AARHUS UNIVERSITY
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

A Gravity-Based Data-Driven Approach for Site Selection of EV Charging Stations

by

Magnus Trøjgård Tang, 201904915

Bastian Aron Kramer, 201908709

Masters Thesis

in

Computer Engineering

Supervisor: Christian Fischer Pedersen

Co-Supervisor: Mads Jensen

Aarhus University

Department of Electrical and Computer Engineering

June 4, 2024

Abstract

The transition towards electric vehicles (EVs) is a critical component of Denmark's strategy to reduce carbon emissions in the transportation sector. This thesis explores the strategic siting of public EV charging stations to support the growing demand for charging infrastructure. Current methods for locating EV charging stations often lack the granularity required to address the specific needs of different areas. To overcome this limitation, a novel approach using a modified gravity-based model is proposed. This model evaluates potential locations based on multiple factors, including geospatial data, traffic density, and socio-demographic information.

The core of the model is the gravity equation, which calculates the attractiveness of a location by assessing the interaction between the location and surrounding amenities. A weighted approach, optimized using a genetic algorithm, determines the significance of each amenity. The model's performance is evaluated by comparing the score assigned to locations with observed hours charged at existing charging stations. The results indicate that the model struggles to capture the actual distribution of observed data, partly due to limitations in the data used in the project.

The model's geospatial features are integrated into a hybrid forcasting model to look into the possibility of predicting the use of charging stations at new locations without historical data. The results show that spatial features can capture parts of the underlying patterns, but struggle to produce accurate predictions without historical data.

Overall, this thesis contributes to the strategic placement of EV charging stations by providing a data-driven, granular evaluation method based on a gravity model.

Resumé

Overgangen til elbiler er vigtig del af Danmarks strategi for at reducere CO2-udledninger i transportsektoren. Denne afhandling udforsker den strategiske placering af offentlige ladestationer for at understøtte den stigende efterspørgsel efter ladeinfrastruktur. Nuværende metoder til placering af ladestationer mangler ofte den præcision, der kræves for at kunne evaluere forskellen på lokationer i nærliggende områder. Derfor presenteres en ny tilgang til at løse dette problem, ved at bruge en modificeret tyngdebaseret model. Denne model evaluerer potentielle placeringer baseret på flere faktorer, herunder geospatiale data, trafikdensitet og socio-demografiske oplysninger.

Kernen i modellen er tyngdeloven, som beregner en lokalitets attraktion ved at vurdere interaktionen mellem lokaliteten og de omkringliggende faciliteter. En vægtet tilgang, optimeret ved hjælp af en genetisk algoritme, bestemmer betydningen af hver facilitet. Modellens nøjagtighed evalueres ved at sammenligne den attraktion, der tildeles lokaliteter, med observerede ladetimer på eksisterende ladestationer. Resultaterne indikerer, at modellen har svært ved at fange den faktiske fordeling af observerede data, delvist på grund af begrænsninger i de anvendte data i projektet.

De geografiske træk ved modellen er integreret i en hybrid forudsigelsesmodel for at undersøge muligheden for at forudsige brugen af ladestationer på nye placeringer uden historiske data. Resultaterne viser at de geografiske træk kan fange dele af de underliggende mønstre, men ikke er i stand til at producere nøjagtige forudsigelser uden historiske data.

Samlet set bidrager denne afhandling til den strategiske placering af EV-ladestationer ved at tilbyde en datadrevet, granulær evalueringsmetode baseret på en tyngdemodel.

Acknowledgements

First and foremost we would like to thank our main supervisor Christian Fischer Pedersen for his guidance and support throughout the project. His expertise and guiding hand have been much appreciated and highly valued. We would also like to thank our co-supervisor, Mads Jensen, for his valuable input and feedback throughout the project. An additional thank you to Norlys for providing data and expertise that were used as a foundation for the project. Lastly we would like to thank Emil Hu, who has worked on a project in parallel with us, and has acted as a sparring partner throughout the project.

Contents

Chapter 1

Introduction

In Denmark, transportation is recognized as a pivotal sector in the national strategy to transition towards a more carbon-efficient economy. The transportation sector accounts for a significant portion of the country's carbon emissions, prompting targeted initiatives aimed at reducing its environmental impact [1]. The Danish government has committed to ambitious carbon reduction targets, with a clear focus on enhancing the sustainability of transport through the accelerated adoption of electric vehicles (EVs). This shift necessitates a corresponding expansion in EV charging infrastructure to support the growing number of EVs on the road. Specifically, public charging stations are essential to ensuring the widespread adoption of EVs, as they facilitate long-distance travel and provide a solution for those who lack the option of private charging facilities. However, the development of such infrastructure cannot rely solely on reactive measures like demand response, as this introduces a 'chicken and egg' dilemma where the availability of charging stations and the adoption of electric vehicles are mutually dependent [2]. To address this challenge, it is essential to investigate the possibility for more strategic approaches through decision support models to aid the placement of charging infrastructure.

This thesis aims to contribute to a strategic solution by exploring the potential of using a gravity based model to evaluate locations for new charging stations at a high granularity. The model will be designed to incorporate geospatial awareness along with other relevant factors, with the goal of using the pre-existing infrastructure and socio-demographics to evaluate new locations for charging stations. Through this research, we aim to contribute to the strategic, greener, and more efficient infrastructure necessary to support Denmark's transition to electric mobility.

1.1 Motivation

The rapid adoption of EVs in Denmark is outpacing the development of necessary charging infrastructure, creating bottlenecks that could slow the transition to sustainable transportation [3]. Using a strategic approach for placing new charging stations allows for a productive expansion of the charging infrastructure, by ensuring geographical coverage and optimal utilization of resources. Yet, there remains a lack of standardized methods for evaluating locations for new charging stations, which leaves the majority of the decision-making process to be based on intuition.

The motivation behind this project is to develop a solution that can support the expansion of EV charging infrastructure in Denmark. By using the gravity-based model, the project aims to explore the possibility of both evaluating locations for new charging stations based on geospatial awareness, and predicting the use of newly placed stations. This approach will allow for a strategic and data-driven expansion of the charging infrastructure in Denmark.

1.2 Structure

The remainder of this thesis is structured in the following manner: First additional background information is provided in Chapter 2, followed by an outline of the problem statement in Chapter 3. Existing literature related to the work in this thesis is reviewed in Chapter 4. The remaining chapters present the proposed solution by detailing the implementation of a gravity-based model in Chapter 7, showing the obtained results in Chapter 8. Finally, the process and results are discussed in Chapter 9, and the thesis is concluded in Chapter 10.

Chapter 2

Background

The background chapter will introduce the context of the project, by providing an overview of the EV market in Denmark, the current landscape of charging infrastructure and the future ambitions within the domain. The chapter will also outline the scope and limitations of the project, to give perspective on the focus areas and the constraints of the analysis. Finally, the chapter will present the main goals of the project, based on the hypothesis and motivation outlined in the introduction.

2.1 Landscape of Electric Vehicles and Public Charging Stations in Denmark

Denmark has set ambitious targets for the transition to EVs and the expansion of charging infrastructure. The government has committed to phasing out the sale of new fossil-fuel vehicles by 2030, and the Danish Energy Agency has set a target of 1 million EVs on the road by 2030 [4]. This is a significant increase from the current number of EVs and plugin-hybrid eletric vehicles (PHEVs) in Denmark, which stands at 11.4% of the total vehicle stock of 2.83 million, equivalent to approximately 322.000 vehicles. The number of EVs has been on an upward trajectory in recent years, with 36% of new vehicles sold in 2023 being EVs, compared to 21% in 2022 [5]. As such, the market for EVs in Denmark is growing rapidly, and the demand for charging infrastructure increases proportionally. The number of publicly available charging stations has risen to 21.000 in the first quarter of 2024, which is nearly a doubling compared to 11.000 in the first quarter of 2023. As a result the amount of EVs per publicly availible charging station has decreased to approximately 10.7 compared to 13.9 in early 2022 [6]. This indicates that the expansion of charging infrastructure is keeping pace with the growing number of EVs, but there is still a need for further development to ensure that the infrastructure is sufficient to support the increasing number of EVs on the road.

Norlys is one of the leading energy companies in Denmark, and has a strong focus on the transition to sustainable energy solutions. The company has set ambitious targets for the expansion of charging infrastructure, with the goal of providing a comprehensive network of charging stations [7]. The work in this thesis is part of a collaboration with Norlys, which have provided valuable insights towards gaining knowledge in the domain of current strategies to implement charging infrastructure. Additionally, Norlys has provided the necessary data to support the evaluation of the model proposed in this thesis.

2.2 Scope and Limitations

The following section will outline the general scope of this project, by presenting the main focus areas aswell as the limitations. The scope is defined to ensure that the project is manageable and that the analysis is focused on the most relevant areas. The limitations are defined to provide perspective on the constraints of the analysis, and to ensure that the results are interpreted within the context of the limitations.

This project focuses on the EV market in Denmark, with a particular emphasis on public charging stations. Therefore the study primarily concerns vehicles which are able to use public charging stations, such as EVs and PHEVs. The project will not consider other types of electric vehicles, such as electric scooters, busses or trucks along with other electric vehicles which are not able to use public charging stations. Note that even though these type of vehicles are not considered in the project, those that are will be referred to with the general term EVs.

The scope of the project is further limited to public charging stations, due to the availability of data. Important to note is that not all public charging stations in Denmark are incldued in the analysis. For further details on the data used in the project see Chapter 5. Private charging stations, such as home and company chargers, are not considered in the analsis, as the data provided by Norlys exclusively concerns public charging stations. Only a subset of the public charging stations in Denmark are included, however the subset contains stations across the country.

The geographical scope of the project is limited to Denmark, and the analysis has been conducted based on data from Danish sources or data that is descriptive of Denmark. While the project is focused on Denmark, the findings may be relevant for other countries with similar characteristics, and the methodology can be applied to other countries with similar availability of data.

Chapter 3

Problem Statement

The problem formulated in this project is to develop a method for evaluating locations for electric vehicle charging stations. The method should be able to evaluate the potential of any proposed location, as the area in which stations are deployed are not solely determined by Norlys. The process typically involves a list of potential locations, which companies within the market are able to bid towards. Therefore it is essential to have a strategic approach when selecting locations to bid on. From a dialog with Norlys, it is apparent that the decision-making process within the field is currently based on intuition and expert knowledge. This poses the need for a method to support the decision on which locations are more profitable than others. Therefore, the problem is not to optimize the placement of charging stations to meet an overall demand criteria, but to place the stations such that they capture the largest possible share of the demand. As a result, the method should be able to evaluate the difference between locations that may be in close proximity to each other. Additionally, the method should be able to provide an estimate of the utilization of a station at the proposed location, such that the decision-makers can make a cost-benefit analysis on the return on investment. As such it is also requested that the method is not a black-box model, but that the results are transparent and informed decisions can be made based on partial knowledge of the model. Finally, it is a wish that the results are presented in a visual manner, such that the results can be easily interpreted by the decision-makers.

3.1 Project Goals

Based on the problem statement, along with the findings from the introduction and background chapters, a set of goals have been defined for the project. The goals are used to guide the development of the project, and to ensure that the analysis is focused on the problem at hand. The goals are presented below:

RQ 1: Investigate the current techniques for determining the optimal location of charging stations, and identify if the techniques are applicable to the problem.

RQ 2: Gain knowledge within the domain, and use that knowledge to develop a model that includes spatial awareness based on available data.

RQ 3: Define a criteria for evaluating the model, and use appropriate optimzation techniques to optimize the model towards the criteria.

RQ 4: Ensure that the model is transparent, and that the results are presented in a visual and interpretable manner.

RQ 5: Compare the method to other techniques within the field, and discuss the advantages and disadvantages of the proposed method.

Chapter 4

Related Studies

The following chapter will present an overview of existing research regarding site selection for electric vehicles. First an outline of a selection of existing research on optimal location for electric vehicles will be presented, followed by an overview of site selection in the domain of facility location research. Finally, a discussion of the relation of the existing research to the work in this thesis will be held.

4.1 Outline of existing research on siting EV charging stations

The field of siting EV charging stations is a relatively new research area, leading to a limited amount of existing research on the topic. However, a broad range of approaches have been used to determine the optimal placement of charging stations. Most studies have formulated the problem as a matter of optimization. The objective of optimzation ranges from minimizing the cost of deployment, optimizing distribution based on grid impact, or minimizing the distance to nearest charging stations for EV drivers. An outline of a selection of existing research on siting EV charging stations is presented below, focusing on studies which have contributed thoughts to the work in this thesis.

A study by Andrenacci et al. [8] conducted a study that explores a demand-side approach for the optimal placement of charging stations. They employed a fuzzy k-means clustering method to minimize the distance of all trips to the nearest centroid, with the final centroids representing the optimal locations for charging stations. The demand for each cluster was calculated based on the number of trips ending in the area, allowing them to determine the necessary infrastructure for each region. Thus, this study uses a data-driven approach to identify both the optimal placement and the estimated demand for charging stations in different regions. Zeb et al. [9] tries to find the optimal placement of charging stations based on the cost associated with deployment, reducing the passive loss of energy, and reducing congestion on transformers. The problem is solved by using Partical Swarm Optimization (PSO) to find the optimal combination of level 1, level 2 and level 3 charging stations. They find that the cost of deployment can be reduced when considering all three types of charging stations, compared to only optimizing the placement of level 3 charging stations. A paper by Rodrigues et al. [10] mixes the use of demand forecasting and demand capture to enhance decision making and resource allocation. The demand forecast uses socio-demographic data to approximate actual sales figures in order to predict demand at specific locations. Two different distance decay functions are used to spread the approximated demand. A gravity model is used to estimate the demand capture, by calculating the likelihood that a demand point will be captured by a location based on the size of the charging station, and the distance to the demand point. The results show that socio-demographics can be used to predict demand, and that the gravity model can be used to estimate the demand capture. It is proposed that future research should focus on the use of regression models to predict demand.

4.2 Inspiration from facility location research

The field of facility location research is closely related to the problem of finding optimal locations for EV charging stations. Optimal facility location is a more visited research area, and has been studied in various contexts such as retail store placement, warehouse location, and healthcare facility locations. Karamshuk et al. [11] uses geospatial- and mobility features to estimate the value of a location for retail stores. The value is based on the number of visitors to the location of three large brand retail stores, Dunkin' Donuts, Starbucks, and McDonald's. The study measures the density of nearby venues, to examine areas where the attractiveness of drawing customers is high. The mobility features is used to estimate the movement of people in and around the geographical area. The information is used to identify the density of people in the area, the movements between different venues and the migration of people into the area. A supervised learning model is applied to estimate the value of a location based on the geospatial and mobility features. The model was tested on two configurations, one using only the geospatial features, and another incorporating the mobility features. The results showed that the model using both types of features outperformed the model using only geospatial features.

Other studies have used GIS-tools to determine optimal site selection across various topics including site selection for solar power plants [12], wind farms [13, 14], and rainwater harvesting systems [15]. The studies use GIS-tools to analyze spatial data, while formulating the problem as a multi-criteria decision-making problem. Different methods are used to rank the importance of the criteria, such that important criteria are weighted more heavily. Erbaş et al. [16] adopted this approach for siting EV charging stations in Ankara, Turkey. A set of criteria was determined with the help of domain experts. In total 15 subcriteria were found and grouped into three main categories: Geographical, Urbanity, and Economic. A significant number of the subcriteria were related to minimizing the distance to other sources, i.e. the distance to roads, substations, petrol stations and maximizing distance to vulnerable areas such as vegetation and water sources. The subcriteria where ranked by the domain experts, and the criteria were weighted using a fuzzy Analytic Hierarchy Process (AHP). The criteria were then used to rank the suitability of locations for

EV charging stations. A similar approach was used by Rane et al. [17] where multi-influence factors (MIF) were used to determine the weight of the criteria. Similar for both of these studies is that the weight of each criteria is calculated on the base of relative importance determined by domain experts, and are not data driven.

In summary, the existing research can be grouped into several categories. First is the demand-side approach, where the location of charging stations are determined based on estimated demands in different areas. Demand can be estimated using a variety of methods, such as looking at the population of traffic in the area, using socio-demographic data or sales figures. Second is the optimization approach, where the location of stations are determined based on deployment cost, grid impact, and minimization of residual energy loss. Finally are the studies using a multi-criteria decision making approach in combination with geospatial data to determine a weighted criteria optimum.

4.3 Discussion of the relation to the work in this thesis

The field of siting EV charging stations is a relatively new research area, leading to a limited amount of existing research on the topic. Most of the studies date only one decade back meaning that alot of the research is still in an exploratory phase where the optimal methods have not yet been discovered or recognized as state-of-the-art. As a result, a lot of time has been spent looking in a broad range of directions, which has lead to much time spent exploring solutions that do not fit directly with the purpose of this thesis. However, a number of the works mentioned above have contributed with methods and domain knowledge that have served as a basis for modelling the presented solution. Given the domain recency most of the data used in the existing literature has been statistically generated, and serve therefore only as an approximation of the real world. Therefore, ongoing research and data collection is important for advancing this field and developing more accurate and effective solutions.

Chapter 5

Data Collection and Processing

This chapter outlines the data collection and processing steps taken to gather the necessary information for the project. The process involves understanding the data sources, collecting relevant datasets, and preparing the data for inclusion in the proposed model. The data collection process is an important step in the project, as the quality and relevance of the data directly impacts the accuracy and reliability of the results, given the data-driven modeling approach of the solution. The chapter is divided into sections, each focusing on a specific dataset or data source, detailing the data collection process, describing the data and the steps taken to prepare the data for analysis. Finally, each section will include the reasoning for including the data in the project and how it contributes to the overall solution. Initially, section 5.1 will present traffic related data, followed by section 5.2 which will present the timeseries data of charging station occupancy provided by Norlys. Section 5.3 will present data regarding the status of EVs across municipalities, and finally section 5.4 will present geospatial data regarding amenities and infrastructure relevant to formulating demand for EV charging infrastructure.

5.1 Traffic Data

To understand the potential demand for electric vehicle charging stations, it is essential to analyze the traffic patterns and the distribution of vehicles across different regions. This analysis is performed to identify areas with high traffic volumes, where the demand for charging infrastructure is likely to be significant. The data used for this purpose has been gathered from an open data source provided by "Vejdirektoratet", which is an instance in the Danish Ministry of Transport [18]. The data is collected from a series of traffic monitoring stations, strategically positioned to capture a comprehensive snapshot of vehicular flow across various locations. The dataset is sampled primarily by "Vejdirektoratet", and is supplemented by local contributions from municipalities across the country. The dataset contains information about the Annual Average Daily Traffic (AADT). AADT is a measure reflecting the average number of vehicles passing a given point daily, over the span of a year. The entire dataset comprises 38.000 records and contains data dating back to 2003. For the purpose of this project only the most recent up-to-date measurements are desired. As a result, the data has been filtered such that only the most recent measurements from 2023 are included. Additionally, the dataset includes both readings of motor vehicles, bicycles and scooters, however, for the purpose of this project only data related to motor vehicles is considered. The filtering has reduced the data used in this project to 5042 records. The data is depicted in Fig. 5.1 which represents the geographic distribution of the sampled traffic measurements across Denmark.

