may be of great help in order to capture the behaviour of electric vehicle owners.

1.1 Background

The European Union has set a goal of reducing emissions by 90% in 2050 (compared to 1990 levels), and having at least 30 million zero-emission vehicles on the road by 2030, with the intent that there will be 3 million public charging poles to support these vehicles is one of the stated measures to accomplish this [5]. The statistics at the time of writing (November 2021) for Sweden shows that there are 284 365 battery electric vehicles (BEV) and plug-in hybrid electric vehicles (PHEV) registered, but the current number of public charging poles is only 14 123 [1]. As EU intends to reach a charging points per electric vehicle (CPEV) value of 0.1, i.e one charging pole per ten BEVs och PHEVs, Sweden is lacking about 14 000 charging poles to meet the demand today, as the current CPEV value is 0.05. A mere 3 407 charging poles has been built in 2021 so far, which results in a 30% reduction in CPEV value from last year, as the value in 2020 was 0.07 [1].

The lack of charging poles is not only a problem pertaining to reducing emissions, as it also affects the phenomenon *range anxiety*, which occurs when a driver worries about the battery running out of power before reaching a charging destination [6]. A driver of a combustion fuel vehicle may feel anxious when the fuel level is low, but once a gas station is found the whole process of refueling takes at most a few minutes (not accounting for waiting in line, buying something at the station shop and so forth). A full charge on an BEV and PHEV depends on battery capacity and available effect on the charging pole, and can take anywhere from 20 minutes (ultra-fast direct current) to 16 hours (single phase alternating current). This generates another type of behaviour than drivers of fuel combustion vehicles have, who can expect to be able to refuel when at a gas station, whereas drivers of electric vehicles will find charging poles to be occupied for longer periods of time.

We have been granted access to a sizable set of anonymized data from the company WirelessCar, a service provider of Volvo Car Corporation. This data set contains information about recent trips made by all connected Volvo brand vehicles. In this context, a *trip* is a representation of a vehicle moving from point A to point B, with the beginning and end of this trip being marked by the engine being turned on and off, respectively. Of note is that the vehicle set, in regards to EV vehicles, contains only PHEV's as the company does not manufacture any BEV's at this time.

This data allows us to consider historical vehicle movement patterns from real-life scenarios, and provides insight into multiple aspects of private vehicle transport. Tying this information into the context of EV charging infrastructure, this work will propose a use of this information with the purpose of aiding planning efforts of where to build new EV charging facilities.

1.2 Goals

The question that this project aims to answer is '*By making use of historical vehicle data, how can we model the demand of charging infrastructure in a specific area at a specific time?*'. The goal is to develop a mathematical model for this, and if time allows, a prototype using the provided data set. The model should be generalized and thereby not representing a single brand of vehicles; while the data available to us is limited to a single brand, we expect most brands are able to provide data points similar to what is available for this work.

Formally, the objectives of the project are as follows:

Model development: Construct a model that presents the electrical charging demand in a specified location.

Prototyping: Develop a prototype using the WirelessCar data set to assess the viability of the proposed model.

Evaluation: Evaluate the model by using the prototype to run simulations and analysing the result.

1.3 Limitations

The data set only contains data from two vehicle manufacturers, Volvo and Polestar, with the vast majority of data entries coming from Volvo vehicles. The variables in the model will be generalized, but any result of prototyping the model will be limited to these two brands of vehicles only simply due to the fact that this is the data available to us.

A few limitations will be present during the prototype implementation. As not all model parameters are known, some need to be simulated. The dataset contains millions of data entries in a vast number of locations, therefore we will select a couple of locations to use in order to not having to run an extreme amount of simulations. Also, since crowd-sourced data will be used in the prototype, the locations of choice will be locations that we personally know and can categorize accordingly – this is to make sure that categorization is not flawed due to human error.

1.3.1 Risks

Out of all passenger vehicles in Sweden in 2020, 10% is manufactured by Volvo. In other words, the other 90% of vehicle data is "missing" in the data set. Missing is put in quotation marks because it for one may not be so that all other car manufacturers collect these type of trip data entries from their vehicles, but even if they did the data is not available to us. Due to a large part of the actual data being unavailable, the risk of bias increases when the data set is used to create a prototype of the model. There could be a risk of Volvo vehicle drivers acting different than other drivers and thereby creating other patterns than what is present in the data set, for example, by having different patterns of parking in different areas. However, as we aim to generalize the model as much as possible and always have a point-of-view of *vehicle behaviour* rather than *Volvo vehicle behaviour*, the model should be sound in theory and hopefully be able to describe electrical charging demand for any type of passenger vehicle, rather than just Volvo vehicles.

Another, similar risk is the fact that the EVs present in the data set is only PHEVs. While drivers may be assumed to be willing to charge their vehicle when possible, there is a large difference between requiring to charge a BEV for it to be operational, and preferring to charge a PHEV to save on fuel cost; it should be expected that this difference leads to differing behaviours between drivers of the respective vehicle category. As described in the previously mentioned risk, the design of the model should not suffer due to this as available data points are shared between PHEVs and BEVS. However, any results of prototyping in particular would of course be affected by a lack of data originating from BEVs, which is unfortunate due to the aforementioned expectation of differing behaviour between drivers.

Some other risks that are involved is that we are working on actual, live data that due to GDPR may change due to requests for data removal or similar. While this risk is small, the best way to mitigate this risk is to discuss this scenario with WirelessCar in order for us to refresh data dumps or prune data that we should no longer have right to use. We can also mitigate the risk of lacking data to test on by simulating data in similar amounts to what would be expected in a real-life scenario.

1.4 Contribution

As a continuous expansion of the charging infrastructure is expected to take place in the coming years, finding relevant locations for new charging facilities is of high importance in order to efficiently spend resources. We argue that due to the sheer scale of this effort, any methods used to aid with determining such locations will be useful, and that a model using historical vehicle data would be highly relevant to include the behaviour of drivers in real world situations in these considerations. Matching charging infrastructure with said behaviour is likely to increase the rate of electrical vehicle adoption overall, as ownership of electrical vehicles will seem more attractive as obstacles to their operation (in this case the need for charging) are eliminated.

1.5 Contents

This thesis is structured as follows: first, we dive into the theory of the dataset and electrical vehicles, and also some background about electrical charging poles and the current state and planned future of the electric charging infrastructure in Sweden. The method section describes the construction of the model and its different parts in detail, as well as the prototype process. The results section presents the finished model and results after prototyping and simulation. A discussion about the model, the methods used, and the result can be found in the final chapter.

2
Theory

In the upcoming sections we provide a further look into the WirelessCar data set that will be used for the project, following with context and background for a couple of APIs that will be utilized in the model. Moreover, information about electrical vehicles, their batteries, the current status and plans for the electric charging infrastructure in Sweden, and different types of charging poles is also provided in this chapter.

2.1 The WirelessCar Dataset

Vehicles of today are able to collect data of their operation and provide this data to relevant parties. The vehicle data used in this work consists of data collected exclusively by the vehicle itself, without any third-party hardware or software, and is arranged on a per-trip basis. A *trip* is defined as a vehicle starting the motor and then moving from point A to point B, where the motor is turned off. From this it follows that there exists a possibility to create a collection of parking events for a vehicle, which is the time between the end of one trip and the beginning of the next trip made by the vehicle. As the exact timestamp for when a trip starts and ends is known, so is also the total time elapsed during the parking event.

While a trip entry does contain some additional data besides the starting/ending location and the corresponding timestamps, such as electrical consumption for the trip, there are also data that is collected by the vehicle but is for several reasons not available in the data set. An example of such data is the State of Charge (SoC) at the time of the trip, which is an EV equivalent to a combustion vehicles fuel level indication. This is data that is simply not saved historically – owners may, however, see current SoC in their phone app or on the vehicle dashboard. Another example is that of trip waypoints – intermediate geographical locations showing the path taken by a vehicle. This data is available for some vehicle models, but is not consistently available for all data.

The dataset that used in this project is provided by the company WirelessCar. As a data processor for Volvo Car Corporation, WirelessCar has access to trip data spanning over the last three months. Earlier data is not available as it is deleted when older than three months due to privacy reasons, in accordance with GDPR. Even though the data provided is solely from Volvo and Polestar brand vehicles, the trip data is saved in a general-purpose format which is common for several other brands also handled by the company. It is, therefore, reasonable to expect that other brands provide similar data points to what is available in this data set, as their respective data is saved in the same format.

It should be noted that the data contained in the data set is not completely raw, vehicle-sourced data. Some enriched data is included with each trip entry. This enrichment is done by WirelessCar in the interest of making the data more human readable and able to be presented to customers (the owners or drivers of the vehicles). This data mainly pertains to the location of the vehicle expressed in a street address, which is gained from reverse geocoding services. In short, we can summarize a trip entry as a description of where and when a vehicle starts and ends a trip, how long the trip was, and how much fuel/electricity was consumed over the course of the trip.

2.2 H3 Hexagonal System

H3 is a grid system developed by Uber to process spatial data. The system divides the earth's surface area into hexagons that are tiled over the surface, which come in different resolutions depending on the need of the user [7]. Every hexagon has a unique, static identifier, allowing the grouping of different but closely located geographical locations (as expressed in the geographic coordinate system) to a single hexagon – with 'close' being a matter of choosing an adequate resolution of said hexagons. Resolutions are defined as 16 different sizes of hexagons, ranging from average hexagon size of 4,250,546.8477000 km^2 (for a total of 122 unique hexagon indices), to 0.0000009 km^2 (for a total of 569,707,381,193,162 unique hexagon indices). Simply put: given an arbitrary coordinate, H3 is able to return an index for a hexagon the coordinate is located in, with there being one unique hexagon for each of the 16 different sizes.

2.3 Open Street Map

Open Street Map (OSM) is a free, crowd-sourced map service that can be used with the feature-rich API Overpass. OSM together Overpass provides information about different elements that exists in an area, such as roads, buildings, and neighbourhoods exist in an area. Furthermore, the elements themselves can have *tags*, a key-value pair that provide more context about them. For example, if an area was tagged as residential, one could expect there to be apartments or houses in the area, and parking for the residents. There might even be private charging poles at these parking spaces, where residents can charge their car overnight.

There is a vast amount of tag key-value pairs available to choose from in OSM, making it possible to describe the elements in an area in great detail. Worth noting is that a tagger can choose to select a more general tag for an element, such as tagging something as simply *building* rather than the, in this example, more correct tag for an apartment building: *building:apartments*.

2.4 Electric Vehicles

Electric vehicles can generally belong to one of three categories: BEVs, PHEVs or Hybrid Electric Vehicles (HEVs). BEVs have no combustion engine whatsoever, relying solely on electrical power to drive the car and power the amenities such as the climate control. The vehicle is charged by plugging in a cable to a charging box or charging pole that is connected the power grid. PHEVs have an electrical engine as well as a internal combustion engine, running on stored electricity until it runs out at which point the car is instead brought forwards with the combustion engine. As the name implies, the battery in a PHEV is charged similarly to BEVs.

HEVs also have two different engines, but unlike the PHEVs, the electrical engine is only possible to recharge by driving the car and mainly by breaking - at the time of breaking, the electrical engine doubles at a generator and stores energy from braking in order to use it for driving the vehicle later. The electrical engine in HEVs is mainly intended to aid the vehicle at low speeds, and using electricity for more extended periods or higher speeds in an HEV is not feasible. As HEVs does not require an external charger, these vehicles should not be included in the demand for public charging poles, and they are also not allowed to park at parking spots that have a charging pole [8].

2.4.1 Batteries

The battery in a BEV or PHEV is typically a lithium-ion rechargeable battery. The capacity of the battery is measured in kWh, and differs between PHEVs, which can be expected to have a battery capacity of around 10 kWh, and BEVs, which battery capacity ranges between about 30 - 100 kWh. With larger capacity comes longer time to fully charge the battery, but also longer driving range.

Vehicle Model	Battery Capacity	Maximum Charge Effect
Volvo V60	11.2 kWh	3.7 kW
Volvo V90	10.4 kWh	3.7 kW
Volvo XC40	9.7 kWh	3.7 kW
Volvo XC60	10.4 kWh	3.7 kW
Volvo XC90	9.2 kWh	3.7 kW
Volvo S60	11.7 kWh	3.7 kW
Volvo S90	10.4 kWh	3.7 kW
Polestar 1	34 kWh	11 kW

Table 2.1: The different PHEV models present in the data set, their battery capacity, and maximum charge effect. Note that battery capacity can differ a bit between older and newer generations of the same vehicle model.