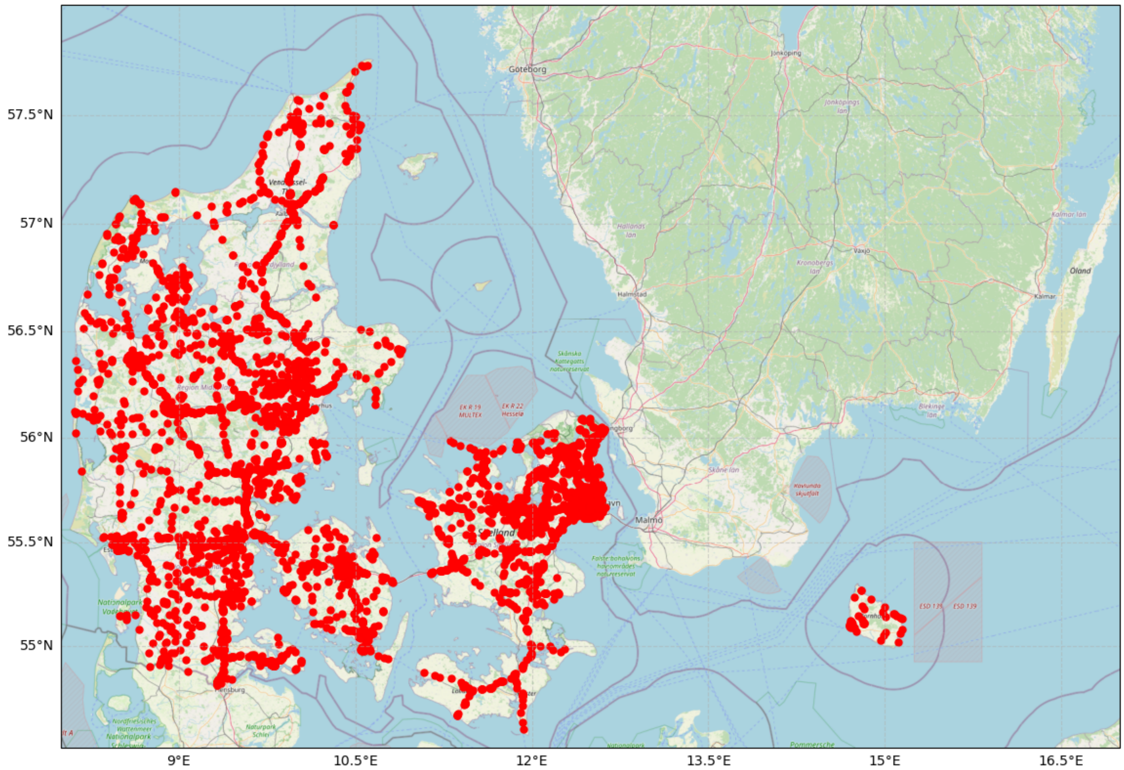


Figure 5.1: Geographical distribution of traffic measurements across Denmark.

The numerical distribution of the AADT values is depicted in Fig. 5.2 displaying a scatterplot of the AADT for each index in the dataset. Each index represents a unique location.

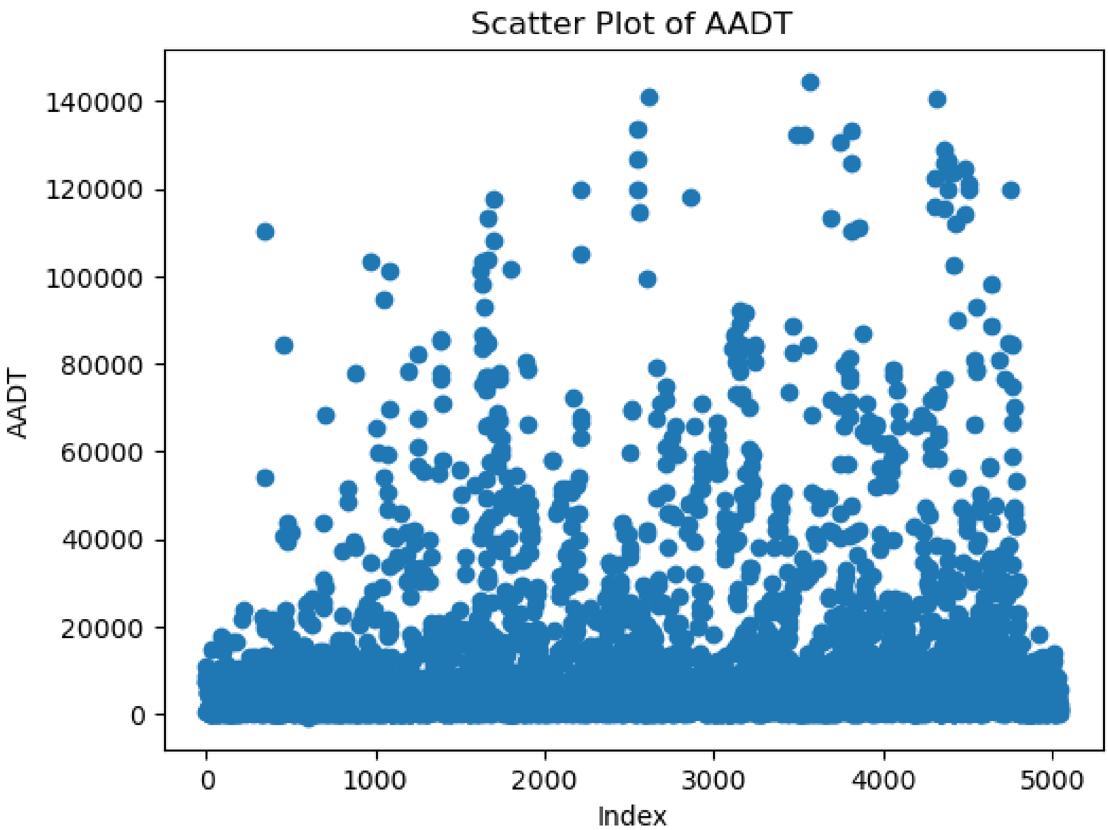


Figure 5.2: Numerical distribution of traffic measurements across Denmark. Each point represents a unique location.

The scatterplot indicates that the majority of roads experience a relatively low amount of daily traffic while few locations have substantially higher traffic counts. This matches expectations, as naturally more significant roads such as highways located at larger cities are central nodes in the traffic infrastructure, and as a consequence will experience a heavier traffic load. To understand the measured values on a statistical basis, Tab. 5.1 presents a number of statistical measures calculated from the dataset. Additionally, the data has been normalized to a range between the values of 0 and 1, as the relational magnitude between the measurements provides interpretable information.

The quartile measures provide a more granular understanding of the traffic distribution. The first quartile value is 1195.25 vehicles, indicating that a quarter of the locations average 1195 vehicles or less on a daily basis. The second quartile value, representing the midpoint of the data, shows 3689 daily vehicles. The third quartile escalates to 10056.5 vehicles, demonstrating that traffic volume increases substantially across higher populated locations. The highest observed AADT value is 144.464 vehicles, which is a drastic increase compared to the proportion between the first three quartiles. This information, including the standard deviation at 18638.41, indicates that the traffic distribution is skewed towards lower traffic volumes with a few locations experiencing significantly higher traffic counts. These AADT values are located along main highways near larger cities, where the traffic

	AADT	$\overline{\textbf{AADT}}$
count	5042.000000	5042.000000
mean	10578.235224	0.077795
std	18638.407663	0.129017
min	0.000000	0.000000
25%	1195.250000	0.008274
50%	3689.000000	0.025536
75%	10056.500000	0.069612
max	144464.000000	1.000000

Table 5.1: Statistical description of the Average Annual Daily Traffic data including a normalized representation.

infrastructure is more dense.

Fig. 5.3 aids the intuitive understanding of traffic density and distribution across different regions. The plot visualizes the magnitude of the AADT values, scaling the size of points according to its corresponding value.

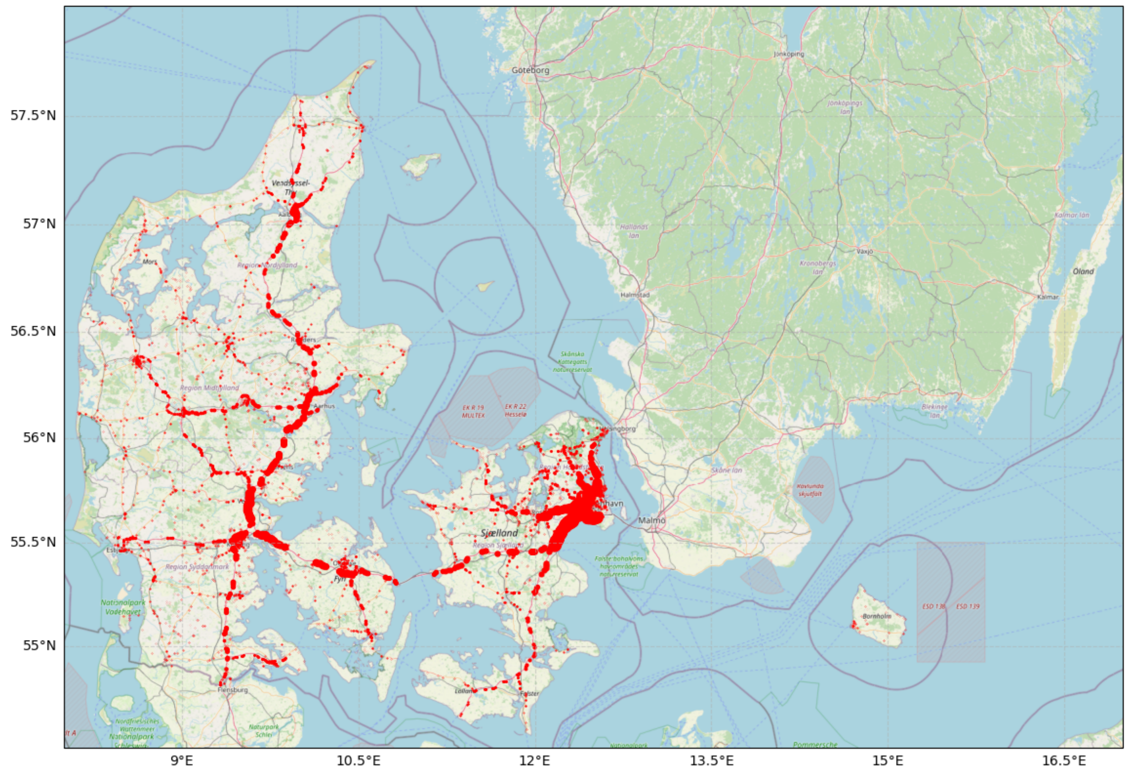


Figure 5.3: Magnitude of traffic measurements accross Denmark. The size of the points represents the magnitude of the AADT values.

The location of each point is stored in the data as UTM coordinates, which are a two-dimensional Cartesian coordinate system used to represent locations on the Earth's surface. A conversion from Universal Transverse Mercator (UTM) coordinates to World Geodesic System (WGS84) coordinates was performed to facilitate the integration of the traffic data for use in later stages of the project. The conversion was done using the Python library "pyproj" [19]. The conversion was necessary as all other data sources included in the project are stored in WGS84 coordinates, which is commonly expressed in latitude and longitude.

The aim of the project is to provide a prognose of the utilization of a charging station based on the location. The traffic data is used to identify traffic volumes accross the country, such that the model considers the volume of passing vehicles in near proximity to the location under observation. Thus, the demand associated to locations situated in an area with high traffic volumes is distinguishable from similar locations in areas with lower traffic volumes.

At a later stage in the project, the model is split into two separate models, with the aim of comparing only charging stations located in a similar geospatial context. This lead to the observation that stations situated along motorways were not able to obtain a value for traffic density. Therefore traffic measurements from motorways were distributed across all points labeled as motorways in the OpenStreetMap data explained later in this chapter. This was performed by using the KDTree algorithm to define a spatial binary tree of all points in the traffic data with an assigned value [20]. Each OSM motorway point was then assigned the value of the nearest point with a known value, essentially performing spatial nearest neighbor interpolation. Thus, every point associated with a motorway then include a value for traffic density.

5.2 Charging station occupancy data

Data which has been used as the target variable for the model is the occupancy of charging stations. A dataset was provided by Norlys, which included timeseries data of the occupancy of a number of charging stations across Denmark. In addition, metadata was provided for each station, which included information about the station's location, along with a number of other attributes. The dataset was provided preemptively to the start of the thesis, meaning that data cleaning and preprocessing was done in a research and development project leading up to this thesis. Therefore, a detailed explanation of the cleaning and processing steps taken will not be explained in this thesis, but the reader is referred to [21] for additional information.

A brief overview of the data is explained below, including an explanation of why and how this data is relevant to this thesis. As mentioned, timeseries data was provided for a number of charging stations across Denmark. The timeseries data was structured such that the occupancy of each station was recorded approximately every 15 minutes and ranges between the period: 13:33:13 August 10th 2023 to 11:16:03 October 11th 2023. Each record consists of a "slug", which is a unique identifier for the station. Each station consists of one or more "plugs", which represent the invidivdual outlets at the station. For each plug a sample includes the timestamp, the status regarding whether the plug was occupied, unoccupied, or out of order as consistent elements. Some stations, typically depending on the provider, include one or more additional attributes for each plug. The list of attributes which are potentially included are a tariff plan, a single pricing, or an attribute labeled "info" which describes the duration of current status, i.e. "Occupied since: 24 minutes". Two additional features have been created from the timeseries data. The first feature is the number of plugs associated with each station, and the second feature is the average daily occupancy of each station. The latter has later been used as an estimation for the average daily hours of use for each station.

The metadata provided included a long list of features, and thus not all will be mentioned here. Therefore only features relevant to understanding the work in this thesis are mentioned and the remaining are described in the same document as the exploratory data analysis. Relevant metadata include first and foremost the same unique identifier slug, such that stations could be mapped across datasets. Secondly, the geographical location of the station was described by latitude and longitude coordinates. Additionally, the minimum and maximum capacity of the station is used, along with the owning company.

5.3 Status of EVs across municipalities

This analysis seeks to quantify the growth of EVs within the broader context of vehicle registrations across Danish municipalities, providing insights into the penetration and growth trajectory of EVs in Denmark. The data is collected from Statistikbanken, which is a database provided by "Danmarks Statistik", the Danish national statistics bureau [22]. The data originates from the Danish Motor Vehicle Register, which records the number of registered vehicles in the country.

The database contains information on the number of vehicles registered in each municipality, segmented by vehicle type and time period. The collected dataset includes data dating back to 2018, and is updated monthly with the latest vehicle registrations. The data is structured such that each record represents the number of vehicles of a certain type registered in a specific municipality at the given time. Two sets of data is collected, one detailing registrations for all vehicle types, and the other focusing exclusively on EVs.

The analysis of EV registration data, involved a series of data processing steps to extract meaningful insights. First, the data was cleaned and structured to facilitate proper analysis. Once cleaned, the growth rate of vehicle registrations was calculated to understand the evolution of registered vehicles and the relation between all vehicle registrations and EV registrations. Fig. 5.4 and Fig. 5.5 illustrate the total number of registered vehicles and EVs, respectively, from January 2018 to January 2024. The plots provide a clear overview of the evolution of vehicle registrations over time. Notably, the graph illustrating registrations of EVs shows a significant upward trajectory, indicating a growing preference for EVs among Danish vehicle owners.

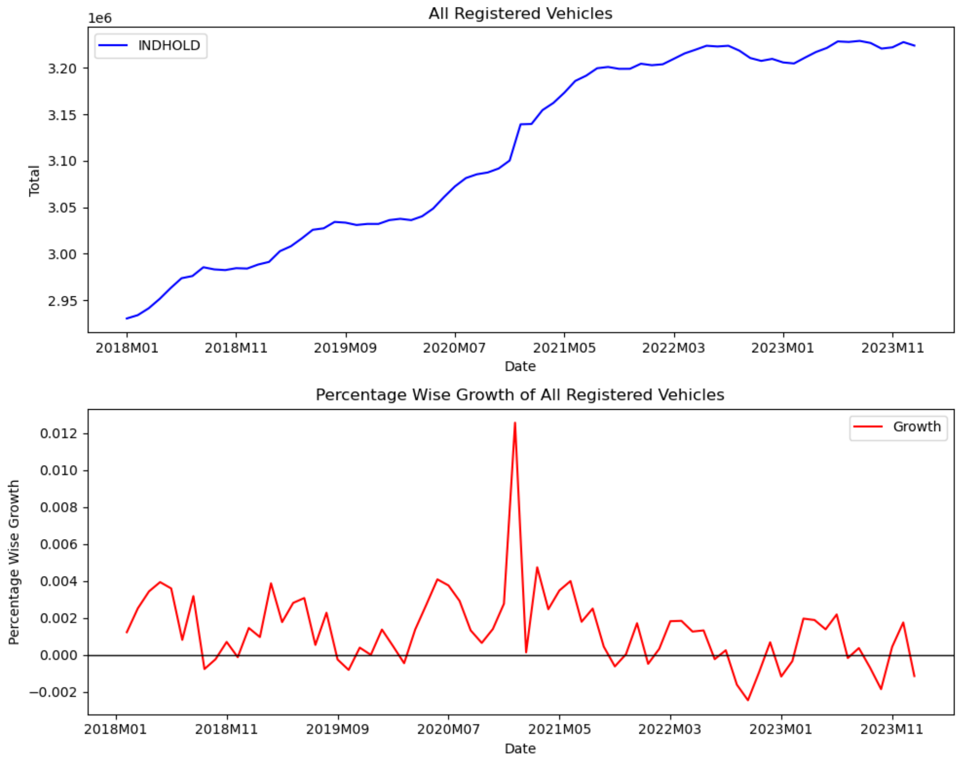


Figure 5.4: The monthly total number of all registered vehicles from January 2018 to January 2024. Below the monthly percentage wise growth.

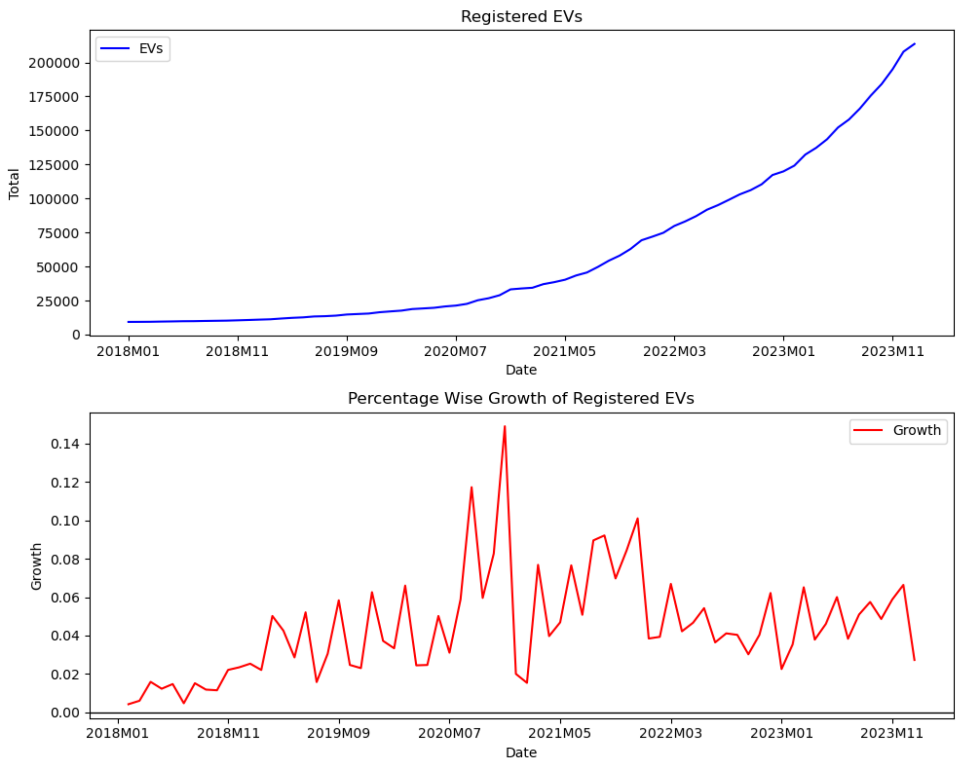


Figure 5.5: The monthly total number of registered EVs from January 2018 to January 2024. Below, the monthly percentage wise growth.

The data was further analyzed at the municipality level to identify regional variations in EV adoption. This was done to understand the distribution of EVs across different municipalities, and to identify if certain regions are more inclined towards EV adoption. Together with the previously mention traffic data, this data provides a foundation for formulating a representation of demand for EV charging infrastructure across different regions.

5.4 OSM Data

The final set of data collected for the project is geospatial information regarding various amenities and additional infrastructure relevant to EV users. This data was gathered from OpenStreetMap (OSM), which is a collaboration project to create a free open source map of the world. In order to only extract information relative to the geographic scope of Denmark, a prefabricated dataset was retrieved from Geofabrik, which is a platform providing OSM data extracts for specific regions. The extraction includes all existing data in the OSM database within a multipolygon reflecting the borders of Denmark [23, 24].

Objects are stored in one of three formats: Nodes, Ways or Relations. Nodes are the most basic elements, represented by a singular geographical point with a unique identifier. Ways are a list of nodes that form a polygonal chain, which can represent linear features or area features such as roads or buildings. Relations have not been used to extract information, and is therefore not explained. As a result of OSM being a user-generated database, the consistency of how certain objects are represented can vary, which will be discussed later in the thesis. The OSM database contains a wide range of objects, from which a subset of categories were selected for this project[25]. The decision of which categories to include was based on their similarity to general urban features used in the existing literature [10, 17, 26]. The list of selected categories are presented in Tab. 5.2. Note that the term "amenities" will be used as a general term for all selected categories further in the thesis.

Category	**Types**
Amenities	Fastfoods, Restaurants, Colleges, Universities, Parkings, Clinics, Dentists, Hospitals, Cinemas, Conference Centres, Events Venues, Community Centres, Theatres, Places of Worship
Buildings	Apartments, Hotels, Sports Halls, Stadiums, Sports Centres
Highways	Motorways, Motorway Links
Shop and Leisure	Fitness Centres, Stadiums, Supermarkets, Malls, Substation Minor Distributions

Table 5.2: List of selected categories from the OSM database.

As mentioned earlier the data is stored as three different types, two of which were relevant for the extraction of above mentioned data. In order to extract the data, the

Python library PyOsmium was used [27]. Data was extracted by iterating over every object stored as Nodes or Ways, and extracting any features matching the selected categories. Nodes were extracted with an ID, a name and the corresponding latitude and longitude coordinates. Ways were extracted in similar fashion, however as ways consist of a combination of points a centroid was calculated and used as the location of the object.

Investigating the extracted data, it was found that objects had the possibility of being tagged as both a node and a way. Additionally, features such as hospitals could consist of multiple nodes or ways. To ensure that each object was only represented once, the data was deduplicated by matching objects within a radius of 50 meters. The radius was chosen based on the assumption that objects within this distance with a matching name label were likely to be the same object. Only motorway objects were treated differently, as it was desirable to distribute a traffic density value to all objects of said type as stated under the traffic data section.

The geographical area of each municipality was determined on the basis of data from GADM [28], which includes a geographical boundary for each municipality in the form of polygons. Each OSM object was then assigned a municipality to facilitate integration with EV registrations across municipalities, and provide deviation in demand for amenities across different regions. The intramunicipality demand is later used to adjust the evaluation of points in one municipality relative to another.

Chapter 6

Theory and Methodology

This chapter presents the theoretical foundations of the core methods used in this thesis. The chapter is divided into three main sections, where section 6.1 introduces the concept of the gravity model, and how it applies to the proposed model. Section 6.2 introduces the general concepts of evolutionary algorithms, specifically genetic algorithms. Finally, section 6.3 discusses the choice of distance calculation method, and explains the Haversine formula used in this study.

6.1 Gravity model

The gravity equation has been adopted by economists to model international trade flows between countries. The model is based on the analogy of Newton's law of gravitation. Newton's law of gravitation states that the force of attraction between two objects is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. Mathematically, the law is expressed as:

$$F = G \cdot \frac{m_1 m_2}{r^2} \tag{6.1}$$

where F represents the force of gravitational attraction between two bodies, m_1 and m_2 are the masses of the two bodies, r is the distance between the centers of the two masses, and G is the gravitational constant, an empirical physical constant involved in the calculation of gravitational effects [29].

The fundamentals of the gravity model has been adapted to model both international trade flows between countries and used for optimization of facility location. The gravity model has been a versatile tool in economics for predicting trade flows and understanding international trade patterns [30]. From the success of the gravity model in economics, it has been adapted to both competitive and noncompetitive facility location problems [31]. The gravity model has been used to model the interaction between facilities and customers, evaluating the attractiveness of a facility based on the distance to the customer. In this study, the gravity model is adapted to evalauate the location of new charging stations based on characteristics of surrounding POIs, traffic flow and the local EV population density. Instead of calculating the interaction between the charging station and the surrounding POIs, the influence of surrounding POIs and traffic flow on the location of a new charging station is calculated. This is analogous to determining the gravitational acceleration at a point in space based on the masses of surrounding objects. The gravitational acceleration is expressed as:

$$a = G \cdot \frac{M}{r^2} \tag{6.2}$$

where a is the gravitational acceleration at a given point, M is the mass of the object, r is the distance between the object and the point, and G is the gravitational constant.

The influence of multiple surrounding objects is calculated by summing the individual accelerations. In physics the total gravitational acceleration under the influence of multiple objects is calculated as a vector sum of the individual accelerations, meaning that the direction of the acceleration is considered. In this study the direction of influence is neglected and only the magnitude of influence is considered. The mathematical representation of the model will be outlined in Chapter 7.

6.2 Evolutionary Algorithms

Evolutionary algorithms (EA) are a large class of optimization algorithms which are inspired by the process of natural selection. There exists extensive literature on EAs which cover a wide range of various algorithms and techniques, why an exhaustive review of EAs is beyond the scope of this study [32–35]. Instead an introduction to the general principles of EAs will be presented, to facilitate how these optimization algorithms can be applied to weight optimization in the model developed in this study.

6.2.1 Theoretical Foundation of Evolutionary Algorithms

The foundation of EAs stem from the well known Darwinian concept of survival of the fittest, where the fittest individuals in a population are more likely to pass their genes, and thus their traits, to the next generation. A population of individuals is evolved over a number of generations, where the individuals are evaluated based on their fitness. The fitness of an individual is a measure of how well the individual solves the optimization problem. Higher scoring individuals are more likely to pass their genes to the next generation. The algorithms searches the solution space and avoids local optima by introducing mutation and recombination of individuals. Mutation introduces random changes to the genes of an individual, while recombination

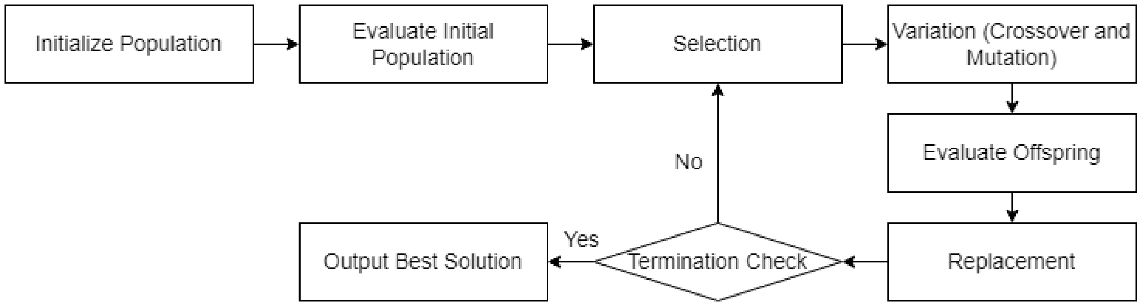


Figure 6.1: Flowchart of an Evolutionary Algorithm. The figure has been adapted from [36].

combines the genes of two individuals, typically higher scoring individuals, to create the next generation of individuals [36]. The flow of EAs is depicted in Fig. 6.1.

There exists a number of different EAs, each with their own variations on how the genetic operators are applied, and thus how individuals evolve through the search space. Some of the most established EAs include Genetic Algorithms (GA), Genetic Programming (GP), Evolutionary Strategies (ES), and Evolutionary Programming (EP). The choice of EA is dependent on the optimization problem and the characteristics of the solution space. The difference between these algorithms are mainly in how mutation and recombination are applied, and how individuals are represented. The general principles of these algorithms are described in the following listing:

- **GA** Individuals are represented as fixed-lenght strings. Pairs of individuals are randomly selected for crossover. Mutation is applied immediately after crossover on the offspring, parents are not mutated. New population is created by selection criteria, typically based on elitism.
- **GP** Individuals are represented as trees. Mutation and recombination are applied directly to the trees. Mutation can be applied to the entire tree, or to a subtree. Recombination is applied by swapping subtrees between two individuals.
- **ES** Individuals are represented as real-valued vectors. Mutation is applied by adding a small random value to the vector. Recombination is not used in ES. Selection is based on the fitness of the individuals.
- **EP** Individuals are represented as real-valued vectors. Mutation is applied by adding a small random value to the vector. Recombination is not used in EP. Selection is based on the fitness of the individuals.

The remainder of this section will focus on GAs, since that is the algorithm used for the optimization problem in this thesis.

6.2.2 Genetic operators: Mutation and Recombination

GAs typically represent an indivdual of the population as a fixed-lenght string or array, which are stored as binary or real-valued numbers. Mutations are applied by flipping bits in the string, or by adding a small random value to the real number. Since the use of GA is used for weight optimization in this study, the individuals will be represented as real valued vectors. Therefore the methods introduced in this section mainly covers real-valued representations of individuals. Mutation for real-valued vectors is typically done by adding a small random value to the vector. This can be done by adding a random value sampled from a gaussian distribution with mean 0 and standard deviation σ. Mutations are applied to each entry in the vector with a certain probability. An example of mutation for a real-valued problem can be thought of as a vector with values $[0.5, 1.2, 3.4, 2.2]$. Suppose the probability of mutation is 0.5 and $\sigma = 0.1$. The mutation operator will add a random value sampled from a gaussian distribution to each entry in the vector with a probability of 0.5. The resulting vector could be $[0.523, 1.175, 3.4, 2.2]$.

Recombinations can be performed in multiple ways. Common methods include using crossover points to swap genes between two individuals. For real-value vectors, crossover can be done by selecting a random point in the vector and swapping the genes between the two individuals, giving the offspring a combination of the parents genes. Recombination for real-valued vectors can also be performed by blending the genes of the parents [37]. Offspring are created by linearly interpolating between parents genes, typically within a range of $[-\alpha, 1 + \alpha]$.

6.2.3 Mating- and Environmental Selection

Mating selection is the process of selecting individuals for mating. The selection process is normally based on the fitness of the individuals. Higher scoring individuals are more likely to be selected for mating. Three common methods are Roulette Wheel selection, Tournament selection and Rank-based selection. In Roulette Wheel selection, indviduals with higher fitness have a larger probability of being selected. Tournament selection randomly forms groups of individuals, where higher scoring individuals are selected for mating. For Rank-based selection individuals are ranked based on their fitness, and selection is based on these ranks rather than raw fitness values.

Environmental selection is the process of selecting individuals for the next generation. There are generally two categories of selection methods, elitist and non-elitist. Elitist methods choose the best individuals from the current generation to be included in the next generation. Non-elitist methods only selects individuals among the offspring of the current generation. The methods mentioned for mating selection may also be used for environmental selection.

6.2.4 Multi-objective Optimization

GAs may also be applied to multi-objective optimization problems. Determining the optimal solution for multi-objective optimization problems is difficult, as the optimum for one objective may be conflicting with the optimum for other objectives. In rare cases the optimal solution achieves higher fitness for all objectives, and thus dominates all other possible solutions. There are several methods for solving multi-objective optimization problems, including weighted sum, and Pareto optimization. For weighted sum optimization, the objectives are weighted and the sum of all objectives is calculated. This method requires knowledge of the relative importance of the objectives, which is not always known in advance. Pareto optimization is a way of delaying the decision of the relative importance of the objectives. A solution is considered pareto optimal if it is better than another solution in at least one objective, and not worse in any other objective. The set of pareto optimal solutions is called the pareto front. The pareto front is the set of solutions which are not dominated by any other solution [38, 39]. Common methods for solving multi-objective optimization problems with GAs include NSGA-II and SPEA2 [40, 41]. However, the optimization problem in this thesis relates to a single objective optimization problem, and thus the methods for multi-objective optimization will not be further discussed.

6.3 Choosing distance calculation method

The choice of method for calculating the distance between points on the surface of the Earth is important in applications which rely on geospatial data. There exists a number of methods to calculate the distance between two points on the Earth's surface, each with its own advantages and limitations. The most commonly used methods include the Haversine formula, the Vincenty formula, and the Great Circle Distance formula [42].

The Haversine formula is a mathematical formula used to calculate the shortest distance between two points on the surface of a sphere given their longitudes and latitudes. It is commonly used in geospatial applications to compute distances between locations on Earth, taking into account the curvature of the planet. The formula is based on the law of haversines, which relates the haversine (half the versed sine) of an angle to the haversine of the angle's complement. The formula is not a perfect representation of the Earth's shape, as the planet is an oblate spheroid rather than a perfect sphere. However, for short distances, the Haversine formula provides accurate results and is computationally efficient [43]. The Haversine formula is expressed as:

$$\Delta\mathrm{lat} = (\mathrm{lat}_2 - \mathrm{lat}_1) \cdot \frac{\pi}{180} \tag{6.3}$$

$$\Delta\mathrm{lon} = (\mathrm{lon}_2 - \mathrm{lon}_1) \cdot \frac{\pi}{180} \tag{6.4}$$

$$a = \sin^2\left(\frac{\Delta \text{lat}}{2}\right) + \cos(\text{lat}_1) \cdot \cos(\text{lat}_2) \cdot \sin^2\left(\frac{\Delta \text{lon}}{2}\right) \tag{6.5}$$

$$c = 2 \cdot \text{asin}\left(\sqrt{a}\right) \tag{6.6}$$

$$d = R \cdot c \tag{6.7}$$

where:

- Δlat & Δlon is the difference in latitude and longitude between the two points in radians
- R is the average radius of the Earth at 6371.0 km
- a is the square of the half-chord length between two points on a sphere
- c is the angular distance in radians between two points on a sphere
- d is the haversine distance between the two points.

Given the satisfactory accuracy and computational efficiency of the Haversine formula, this method is chosen for calculating the distance between locations throughout this study.

Chapter 7

Implementation

The implementation chapter describes the modelling choices along with the foundation of the adjusted gravity model in section 7.1. Subsequently, a method for is defined for adjusting the weights of the model, such that the modal is able to approximate actual demand associated with existing charging stations. Section 7.2 defines a set of criteria for determining when the model is performing well, and presents a justification for the use of a genetic algorithm optimization approach. The chapter concludes with a description of the initial implementation of the GA in section 7.2.3.

7.1 Gravity Model

The gravity based model has been adopted to fit the research questions formulated for this project. The model considers the proximity of relevant data to candidate locations for charging stations. To ensure high performance and scalability C++ was chosen, and the model was implemented to facilitate integration with the data processing pipeline. As mentioned in Chapter 6 the model is built by the concept of gravitational acceleration. The core idea is that a points acceleration or influence is determined by the "mass" of objects in its vicinity. The mass of the objects is determined by a two factor equation. The first factor is a weight of the type of object, and the second factor is determined by the municipality the object lies within. The impact of each object is determined based of the distance to the point under consideration. Object may consist of two main categories. The first category is a list of amenities. The second category is simply the density of vehicles travelling on nearby roads. The model is then designed to output a "gravity score" which is a sum of the influence of all the objects in the vicinity. For transparency and interpretability the model is also designed to save the individual scores of each object, such that the influence of each object can be manually inspected. An overview of the flow of the model is shown in Fig. 7.1.

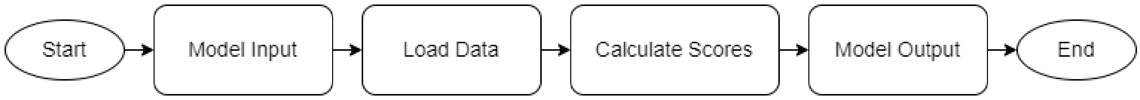


Figure 7.1: Simplified flow diagram of the gravity based model

7.1.1 Model Input and Data Loading

The model takes in a file containing a list of the candidate locations that are to be evaluated. The file should consist of the latitude and longitude coordinates of each location, as well as a 0 valued entry for representing the gravity score at model initialization.

Next the model loads all the preprocessed data into a set of vectors. As the model is developed in C++ custom data structures are established to encapsulate the data. The data is originally stored in CSV files containing multiple data types, and thus require custom data structures for encapsulation. Three main data structures are defined, namely **LocationTraffic**, **LocationAmenity** and **LocationEV**.

The first data structure, **LocationWithValues**, is designed to represent points with numerical values, in this case traffic density. The structure is defined as follows:

```
    struct LocationTraffic {
        double longitude;
        double latitude;
        double value;
    };
```

Listing 7.1: Data structure used for handling traffic related data

The structure takes in three parameters including the geospatial coordinates of the point in terms of latitude and longitude, and a numerical value representing the traffic density allocated to the point. The data structure is also reused to store the gravity score of candidate locations.

The second data structure, **LocationAmenity**, is developed to determine the coordinates of an amenity and stores the associated municipality.

```
    struct LocationWithCommune {
        double longitude;
        double latitude;
        string municipality;
    };
```

Listing 7.2: Data structure that reads the location and municipality of each amenity

The third and final costum data structure **LocationEV** stores the percentage growth of EVs in a given municipality.

```
    struct LocationEV {
        string municipality;
```

```
        double EVGrowth;
    };
```

Listing 7.3: Data structure used to store the local trajectory of EVs

The vectors are populated with all the preexisting data necessary for calculating the gravity score through the use of the listed data structures. Once candidate locations as well as the data vectors are loaded the model is able to progress to calculating the gravity score of each candidate location.

7.1.2 Gravity Score Components

The calculation process is the core logic of the model. This section will provide the mathematical foundation of the model, and thus explain how a gravity score is calculated based on the concept of gravitational acceleration. The score calculation consists of three main components, a "mass" of the object, a distance to the object, and a regularization factor.