The battery capacity is permanently affected by the age and usage of the vehicle. The exact range of degradation differs between vehicle models, but an average annual decrease was found to be 2.3% by the Amerian tech company GeoTab [9]. A common warranty given by vehicle manufacturers is that the battery should keep 70% of its original capacity up to 8 years or 160 000 km.

The battery capacity may also temporarily decrease with temperature. According to GeoTab, battery capacity was found to be at its best when the ambient temperature is 21°C [10]. The decrease in capacity is mainly due to the electrochemical reaction in the battery being slower in colder weather.

2.5 Electric Charging Infrastructure

As previously mentioned in the introductory chapter, the EU has decided on attaining a goal to have at minimum 30 million zero-emission vehicles and 3 million public charging poles to support them by the year 2030. In 2014, the Alternative Fuels Infrastructure Directive (AFID) was adopted by the EU, which states that member states should have an "appropriate" amount of charging stations accessible to the public by the end of 2020 [11]. The appropriate amount was to be decided by each member state and influenced by the projected number of electric vehicles by 2020, the states national framework, as well as recommendations set by the Commission, such as the CPEV value. Most member states, therefore, set a goal for the number of charging stations that should be ready in 2020, and the union as a whole reached and surpassed the plan with the number of charging stations being at almost 140% of the set amount [12].

Sweden, however, did not set a goal for number of charging stations and is therefore not a part of the reason the union as a whole reached the goal. As already mentioned in the introduction, the CPEV for Sweden today is 0.05 while the recommended value by the EU is 0.1, Sweden is thereby missing about 14 000 charging poles to meet the demand of today.

The Swedish government published an official report in June 2021 on policy instruments for electric charging infrastructure, and while the report mentions the CPEV ratio recommended by the EU, the primary strategy seems to be to ensure that there is a sufficient amount of private charging poles, i.e. by reducing the cost to install charging devices, as well as implementing charging infrastructure requirements for new constructions and extensive renovations [13]. Following this strategy, in January 2021, the Swedish Tax Agency implemented the current 50% cost reduction on material and labour for owners of single-household houses that install either a charging pole or a charging box for their own private EV on their property [14]. Another measure that the report deemed necessary is to ensure a sufficient number of charging poles with fast charging along the Swedish road network, where a shortage could be found mainly in the northern parts when the report was published.

The CPEV metric is not the only available metric that is used for measuring the sufficient number of supporting charging poles for all EVs in a state; another example is presented in the *Recharge EU* report by the organization Transport & Environment [15]. Their model uses weighted values when calculating the sufficient amount, where BEVs have twice the weight of PHEVs, and charging poles with faster charging speeds have a greater weight than charging poles with slower speeds. Individual charging poles are also weighted based on if they are public or semi-public.

According to the T&E model, the sufficient amount of public charging poles in Sweden in 2030 would be 90 000. A report published in 2020 by Stockholm's Chamber of Commerce presented three different scenarios for the number of EVs in Sweden 2030, where the low-value scenario was 1.4 million EVs, the mid-value scenario was at 2.6 million EVs, and the high-value scenario was 3.4 million EVs [16]. Using the mid-value at 2.6 million expected EVs in 2030 and the EU model with a 0.1 ratio, the number of charging poles in 2030 should be at 260 000 to be sufficient. If instead using the T&E value with 90 000 charging poles, the CPEV value would be about 0.03, which is worse than today's value of 0.05.

Regardless of which one of the two models is considered, somewhere between 76 000 and 246 000 public charging poles needs to be built in order to have a sufficient coverage for all EVs in Sweden in the 8 years leading up to 2030. This demonstrates a need for considering new locations to provide with charging infrastructure, as well as evaluating current locations for possible extensions.

2.5.1 Charging Poles

While BEVs and PHEVs may be charged from standard wall outlets, there also exists stations dedicated for the charging of vehicles in different places around the country. In Sweden, there is roughly 2600 charging stations with about 14 000 unique charging poles dedicated for this purpose [2]. Stations may differ in both plugs and wattage, but are generally able to serve any vehicle when taking adapters into account.

Type of Charger	Effect	Time to Charge
Single Phase AC	3-7 kW	7-16 hours
Tri-Phase AC	11-22 kW	2-4 hours
Fast DC	50-100 kW	30-40 minutes
Ultra-Fast DC	100+ kW	10-20 minutes

Table 2.2: Different type of chargers for EVs, their effect and the time it takes to fully charge an EV battery using each charger.

While public charging poles may be equipped with fast and ultra-fast DC chargers, the chargers installed in homes are single-phase or tri-phase AC chargers. The effect is further decided by the available ampere, where a lower ampere amount (16 A) gives the lower effect output, 3.7 kW for single phase and 11 kW for tri-phase, while a higher ampere (32 A) gives the higher effect output at 7.4 kW for single phase and 22 kW for tri-phase, respectively. Worth to note is that many PHEVs in particular are unable to receive more effect than 3.7 kW, even if using a charger that is able to output higher effect [17].

3
Methods

This chapter defines and outlines the parameters required to create the model including motivations of why the parameter is relevant for our model and how to feasibly construct it with the help of either our provided data, external data, previous studies or the combination thereof.

3.1 Model Overview

At the most abstract level, the model aims to answer a single question for an arbitrary geographical boundary – *is the current number and capacity of the electrical charging poles currently in the area matching the demand of electrical vehicles in said area?* From the questions posed, one may possibly spot a well-known general concepts: supply and demand. With only this level of abstraction, the answer to the question becomes simple; if the available demand is higher than the available supply in an area, then the area should be a candidate for extension of charging infrastructure. If not, resources for this purpose are better spent elsewhere.

The follow-up question then becomes how to, in this context, express supply and demand in a way that would allow us to answer the question? The following sections aims to provide a more in-depth argument for how to model these concepts.

3.2 Model Parameters and Concepts

The following section lists parameters that will be used in the model, some only using the help of the base data, and some with the help of APIs to fetch and use external data.

3.2.1 Geospatial Indexing

As the end goal of the model is to answer the question if the electrical charging poles present in an area is adequate, a key point is to identify what constitutes an area. For the proposed model we will be using H3 indexing, and as such, the terms *hex* and *area* will be used interchangeably going forward. The choice of this indexing in particular is that it provides complete geographical coverage, indexed areas are uniform and easily adjustable in size. The hex resolution used for the indexing is 9, but could be changed to either a smaller or larger resolution if that would fit the need better.

However, assuming that movement data uses typical longitude and latitude positioning, and expresses end-of-drive sessions on this format, any form of indexing can be used depending on need. Formally, what is required is some indexing function that given a coordinate returns a unique index for a corresponding geographical area. The requirements for a geospatial indexing usable in the model are forgiving: Any geospatial indexing system that given a latitude and longitude returns a single, unique index of the area containing the given coordinate is usable for this purpose. By definition, the model will be able to be applied to any area included in the chosen geospatial index, but results will likely be more useful when working with well-bounded areas of comparable sizes.

3.2.2 Area Categorization

In order to get a better understanding of the area in question, the next step after H3 indexing is to categorize the individual hexes in further detail. The categorization allows more relevant assumptions about the data, which in turn benefits the model. Indexing by hexes specifically is not needed to implement the categorization as it could be done with any arbitrary indexing, however, as H3 hex indexing is the method of choice in this project this is what will be explained in this section.

To implement the categorization, we use the Overpass API to fetch data from OSM. The following tags are deemed relevant to use in the process:

Key	**Value(s)**
Amenity	Parking.
Landuse	Commercial, Industrical, Retail.
Highway*	Residential.
Building	Apartment, House (Single Household), Office.

Table 3.1: Key-Value pairs of Open Street Map tags that are used for area categorization in the model. **Note: Highway is the tag used for any type of road, street or path. Examples range from motorway to sidewalk.*

Assuming that we can query Overpass for an area corresponding to the target area for our model, we are able to categorize each hex within the total area. In order to do so, we begin by sending an API query to Overpass that requests all elements with any of the tags listed above within the total target area. The API returns a JSON object with a list of each element. Some elements are single points (nodes), while a combination of nodes creates others (ways or areas). No matter the element type, each has a centre point represented by a latitude and longitude value. The centre points are extracted, and their corresponding H3 hex is found. If a corresponding hex exists in our list of hexes, i.e. historical trip data is reported there, a point for the element tag is given to the hex. A hex will be assigned the same category as the one for which most centre points reside in that hex – as an example, a hex containing the centre point of two parking lots, three apartments and one office space, will be categorized as an apartment area.

When a hex has been categorized, assumptions about the electrical charging supply can be made for where there is no data about public charging stations, which mainly is the case for residential areas and office areas. This is explained in more detail in Section 3.2.3.1.

3.2.3 Area Data

Other than what is directly available in the data set, some auxiliary data is required for the model. For any given indexed area, there is a need to add a set of parameters which will be connected to each such area and serve to model behaviours that are specific to certain areas. Areas must be considered separately for natural reasons – as an example, considering two different parking lots, it should be expected for vehicles to generally stay longer on a parking lot close to office areas than a lot closer to a supermarket. The following sections outline parameters used by the model and how they may be determined.

3.2.3.1 Charging Stations

The amount of charging stations currently in a given area is, not surprisingly, the base factor of the charging capabilities of said area. For commercial and public areas, data regarding the location and capacity of charging stations is readily available, both as open data and as proprietary service offerings. In this project we will use the NOBIL dataset API for public charging poles, which is free to use for anyone after applying for and receiving an API key. Important to note is that electric vehicles may also be charged by other means than public stations; a driver charging their vehicle at their residence may come to mind.

For charging stations not covered by available data, determining the available points of charge is not as straight forward but may be estimated with the help of the previously mentioned area categorization. As an example, in a residential area consisting of exclusively single-household houses we could assume that any EV's parked overnight in the area belongs to a residence owner and would at the very least have a outlet available for low-effect charging. Office parking areas are harder to make clear-cut assumptions about, but charging outlets may be available for employees or similar.

3.2.3.2 Average Parking Duration

For a given area, the average parking duration in it may determine how efficient use of resources a potential addition of charging poles would be. For each trip ending in a certain hex, the total parking time may be determined by looking at the average time until the same vehicle begins a trip anew. Not all trip starts may signify the vehicle leaving however – if a trip is made with a length that is less than some value ϵ, it will not be considered and the next significant trip is considered to be trip concluding the parking duration. For each hex, it is straight-forward to find the average parking duration by simply sorting starting and leaving events by timestamp and averaging the time between a vehicle ending a trip in a hex and the following significant trip.

The choice of area separation could give rise to a potential edge case that could occur whenever any uniform division of area is used, such as H3 indexing. Should a vehicle be parked very near the boundaries of an area, it is possible that any movement less than ϵ would result in the vehicle ending up in a different area. A more categorical division of areas would alleviate this issue as we would then expect a single parking lot to be considered a single area – assuming that the GPS is accurate enough. Currently, we will move forward accepting this risk.

3.2.3.3 Charging Redundancy

When parked and plugged in to a charging pole, it is not a given that vehicles leave shortly after they have been charged to full. Rather, it may be assumed that a vehicle parked and plugged in will occupy that space until the driver is leaving the area. This may affect charging efficiency. The time difference between how long a vehicle needs to stay at a certain charging pole to be fully charged and the time it actually stays there results in loss of potential charging, something that needs to be taken into account when considering the supply of vehicle charging in an area.

Given that the parking duration is available with the historical trip data, it is possible to model this redundancy. Under the assumption that a vehicle is plugged in over the full duration of a parking – which is very feasible as the driver is most likely doing something else in the vicinity – we are able to calculate the theoretical time to full charge depending on the effect of the charging pole and the capacity of the vehicle. Any time spent parked exceeding that time will be considered to be *charging redundancy*.

Charging redundancy could be calculated on a vehicle-to-vehicle basis, but we argue that the location plays a larger role in the charging redundancy than the driver – it is likely that charging poles placed in office area parking lots would see more redundancy than those placed outside a supermarket due to the expected time for someone to be standing in each place. Furthermore, given that the question is where charging poles should be placed based on historical data, expressing charging redundancy by certain areas would certainly help in determining where new charging poles are both desirable and less likely to be occupied by a single vehicle for longer periods, which as a metric complements demand which is not as likely to cover this particular aspect of charging infrastructure.

3.2.4 Vehicle Data

Assuming vehicle movement data in general can be expected to include the data points as described in Section 2.1, each trip entry includes information of the electrical consumption for the trip. This data does not provide much information in itself, and it is unknown if the vehicle uses charging poles, the current SoC, and to find out anything regarding the charging habits of the drivers. Nevertheless, using the vehicle movement data together with other available information and introducing variables based on findings in previous studies enables the definition of a model for the demand of a vehicle.