Distance calculation and constraints

The first component handled is the distance between the candidate locations and the objects. The distance is calculated using the Haversine formula. Two constraints are applied to determine which objects are considered to have an influence on a given location. The first constraint is a minimum distance threshold, which ensures that objects located extremely close to a candidate location do not receive an exaggerated score. The reasoning behind this constraint is that all the considered objects are stored as a point rather than the area they cover. Therefore cases where a candidate location would be within the area of an object would result in an exaggerated score, also avoiding division by 0. Additionally, the feasibility of deploying a charging station should be considered, why a minimum distance threshhold of 100 meters has been decided. This results in an assumption that any objects within the range of 100m do not have a increased impact on the location. Translated to a domain specific expresion: It is assumed that people do not distinguish between walking distances less than 100m from the charging station. The second constraint is a maximum distance threshold, limiting the amount of objects that are able to influence a candidate location. This constraint is applied to ensure that the model is not overly influenced by objects located excessively far from the candidate location. Translated to a domain specific expression: It is assumed that people do not walk more than a certain distance from the charging stations to reach an amenity. The original maximum distance threshold was set to 1000 meters as around 90% of movements to nearby venues occur within said range [11]. In summary, the constraints are applied to ensure that the model is not overly influenced by objects located too close or too far from the candidate location, while also limiting the computational complexity of the model.

Mass representation of amenities

The "mass" of objects can be further divided into two categories, amenities and traffic density. The mass of amenities is determined by a weight and a variable called weightAdjustmentFactor. The weightAdjustmentFactor describes the evolution of EVs and total number of EVs in a municipality. It is incorporated into the model by determining a weight adjustment factor. This factor is used to add dynamism to the model, and to distinguish between the impact of similar amenities located in different municipalities. Originally the idea was to use data such as "Telia Crowd Insights" to determine the activity around each amenity, however unfortunately the data was not available [44]. Instead the adjustment is based on a combination of the total number of EVs in the municipality and the percentage growth of EVs. The total number of EVs is used to estimate a base level of activity, while the percentage growth is used to ensure that the model is adaptable to future changes in EV adoption. A formula for the weight adjustment factor is shown in Eq. 7.1.

$$\text{weightAdjustmentFactor} = (1 + \frac{EV_{increase}}{100}) \tag{7.1}$$

Where weightAdjustmentFactor is a factor used in the calculation of the mass of an object and $EV_{increase}$ is the percentage growth of EVs in the municipality. As the weightAdjustmentFactor is used in a multiplicative relationship, one is added to the percentage growth to ensure that the factor has a positive impact on the mass, and then divided by 100 to get a decimal representation.

The weight term of the mass representation describes the relative influence different types of amenities have on the score. Determining the weights is a crucial as it directly impacts the outcome of the model. How the weights are determined is described in Section 8.2. Thereby the combined mass representation for amenities is defined by Eq. 7.2.

$$m_j = w_j \cdot weightAdjustmentFactor \tag{7.2}$$

Mass representation of traffic density

The mass of traffic density is determined quite similarly to amenities, however the weight adjustment factor is replaced by a normalized value of the average annual daily traffic (AADT). The AADT is normalized to ensure that the mass of traffic density is comparable to the mass of amenities. The normalized AADT is then multiplied by a weight to determine the mass of traffic density. The formula for the mass of traffic density is shown in Eq. 7.3.

$$m_{traffic} = w_{traffic} \cdot \overline{AADT} \tag{7.3}$$

Regularization factor

The final component of the gravity based model is a regularization factor. The objective of the regularization term is to regulate the influence of multiple similar amenities. The regularization factor is calculated as an exponential decay function, where the influence of each similar amenity decreases exponentially. Eq. 7.4 shows the regularization term.

$$r_i = e^{-0.5\cdot(i-1)} \tag{7.4}$$

Where r_i is the regularization factor, and i is the index of the amenity. Amenities are indexed from closest to furthest.

7.1.3 Gravity Score Calculation

Now that each individual component is defined, the gravity score can be calculated. The flow of the calculation is described by the following equations:

Let ϕ and λ be the latitude and longitude of a candidate location respectively.
Let $A = \{(lon, lat, municipality)\}$ be a vector of vectors containing latitude, longitude and municipality of an amenity.
Let w be the weight parameter of the amenity type.
Let $d_{min} = 0.1$ be the minimum distance threshold in kilometers.
Let d_i be the distance to the i'th amenity.

The distance d_i is determined based on the conditions:

$$d_i = \begin{cases} \max(d_{\min}, d(\phi, \lambda, lat_i, lon_i)) & \text{if } d(\phi, \lambda, \mathrm{lat}_i, \mathrm{lon}_i) < 1 \\ \text{undefined} & \text{otherwise} \end{cases} \tag{7.5}$$

The conditions ensure that the distance is not less than the minimum distance threshold, and that amenities located further than 1 kilometer away are not considered.

For all amenities satisfying the conditions, the weight adjustment factor is described by Eq. 7.1.

Let $EV_{increase}$ be denoted as E.
Let $x = (municipality_i, E_i)$ be a vector mapping the municipality and percentage growth of EVs.

The weight adjustment factor can then be defined as:

$$weightAdjustmentFactor_i = 1 + \begin{cases} \frac{E_i}{100.0} & \text{if } A_{municipality_i} \in x \\ 0 & \text{otherwise} \end{cases} \tag{7.6}$$

where $weightAdjustmentFactor_i$ is the weight adjustment factor of the i'th element in A.

With the weight adjustment factor determined, the influence of one amenity is described by Eq. 7.7.

Let s be the vector of scores for all elements in A.

$$s_i = \frac{w \times \text{weightAdjustmentFactor}_i}{d_i^2} \tag{7.7}$$

The scores for all amenities satisfying the conditions are collected and sorted in descending order in s. A new condition is then applied such that only the 5 highest scores are considered.

Let $N = min(||s||, 5)$ be the number of similar amenities considered.

The regularization factors for the considered amenities are then calculated:

$$r_i = e^{(-0.5 \cdot (i-1))}, \quad \text{for } i = 1, 2, \ldots, N \tag{7.8}$$

The gravity score is then calculated by summing the scores of the considered amenities:

$$GravityScore_{amenity} = \sum_{i=1}^{N} s_i \times r_i \tag{7.9}$$

The gravity score for the traffic density is then calculated:

Let $T = \{(lat, lon, value)\}$ be a vector of vectors containing latitude, longitude and value of traffic density. Let w_t be the weight parameter for traffic.

The distance d_i is determined similarly to amenities, however the maximum distance threshold is set to 2 kilometers. The distinction is made such that the majority of points are affected by traffic density. Thereby, the distance is determined by the conditions:

$$d_i = \begin{cases} \max(0.1, d(\phi, \lambda, lat_i, lon_i)) & \text{if } d(\phi, \lambda, lat_i, lon_i) < 2 \\ \text{undefined} & \text{otherwise} \end{cases} \tag{7.10}$$

The traffic portion of the gravity score is then calculated, given that d_i is defined:

Let t be the vector of scores for all elements in T.

$$t_i = \frac{w_t \cdot value_i}{d_i^2} \tag{7.11}$$

This is repeated for motorway points and the combined gravity score for traffic density is defined by Eq. 7.12.

The highest score is then selected from the vectors t:

$$GravityScore_{traffic} = \max(t_{traffic}) + \max(t_{motorway}) \tag{7.12}$$

Let M be the total number of amenities influencing the candidate location, then the total gravity score for the candidate location can be described by Eq. 7.13.

$$GravityScore_{total} = (\sum_{j=1}^{M} \sum_{i=1}^{N} \frac{w_j \cdot weightAdjustmentFactor}{d_i^2} \cdot r_i) + GravityScore_{traffic} \tag{7.13}$$

7.1.4 Model Representation

In order to facilitate a clearer understanding of the mathematical foundations and the flow of the proposed model, the following diagrams illustrate the key components and processes involved. The overall structure of the model is described in Fig. 7.2.

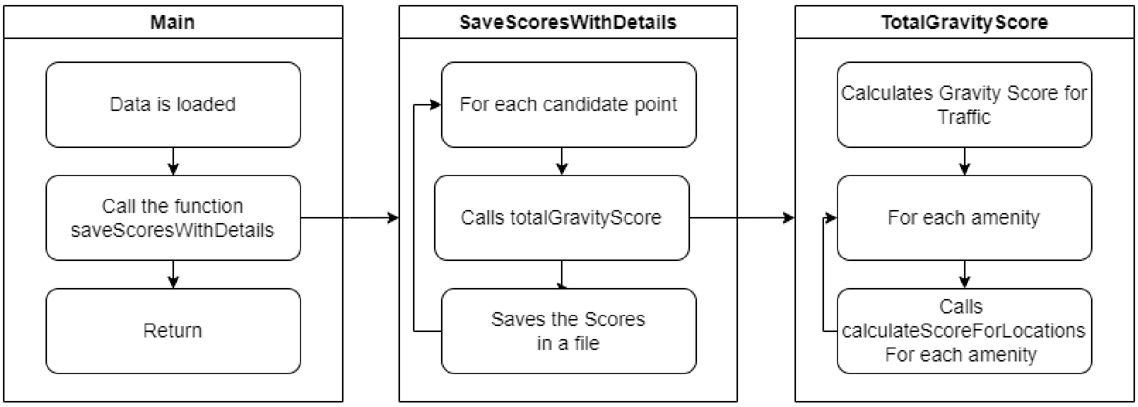


Figure 7.2: Overall structure of C++ implementation

The figure represents how the model is structured, and how different components interact. The leftmost box shows that the model is initiated by loading all the necessary data, including amenity locations, traffic locations and values, EV growth in municipalities and candidate locations. Once the data is loaded, a score is calculated for each candidate location. This is described by the middle section, which notes that a gravity score is calculated seperately for each candidate location. The actual computation of the gravity score is described in the rightmost box, which shows the components that make up the gravity score. The gravity score of a candidate location is comprised of two main components, the score of amenities and the score from traffic densities. The score of each amenity is calculated seperately, and repeated for all amenities. In order to provide a more detailed understanding of the actual computation of the gravity score, the flow of the calculation is described in Fig. 7.3.

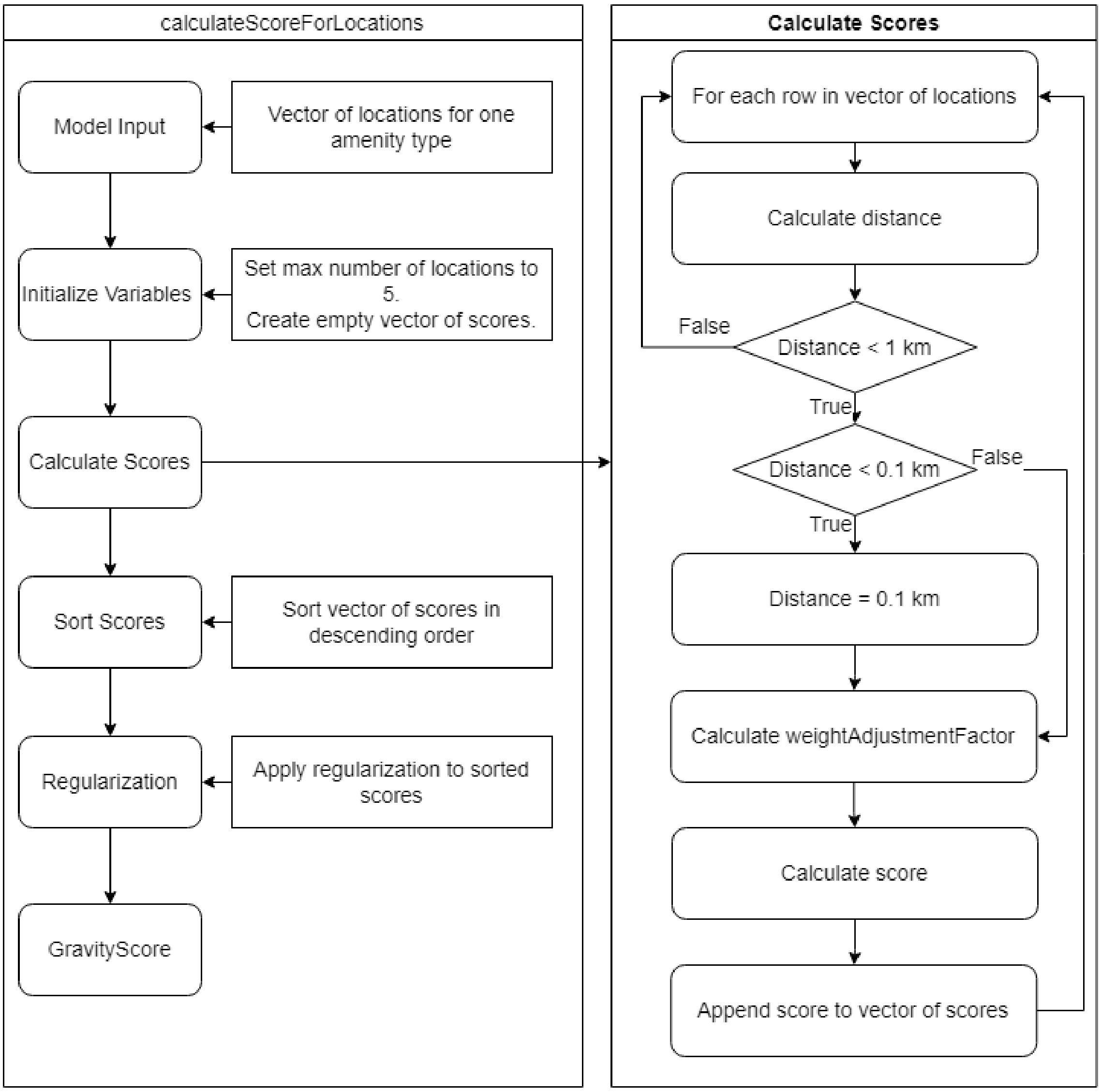


Figure 7.3: Flowchart of the gravity score calculation process.

The process is initiated by instantiating a vector, which holds the locations of all instances of one type of amenity. Additionally, the number of instances of one amenity type able to influence a candidate location is defined. An empty vector is created to store the scores that each amenity object contributes to the candidate location. Once everything is initialized, the actual computation is invoked. This process is shown in the right hand side of the figure. The score is calculated as an iterative process for each amenity object. First the distance between the candidate location and the object is calculated. Once the distance is obtained, a distance constraint is checked to determine if the object should influence the candidate location. If the constraint is satisfied, another constraint is checked to ensure that the score is not overly inflated by the distance being to small. When the final distance is determined, the weight adjustment factor for the object is retrieved from a lookup table. Once all the necessary parameters are obtained, the score of the object is calculated as described by Eq. 7.7. This score is then appended to the vector, and the process is repeated first for all objects of one type of amenity. The vector containing the scores for each object is then sorted to ensure that the objects lying closer to the candidate location are the most influential. The at most 5 best scores are selected and regularization is applied. After regularization is applied the scores are summed to determine the combined influence of the evaluated amenity type.

7.1.5 Model Case Study

The models behaviour is further explained by the example shown in Fig. 7.4.

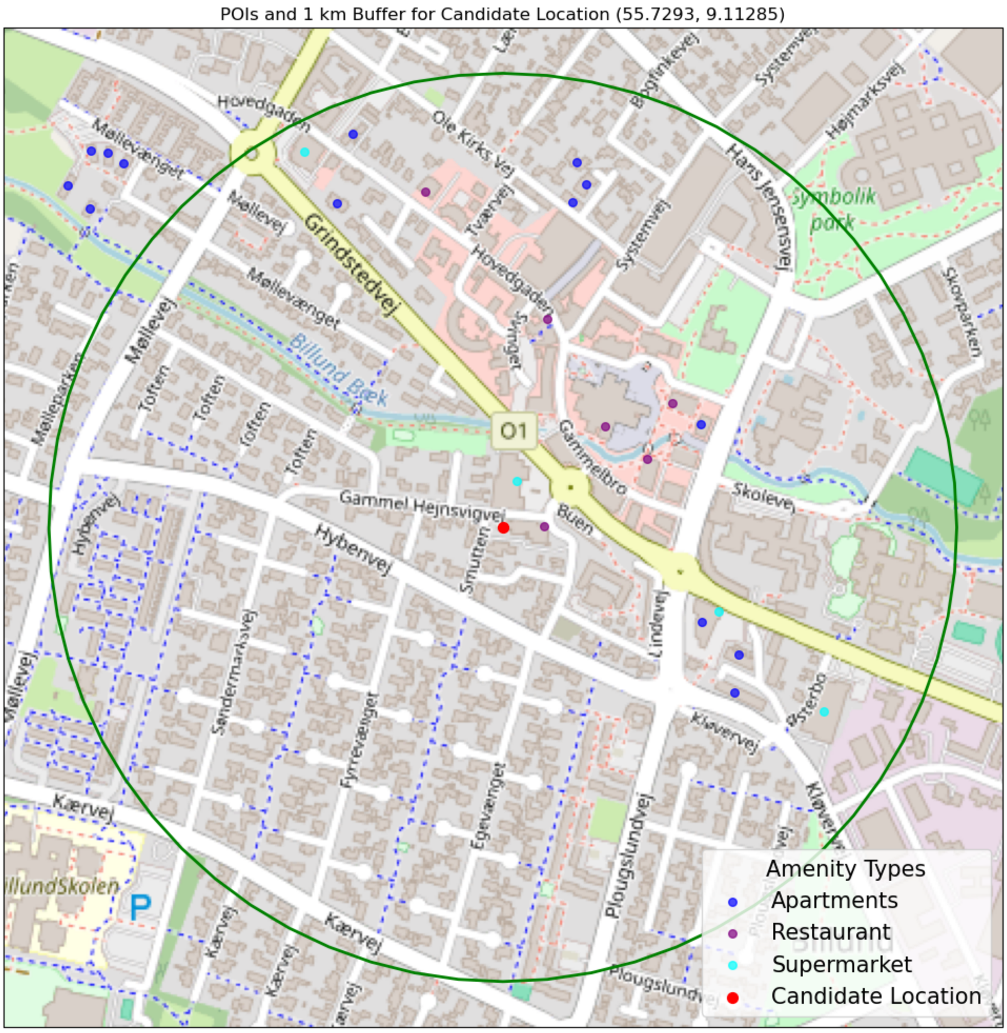


Figure 7.4: An example location (latitude 55.7293, longitude 9.11285). The green circle is the distance constraint, the red point is the candidate location and remaning points are amenities in the vicinity.

The example shows a candidate location, denoted by the red point, and the amenities that are able to influence the gravity score of the candidate location. In this case the candidate location is influenced by apartments, restaurants and supermarkets. Traffic density and the remaining amenities are not considered in this case, as no objects of the remaining types are within the distance constraints. Assuming the amenities are examined alphabetically, the influence of apartments is considered first. The distance to all apartments is calculated, and all apartments satisfying the distance constraint, depicted by the green circle, are considered. In total 9 apartments satisfy the constraints, yet the gravity score is determined by the 5 highest scoring apartments. This is ensured by sorting the vector, and selecting the 5 best scoring apartments. Finally, regularization is applied andt the scores of the apartments are summed to a combined score. This process is repeated for restaurants and supermarkets. After iterating through all amenities, the scores across amenity types are summed to determine the total influence of all amenities on the candidate location.

7.1.6 Model Output

The model outputs a file containing the assigned scores for each candidate location. Additionally, the combined score for each type of amenity are included in the output file for transparency and interpretability. The information can then be used for further analysis. The scores of individual objects are not included in the output file, as including scores for each object would result in increased complexity.