The following sections outlines the parameters used in the model that originate from the vehicle data in the provided data set, as well as auxiliary data directly connected to vehicle use.

3.2.4.1 State of Charge

SoC of the vehicle refers to the remaining charge in the battery, usually expressed as a percentage of the total capacity. While the SoC at time of a completed trip is not reported, it is, of course, visible in the vehicle for the driver. While SoC could be estimated by taking into account the capacity of the vehicle model and the electrical consumption historically, there would have to be some assumptions made regarding the SoC at some point. For instance, it might be a reasonable assumption that a vehicle leaving a residential area with mostly houses at morning would have been charged overnight.

Although estimates would be possible, the model will assume that the vehicle reports SoC at the time of a completed trip. For this work it is motivated by the fact that it is a readily available value in the car which would only slightly increase the volume of data sent from each car – while the raw amount of data from a large number of vehicles may not be negligible, it would serve as a very minor increase in relation to the volume already being sent.

As data is available regarding which vehicle model is performing each trip, a baseline model for demand can be built as a foundation before looking at other factors that may affect total demand. As the demand for charging must at some point originate from the current SoC of a vehicle, there is a need to include some estimation of this value as it is not reported from the vehicle itself.

While SoC would be reasonable data to include from the vehicle as part of any trip entry, the current lack thereof requires some way to make an estimate. For the purposes of our prototype, we will make some very broad assumptions and let the SoC of a vehicle at any given time follow a normal distribution $\mu = 50, \sigma^2 = 20$, as drivers would tend to avoid sinking to low charge due to range anxiety, but also tend not to look to charge their vehicle when it is not necessary. Some proposals for more elaborate methods for estimating the SoC are listed in Section 5.8.1.

3.2.4.2 Charging Probability Density

When a vehicle stops by an area with a charging pole, it is not certain that the driver will use it – all available poles could be occupied, or the driver might deem the current SoC sufficient for the time being and opt not to charge the vehicle even when given the opportunity. Previous work shows that in general, the majority of charging events take place when the SoC is between 25% and 75% [18], indicating that drivers tend generally tend not to let the charging level of the vehicle decrease too much. It should be noted that these findings are based on BEV's, and that the behaviour of PHEV drivers are likely to differ as while driving electrical is economically sound, it is not a necessity making it possible for drivers to skip charging in favor of convenience.

For the purposes our model, the SoC will use a simple table based on earlier findings, in which we model the probability of a driver opting to charge their vehicle to be dependent on the SoC when stopping the vehicle.

3.2.4.3 Battery Capacity

All of the different vehicles in the dataset are not equipped with the same drive train, nor the same battery. Information about exactly which battery the vehicle has is not available in the dataset, and has thereby been added manually by matching vehicle model to its respective battery according to information from Volvo. This is a time consuming task, but it ensures that the model final represents each vehicles electrical demand with higher accuracy.

According to several studies made on battery capacity in electric vehicles, the battery capacity decreases with both age and use [19, 20]. The exact range of degradation differs between vehicle models, but an average was found to be 2.3% decrease annually by the American tech company GeoTab. Therefore, a variable b_{age} is introduced to represent this decrease in capacity:

$$b_{age} = (1 - 0.023)^y$$

where y is the age of the vehicle in years.

The battery capacity also decreases with temperature. According to GeoTab, battery capacity was found to be at its best when the ambient temperature is 21°C [10]. The decrease in capacity is mainly due to the electrochemical reaction in the lithium ion battery being slower in cold temperatures. The previously mentioned study presents a range curve of battery capacity over temperature, and a curve fitting was done to find the following equation:

$$b_{temp} = 0.65 + 0.02263276 * x + 0.0001167148 * x^2 - 0.00001703223 * x^3 - 1.856662e - 7 * x^4 + 5.194805e - 9 * x^5$$

Where x represents degrees in Celsius and is limited to the range between -25 and 21.

We will assume that a PHEV behaves in the same way as a BEV, i.e. uses 100% electricity when it is available, although the reality might be more complex. For example, the air conditioning in a BEV will be powered by the battery, when a PEHV may use the combustion engine for it instead.

3.2.5 Modelling Demand Per Area

In order to determine the current demand in a defined area, an aggregate of the vehicles in said area needs to be considered. For the purposes of our modelling, the demand in an area d at time t is completely reliant on the demand of the charging vehicles parked in the area at time t.

3.3 Creating a Prototype

The model as described above will be implemented as a prototype using Python and suitable libraries, where a few chosen areas are selected for simulations and key metrics are extracted. Results should be compared with known and existing data in order to have some sort of baseline towards which to evaluate the model – it should be expected for charging demand in an area to somewhat follow the expected traffic in the area (assuming that it is an area in which drivers tend to park their vehicle).

To evaluate the prototype we will run simulations with the WirelessCar dataset. The maximum charging demand and redundancy, as well as the mean, standard deviation and standard error, will be calculated for 20 samples. In order to see how the demand value changes over times, we will present a demand over time graph showing the mean demand for a selected hex for each hour during 24 hours. Finally, a sensitivity analysis will be made to understand how the different variables effect the result of the prototype.

4
Results

In this chapter we present the proposed model, using the factors and parameters described in the previous chapter. There are two primary factors in the model – the available supply in an area, and the current demand of the vehicle(s) that are parked there. These two factors will be described separately before presenting the results of prototyping.

4.1 Area Supply Modelling

The area supply is expressed as the couple (e, n) where e is the total effect of charging poles in the area expressed in kWh, and n is the number of available charging stations. n is the sum of two terms, $n = n_{pub} + n_{priv}$, where the former is the amount of charging stations represented in relevant data sets. n_{priv} is an estimate of privately available charging poles based on the area categorization. The estimate may vary on how the categorization is done and what categories are considered. For some simple assumptions on different area categories, refer to Appendix B.

For an area, we define two sets of charging poles, N_{pub} and N_{priv}:

$$N_{pub} = [pub_1, pub_2...pub_k]$$

$$N_{priv} = [priv_1, priv_2...priv_n]$$

For N_{priv}, private charging poles, we assume them have the effect of 3.7 kWh. The effect of public charging poles, N_{pub} is known. The total effect in an area is thereby gained through:

$$e = \sum_{i \in N_{pub}} f(pub_i) + \sum_{j \in N_{priv}} 3.7$$

Where the function $f(pub)$ given a public charging pole pub returns the effect value of that charging pole.

4.2 Area Demand Modelling

Vehicle demand in an area is expressed as a couple (d, r) where d is the energy demand expressed in kWh and r is the redundancy, i.e. the excess charging of vehicles parked in the area expressed in kWh. Demand is defined as the sum of demand of the individual vehicles in the area at the input time t. Redundancy is similarly defined as the sum of redundancy of individual vehicles in the area at the time t.

4.2.1 Vehicle Demand Modelling

A vehicle v stopping in a bounded area A will contribute its demand at that point at time t to the total demand of the area. The demand d_v of a vehicle is the difference between the current charge left at the time of stopping in the area and the capacity c_{max} of the battery, adjusted for decay of age and temperature at that time. SoC is assumed to be given as a percentage of full capacity.

Battery capacity is influenced by temperature and age of battery, which may be expressed as a coefficient to the maximum capacity, c_{max} which is the theoretical capacity as stated by the vehicle manufacturer. The effective capacity c is given by

$$c = c_{max} * b_{age} * b_{temp}$$

where

$$b_{age} = (1 - 0.023)^y$$

where y is the age of the vehicle in years, and

$$b_{temp} = 0.65 + 0.02263276 * x + 0.0001167148 * x^2 - 0.00001703223 * x^3 - \\ 1.856662e - 7 * x^4 + 5.194805e - 9 * x^5$$

where x is the temperature at the time of the parking event expressed in degrees Celsius, $-25 < x < 40$.

The energy necessary m to completely recharge the battery for vehicle v at the time of the parking event may then simply be expressed as:

$$m = c - SoC * c$$

4.2.2 Average Area Parking Duration

Average area parking duration will mainly be used to model the redundancy in an area, and represents a simple estimate of how long a vehicle can be expected to stay in a particular area.

To begin with, fetch all electrical vehicle trips E from the data set of trips D:

$$E = \{t \in D : electrical\_consumption(t) > 0\}$$

Build the set of unique vehicles $V_{electric}$ in E, fetch all trips made by vehicles with electric engine T_e :

$$T_e = \{t \in D : vehicle(t) \in V_{electric}\}$$

We consider a specific area, in our case hex h, and find all trips $t \in T_e$ where all t_{stop}, the timestamp for when a trip ended, and the next t_{start}, the timestamp for when a trip started, are in h.

$$T_h = \{t \in T_e : t_{stop} == h \vee t_{start} == h\}$$

We define a function $time(t_{stop}, t_{start})$ that accepts a pairwise stop and start time and returns the length of this time span.

Lastly, the average time difference between all t_{stop} and t_{start} in hex h will be found and set as a variable o_h.

$$o_h = \frac{\sum_{i \in T_h} time(t_{i,stop}, t_{i,start})}{|T_h|}$$

4.2.3 Vehicle Charging Probability

The probability of a driver looking to charge a vehicle may be modelled as a function of the current SoC of a vehicle. Building this model on the findings of Smart and Shey [21], we express the final demand of a single vehicle to be $d_v = m * u(SoC)$, with $u(SoC)$ being the probability of a charging event taking place depending on the SoC at the time of the stop according to Table 4.1 below.

SoC %	Probability of Charging	SoC %	Probability of Charging
0-10%	0.04	50-60%	0.15
10-20%	0.06	60-70%	0.06
20-30%	0.14	70-80%	0.03
30-40%	0.23	80-90%	0.02
40-50%	0.25	90-100%	0.01

Table 4.1: Probabilities for a charge event to take place depending on the vehicle's current State of Charge.

4.2.4 Redundancy per Vehicle

Charging redundancy may be calculated in an area at any time a vehicle leaves an area. By knowing current energy missing for a full charge m, we can use the $time(t_{stop}, t_{start})$ function to get the length of a parking occurrence, and define redundancy w as difference between the theoretical energy a station could supply during the full parking duration and the missing capacity from the battery:

$$w = u(SoC) * (t_{park} * e - (1 - SoC) * c)$$

where $t_{park} = time(t_{stop}, t_{start})$ for the timestamps which mark the start and stop of the parking occurrence. Note that we use the probability of charge here as well.

For simplicity, we assume that all vehicles support all charging effects as redundancy is intended to model the theoretical charge that could have been gained from vehicles in the area if charging poles were unoccupied.

4.2.5 Area Specific Demand Reduction

For the purposes of this model, namely to estimate the current use and need of charging stations in specific areas, certain vehicles may not exert any general demand under some circumstances. The most common example would be residential areas with single-household housing – vehicle owners parked here overnight is likely to have a private charging facility for their own use, which would result in the demand of their vehicle in that particular area being 0. Formally, we can define a function for each area that given a vehicle returns either the standard demand d_v of the vehicle as described above, or for some specific vehicles returns some other value. For some example functions, refer to Appendix C.

4.3 The Model

Read previous sections in Chapter 3 for details on each step.

Algorithm 1: Pre-Process Area Data

Data:
Area which to apply model to.
divide_areas(): An function that accepts a total area as argument and returns a set of areas A after arbitrary division.
categorize_areas(): An function that accepts an area a and returns the arbitrary categorization of that area.
Result:
A: The set of all categorized areas
a: A single categorized area $a \in A$
begin
  Apply *divide_areas*(A) for division of total area into the set of arbitrary areas A.
end
for $a \in A$ **do**
  Apply *categorize_areas*(a) to give area a a categorization.
end

The result after running the full list of algorithms is a list of hexes with a supply and demand at a specific time t, as well as a redundancy value r. Areas with high difference $d - e$ and low redundancy r would be more likely to see efficient use of new charging stations for the given time t. Re-run model using different times t for making statistics over time.

Algorithm 2: Model Area Supply

Data:

A: The set of all categorized areas
N_a: The set of all public charging poles in area $a \in A$
e_n: The energy supply in a single charging pole $n \in N_a$

Result:

e_a: The total energy supply in area $a \in A$
$\bar{e_a}$: Average energy supply in a single charging pole in area $a \in A$

for $a \in A$ **do**
  $e_a \leftarrow \sum_{n \in N_a} e_n$
  $\bar{e_a} \leftarrow \frac{e_a}{|N_a|}$
end

Algorithm 3: Create List of Parking Events

Data:

A: The set of all categorized areas
$time(start_p, stop_p)$: A function that returns the total time between parking event p's start and stop time.
$u(SoC)$: A function modelling if a vehicle is in a charging state or not, returning 1 and 0 with probability dependent on SoC.