7.2 Model Weight Optimization

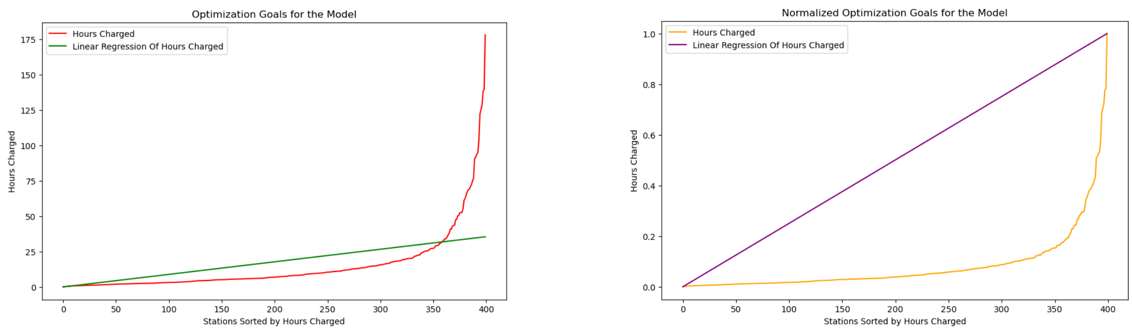
The formulation of demand is represented by the weights of the amenities and traffic density measures. Thus, the weights should be approximated to represent the actual demand in the area. In order to determine the weights, a representation of the actual values must be obtained. The following section will describe the optimization objectives used to approximate the weights, followed by the implementation of the genetic algorithm which is later used to approximate the weights.

7.2.1 Determining the optimization objective

Determining the objective function for predicting the usage of charging stations at new locations is a complex task. This function must effectively assess how well the model predicts actual demand for charging stations. Data from Norlys provided an approximation of the actual demand at 400 stations, enabling the model to be trained accordingly. Initially, the objective function aimed to minimize the RMSE between the estimated and actual demand. However, given the non-linear data distribution, an additional objective was introduced to approximate a linear representation of the actual demand. It was also decided that the model should not strictly capture the exact numerical distribution of the data. Therefore, two more objectives were added, with values normalized between 0 and 1. Ultimately, four optimization objectives were established:

The optimization objectives fall into two categories, each containing two objectives. The first category uses the numerical value of hours charged at the stations as the optimization target. The data distribution is modeled in two ways: the first objective employs the actual distribution of values as a training goal, while the second uses a linear regression of the actual values to create a linear representation. This was done to explore the model's performance with a linear target.

The second category of objectives is based on normalized values. The first objective mirrors the original distribution but normalizes the values to a scale of 0 to 1. The second objective applies normalization to a linear representation of the actual values after regression, ensuring a comparable scale between the predictions and the optimization objective values. The objectives are summarized in the illustration shown in Fig. 7.5.



(a) Non-normalized objectives

(b) Normalized objectives

Figure 7.5: Illustration of the optimization objectives

Thus, the optimization objectives are formulated as follows:

1. Minimize the RMSE between the actual hours charged and the predicted hours charged
2. Minimize the RMSE between the linear representation of the actual hours charged and the predicted hours charged
3. Minimize the RMSE between the normalized actual hours charged and the normalized predicted hours charged
4. Minimize the RMSE between the normalized linear representation of the actual hours charged and the normalized predicted hours charged

7.2.2 Brute Force Optimization

Brute force optimization is an exhaustive approach to optimization, where all possible solutions are evaluated. The approach ensures finding the optimal solution, if the solution space is discrete and finite. For the case of optimizing the weights of the model, the solution space is continuos as weights are represented as floating point values in the range of 0 to 100. Say that the solution space is discretized to a 0.1 resolution, given that there are 27 weights to optimize, the brute force approach would require 1001^{27} evaluations. This is an infeasible number of evaluations, and thus the brute force approach is not suitable for this case. Instead the use of a genetic algorithm is proposed, as it is able to handle the large solution space and provide a good approximation of the optimal solution with a lower computational complexity.

7.2.3 Optimization through Genetic Algorithm

This section describes the implementation of a GA designed to optimize the weights of each type of POI. Weighing the parameters has a heavy influence on the model performance, why finding the optimal set of weights heavily determines the accuracy of the model. To avoid an exhaustive search of the solution space, a GA is used to approximate the optimal solution. To aid the implementation of a GA the DEAP python library is used [45]. The library provides a set of tools for implementing evolutionary algorithms, while allowing for a high degree of customization. As the optimization objectives are single-objective optimization problems, a general purpose genetic algorithm is used.

Three main components are required to implement a genetic algorithm. The first component is the evaluation function, which evaluates the fitness of a given individual. The fitness function was customized to evaluate the performance of the model based on the optimization objectives. This required the implementation of a function that converts the individuals from the GA population into a vector of weights, which are then used to calculate the gravity score. Initially, the weights are set to a random value with an upper and lower limit depending on the optimization objective, then the fitness is evaluated. Fig. 7.6 shows the flow of the genetic algorithm setup.

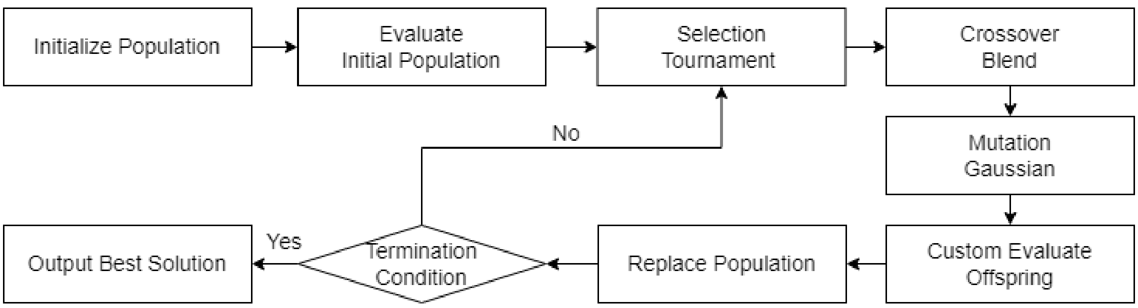


Figure 7.6: Flow diagram of the implemented genetic algorithm. The diagram is a modification of the diagram presented earlier in Fig. 6.1.

The second component involves customizing the crossover and mutation operators. Since weights are represented as floating point values, the blend crossover and gaussian mutation operators are used. However, a constraint is applied, to ensure compatibility with the intent of the gravity model. The constraint ensures that after crossover and mutation, the weights maintain a value higher than 0, thus negating the possibility for negative weights. The weights are constrained to a positive value since each amenity included in the model should have either a positive impact on the gravity model or none at all. Allowing negative weights could lead to trade-offs where the model sacrifices the influence of some amenities to improve the overall accuracy, which is undesirable when all amenities are intended to add value. By enforcing positive weights, the integrity and interpretability of the model is maintained, ensuring that each amenity contributes positively, aligning with real-world expectations that these factors are inherently beneficial.

The third component is tuning the hyperparameters of the genetic algorithm. These include the population size, crossover probability, mutation probability, and the number of generations. The initial values for the hyperparameters were set to common values used in the literature, more specifically based on the findings in [46]. The initial population size was set to 50, the crossover probability to 0.7, the mutation probability to 0.2, and the number of generations to 30. The effect of tuning the parameters is further discussed in Chapter 8.

7.3 Complete Overview of the Implementation

To provide a complete overview of how the solution is structured, a flow diagram is presented in Fig. 7.7. First all data sources are loaded and formatted to interact with the gravity model. A optimization objective is then assumed in order to approximate the weights of the model. The genetic algorithm fits the weights of the model parameters to fit the optimization objective. In each iteration, new gravity scores are calculated for the fitness evaluation. The process is repeated for 30 generations, and the best representation of the weights are saved and assigned to the gravity model. The output of the model is then saved to a file, consisting of the total gravity score for each of the candidate locations, as well as the individual score contribution of each type of amenity.

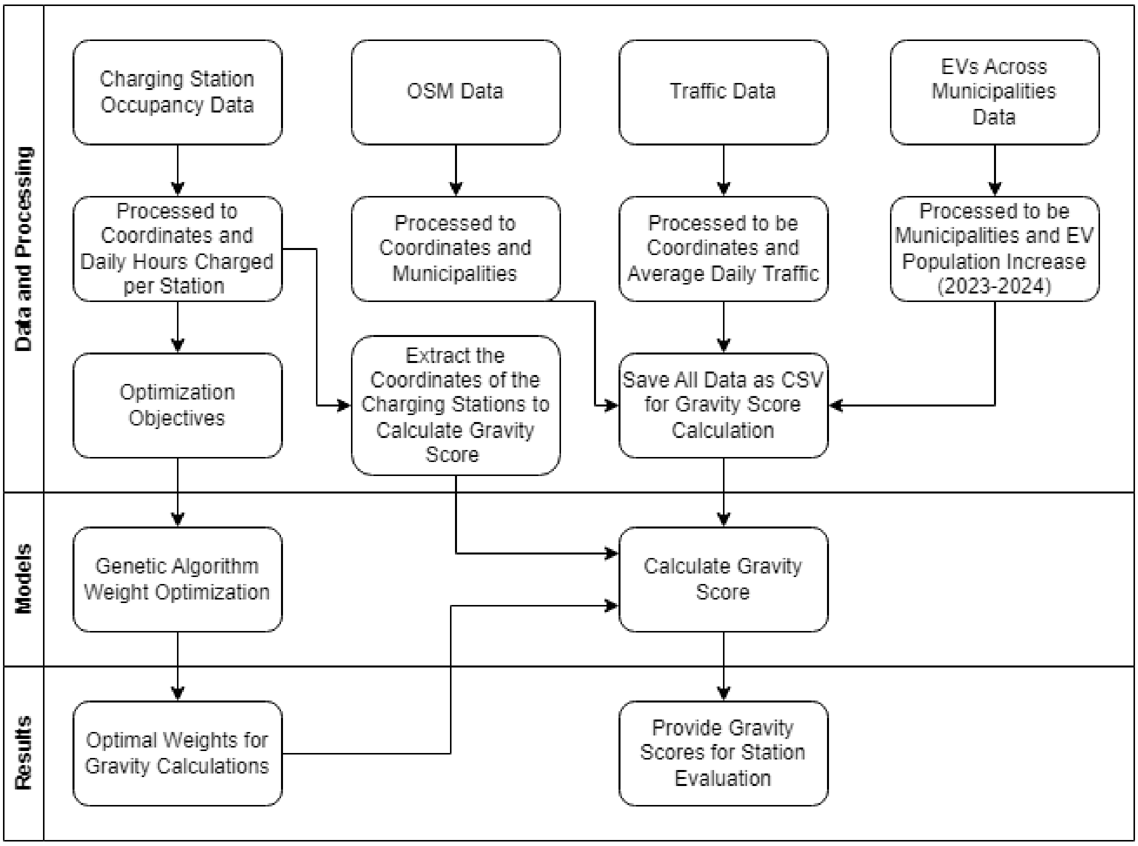


Figure 7.7: Illustration of the complete implementation flow.

Examples of model input and output are provided in the appendix.

Chapter 8

Experiments and Results

The following chapter presents the experiments conducted to evaluate both parts of the proposed model, and the proposed model as a whole. Additionally, the results of the experiments are presented. Section 8.1 presents the impact of applying regularization in the gravity model. Following, section 8.2 presents the results for determining the weights of the gravity model using both a statistical approach and through automated optimazation using a GA. After looking into the weights of the gravity model as a whole, the model is split into urban and motorway stations to evaluate the performance of the model in different environments. The results of this evaluation are presented in section 8.3. Section 8.4 presents the results of using the gravity score to predict the average daily hours charged at new stations. A visual representation of the model predictions are presented in section 8.5 and finally a proof of concept of a hybrid forecasting model is presented in section 8.6.

8.1 Regularization and Sensitivity in the Gravity Model

As the proposed model is heavily related to the spatial distribution of amenities, it is important to ensure that the model is not overly sensitive to the presence of specific amenities. Especially in the case where data is sampled from an open source database. As described back in section 5.4, the amount of data available for each amenity can vary greatly. Regularization is a common technique used in machine learning to prevent overfitting, and to ensure that the model is not overly sensitive to the presence of specific data points. In this case, regularization is used to ensure that the gravity model is not overly sensitive to the presence of specific amenities. To evaluate the impact of regularization on the gravity model, a gravity score for the 400 locations included in the dataset was calculated with four different configurations. The configurations for the four runs are described in the following list:

- **No Limit:** No regularization factor was applied to get the baseline variance for each amenity type.

- **Max 5 amenities of similar type:** An upper constraint on the number of the amenities of the same type allowed to influence each location. This method is assumed effective to reduce model sensitivity to the varying sample size of each amenity type.
- **Regularization by exponential decay function:** An exponential decay factor was applied based on the distance of each amenity from the locations, progressively reducing the influence of subsequent amenities of the same type as their distance increased. The regularization factor is added on top of the previous constraint of only considering the five closest amenities of the same type.
- **Max 1 amenity of similar type:** Only the nearest amenity of each type was considered, effectively disregarding the presence of multiple amenities of the same type. This method is assumed to provide the minimum variance of the model which relates to the difference in distances from stations to amenities.

The primary evaluation metric was model variance, which was calculated for each amenity type using the four regularization methods. The variance was used as a measure of the model's sensitivity to the presence of multiple amenities of the same type, with lower variance indicating a more balanced influence of amenities on the model. The results of the regularization tests are presented in the following section.

8.1.1 Impact of Regularization on the Gravity Model

The impact of regularization on the gravity model was evaluated by testing the four approaches mentioned above. The primary goal was to minimize variance, ensuring that amenities with a higher count did not disproportionately affect the model. The results of these tests are summarized in Tab. 8.1.

The table shows the variance and number of the 400 stations each type of amenity is in the vicinity of, e.g. 237 stations have one or more apartments within the distance threshold of 1 kilometer. The variance is shown for each amenity type, for each of the four evaluations. The evaluation without any constraints or regularization applied serves as the baseline variance of the model when all data points are considered. Generally, the variance of amenities with a high count is large, while lower counts tend to have lower variance. This is expected, as the presence of multiple amenities of the same type at one station and only a single amenity of the same type at another station will result in increased variance. Certain exceptions exist, such as the "placesOfWorship" amenity. The exception can be explained by a combination of the distance between placesOfWorship, which is generally large enough to prevent multiple instances in the vicinity of the same station. This is shown in the close to stable variance across all four evaluations. On the other hand, amenities with a high count and where amenities of similar type may lie within shorter distance of eachother, such as apartments and fastfood restaurants have higher variance. The variance of amenities with such characteristics experience a significant drop when constraints or regularization is applied. For instance, the variance of

Type	Count	Variance			
		No Limit	Max 5	Regularized	Max 1
Apartments	237	3.92e+09	2.34e+08	6.36e+07	1.67e+07
Cinemas	105	1.19e+07	1.15e+07	7.92e+06	5.84e+06
Clinics	91	6.09e+06	6.08e+06	5.45e+06	4.73e+06
Colleges	66	1.21e+07	1.21e+07	7.85e+06	5.60e+06
CommunityCentres	143	5.35e+06	5.32e+06	4.73e+06	4.30e+06
ConferenceCentres	13	8.35e+06	8.35e+06	8.35e+06	8.35e+06
Dentists	83	1.07e+07	1.06e+07	9.67e+06	9.17e+06
EventsVenues	111	5.18e+06	4.79e+06	3.86e+06	3.21e+06
FastFoods	170	3.92e+07	3.93e+07	2.53e+07	1.66e+07
Hospitals	64	1.05e+07	1.05e+07	5.98e+06	4.49e+06
Hotels	53	2.48e+07	2.11e+07	1.44e+07	1.01e+07
LeisureFitnessCentres	170	1.21e+07	1.18e+07	1.03e+07	8.98e+06
LeisureStadiums	39	2.07e+06	2.07e+06	1.51e+06	8.45e+05
MotorwayLinks	152	8.79e+07	6.03e+07	2.39e+07	9.27e+06
Parkings	393	2.70e+08	1.13e+08	3.52e+07	9.78e+06
PlacesOfWorship	256	8.51e+06	8.04e+06	6.56e+06	5.58e+06
Restaurants	252	1.83e+08	7.08e+07	3.10e+07	1.47e+07
ShopMalls	87	1.31e+07	1.31e+07	1.24e+07	1.15e+07
ShopSupermarkets	299	6.04e+07	5.81e+07	3.38e+07	1.86e+07
SportsCentres	30	4.61e+06	4.61e+06	4.52e+06	4.45e+06
SportsHalls	43	3.10e+06	2.94e+06	1.52e+06	1.22e+06
Stadiums	4	3.68e+05	3.68e+05	2.60e+05	1.31e+05
SubstationMinorDistributions	113	2.31e+08	8.91e+07	3.69e+07	1.56e+07
Theatres	89	1.31e+07	1.21e+07	8.77e+06	5.67e+06
Universities	35	4.88e+05	4.67e+05	3.83e+05	3.42e+05

Table 8.1: Variance of constraints and regularization for each amenity type.

apartments drops from $3.92 \cdot 10^9$ without regularization or constraints to $6.36 \cdot 10^7$ after regularization is applied.

The variability in sample size for each amenity is apparent in the count column, where the number of amenities within the 1 kilometer threshold ranges from 4 to 393. These findings highlight general trends stating how existing stations are placed in relation to amenities. For instance, parking spots and supermarkets are represented in the vicinity in the vast majority of stations, even through a user generated database, indicating that stations are typically placed near these amenities.

An important consideration when regularizing the model is the balance between reducing variance without intoducing too much bias. Thus, comparing the variance between the different evaluations shows that the regularized evaluation achieves the lowest variance difference between amenity types, while maintaining a discrepancy from the minimum variance shown in the max 1 evaluation. This indicates that the regularization method is effective in balancing the influence of multiple amenities while still considering their presence.

Overall, the regularization approach successfully minimized variance and provided a more balanced influence of amenities on the gravity model. This supports the hypothesis that regularization can lessen the impact of having numerous similar data points around a location, leading to more stable and reliable model predictions while maintaining the necessary influence of multiple amenities. The findings suggest that the regularized approach offers a practical and effective solution for improving the gravity model, ensuring balanced representation and minimizing disproportionate effects from high-count amenities.

8.2 Determining the weights of the gravity model

Determining the mass of each amenity in the gravity model has a significant impact on the influence each amenity has on the model output. Therefore, it is important to evaluate different methods for determining these weights. This section will compare the results of setting model weights through statistical analysis and automated optimization using a genetic algorithm. The statistical analysis will be explained in section 8.2.1, while the genetic algorithm optimization will be explained in section 8.2.2.

8.2.1 Statistical Weight Analysis

A series of statistical analysis was conducted to determine the importance of amenities near each station by analyzing their proximity and its correlation with the average daily use of the station. An analysis was conducted to determine the importance of amenities at multiple distances. Examined distances were 50m, 100m, 200m, 500m, and 1000m. To facilitate an examination of which amenity types are represented near stations with high- and low usage ratios, a ratio of the average daily use of each charging station was calculated from the timeseries data. The average daily use ratio was then converted to the total number of hours charged at each station, meaning that stations with a large number of plugs with the same ratio as stations with fewer plugs would have a higher total number of hours charged. For each station, the distance was calculated to each amenity type to mark the presence of each type of amenity within the distance threshold. If an object of the type was present within the treshold, the hours charged at the station was added to the total hours charged for the corresponding amenity. Thus, a representation of the total hours charged at stations where each amenity type was present was created, to determine which amenities where present for stations of either high or low usage. Tab. 8.2 shows the average hours charged at stations where each amenity type was present, as well as the statistical analysis of the data with the distance threshold set to 1 kilometer.