Result:

P_a: A sorted list of parking events and their properties for each area $a \in A$.
p: A single parking event in P_a that includes: {
  v_p: The vehicle that created parking event p.
  c_p: The battery capacity for vehicle v_p at time of parking event p.
  m_p: The energy needed to recharge battery in vehicle v_p at time of parking event p.
  d_p: The demand of vehicle v_p during parking event p.
  w_p: Redundancy
}
o_a: The average parking time in area $a \in A$.

for $a \in A$ **do**
  Create list P_a that contains every parking event in a.
  Sort P_a in ascending order based on time of event.
  for $p \in P_a$ **do**
    $c_p \leftarrow c_{max} * b_{age} * b_{temp}$
    $m_p \leftarrow c_p - (SoC * c_p)$
    $d_p \leftarrow u(SoC) * m_p$
    $w_p \leftarrow u(SoC) * (time(p_{stop}, p_{start}) * \bar{e_a} - (1 - SoC) * c_p)$
  end
  $o_a \leftarrow \frac{\sum_{p \in P_a}(time(p_{stop}, p_{start}))}{|P_a|}$
end

Algorithm 4: Model Area Demand

Data:

a: A single categorized area $a \in A$

P_a: A sorted list of parking events and their properties for area a.

t: A specific time.

$has\_relation(a, v_p)$: A function that returns true or false if a vehicle v_p has a relation to area a.

$specific\_demand(a, v_p)$: A function that returns a demand value for vehicle v_p that is tailored to area a based on the relation between them.

Result:

d_a: The energy demand in area a.

r_a: The charging redundancy in area a.

```
begin
    for p ∈ P_a do
        r_a ← r_a += w_p
    end
    P_t ← p ∈ P_a : p_start < t < p_stop
    for p ∈ P_t do
        if v_p not has_relation(a, v_p) then
            d_a ← d_a += d_p
        else
            d_a ← d_a += specific_demand(a, v_p)
        end
    end
end
```

4.4 Prototype

The above algorithms were implemented in a prototype using Python, alongside the pandas library for data processing. The prototype was limited to a few, cherry-picked areas in an effort to decrease the unknown variables, namely the efficiency and correctness of any area categorization method.

As such, the prototype was applied to three areas which the authors are able to manually categorize. The three areas are H3 hexagons which contain, respectively: a residential area in Askim, Göteborg (hex id 891f2504e13ffff), a supermarket parking lot (hex id 891f250693bffff), and the parking lot of an commercial office area (hex id 891f2515923ffff). These areas have been cherry picked for a couple of reasons, most importantly that the authors have prior knowledge of them, as in knowledge of expected amounts of traffic during different hours of the day and the currently available charging infrastructure. As the authors have easy access to the areas distance wise, it would be possible to visit them to collect real-life data if needed. The chosen areas are visualized on a map in figure 4.1. More details about the selection can be found in Section 5.5.

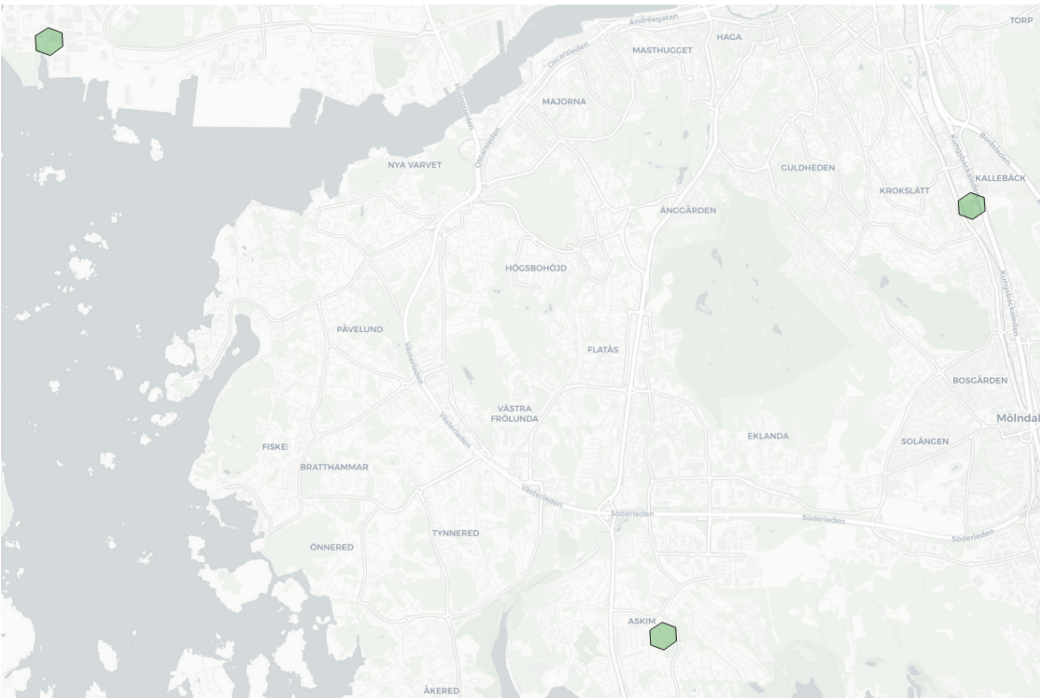

Figure 4.1: Map of the areas chosen for prototyping. The northwestern area is an office area, eastern area is a supermarket parking lot, and southern is a residential area.

The chosen areas are of H3 resolution 9, which translates to a size of approximately 0.1 km^2. Due to the nature of H3 hexagons, the areas do not exclusively contain the aforementioned parking lots, but the overwhelming majority of the area in which vehicles may be expected to stop is covered by said lots. Conversely, the residential area is big enough to require several H3 hexagons to cover, of which one completely nested in this area was picked arbitrarily.

As charging poles available in a certain area tend to be static compared to the vehicle traffic, their supply is calculated once based on the information found on public APIs, which in this case applies only to the supermarket area. The data regarding existing charging poles here are taken from the crowd-sourced API hosted att uppladdning.nu – the NOBIL API did at time of prototpying not contain any information regarding these charging stations, which were known to exist by the authors beforehand. Any implementation looking to either increase the covered area or look over longer periods of time should consume the API programmatically to check the supply at the relevant time, keeping in mind that charging pole APIs are not guaranteed to be complete.

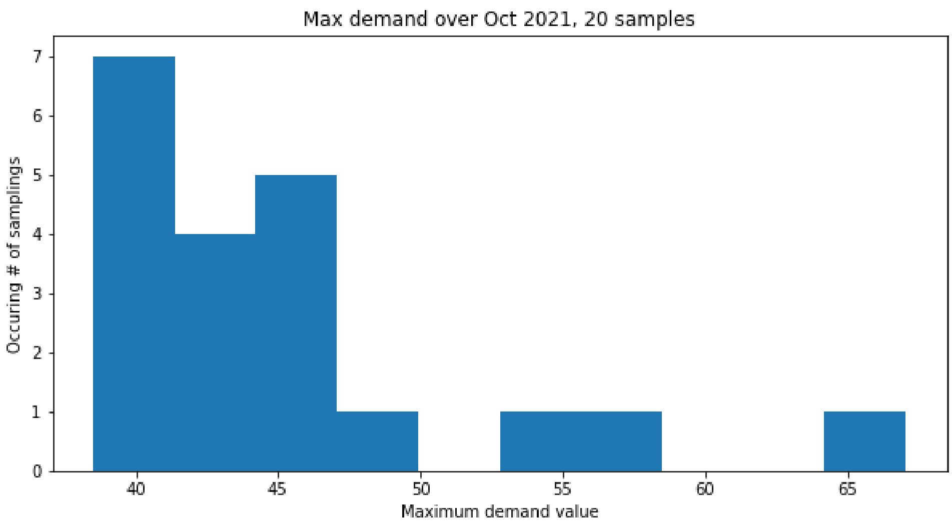

Figure 4.2: Histogram graph showing maximum electrical charging demand for the month of October 2021 in a supermarket parking area, 20 samples.

4.4.1 Maximum Calculated Demand per Sampling

The following histograms show the maximum value of demand as calculated by the model for 20 random samplings of SoC and the subsequent sampling of *u(SoC)* for the three chosen areas; a supermarket parking lot area, a residential area, and a commercial office area in Figures 4.2. 4.3 and 4.4, respectively.

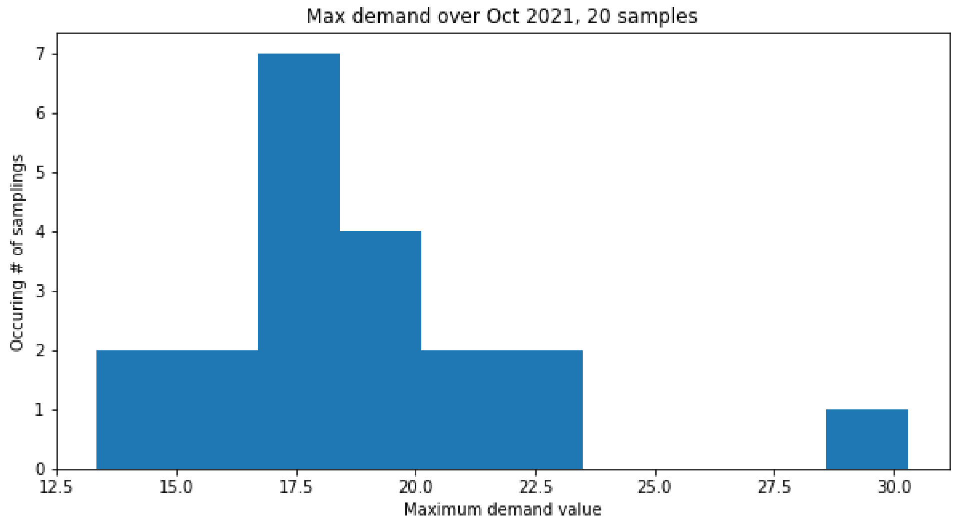

Figure 4.3: Histogram graph showing maximum electrical charging demand for the month of October 2021 in a residential area, 20 samples.

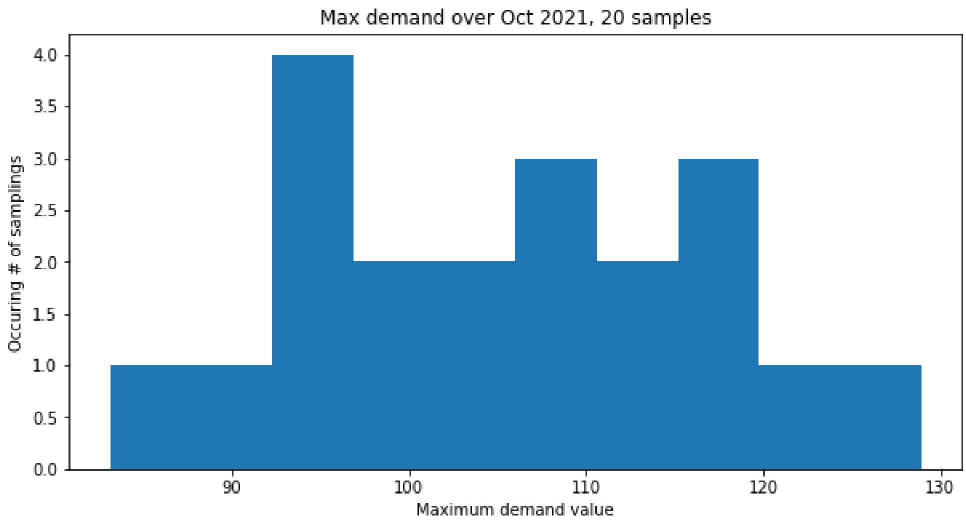

Figure 4.4: Histogram graph showing maximum electrical charging demand for the month of October 2021 in an office parking area, 20 samples.

Metric	Supermarket Area	Residential Area	Office Area
Mean	45.17	18.88	105.19
Standard Deviation	7.11	3.63	11.71
Standard Error	1.59	0.81	2.62

Table 4.2: Statistical measures (mean, standard deviation, and standard error) from sampling of maximum demand in three different areas (expressed in kWh).

For the samples visualized in the histograms, the mean, standard deviation and standard error were calculated and is presented in Table 4.2. The maximum value of demand tend to fluctuate between samplings, but all areas show consistency in terms of overall demand in relation to other areas in that demand tends to be highest in the office parking area and lowest in the residential area.

In the resulting graphs above, area supply is not included in the final result. This is for comparative purposes, as the evaluation of the output is made easier when considering the more fluctuating values of demand. For the chosen locations, the only one with existing charging poles (according to public information) is the supermarket area. These charging poles would total an effect 14.8kWh, meaning that the maximum demand in the supermarket area when also considering area supply should be decreased accordingly.