Category	Sum	Min	Max	Mean	Median	Std Dev	Q1	Q3	IQR	TC
Status	5882.42	0.08	177.98	14.71	6.94	22.72	3.16	15.56	12.40	400
Apartments	3415.74	0.44	138.28	14.06	8.27	18.17	4.48	15.74	11.26	243
Cinema	1517.37	0.44	93.89	13.31	8.20	15.67	4.69	14.69	10.00	114
Clinic	1424.99	0.53	94.53	13.97	8.18	16.51	5.50	15.45	9.95	102
College	1448.55	0.44	177.98	20.12	11.27	29.37	5.75	18.18	12.43	72
Community Centre	2340.00	0.08	177.98	15.00	8.18	22.36	4.65	15.39	10.73	156
Conference Centre	163.90	1.36	34.07	11.71	10.02	9.58	5.18	15.35	10.17	14
Dentist	1291.33	0.69	103.28	14.04	7.00	17.50	4.54	15.09	10.54	92
Events Venue	1589.34	0.16	103.28	12.82	8.44	14.90	4.68	14.68	10.00	124
Fast Food	5582.93	0.16	177.98	15.73	7.91	23.66	3.75	16.07	12.32	355
Hospital	793.55	0.99	55.16	11.34	7.59	10.99	4.45	14.64	10.19	70
Hotel	789.92	0.98	90.90	13.17	7.58	16.88	4.02	14.97	10.95	60
Leisure Fitness Centre	2489.40	0.44	103.28	13.99	9.46	15.60	5.23	15.61	10.37	178
Leisure Stadium	574.45	1.03	90.90	12.22	9.40	14.37	4.91	13.38	8.47	47
Motorway Link	3295.18	0.63	177.98	20.47	7.39	31.09	3.22	20.08	16.87	161
Motorway	3288.78	0.16	177.98	20.43	7.53	30.93	3.22	20.65	17.43	161
Parking	5822.70	0.08	177.98	14.82	6.99	22.79	3.20	15.56	12.36	393
Place of Worship	3930.94	0.08	128.49	14.19	7.88	19.53	3.53	15.56	12.03	277
Rest Area	178.96	0.26	52.57	7.78	2.56	14.27	1.35	5.41	4.06	23
Restaurant	3487.52	0.37	103.28	13.21	8.08	16.19	3.58	15.39	11.81	264
Shop Mall	1708.59	0.44	103.28	17.43	9.23	21.95	5.28	18.15	12.88	98
Shop Supermarket	4039.03	0.08	128.49	13.20	7.48	18.12	3.53	14.73	11.20	306
Sports Centre	441.04	0.99	55.16	13.78	9.06	13.17	5.40	15.16	9.76	32
Sports Hall	746.66	1.48	125.19	15.24	8.87	21.81	5.58	15.85	10.27	49
Stadium	40.98	1.36	14.74	8.20	6.05	5.72	5.29	13.54	8.24	5
Substation Minor Distribution	2179.70	0.53	177.98	18.63	10.41	27.04	5.58	19.50	13.92	117
Theatre	1409.31	0.44	125.19	13.82	8.25	15.94	4.58	17.57	12.99	102
University	577.11	0.44	55.16	14.43	10.17	13.49	4.86	18.26	13.39	40

Table 8.2: Statistical analysis of amenities at a distance threshold of 1000m.

The table displays an overview of the combined number of hours charged at stations where each type of amenity is present. Additional statistical measures are included to provide insights into the distribution and significance of amenities around high- or low usage stations. The first takeaway is to look at the sum of hours charged for each amenity. It is apparent that amenities such as parking spots, fast food restaurant and supermarkets are present at a high number of stations, while amenities which are not as common, such as stadiums, universities and conference centres do not accumulate as many hours charged. To evaluate whether the presence of an amenity is correlated with the usage of a station, a number of statistical measurements are considered. The mean of hours charged gives an indication towards whether the amenities are present at stations with high or low usage. The mean across all stations usage is 14.71 hours, which serves as a baseline for comparison. The mean hours charged i.e. for apartments is 14.06 hours, which is slightly below the baseline mean, indicating that apartments on average are more present at stations with lower usage. However, apartments are typically only present in urban areas, where the probability of deploying large charging stations is not as high as at more rural areas, such as at a rest area where the geographical area needed to deploy large stations is more likely to be available. As such, the presence of apartments at stations with lower usage is not necessarily an indication of a negative correlation. This also indicates that a future improvement to the model could be to split the model into urban and rural areas, to better account for the different types of stations. Opposingly, the maximum value of hours charged at a station with apartments present is 138.28 hours, which is significantly higher than the baseline mean, indicating that apartments are also present at stations with high usage. Therefore, determining weights of amenities for the general model is further evaluated.

Variable	100m	200m	500m	1000m	Proposed Weight
Apartments	12.75	13.06	14.42	14.06	40
Charging stations	17.56	17.08	15.95	15.75	Special case
Cinema	9.76	12.82	14.14	13.31	40
Clinic	14.24	13.93	14.93	13.97	45
College	13.42	11.62	20.96	20.12	50
Community centre	11.71	16.60	13.38	15.00	50
Conference centre	13.15	7.25	9.60	11.71	25
Dentist	14.07	12.49	13.54	14.04	40
Events venue	5.31	8.16	14.19	12.82	40
Fast food	18.93	18.55	15.96	15.73	75
Hospital	12.83	6.30	11.18	11.34	35
Hotel	15.20	11.21	14.11	13.17	45
Fitness centre	9.85	13.42	14.12	13.99	45
Leisure stadium		11.39	11.07	12.22	25
Motorway	16.87	16.20	22.83	20.43	60
Motorway link	22.79	25.85	21.64	20.47	80
Parking	13.35	14.92	14.84	14.82	Special case
Place of worship	14.19	13.19	12.43	14.19	40
Restaurant	13.23	14.81	13.49	13.21	45
Mall	17.87	22.53	18.54	17.43	60
Supermarket	13.01	14.26	13.57	13.20	35
Sports centre	5.44	16.32	14.31	13.78	50
Sports hall		5.90	18.51	15.24	55
Stadium			6.56	8.20	25
Substation minor distribution	15.79	17.90	16.38	18.63	25
Theatre	7.40	13.04	12.66	13.82	35
University		17.84	12.50	14.43	45

Table 8.3: Proposed weights based on multiple distance statistics.

To determine the weights, mostly the mean usage of stations where the amenity is present is considered. Amenities with higher mean than the baseline mean will generally be present at stations which are used more than average, and thus have a positive correlation with the usage of the station. To ensure the mean is representative, the minumum and maximum values are included in the decisionmaking. Exploring the extreme points in combination with the mean provides a more comprehensive understanding of whether the correlation between the amenity and the usage of a station is exclusively positive or negative. The process is repeated for each distance threshold, to explore the impact of the distance threshold on the correlation between the amenity and the usage of the station. A list of tables for each distance threshold is included in the appendix. The proposed weights, along with the mean value for each distance treshold is presented in Tab. 8.3.

Remember the baseline mean is 14.71 hours. A base assumption that all amenities are equally important would be equal weights for all amenities. For simplicity the weight is set to 50. The exact weight is irrelevant as higher weights will simply inflate the gravity score for a location, but the relative difference between scores at the same locations will remain the same. The proposed weights are determined based on a combinatorial evalutaion of amenities statistical measures across distances. For example, the mean of apartments is lower across all distances compared to the baseline model, resulting in a slight below average weight of 40. Another example could be colleges, where the mean for higher distances, 500m and 1000m is significantly higher than the baseline mean. However, at shorter distances the mean is significantly lower, indicating that the charging stations location in immediate vicinity of colleges are not as used as stations further away. An explanation could be that colleges are from manual examination located near motorways, which have stations with higher usage patterns. Intercorrelations between amenities should therefore be considered when manually determining the weights of each amenity. However, such intercorrelations are not considered in the manual weight determination, as it was quickly discovered that an automated approach lead to increased accuracy. The trade-offs between using a manual approach and an automated approach are discussed in Chapter 9. The discovered pattern for colleges lead to an estimated weight equal to the baseline weight of 50. The process was repeated for all amenities, where amenities with a high mean across all distances were assigned a higher weight, with a similar process for amenities with a low mean. Parking was assigned a special case, as it is represented at nearly all stations, and in the real world parking spots are mandatory for establishing charging stations. For those reasons, the weight of parking spots was set to 0, as the presence of parking spots should exist across all stations.

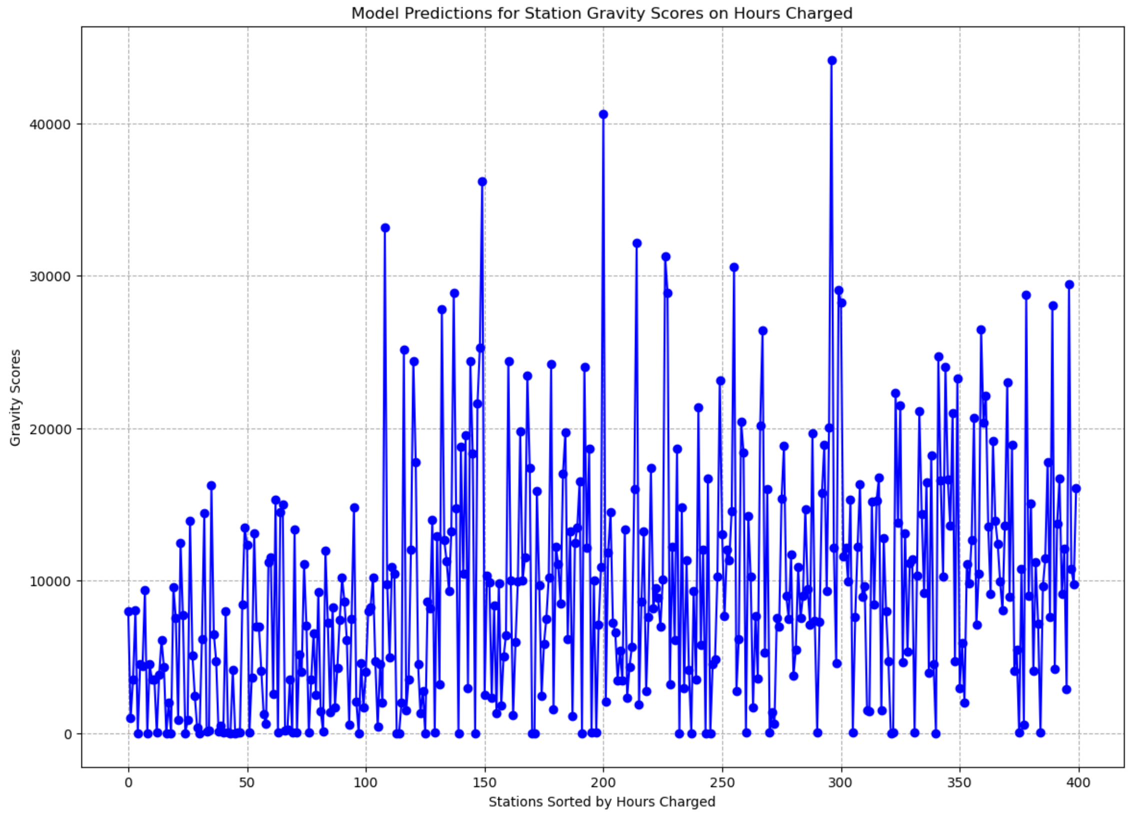


Figure 8.1: Gravity scores for each station based on the manual weighing scheme.

After estimating the weights for each amenity, the models ability to predict the actual use of stations is evaluated. Fig 8.1 shows the gravity scores calculated for each station based on the manual weighing scheme. It is immediately noticeable that the gravity scores do not follow a clear pattern when sorting the stations based on hours charged. This indicates that the manual weighing scheme may not be the most effective method for determining the weights of the gravity model. Anyway, the predictions are evaluated in comparison to the two normalized distributions of the actual values. Since the weights have been determined on a specific scale with a baseline weight of 50, the gravity scores have significantly different values from the scale of the actual usage of the stations based on hours used. Therefore, when evaluating the model based on the statistical weighing of amenities, both the predicted scores and the actual values are normalized aligning the scales. The results of the two predictions are depicted in Fig. 8.2 and Fig. 8.3,

Fig. 8.2 shows the predicted values compared to a normalized linear regression of the actual values while Fig. 8.3 shows the comparison with a normalized representation of the actual distribution. The errors are calculated using two different metrics, MSE and RMSE. The errors are shown in Tab. 8.4.

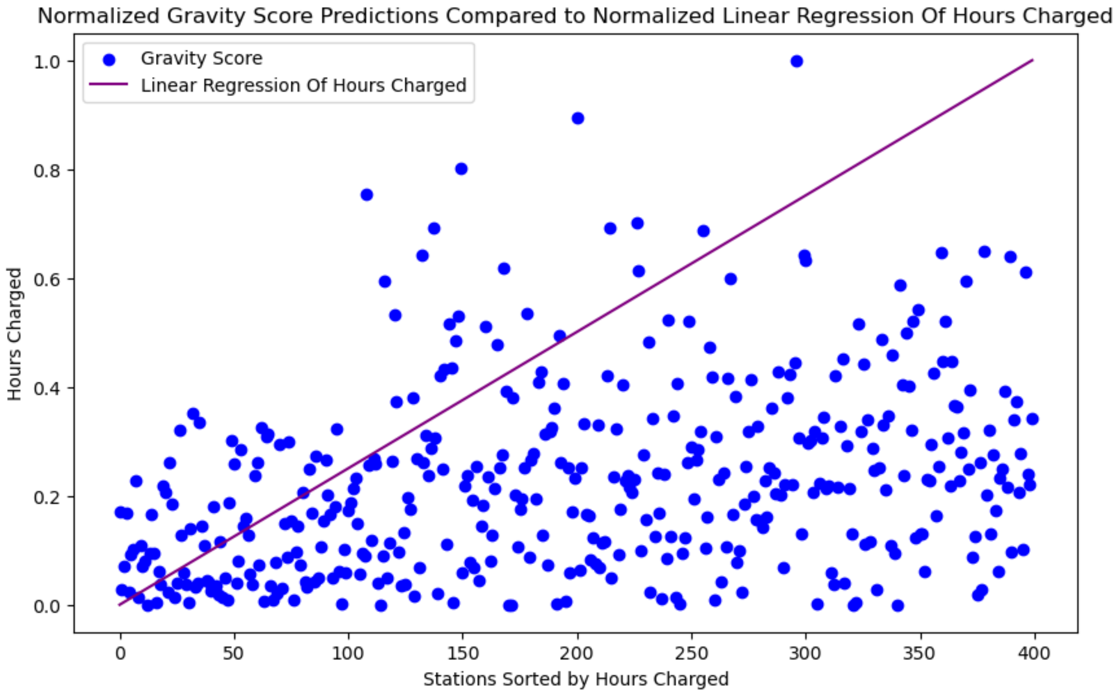


Figure 8.2: Normalized statistical weight prediction compared to the normalized linear regression of the actual distribution.

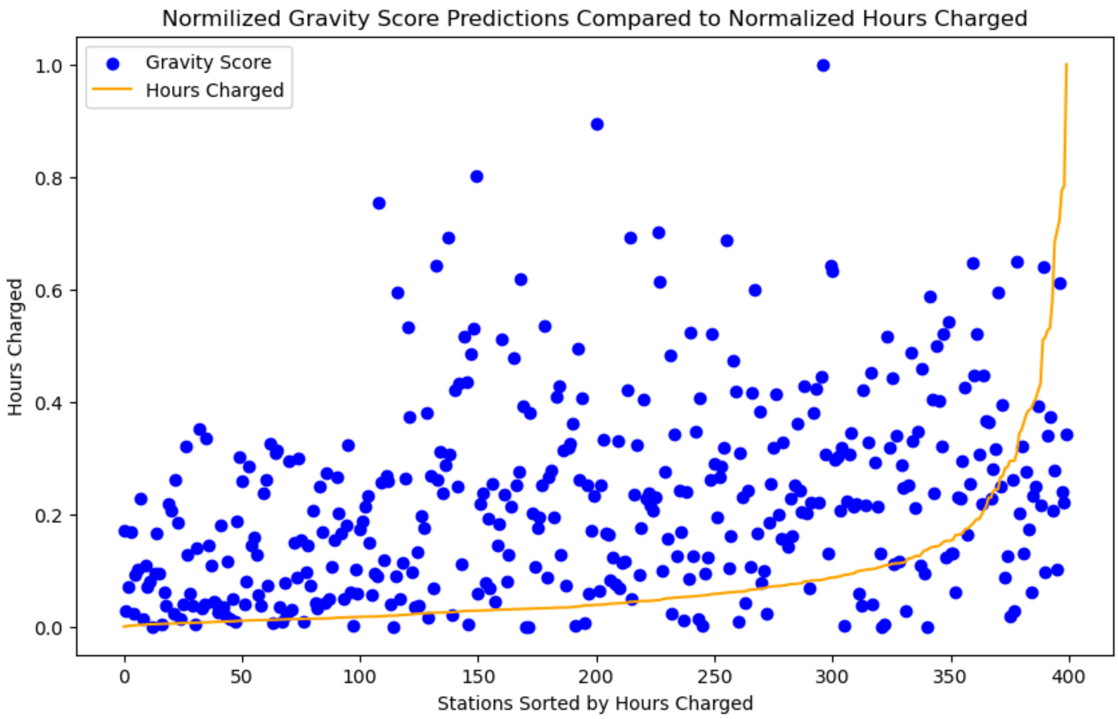


Figure 8.3: Normalized statistical weight prediction compared to the normalized actual distribution.

Distribution	MSE	RMSE
Normalized Linear Regression	0.16612	0.40758
Normalized Real	0.05816	0.24115

Table 8.4: Error of the model with manually determined weights.

Clearly the predictions do not follow the actual distribution which is also apparent in the relatively high error values. However the results are interpreted as a baseline for comparison with the automated optimization of weights using a genetic algorithm. Before looking at the results of the genetic algorithm, the results of the statistical analysis are interpreted. The main takeaway from the scores generated by the gravity model is that they have a small linear correlation with the actual usage of charging stations, however the model fluctuates heavily around the actual values. The next step was to examine whether the predictions of the gravity model could be improved by manipulating the weights of the amenities. To explore whether the model predictions could be optimized through changing the weights of the amenities, a genetic algorithm was used to manipulate the weights such that the model predictions would approximate the actual distribution of the stations.

8.2.2 Genetic Algorithm Weight Optimization

The automated weight analysis was conducted using a genetic algorithm. The genetic algorithm was configured with the hyperparameters described in section 7.2.3. Adjusting the weights through a genetic algorithm allows for evaluating the gravity model scores directly against the actual usage of the stations, as the algorithm is able to adjust the weights to fit the scale of hours charged. Therefore, the algorithm could be evaluated for all 4 optimization objectives. The algorithm is then trained to minimize the error of the gravity model predictions compared to the 4 different distributions. The weights assigned to each amenity for all 4 optimization objectives are shown in Tab. 8.6, and the loss of the best individual for each optimization objective is shown in Tab. 8.5. To compare the error values of the objectives with non-normalized values with errors from the objectives with normalized values, the normalized RMSE (nRMSE) is calculated for the non-normalized scenarios. The nRMSE is calculated by dividing the RMSE of the non-normalized scenario with the range of the actual values [47].