4.4.2 Maximum Calculated Redundancy per Sampling

The following histograms show the values of redundancy for the same sampling as for the demand results above. We can see that the office parking area shows the highest redundancy values, while lower redundancy values are found in the supermarket parking area and residential area. Both the office parking area and supermarket parking area there are a few outliers but mostly a consistent output, where the residential area have a larger spread of values with similar amount of occurrences.

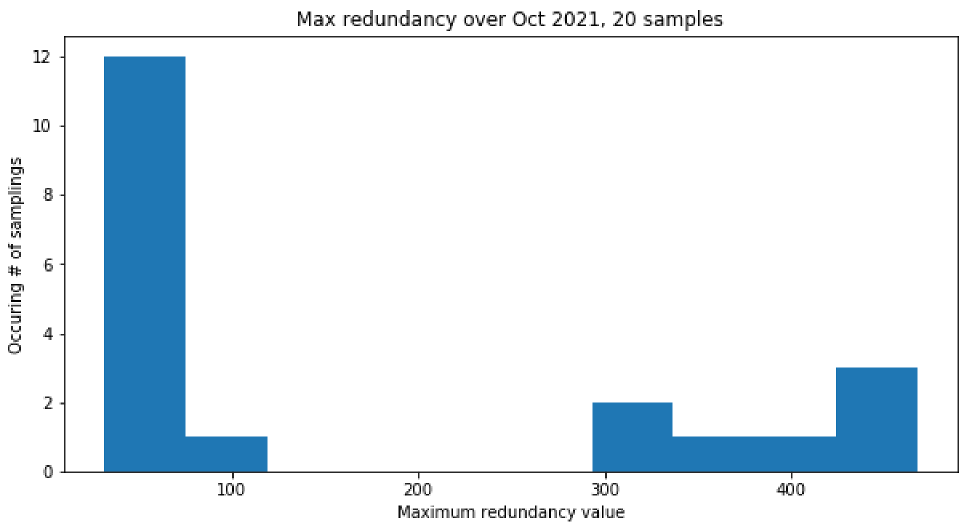

Figure 4.5: Histogram graph showing maximum electrical charging redundancy for the month of October 2021 in a supermarket parking area, 20 samples.

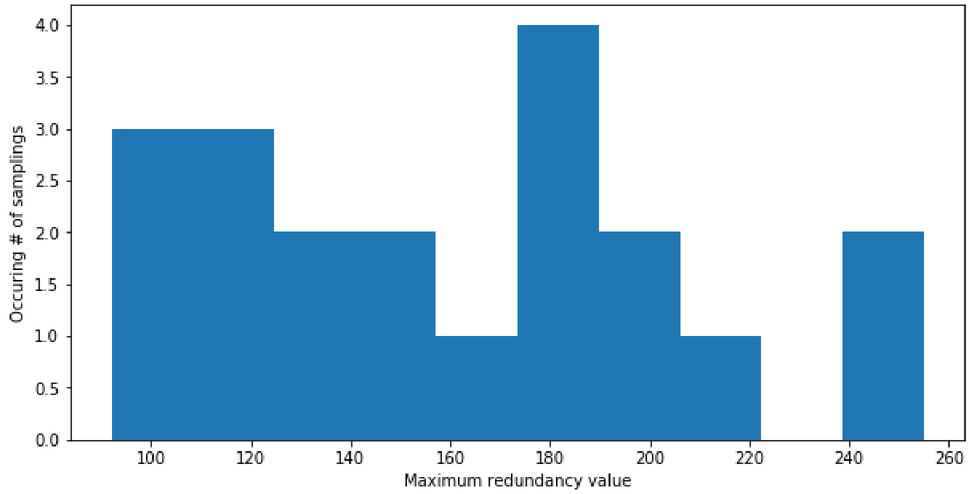

Figure 4.6: Histogram graph showing maximum electrical charging redundancy for the month of October 2021 in a residential area, 20 samples.

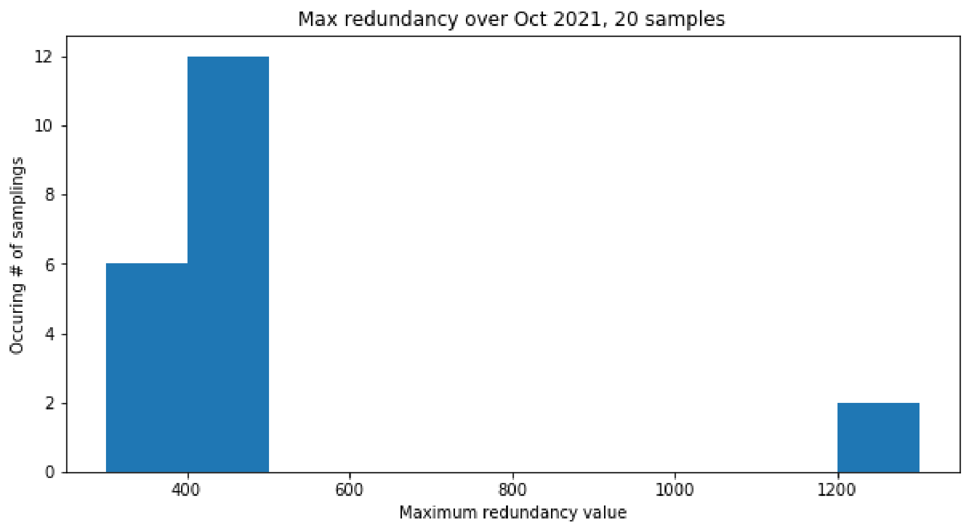

Figure 4.7: Histogram graph showing maximum electrical charging redundancy for the month of October 2021 in an office parking area, 20 samples.

Metric	Supermarket Area	Residential Area	Office Area
Mean	174.56	159.40	493.96
Standard Deviation	172.54	46.68	275.27
Standard Error	38.58	10.44	61.55

Table 4.3: Statistical measures (mean, standard deviation, and standard error) from sampling of maximum redundancy in three different areas (expressed in kWh).

For the samples visualized in the histograms, the mean, standard deviation and standard error were calculated and is presented in Table 4.3. For redundancy, the consistent ordering as seen in the demand is not present. Moreover, the standard error for redundancy results are also higher.

4.4.3 Demand and Redundancy over Time

As the previous histograms only show the maximum demand over the whole month of October 2021, we have chosen another option to present demand for the three areas. The three following graphs show the mean demand during every hour of October 20th 2021 for a supermarket parking area, a residential area and an office area.

The mean demand for the supermarket parking area can be seen to follow the opening hours 6.00-23.00, with some variances in demand over the day. For the residential area, the highest demand is during the evening and night, times when most people are expected to be at home. The demand is instead low during working hours, beginning at 6 and ending at 18. For the office area, demand is at its peak during working hours, here between 6 and 16, and very low during the hours of the night.

We find that the graph for the residential parking area and the office parking area resembles opposites of each other, which seems to be reasonable as most people leave their home to go to work and vice versa. Note that time zone is UTC format, resulting in a 1 hour difference from local time.

If we were to consider supply in the context of these results, one interesting observation to make is that the existing charging poles in the supermarket area would be perfectly sufficient to decrease the mean demand to 0 for the whole period. Again, the supply is omitted when visualizing the results in order to provide better grounds for evaluation.

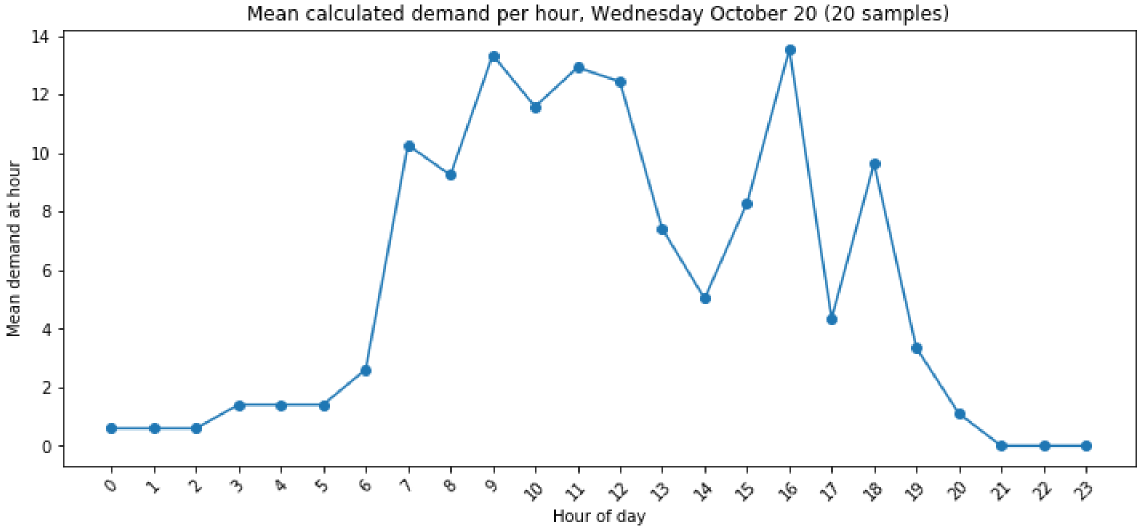

Figure 4.8: Line graph showing mean demand per hour for October 20 2021 in a supermarket parking area, 20 samples.

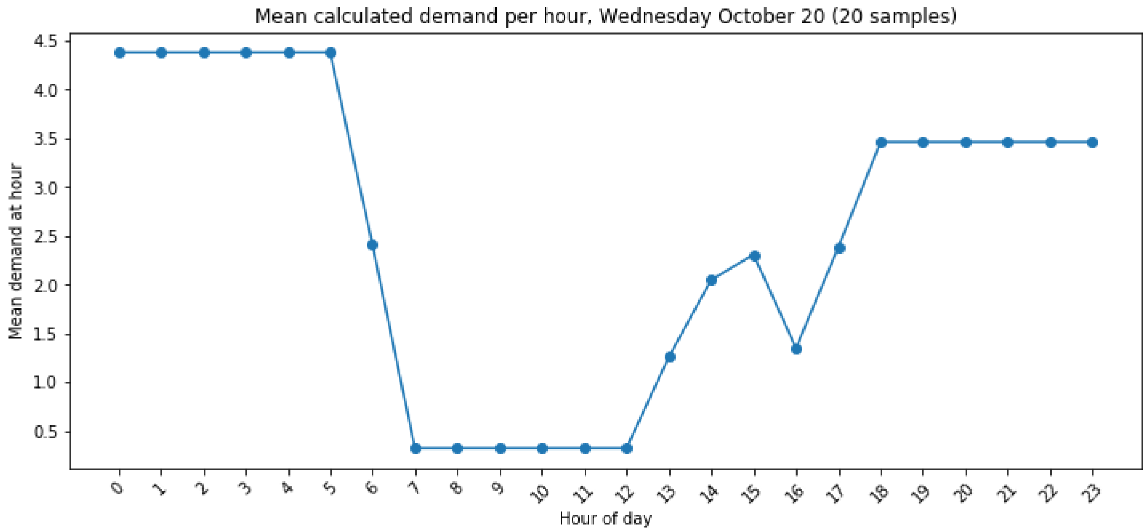

Figure 4.9: Line graph showing mean demand per hour for October 20 2021 in a residential area, 20 samples.

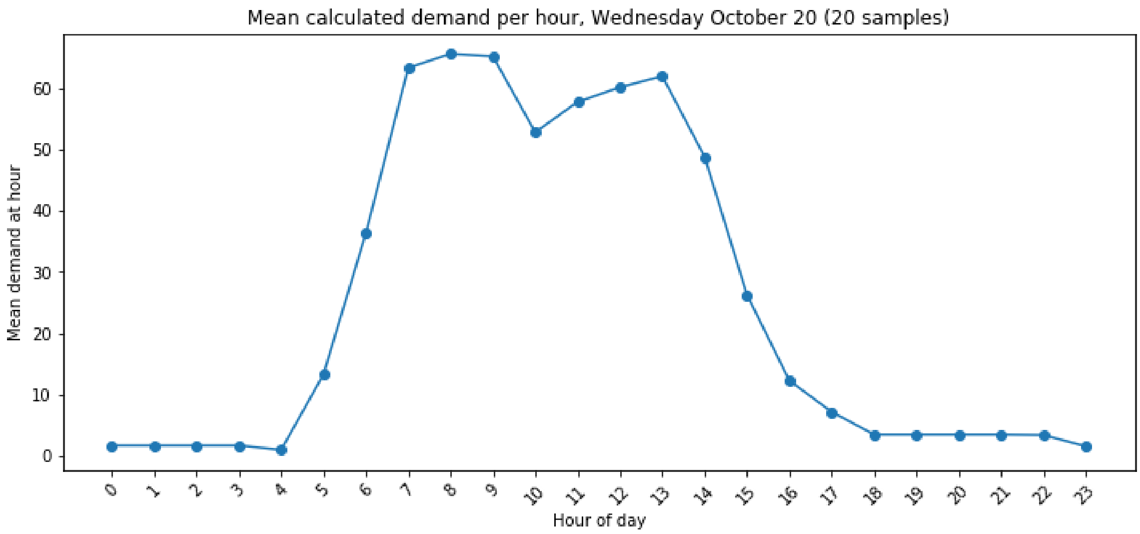

Figure 4.10: Line graph showing mean demand per hour for October 20 2021 in a office parking area, 20 samples.