	Real	Real Lin	Norm Real	Norm Lin
MSE	466.779704	111.515884	0.020901	0.089106
RMSE	21.605085	10.560108	0.144572	0.298506
nRMSE	0.121472	0.298193	0.144572	0.298506

Table 8.5: MSE, RMSE and nRMSE values for four tests

The first optimization objective was to minimize RMSE when trying to predict the hours charged on the scale ranging from 0.08 to 177.98, which reflects the average daily hours charged at the stations. The GA was initialized to approximate the real distribution, and the resulting predictions are shown in Fig. 8.4.

Category	Real	Real Lin	Norm Real	Norm Lin
Traffic_weight	0	0	0	11.3581
Apartments_weight	0.000917	0.013021	23.7819	14.9822
Cinemas_weight	0.000370	0.000089	336.9063	0
Clinics_weight	0.068164	0.050440	3.0066	76.7804
Colleges_weight	0.045163	0.023629	9.2619	49.4167
CommunityCentres_weight	0.000004	0.060530	0	60.5257
ConferenceCentres_weight	0	0.002229	35.6065	76.9799
Dentists_weight	0.044545	0.012061	16.5360	0.2299
EventsVenues_weight	0.004641	0.007825	66.9387	11.3385
FastFoods_weight	0.178469	0.084388	126.7690	118.3753
Hospitals_weight	0.012238	0.006204	31.5542	7.7165
Hotels_weight	0.003030	0.000318	303.7520	7.3018
LeisureFitnessCentres_weight	0.002954	0.005111	16.9446	6.8645
LeisureStadiums_weight	0.000799	0.003152	0	19.0356
MotorwayLinks_weight	0.107890	0.044318	60.5025	108.8809
Motorways_weight	0.026636	0.029216	2.5937	5.4255
Parkings_weight	0.032387	0.081620	0	107.4836
PlacesOfWorship_weight	0.023493	0.001080	0.0104	25.1976
Restaurants_weight	0.023591	0.002206	0	0
ShopMalls_weight	0.047810	0	16.2061	15.1741
ShopSupermarkets_weight	0.000021	0.011644	15.8491	28.9283
SportsCentres_weight	0	0.001114	60.9115	50.4235
SportsHalls_weight	0.003479	0.038618	35.6527	38.1968
Stadiums_weight	0.101886	0.005514	45.1115	57.5513
SubstationMinorDistributions_weight	0.056027	0.053280	18.3372	73.6809
Theatres_weight	0.005573	0.006458	110.1540	0.1687
Universities_weight	0	0	73.3312	52.8279

Table 8.6: Assigned weights for each amenity for each optimization objective.

Large variations still exist in the predicted values, however, the predictions follow a trend somewhat similar to the actual distribution where less used station receive a low score, and locations related to stations towards the upper end receive higher scores. It is noted that the amount of stations with a prediction of exactly 0 is significantly reduced, which at first thought was a positive outcome. However, it was then realised that the genetic algorithm naturally adjusted the weight of parking spots, why stations with no amenities in the vicinity were assigned a non-zero value. Fig. 8.5 shows the evolution of the errors throughout each generation. The model converges towards a minimum error relatively quickly, with the error stabilizing between generation 5 and 10.

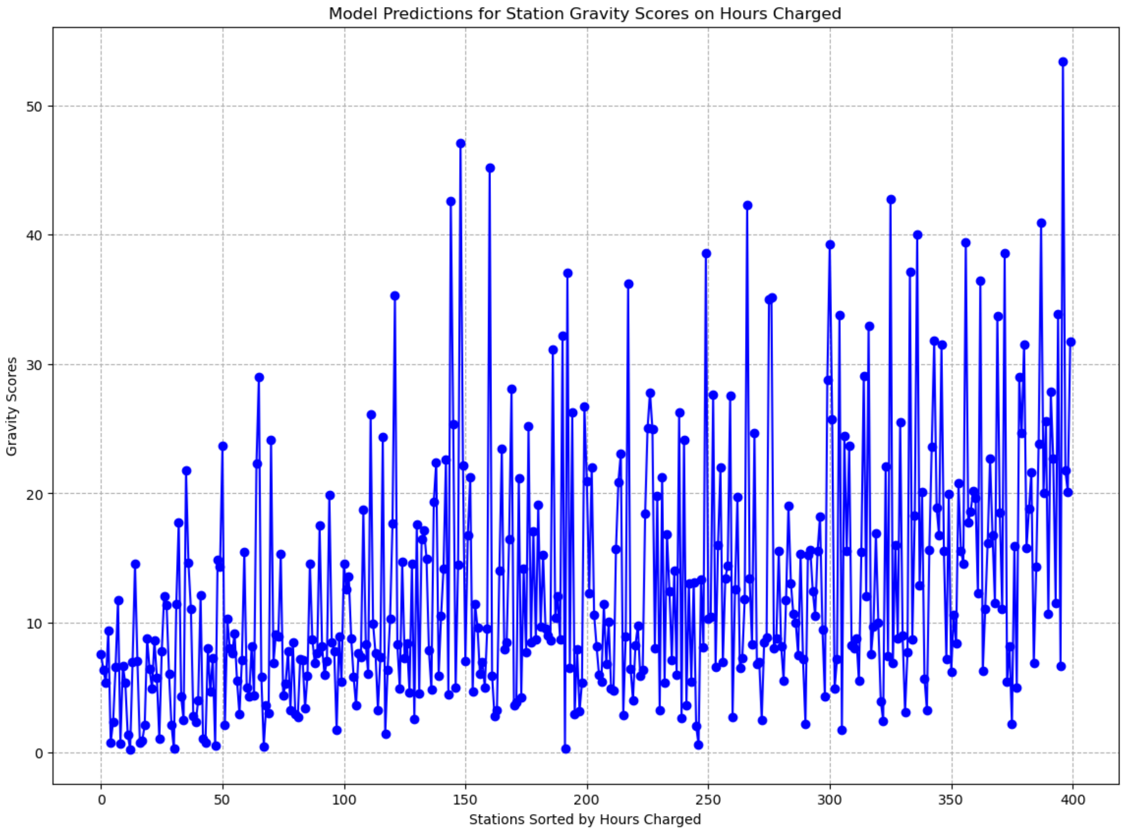


Figure 8.4: Predicted gravity scores for optimization objective 1.

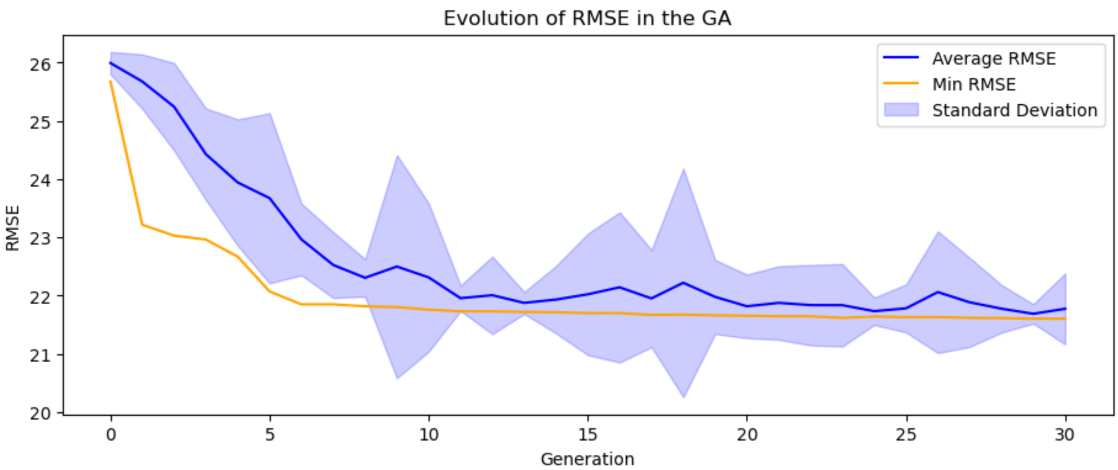


Figure 8.5: Training loss for optimization objective 1.

Fig. 8.6 shows the model predictions compared to the actual distribution. The model predictions appear to follow a linear tendency, and is not able to capture the exponential growth appearing towards the upper end of the distribution.

The second optimization objective tried to fit the model prediction towards the optimal linear representation of the actual distribution. The resulting predictions are shown in Fig. 8.7.

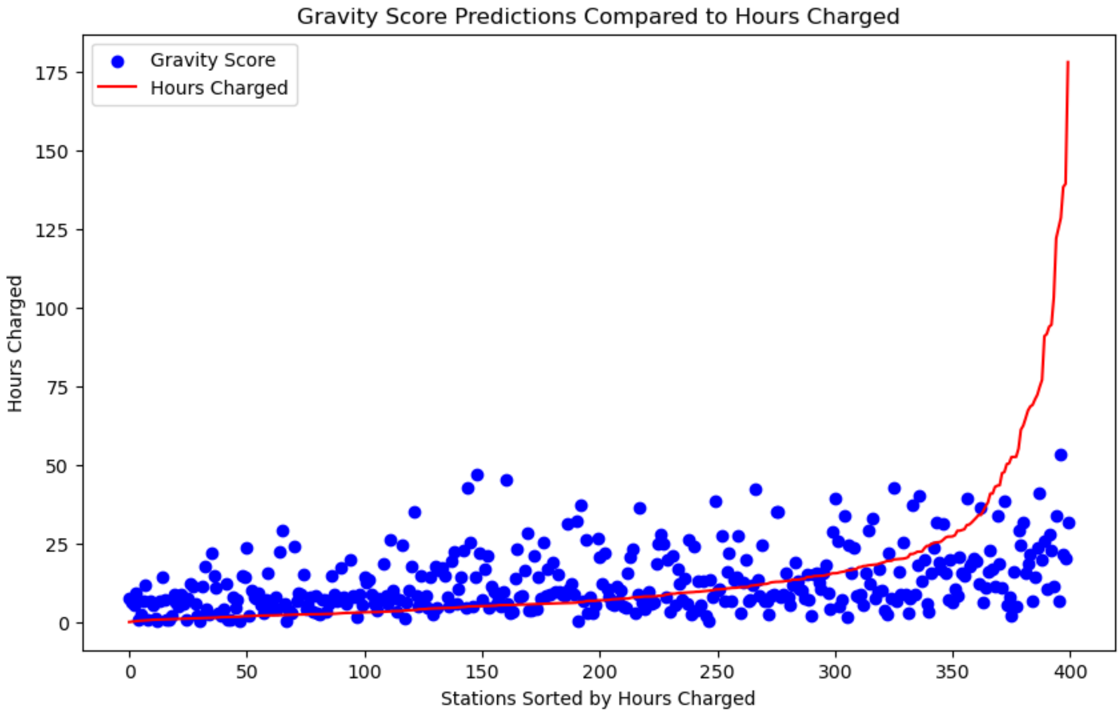


Figure 8.6: Predictions compared to the actual distribution for optimization objective 1.

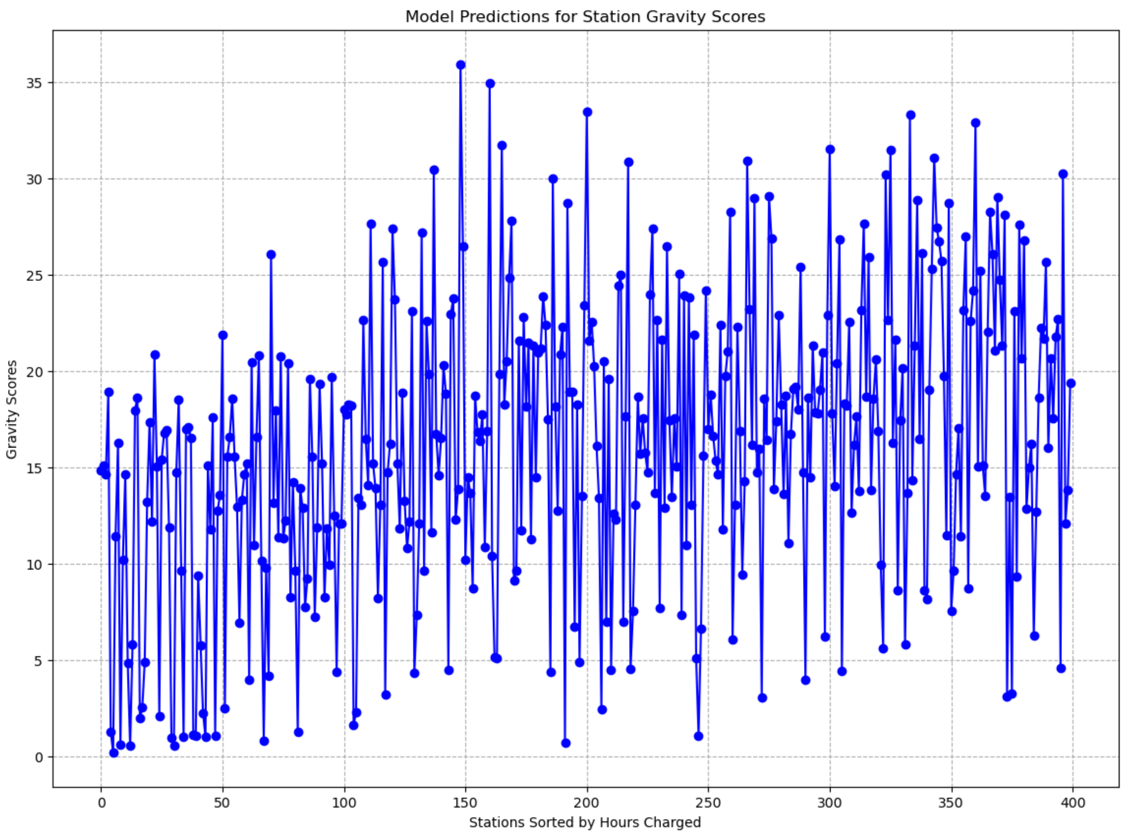


Figure 8.7: Predicted gravity scores for optimization objective 2.

The gravity scores obtain a more linear distribution in comparison to the first optimization objective. The predictions are more evenly distributed across the range of hours charged, where not as many stations receive an very low score in the range of 0-10 hours charged. Instead the majority of scores are distributed between 10-30 hours, with a increasing linear trend. The evaluation of the model performance is shown in Fig. 8.8.

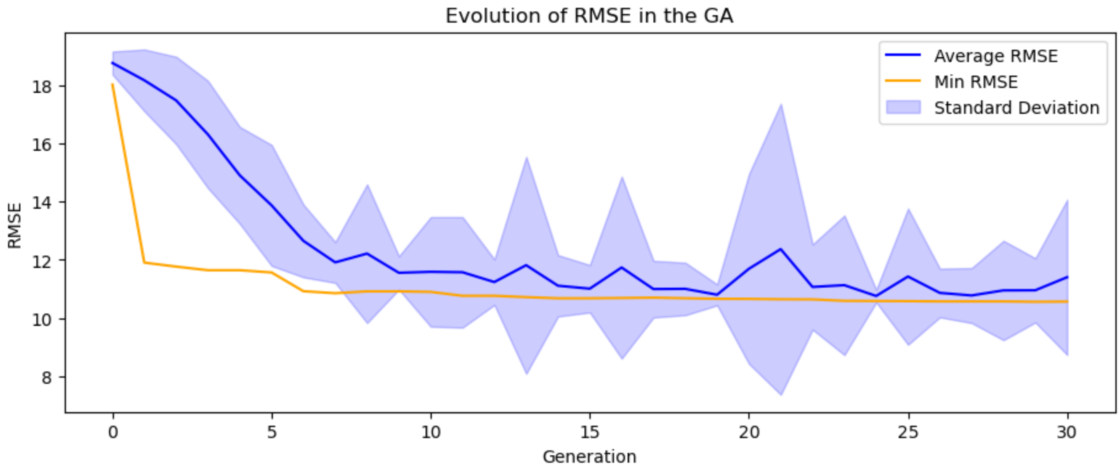


Figure 8.8: Training loss for optimization objective 2.

The model converges even quicker than the previous objective towards a minimum error, with the error stabilizing around the 7th generation. The graph illustrates that the GA tries to explore the solution, by applying mutation to the population, however the solutions seems to continuously converge towards the same minima. The predictions in contrast to the optimization goal is depicted in Fig. 8.9.

An interesting observation is the early tendency of overshooting, while predictions towards well visited stations are underestimated. These tendencies imply that there are limitations to the model, which might be explained by stations located near motorways, typically have high usage patterns, however a low number of amenities in the vicinity. This observation is further discussed later in the thesis.

The third objective was to approximate the actual distribution on a normalized scale. This involved normalizing both the actual values and the predicted values. The resulting predictions are shown in Fig. 8.10.

The results show a tendency that most predictions are pushed towards a low value, at the cost of sacrificing the prediction for certain locations. This is possible due to the normalization of the values, where increasing the weight of a single amenity can push all other scores towards a low value if the amenity is only present at a single or few stations. As a result the remaining associated figures are included in the github found in the appendix.

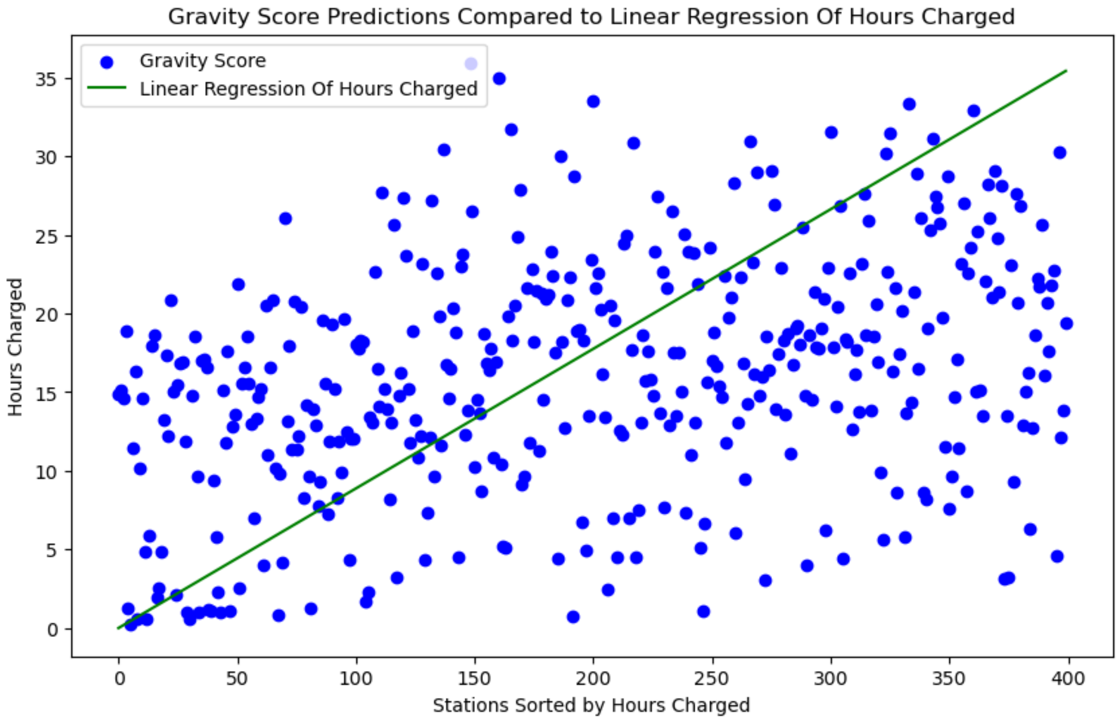


Figure 8.9: Predictions compared to the actual distribution for optimization objective 2.