4.5 Sensitivity Analysis

Other than the above results, a simple sensitivity analysis was performed, in which we observed the impact of sampled/estimated values on the final result of the model. Table 4.4 below outlines the difference of mean output when adjusting different input, one value at a time, in order to determine the influence of the unknown variables on the final result, with the result in this case being the mean max demand $\overline{d_{max}}$ and mean max redundancy $\overline{r_{max}}$ over 20 samples. Values for SoC are for the sake of the analysis given and manually adjusted, while *u(SoC)* will estimate a charging event taking place twice as often when under test. Trip records for the commercial space was used for this analysis.

Tested Input	**SoC**	***u(SoC)* Factor**	$\overline{\mathbf{d_{max}}}$	$\overline{\mathbf{r_{max}}}$
Baseline	60%	1	38.96	279.08
SOC	30%	1	194.5	866.75
u(SOC)	60%	2	173.28	1202.45

Table 4.4: Resulting outputs from the sensitivity analysis, done by artificially adjusting the inputs SoC and the *u(SoC)* factor one-by-one.

Both demand and redundancy outputs are, as shown in Table 4.4, heavily influenced by changes to SoC and *u(SoC)*.

5

Discussion

In this chapter we present a discussion about the presented results, the chosen methods, and the different data that has been used in the simulations. The choice of auxiliary functions, as well as ethical considerations are also considered. The chapter ends with a summary of the work done in this thesis.

5.1 Evaluation of Prototype Simulations

First and foremost, one can see that the demand follows the general flow of traffic one could expect to see in each area, in that commercial areas see employees arrive at morning and leave later during the afternoon, supermarkets being visited mainly during opening hours and increase after working hours, and people returning to home in residential areas. The output of the model does correspond to these traffic patterns, indicating some correlation with real-life scenarios – not completely unexpected as the trip data is sourced from real vehicles.

Moreover, demand can be seen (as expected) to be lowest in residential areas, and other areas to have a higher demand. This result is in line with expectations – partly because the model is intended to disregard vehicles whose owners live in the area and are expected to have private charging solutions, and partly because the amount of parked vehicles in a residential area likely is lower than on equally sized well-visited parking lots. In the same manner, one would expect people to be parked longer in office areas when disregarding overnight parking, which is reflected in the redundancy output.

The output redundancy in supermarket areas and office areas is somewhat more consistent than the residential areas when looking at the histograms in Section 4.4.2. Since many vehicles are being excluded in the latter area due to area-specific considerations, we believe that the redundancy variations are larger in the residential area because of the relatively few vehicles contributing to the demand and redundancy there, meaning that the sampling of *u(SoC)* heavily impacts the resulting output.

The sensitivity analysis indicates that both SoC and *u(SoC)* has a large impact on the output of the model. We argue that this shows that the inclusion of charge levels and charging behaviour are of high relevance and therefore a sound inclusion in any model using historical trip data in this manner. On the other hand, the large differences in output also tells us that the current assumptions and methods for estimating these values are likely to require further refinement in future work. We can imagine a few approaches for this purpose, which are discussed more extensively in Section 5.8.

Finally, an interesting result is that of the demand in the supermarket parking lot over the course of a single day. The demand is almost perfectly matched by the supply, meaning that according to model output, the area is currently sufficiently supplied with charging poles without any excess.

5.2 Reliance on PHEV Data

As mentioned in Section 1.3.1, only PHEV data is present in the data set. We believe that this does not affect the general theory behind the basic model, however – all parameters used in the model are data points that could easily be expected to be present in data sets sourced from BEVs as well as PHEVs. The data structure used by WirelessCar is expected to be relevant for combustion vehicles, fully electric vehicles, and everything in between, and data from fully electric vehicles are stored and utilized in this format already today. In other words, nothing from a technical standpoints restricts the model from being applied to BEVs, PHEVs, or a mix of both.

However, it would be prudent to look into possible ways to expand the model to consider PHEVs and BEVs differently. Right now, the model does assume all vehicles, or drivers rather, to be equal in terms of when they look to charge the vehicle given the same SoC. For BEVs, this may be somewhat true – the vehicle needs electricity to run, after all. For PHEVs, this is not necessarily the case. While owning a PHEV without using the on-board battery might not be a common occurrence, it is very simple to imagine different drivers may value charging opportunities differently. With the given format of the trip data, it would be possible to take this into account - given that a vehicle is a PHEV, taking a look at previous trips may give insight about how often the vehicle tends to be driven on electricity in contrast to combustion fuel.

One way to do this could be to consider the electricity consumption in relation to fuel consumption and model the proportions of the total distance driven using each. While feasible to do with the data present in the data set, there were also some caveats that resulted in the exclusion of the value. Longer trips made which by necessity required use of fuel could skew results heavily, for instance. While it is possible to consider such factors in the model, time restraints and computational times did limit us from finalizing a satisfactory method of PHEV consideration. Exclusion of this factor was thus deemed the reasonable, if somewhat unfortunate, choice.

To summarize the matter of PHEV data being the only available input, we did make a conscious decision to construct a model that was general enough to apply to all electrical vehicles, regardless of the presence of a combution engine or not. We would like to stress yet again that we do expect a different behaviour between drivers of the two variants, and that expansion and improvement of the model should take this into account; perhaps by looking into our proposed method above, a refinement thereof, or by some completely different method.

5.3 Area Categorization

As users worldwide work together to build the OSM database, it can lack completeness in some areas, while larger cities, such as New York City, have more accurate tags. An excellent example of this is residential areas, identified by their roads (*highway: residential*) and the buildings (*building: apartment, house*). When manually sampling data in known places in Gothenburg on OSM, there were several occurrences where apartment and house buildings had no tags that indicated this, instead only a general building tag (*building: yes*). Only having general tags makes the categorization of hexes more difficult, as there are significant differences between residential areas that could be of use in the model. An example is the number of available charging poles, as we assume that a driver in a single household house would install a personal charging pole while a driver living in an apartment would have to make do with the number of charging poles that the property owner decided to install.

If the tags in OSM were more complete (or if there was another data set for this purpose available), the area categorization could be given more thought and made more complex. The current model simply categorizes areas based on the most number of buildings present in that area, but a improvement would be to extend categorization even further. An example would be residential areas with single-household housing, where we expect that if a vehicle belongs to a hex, i.e. it is parked there more than 50% of the hours in a day, then the vehicle has a private charging pole that supports it. However, as the tags for buildings, especially residential, are lacking in quality, it is not possible to tag an area in more depth than simply residential without manually looking at the buildings in the area or having a personal knowledge about it.

Another possibility would be to give different type of tags different weights. The benefit would pertain to parking spaces in particular, as some parking spaces are tagged individually. Imagine an area where you have three large apartment buildings that each are ten stories high, and withing the same area lies the accompanying parking lots with one parking spot per two apartments. If these parking spaces were individually tagged, the total amount would far surpass the three apartment building tags, which would skew the categorization for the area. Given a situation like described, giving individual parking spaces a lower value than 1 could help combat this and ensure a better categorization.

Nevertheless, we have chosen to not put any extra focus on making the area categorization better as this is not the primary objective of the project but rather an aid to the proposed model. Since we are manually picking out areas to run simulations on, we ensure that the areas picked we have personal knowledge of and can vouch for their categorization to be correct. If the model was to be used in a real-life scenario, the area categorization would be a component requiring some care in order to be useful; improvements could be done either by applying a more correct categorization method or improving the data set, or by limiting the use of the model to in areas which could be categorized manually as in the simulation.

Area categorization does not need to be limited to the different categorized as shown in this work – the chosen variations of areas is mainly used to provide an example of how driver behaviour may differ between certain areas and how this feature of the model attempts to include these differences. It would be reasonable for any application of this model to adjust area categorization based on prior observations and data of the modelled area when striving for higher accuracy.

5.4 Area Division

The usage of H3 hexes in our prototype and model is not a necessity – much like the area categorization, area division may be done in a way that fits the applications. In our case, H3 hexes are an easy and portable alternative to more complex geofences, which could be used to more accurately covering areas of interest and, if done with enough detail, simplify area categorization by virtue of dividing an area into a single-purpose area, such as enclosing a residential area in a geofence and leaving out areas used for any other purpose.

The area division does come with some caveats, however – in particular, we noticed that it is not uncommon for vehicles to report differing positions for two sequential trips, something that is may very likely be attributed to GPS inaccuracy. This causes some problems when a starting trip positions happens to be reported outside the boundary of a modelled area, as a single vehicle would be expected to start from the same area as it previously stopped, barring some very special circumstances. There are some ways to get around this – for our simulation, we opted to filter out any sequences of start/stop actions so that any given vehicle "stop" event in a area would also be followed by an corresponding "start" event. Another option could be to not consider the location of a starting position for a vehicle and simply use the latest stopping positions, assuming no movement at time of engine off. Whichever method is chosen to alleviate the issue, it is likely that any complex area division needs to consider erroneous GPS positioning when defining boundaries.

5.5 Prototype Area Selection

As previously mentioned, 3 areas were cherry picked to use in the implementation of the prototype. This was mainly done to be able to have an understanding if the results were feasible, as choosing an area of size 0.1 km^2 in Sweden at random would leave us with little room for analyzing the results without first learning more about the area itself. As only about 3% of the total area in Sweden is measured to be built-up land (i.e. not forests, agricultural land and similar), the chance of choosing a urban area at random is very small [22]. We decided to omit the previously mentioned specific area categorization from this process, as any method of area categorization could be used, and the model itself was the main thing to be tested. Therefore, we proceeded to look at the whole city of Gothenburg to find suitable areas.

Before selecting specific areas, we decided that we wanted to include 3 different types of areas in the selection, as we expected the traffic patterns in these to differ from each other. First, we wanted a residential area and a workplace area, to see if the traffic in these would mirror each other, as people leave home to go to work and vice versa. We also wanted to find a commercial area, the primary challenge here was selecting an area with a broad scope of visitors. The area chosen was a supermarket parking area, which was deemed as a good choice as everyone needs to buy food, while a spa, for example, would have a much smaller clientele. It should be mentioned that the supermarket is a large store located in central Gothenburg, which may mean that most of the shoppers there live in the vicinity or passes by on their way home or to work. This would be true for most areas, and could be a factor that should be taken into account if wanting to consider socio-economic differences between areas.

To make the final selection of areas, we looked for hexes where the bounded area consisted of as much of the type of relevant area as possible. Some of the other areas considered were not chosen due to them being split into two or more hexes. This could of course affect the results of prototyping, as the intended target areas did not perfectly match the hexagon boundaries, but we considered the areas to be relevant and well-defined enough for our purpose. Constructing our own geofences could certainly help alleviate this issue.

If there was more time available, we would like to have several different areas of the same type and make comparisons between them. Another possibility would be to "build" areas from sets of higher resolution hexagons. In the interest of maintaining the same size of compared areas, we opted to simply pick single hexagons – as a part of evaluating the results, we wanted to see if areas of similar size would result in differing demand dependent on the expected level of traffic in each such area, without the need of considering size differences.

Overall, there are a set of improvements and considerations to be made for further prototyping. We deemed the few handpicked areas a good start to begin looking at the result, and argue that the chosen areas allow us to compare traffic over time and intensity of said traffic as described in the beginning of the chapter. Furthermore, these areas allowed us to verify the existing supply and, if necessary, visit them.

5.6 Available Data

In this section, we discuss the merits of the data available to us, as well as other data points that could prove useful but are non-existent within the provided data set. Refer to the appendix for complete lists of available data used for prototyping.

While the data set is proprietary data, the company handles several brands and opts to use the format described in the appendix to store vehicle trip data in a general format. This indicated that it is not unreasonable to assume that vehicles tend to be able to report their position and the time of which such readings are made – the data required is in that regard quite basic and the model should therefore be usable for any actor with access historical vehicle data, as the non-vehicle related data is either open or available as a service offering.

We have previously mentioned that the data set contains data from the last three months only, as older data is removed due to privacy reasons. Having only recent data to analyze provides both benefits and disadvantages. A benefit with using recent data is that it raises the confidence about the data correctly representing the current state of the areas in question, as many changes can happen in three months time. Owner of vehicles can change jobs, or they can move, and thereby change their vehicle movement patterns and areas where they add to the demand.

Using solely recent data could also be a disadvantage; by only considering the vehicle movements during the last three months we increase the risk of presenting a model that is accurate during one part of the year, but highly inaccurate during the rest. An example of this could be a parking area near a popular beach. During the summer months, the parking spaces can be expected to be filled to the brim with vehicles during the whole day. The months leading up to summer and the early fall months may also have a higher frequency of parking events. But during the winter months, we can probably expect there to be close to no vehicles parked there at all.

With this in mind, an idea to possibly work around this is to run the model as often you see fit, with the latest data as input, and save each result in order to collect historical data. This could, for example, be done once a month with the result being saved for a year until it is updated with the new result from the latest data. This ensures that there is historical data over all the months of the year, even though the input data from each of the months may not be available, which in turn will take seasonal differences in parking events into account.

5.7 Public Charging Infrastructure Data

A difficulty that we encountered early on in the project was the lack of official Swedish data on publicly available charging poles. We opted to use the public NOBIL database since it had an API that the data could be accessed from. NOBIL is the product of a cooperation between Norwegian Electric Vehicle Association and Enova, a state-owned norwegian company that brings financial aid to environmentally friendly actions. Another reasonable data source could be Open Charge Map, a crowd-sourced service in the like of Open Street Map, which actually lists NOBIL as a data provider. However, the problem with these sources are that they lack data about some existing charging poles, and that they depend on the public to add this data. To add a charging pole, you need to supply information about the exact location in coordinates, as well as the available charge types and more, meaning that the person needs to put in some effort.

We experienced the issue about missing charging poles for one of our three chosen areas, the supermarket parking lot, where we personally know that there are a number of charging poles installed and working, and has been for at least a year prior. The supermarket is one of the largest in the Gothenburg area, so if these charging poles have not been registered in the named databases, how can we expect there to be a complete registry of available charging poles in less frequently visited places? We did, however, find that at least one provider (uppladdning.nu) had the charging poles at the supermarket listed, and opted to use this data instead. This issue exemplifies the importance of having complete and correct data for the model, but as the adoption rate of EV's increase, it should be expected the quality of these API's to improve with time. It might also be possible that there exists other sources that are more complete, publicly accessible or otherwise.

All in all, for a more extensive use of the model, the source of data for public charging infrastructure needs to be thoroughly evaluated to ensure that the data is correct and complete.

5.8 Choice of Auxiliary Functions

As one might notice, there are several occurrences of simplified functions and assumptions used in the model. In general, these functions and features are handled this way as the scope of the project does not grant the time required for complex modeling of all these aspects. With the main goal being to explore and propose uses of the vehicle data for the purpose of evaluating charging infrastructure, these auxiliary functions are one of the ways the model may see improvement.

5.8.1 State of Charge

As we have mentioned in Section 2.1, there are several data points that are collected by the vehicle but not available in the dataset. The most important example of this is the vehicles SoC. For an EV, the SoC metric is equivalent to a combustion vehicles fuel level metric. As the Soc value is not available, we are currently giving vehicles a value by random sampling from a normal distribution. Since there are important components of the model that are dependent on the SoC value, having the true SoC value would be a significant improvement for the model.

With the SoC value unavailable and requiring an estimate, we would like to propose some future options for improving said estimation; the normal distribution is after all a very naive solution, but more elaborate methods would most likely require a fairly intricate model in and of itself which put these options out of scope for this project. We can suggest two different approaches; SoC could be more reasonable estimated either through the development of a more intricate distribution for this purpose, which could be used to more accurately represent the patterns of EV drivers, or through the use of the electricity consumption data available for each trip.

In the first case, looking at the work of Quirós-Tortós et al., we see the development of a statistical model for some key metrics of EV use [18], creating Gaussian Mixture Models (essentially combinations of several normal distribution) in order to represent the charging seen in real-life scenarios. The obvious caveat in attempting this method is that it would most likely require correct SoC readings to begin with to allow for its creation. Regardless, if such a representation was to be developed, the naive normal distribution used for sampling a SoC in our prototype could be replaced with a (hopefully) more accurate distribution.

A second possible way of more accurately estimating the SoC of a vehicle would be to use the historical data for this purpose, as the electrical consumption per trip of a vehicle is included in the data. The major disadvantage in doing this is of course that the historical trip data in itself does not grant any knowledge of when the vehicle recharges the batteries, still requiring some baseline estimates to be made about the SoC. However, these estimates could be made more limited. An example of how to do so would be to assume that any PHEV parked overnight in a residential area is fully charged in the morning; the assumption is that the driver parks on private property with access to a outlet, and that the capacity of the PHEV is low enough for it to be likely to be fully charged overnight even with slow charging. By looking at the electrical consumption over the course of the day, it might be possible to draw some conclusions based on the actual data; if the total consumption of electricity for a given day is larger than the capacity, then the vehicle must have been charged at least one, for instance. Similarly, if a vehicle were to report no electrical consumption for a trip, it would be reasonable to believe that the battery was empty when starting the trip.

Finally, the most viable solution would be to just include SoC in the historical data. For this particular data set, we know that SoC is actually a metric that is sent to WirelessCar today. It is not included in this work as it is saved as a snapshot value which is not put in a historical context (i.e. the trip data). So in theory, the data is present to include SoC in trip entries going forward, which would eliminate the need for assumptions and estimations of SoC. It would not be unreasonable to assume that vehicles using electrical propulsion in some manner are able to report their SoC in the context of trips, regardless of manufacturer – the suggested way forward for producers and data processors alike would be to simply arrange for this addition.

5.8.2 Charging Probability Density

While the probability of a vehicle beginning a charging session given a certain SoC tends to resemble a normal distribution, the main points of improvement regarding this function is twofold: the resolution of the discrete probability used in the prototype could be higher, and there could be separate distributions depending on if the vehicle in question is a BEV and PHEV. It would be reasonable to assume that a driver of a PHEV could at times forgo charging for convenience while this is not a possibility for a BEV driver. It would then be reasonable to believe that PHEVs have an increased probability of charge at lower SoCs rather than the middling values. Given available data for this disparity, the model could be adjusted to account for this. Even more so, this could be considered on a vehicle-per-vehicle basis.

Similarly to SoC, *u(SoC)* would not require estimation with the right data available – both vehicles and charging stations alike are able to report their status given a time, but unlike SoC the relation to trip data is not as apparent. As the parking events we use are a derivative from trip data which could be enriched with a starting SoC or ending SoC, charging events happen when the vehicle is not in motion and would either require some work to pair with trip data as it is saved today, or beginning to log parking events as well.

If the SoC is made available there is another option – inferring charging events using the difference in SoC between trips. If vehicles were to report their SoC at both start and end of trips, this could generally indicate if charging takes place or not. There could be some edge cases where the vehicle has only charged for some part of the parking duration and used battery for e.g. heating after that which would result in false negatives, but in general a trip starting with higher SoC than the same vehicle had at the last stop would indicate that charging took place when the vehicle was parked.

5.8.3 Alternate Charging Probability Considerations

After evaluation of the prototype, a need for a different model for charging probability has become apparent. The main issue is that the *u(SoC)* as it is used on the model is deriving a probability of charging from trends of driver charging behaviour. The issue becomes clear when we consider an example case. Assume a driver of a BEV with a SoC of 5% SoC. The probability of the driver charging the vehicle should be expected to approach 1 as the SoC continues to approach 0, as the vehicle is simply inoperable at 0% SoC.

By looking at trends of when drivers are most likely to charge, we do not consider the probability of an individual driver who is running low on charge. The charging probability distribution should therefore be changed to reflect this behaviour. We may argue that such a distribution should approach a charging probability of 1 as the SoC approached 0 as alluded to above, and that the probability of charge at high SoC should be significantly less, even more so when taking into account that good battery maintenance involves avoiding unnecessary charging. The issue is further exacerbated when we consider that PHEVs and BEVs likely experience different charging patterns. While studies exist that show general trends of EV usage (like the one that the original *u(SoC)* is based on), we have been unable to find data that accounts for individual drivers. Any alternate probability distribution for this purpose should be based on such data. Such data may be collected by surveying PHEV and BEV owners. Nevertheless, we now consider this to be one of the main usages of reporting SoC on trip beginning and end – it would allow for this problem to be bypassed as there would no longer be a need for a probabilistic model at all, instead inferring charging events from increased SoC after a parking event.

5.9 Ethical Considerations

The primary ethical consideration of this work and the potential outcome is that of privacy. While consent of data collection is given by owners already today, that consent likely does not extend to the usage of historical vehicle movement analysis, and as such, any future implementation of the developed method should require consent from drivers to have their data collected. Of course, the development of the method should be done with the intent that all data used should consist of anonymous data aggregates and should not require the exposure of any particular individual.

Similarly, we have worked with anonymised data for the development of the prototype and simulations as well. While the data has been anonymised, we have kept the possibility to differentiate unique vehicles. This is solely due to the functions *specific_demand()* and *has_relation()*, which both require the ability to identify a unique vehicle. These functions are tied to residential areas, and does not currently affect any other types of areas. This also means that if area categorization was done differently, or if the model was used to investigate commercial areas only, the need to differentiate unique vehicles would not exist.

Since aggregate data is preferable from a privacy standpoint, some care should also be taken in chosen which area to apply the model to – it is desirable to have a reasonable sample size to work with, even if vehicles are not able to be identified directly. This is due to the fact that drivers habits and schedules should not be able to be identified, which means that in order to apply the model one should strive to include at least some variety in order to obscure individual drivers.

5.10 Summary

In this thesis we have developed a mathematical model for calculating the electrical demand and redundancy in a given area and time. A prototype of the mathematical model was created and historical vehicle movement data provided by WirelessCar was used as input for the relevant areas. In prototyping the proposed model for three different areas, a residential area, a office area, and a supermarket parking area, the results show that the output follows expected traffic patterns in said areas, while still being heavily impacted by our proposed additions to the existing data. The most important additions of auxiliary data was State of Charge and Charging Probability Density, and results from a sensitivity analysis on these parameters showed that changing these input values had large effect on the output.

As the two aforementioned values pertaining to the state of the battery are estimated using stochastic variables, the resulting output between samplings varies. Because of the impact these parameters have, more sophisticated methods of estimating them would be the best way to further improve on the model. There is in fact feasible methods for determining these values without the need to estimate, making the acquisition of the true State of Charge value the best way forward for any future prospects of improving the model.

Bibliography

[1] Power Circle. *Elbilsstatistik*. 2021. URL: https://www.elbilsstatistik.se/elbilsstatistik (visited on 11/30/2021).

[2] Power Circle. *Laddinfrastrukturstatistik*. 2021. URL: https://www.elbilsstatistik.se/laddinfrastatistik (visited on 11/30/2021).

[3] Yagang Zhang et al. "Optimal Allocation of Changing Station for Electric Vehicle Based on Queuing Theory". In: *PROMET - TrafficTransportation* 28 (Oct. 2016). DOI: 10.7307/ptt.v28i5.1974.

[4] Bo He and Yushuo Hou. "Research on Estimation Method of the Balance Density of Electric Vehicles Charging Points". In: *2017 IEEE International Conference on Energy Internet (ICEI)*. 2017, pp. 273–278. DOI: 10.1109/ICEI.2017.55.

[5] The European Commission. *Communication from the Commission on the EU Strategy for a Sustainable and Smart Mobility*. 2020. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0789.

[6] Dario Pevec et al. "Electric Vehicle Range Anxiety: An Obstacle for the Personal Transportation (R)evolution?" In: *2019 4th International Conference on Smart and Sustainable Technologies (SpliTech)*. 2019, pp. 1–8. DOI: 10.23919/SpliTech.2019.8783178.

[7] Uber. *H3: A Hexagonal Hierarchical Geospatial Indexing System*. 2021. URL: https://github.com/uber/h3 (visited on 11/17/2021).

[8] Sveriges Riksdag. *Trafikförordning (1998:1276)*. 1998. URL: https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/trafikforordning-19981276_sfs-1998-1276.

[9] GeoTab Charlotte Argue. *What can 6,000 electric vehicles tell us about EV battery health?* 2020. URL: https://www.geotab.com/blog/ev-battery-health/ (visited on 12/29/2021).

[10] GeoTab Charlotte Argue. *To what degree does temperature impact EV range?* 2020. URL: https://www.geotab.com/blog/ev-range/ (visited on 11/27/2021).

[11] EUR-Lex. *Directive 2014/94/EU of the European Parliament and of the Council of 22 October 2014 on the deployment of alternative fuels infrastructure*. 2014. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:02014L0094-20211112&qid=1640638937702&from=EN.

[12] European Alternative Fuels Observatory. *Ladda din elbil hemma*. URL: https://www.eafo.eu/countries/european-union/23640/npf-targets-and-progress (visited on 12/28/2021).

[13] Sveriges Regering. *Styrmedel för laddinfrastruktur*. 2021. URL: https://www.regeringen.se/49bb4b/contentassets/3c895fca1e1641ff8591e6ec1d6ad996/sou_2021_48_del_2.pdf (visited on 12/28/2021).

[14] Skatteverket. *Godkända arbeten – grön teknik*. 2021. URL: https://www.skatteverket.se/privat/fastigheterochbostad/gronteknik/godkandaarbetengronteknik.html (visited on 12/27/2021).

[15] Lucien Mathieu, European Federation for Transport and Environment AISBL. *Recharge EU: how many charging points will EU and its Member states need in the 2020s*. 2020. URL: https://www.transportenvironment.org/discover/recharge-eu-how-many-charge-points-will-eu-countries-need-2030/.

[16] Stockholms Handelskammare. *Elbilar på frammarsch*. 2020. URL: https://stockholmshandelskammare.se/sites/default/files/2021-04/Rapport_Elbilar%20%281%29.pdf (visited on 12/28/2021).

[17] Energi- och Klimatrådgivningen i Stockholmsregionen. *Ladda din elbil hemma*. URL: https://energiradgivningen.se/ladda-din-elbil-hemma/ (visited on 12/27/2021).

[18] Jairo Quirós-Tortós et al. "Statistical Representation of EV Charging: Real Data Analysis and Applications". In: *2018 Power Systems Computation Conference (PSCC)*. 2018, pp. 1–7. DOI: 10.23919/PSCC.2018.8442988.

[19] Alan Millner. "Modeling Lithium Ion battery degradation in electric vehicles". In: *2010 IEEE Conference on Innovative Technologies for an Efficient and Reliable Electricity Supply*. 2010, pp. 349–356. DOI: 10.1109/CITRES.2010.5619782.

[20] Arijit Guha and Amit Patra. "State of Health Estimation of Lithium-Ion Batteries Using Capacity Fade and Internal Resistance Growth Models". In: *IEEE Transactions on Transportation Electrification* 4.1 (2018), pp. 135–146. DOI: 10.1109/TTE.2017.2776558.

[21] John Smart and Stephen Schey. "Battery Electric Vehicle Driving and Charging Behavior Observed Early in The EV Project". In: *SAE International Journal of Alternative Powertrains* 1 (July 2012). DOI: 10.4271/2012-01-0199.

[22] Statistikmyndigheten. *Marken i Sverige*. 2021. URL: https://www.scb.se/hitta-statistik/sverige-i-siffror/miljo/marken-i-sverige/ (visited on 03/06/2022).

A

Appendix 1

This appendix describes the format of the data used in this thesis, in the same groupings as it is gathered and stored today. It follows no particular standard as it is proprietary data.

A.1 Datapoints of Vehicle Configurations

A *vehicle configuration* data object contains some basic information about the vehicles performing the trips in the data set.

A.1.1 id

A unique identifier for the vehicle.

A.1.2 yearmodel

The year of make of the vehicle. Not necessarily the year it was made - model years start early and are often one year ahead of the current year (meaning year model 2022 may be manufactured already in 2021, for instance).

A.1.3 gearboxid

A simple id for gearbox type. Not used in the model.

A.1.4 gearboxtype

Readable description of gearbox type, e.g. 'Automatic'. Not used in the model.

A.1.5 colourid

Internal identification number of colour of vehicle. Irrelevant for this work.

A.1.6 colourname

The name of the colour denoted by *colourid*. Irrelevant for this work.

A.1.7 vehiclehandle

An identifier for the vehicle. Usually equal to the VIN (vehicle identification number), but for this work the data is hashed for privacy purposes.

A.1.8 country_iso2

Two-letter country code of the vehicle's country of registration.

A.2 Datapoints of Trips

A *trip* data object is a top-level data container that contains some very basic information about the trip. The data in these objects are not overly informative, but serve as a root of trip-related data. The data points in these object are as follows:

A.2.1 id

A unique identifier for the trip.

A.2.2 category

A category of the trip - trips may be classified as private trips or business trips, or as "unassigned" if not specified.

A.2.3 starttime

UTC Timestamp denoting start of trip. In practice, the point of time in which the engine is turned on.

A.2.4 endtime

UTC Timestamp denoting end of trip. In practice, the point of time in which the engine is turned off.

A.2.5 vehiclehandle

An identifier for the vehicle. Usually equal to the VIN (vehicle identification number), but for this work the data is hashed for privacy purposes.

A.3 Datapoints of Trip Details

Trip details are supplementary data points connected to a trip. It contains several details regarding the location and fuel consumption of the trip and is the main data object used for analysis in this project.

A.3.1 id

A unique identifier for the trip details. This is not the same id as for the trip data outlined above (see trip_id).

A.3.2 trip_id

The identifier of the trip which this data object pertains to.

A.3.3 startlongitude

The longitude coordinate of the trip starting point, with 6 decimal precision.

A.3.4 startlatitude

The latitude coordinate of the trip starting point, with 6 decimal precision.

A.3.5 endlongitude

The longitude coordinate of the trip ending point, with 6 decimal precision.

A.3.6 endlatitude

The latitude coordinate of the trip ending point, with 6 decimal precision.

A.3.7 starttime

UTC timestamp denoting starting time of trip, precise to the second.

A.3.8 endtime

UTC timestamp denoting starting time of trip, precise to the second.

A.3.9 startodometer

Odometer reading from the vehicle at time of departure. Expressed in meters.

A.3.10 endodometer

Odometer reading from the vehicle at time of arrival. Expressed in meters.

A.3.11 fuelconsumption

Total consumption of liquid fuel over the course of the trip. Diesel or gasoline. Expressed in centiliters.

A.3.12 electricalconsumption

Total onsumption of electricity (for propulsion purposes) over the course of the trip. Expressed in Watt/hours

A.3.13 startstreetaddress

The address from which the trip started, as given by the Google Reverse Geocode API when provided with starting coordinates.

A.3.14 endstreetaddress

The address on which the trip ends, as given by the Google Reverse Geocode API when provided with starting coordinates.

A.3.15 startregion

The region in which the trip started, as given by the Google Reverse Geocode API when provided with starting coordinates.

A.3.16 endregion

The region in which the trip started, as given by the Google Reverse Geocode API when provided with ending coordinates.

A.3.17 startcity

The city in which the trip started, as given by the Google Reverse Geocode API when provided with starting coordinates.

A.3.18 endcity

The city in which the trip started, as given by the Google Reverse Geocode API when provided with ending coordinates.

A.3.19 starttime

UTC Timestamp denoting start of trip. In practice, the point of time in which the engine is turned on. Same timestamp as in the trip data object.

A.3.20 endtime

UTC Timestamp denoting end of trip. In practice, the point of time in which the engine is turned off. Same timestamp as in the trip data object.

A.3.21 startpostalcode

The postal code of the area from which the trip begins.

A.3.22 endpostalcode

The postal code of the area in which the trip ends.

A.3.23 startregion

The region in which the trip started, as calculated by the Google Reverse Geocode API when provided with trip starting coordinates.

A.3.24 endregion

The region in which the trip started, as calculated by the Google Reverse Geocode API when provided with trip ending coordinates.

A.3.25 startiso2countrycode

Two-letter ISO country code of the country in which the trip begins.

A.3.26 endiso2countrycode

Two-letter ISO country code of the country in which the trip ends.

A.3.27 electricalregeneration

How much electricity is regenerated by vehicles over the course of the trip. Value is 0 for all non-HEVs.

A.3.28 tripwaypointsnbrof

Number of saved waypoints as part of the trip. A waypoint is a timestamped longitude-latitude pair that corresponds to a location the vehicle passed at some point over the course of the trip. Used for re-tracing the route taken. Actual waypoint data not provided as part of this project and stated to not be used for all vehicle models.

A.3.29 tripwaypointsminlongitude

The minimum longitude of all trip waypoints, if any.

A.3.30 tripwaypointsmaxlongitude

The maximum longitude of all trip waypoints, if any.

A.3.31 tripwaypointsminlatitude

The minimum latitude of all trip waypoints, if any.

A.3.32 tripwaypointsmaxlatitude

The maximum latitude of all trip waypoints, if any.

A.3.33 tripwaypointsnbrof

The number of waypoints logged as part of the trip.

B

Appendix 2

This appendix lists a few naive suggestions on how to estimate the number of private charging stations in an area depending on what category the area belongs to. In this case, the OpenStreetMap Overpass API was used to categorize areas and the suggestions are thus based on the features used in this data set. More sophisticated or complex estimations may be used in the model, given that as there for each category exists some function that estimates the number of non-public charging stations in the area. If the modelled area is known, then this modelling may be superfluous and adjusting according to knowledge of the area instead.

B.1 Amenity: Parking

An area designated as a parking area is assumed to mainly consist of (public) parking lots. The estimated number of private charging stations in these areas are 0 - it is assumed that charging poles in these areas are public and exists in the data set of public charging stations.

B.2 Land use: Commercial

An area is designated as commercial if there mainly exists offices and non-consumer company facilities within the area. The number of private charging stations in these areas are likely to vary, but a naive way to estimate charging stations is to start with estimating the number of parking spots in the area. This may be done by calculating the average number of *all* vehicles parked in the area at weekdays at 10.00; it is assumed employees have arrived at this time and have yet to leave for lunch. Then, we can make a very rough estimate of the proportion of parking spots that have charging stations installed. In Sweden, a resolution from 2020 states that newly produced non-residential buildings with connected parking lots require charging outlets for at least one fifth of the spaces, suggesting that the current proportion of such spaces is significantly smaller than 20%. A naive guess would put our suggested estimate at 5%, meaning that we would assume there to be charging opportunities for the same number of vehicles. Of

B.3 Land use: Industrial

An area is designated as industrial area if the area mostly consists of factories and warehouses. The same estimate as for commercial land is used here.

B.4 Land use: Retail

An area with mainly shops and stores. For these areas, we assume 0 non-public charging poles – we expect parking areas for customer use to be present in nearby areas, and any charging stations in these areas to be public.

B.5 Highway: Residential

Residential areas are where people generally live – for the areas covered by the prototype, they generally correspond to single household building and will be used as such in this example. The amount of non-public charging stations in this area is estimated as 0 - not because no charging opportunities exist, but because the charging possibilities of house owners in the area is better modelled as a decrease in demand for some particular vehicles rather than an increase in supply, as the existing supply is tied to the respective houses (and the vehicles that tend to be parked there) and not expected to be available to anyone in the area.

B.6 Building: Apartment

Apartment areas (or rather, areas with mainly apartment buildings) and their multiple households are difficult to estimate, as parking spaces may be either dedicated or public. For public spaces, it should be expected that charging poles are available in place. Similarly to residential areas, dedicated spaces are not expected to be accessible to everyone, and thus any existing charging points on those spaces are better modelled as a demand decrease for specific vehicles, described more in appendix C.

B.7 Building: House

An area mainly filled with this feature is a residential single household area, see section B.6.

B.8 Building: Office

An area mainly filled with this feature is a commercial space, see B.2.

C

Appendix 3

This appendix list a few suggestions on how to model a vehicle-specific reduction of demand in specific areas. Functions could be specific to particular areas, or apply generally to a category of areas as exemplified in B. In the following examples, we define functions that apply to areas with certain categories only, along with a default function for any areas not of the aforementioned categories.

C.1 Residential Areas

In areas categorized as residential (both for single houses and apartment), vehicle demand will be using the default model for demand unless the vehicle spends more than 50% of its total parked time in this particular area, indicating that the owner lives in a building in the area. It may then be assumed that the owner has access to private charging station which is likely dedicated to that particular vehicle, with the motivation that it is not economically feasible to buy an EV without access to reliable charging today. 50% is a rough estimate of the time an average person spends at home in general. For vehicles matching these criteria, their demand d_v for these particular areas are 0 for the purpose of the model.

C.2 Default Demand by Area

Default demand in an area applies to all vehicles in area other than residential areas as described above, and to vehicles in residential areas spending less than 50% of their parked time there; in other words, simply $d_v = m * p(SoC)$.