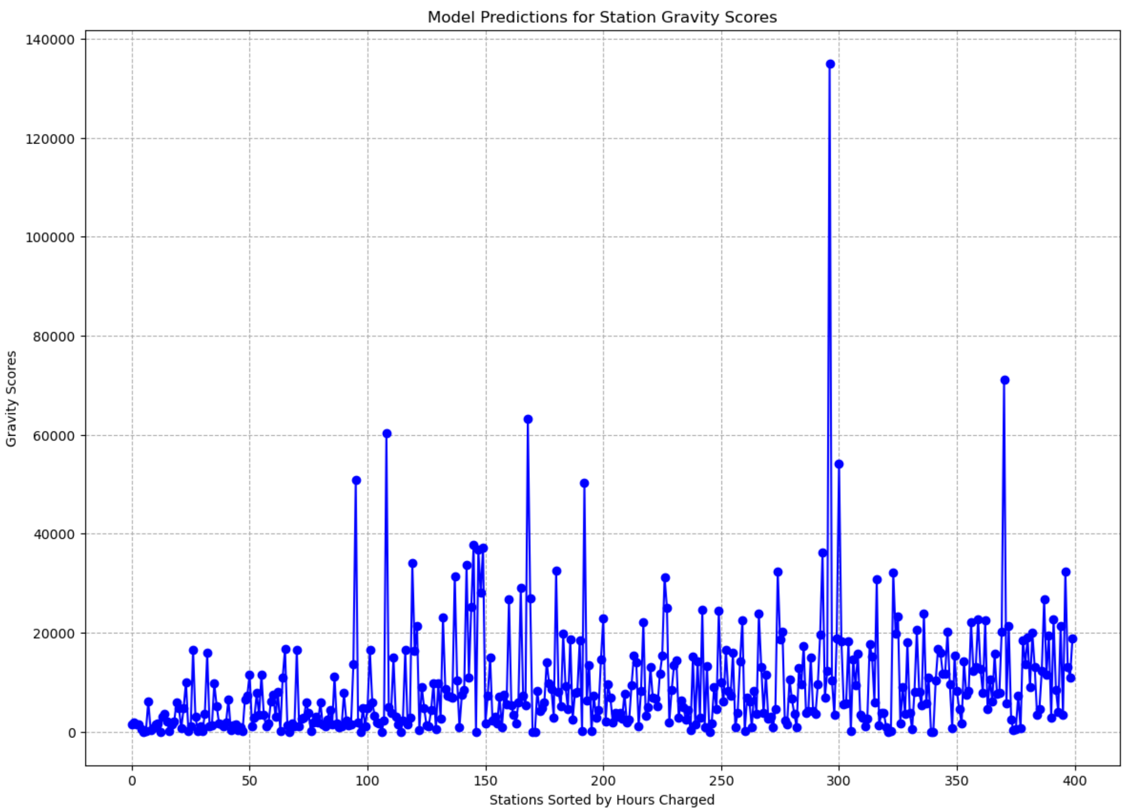


Figure 8.10: Predicted gravity scores for optimization objective 3.

The final optimization objective was to approximate the actual distribution on a normalized linear scale. The resulting predictions are shown in Fig. 8.11.

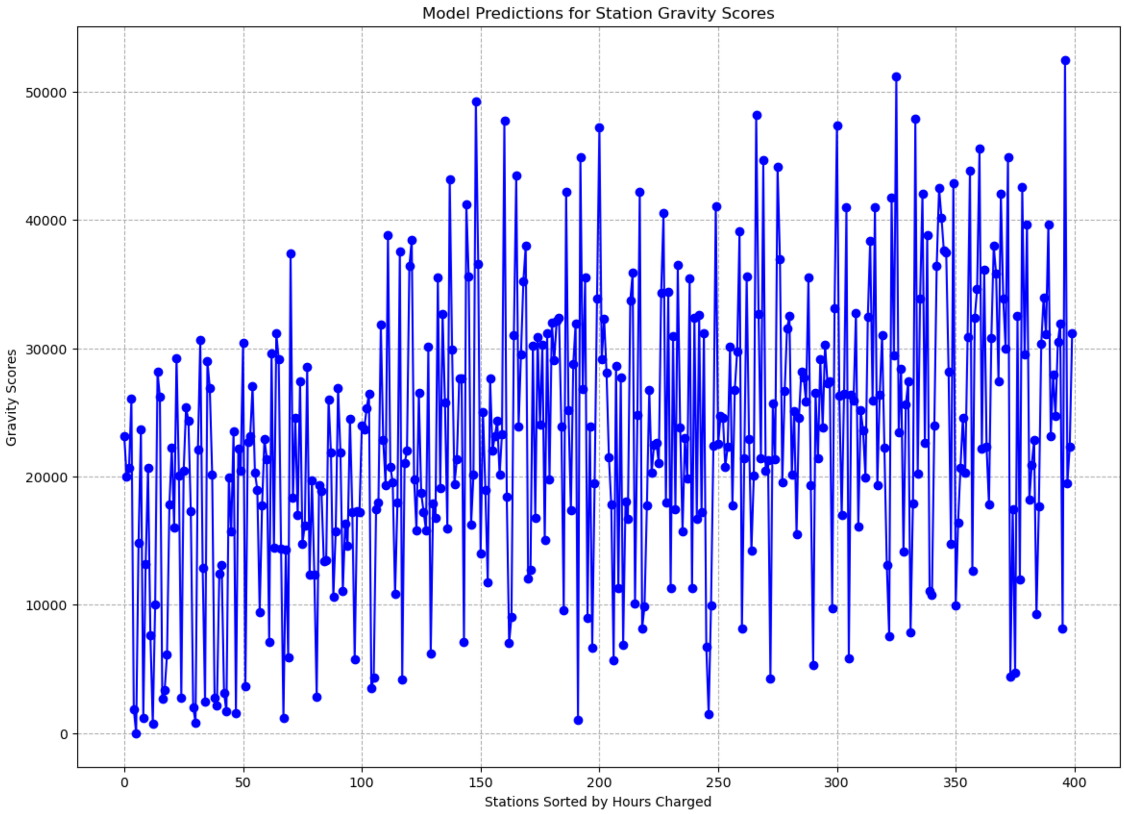


Figure 8.11: Predicted gravity scores for optimization objective 4.

The distribution of predictions is quite similar compared to the non-normalized linear distribution. Only small variations are observed, especially towards the upper end of the distribution, where predictions tend to have a slightly higher relative value. Evaluation of the convergence of the model is shown in Fig. 8.12.

The training shows that the model has an individual in the initial random population

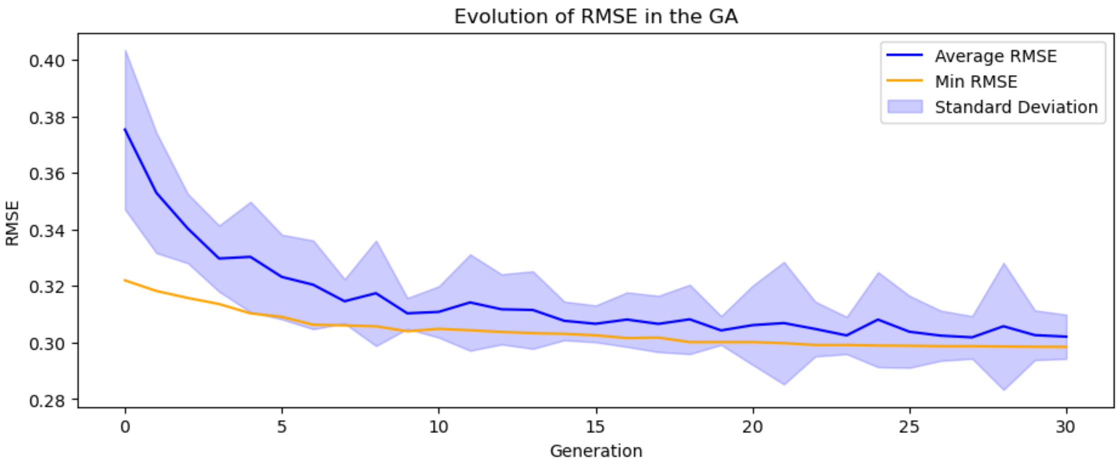


Figure 8.12: Training loss for optimization objective 4.

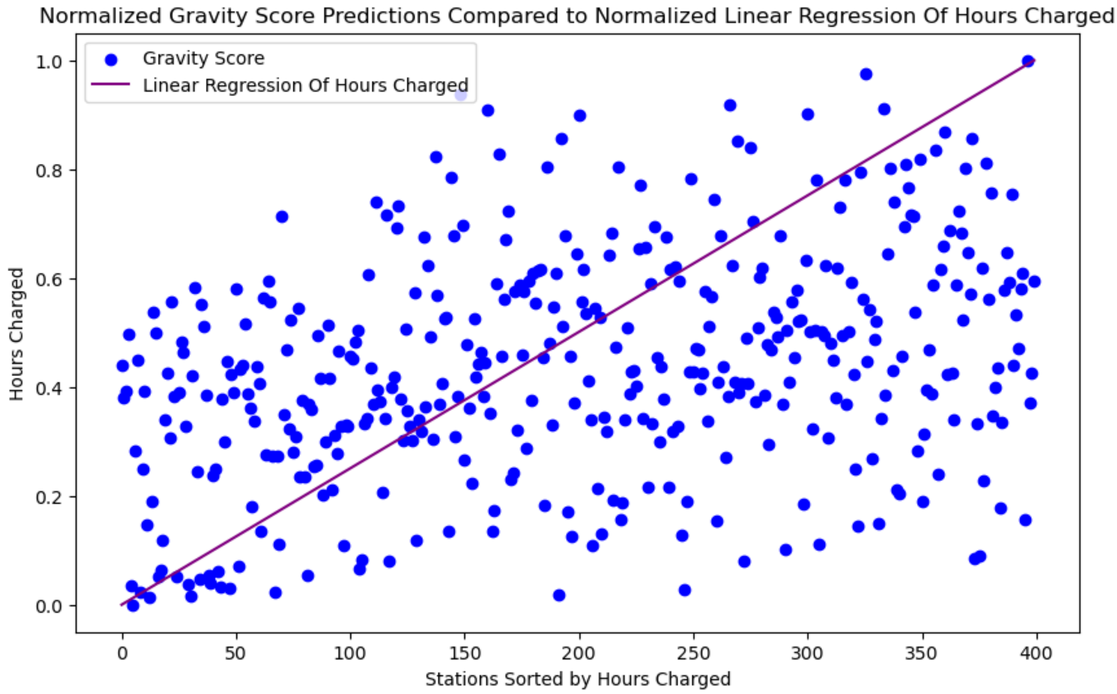


Figure 8.13: Predictions compared to the actual distribution for optimization objective 4.

which performs well, however the overall convergence is slightly slower than previous objectives. The comparison of the model predictions with the actual distribution is shown in Fig. 8.13.

The relation between predictions and actual values are similar to the previous linear objective, where especially predictions at either end of the distribution are far from the actual values. The tendency that the model underestimates the scores for stations with high usage and overestimates the scores for station with low usage is also consistent.

While looking at the accuracy of the model is important, the weights assigned to each amenity at each optimization objective is also of relevance. Looking at Tab. 8.6. Cases where the model tries to predict non-normalized values causes the weights to shrink. As such the weights are less interpretable, and not directly comparable across all 4 objectives. Despite weights not being directly comparable, it is obvious that the relative magnitude between amenities differs significantly across the 4 optimization objectives. The two most similar objectives in terms of prediction distribution were the linear and normalized linear objectives, which could indicate that the weights for these objectives were comparable in the sense that the magnitude of weights between each amenity would be similar. While there is some similarity, the solutions are not near identical. As a result, the solution space appears as complex, and a definitive singular solution can be hard to obtain. The combination of a complex solution space and the observed patterns in the predictions leads to the conclusion that the model is not able to accurately predict the usage of stations based on the entire set of amenities. Therefore, the model is split into two separate models, where one model tries to explain the correlation in an urban setting, while the other model aims to explain the correlation of stations located along motorways. Further explanation of the split model is provided in section 8.3.

8.2.3 Comparing Manual and Automated Weight Assignment

Comparing the results obtained through the statistically based manual weight assignment and the automated correspondant genetic algorithm, it is clear that the manual assignment of weights is not as effective as the automated assignment. However, the manual assignment included rational reasoning for the weights, ensuring that amenities that are present near well visited stations achieve a significant weight. As a result, the accuracy was lower than the automated process. As mentioned before, the predictions based of the automated process had tendencies to under- and overshoot at the high and low usage ends of the distribution respectively. Combined with the variation in the weight assignment, the complexity of the solution space was too large to obtain a definitive answer. To reduce the complexity of the solution space, stations are split into two categories, and the models are further evaluated.

8.3 Reducing Solution Space Complexity

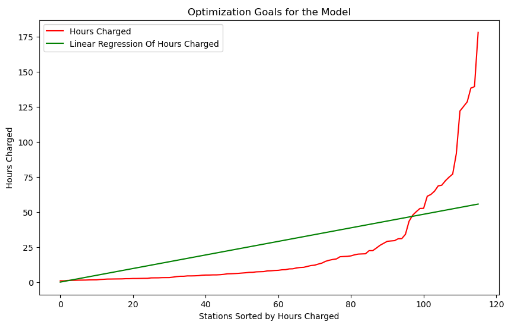
Reducing the complexity of the solution space is necessary to obtain a more accurate model, and to eliminate as many unknowns as possible. To achieve this goal, the stations were split into two main categories, and a small number of stations were removed from the dataset. The two categories are stations located along motorways, and stations located in urban areas. The easiest way to ensure a correct categorization of stations was to perform a manual geografical examination of each station, as categorizing stations based on amenity proximity is not a straightforward task due to motorways intersecting the urban areas. From the categorization, 266 stations were identified within urban areas, and 116 stations were identified along motorways. Stations located at the intersection of urban areas and motorways were included in both categories, which was the case for 6 stations. The categorization reduced the number of stations from 400 to 372 stations, where 6 stations were represented in both categories. The stations which were filtered out were stations located at locations where the the geographic area did not match with any of the two categories. An example could be a station located similarly to the station in Fig. 8.14, which is located in a rural area far from any urban area or motorways.

8.3.1 Evaluation of Motorway Model

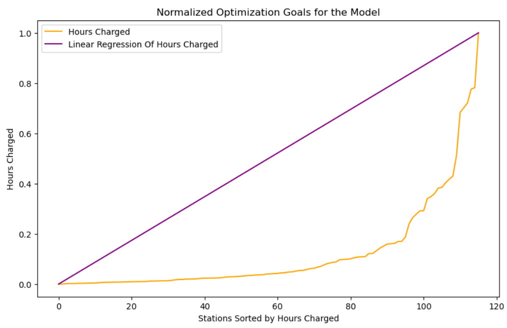
A new baseline model was trained on the stations located along motorways. The model was trained using an identical configuration as earlier, to test whether the model could capture the correlation between all the amenities and the motorway

Figure 8.14: Example of a station which was removed from the dataset during categorization.

stations. The distribution of stations along motorways, including the optimization objectives are shown in Fig. 8.15.



(a) Distribution and optimization goals.



(b) Normalized optimization goals.

Figure 8.15: Distribution of data and optimization goals for the optimized motorway model.

Generally, the model is mostly capable to capture linear tendencies. Therefore, the evaluation of the split models will primarily be based on the performance of the linear objectives. The difference in performance between the two linear objectives are negligible, while the proposed weights are more interpretable for the normalized linear objective. Thus, the evaluation of the motorway and urban models will be based on the normalized linear objective. The figures explaining the other objectives are included in the github. The predictions for the chosen optimization goal and accuracy for each optimization objective of the motorway model are shown in Fig. 8.16 and Tab. 8.7.

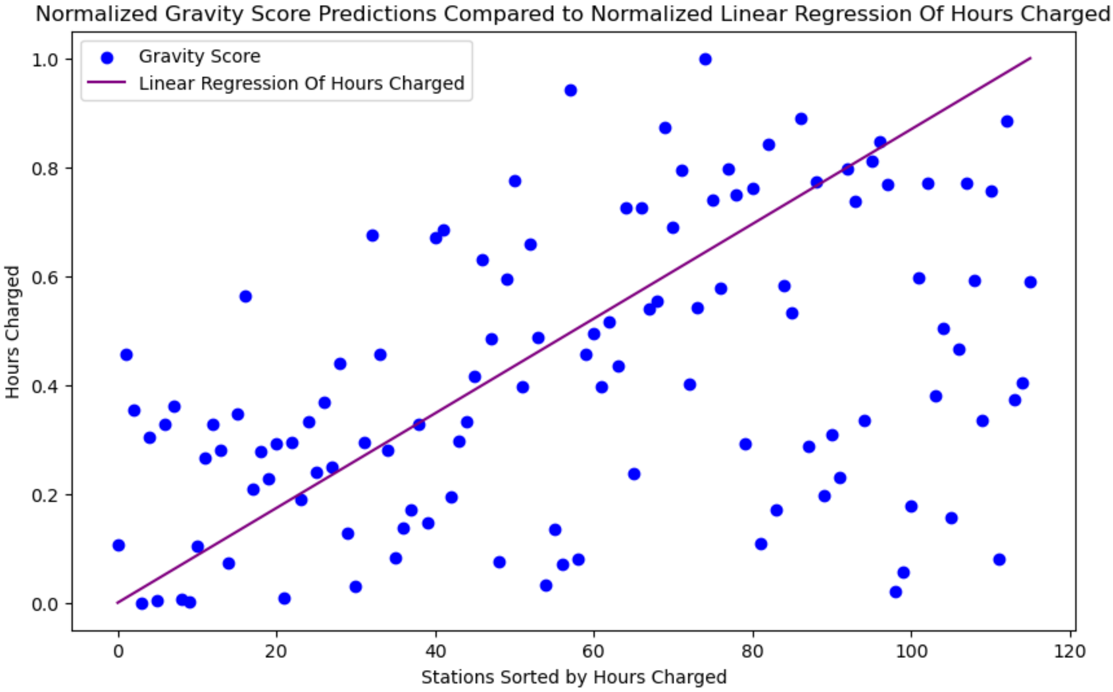


Figure 8.16: Baseline model predictions for motorway stations.

	Real	**Real Lin**	**Norm Real**	**Norm Lin**
MSE	1059.0819	284.2189	0.0418	0.0955
RMSE	32.5435	16.8588	0.2044	0.3091
NRMSE	0.1838	0.3032	-	-

Table 8.7: Baseline model evaluation for motorway stations.

With the data split into categories, a selection of amenities was removed to tailor the model to aminities relevant to stations located near motorways. This was also done in a manual process. Each station was examined to determine which amenities were present in the vicinity. Generally there were fewer amenities, and three types of stations were identified. The first type was located at rest areas where nothin but the station itself was present. The second type was located at gas stations which typically also have some sort of fast food restaurant and in some cases shops registered as supermarkets. The third type were stations located at the intersection of motorways and cities where a larger variety of amenities were present. To generalize the model towards stations relevant to motorways, only a few amenities were considered in the model. The original list of amenities considered, including traffic, consisted of 27 amenities. The list was reduced to 6 amenities, which were traffic, fast food places, supermarkets, motorway links, restaurants and motorways. The final modification was to shorten the distance which amenities were considered. An assumption was made based on personal experience, that the distance to relevant amenities would be shorter at stations along the motorway. The justification

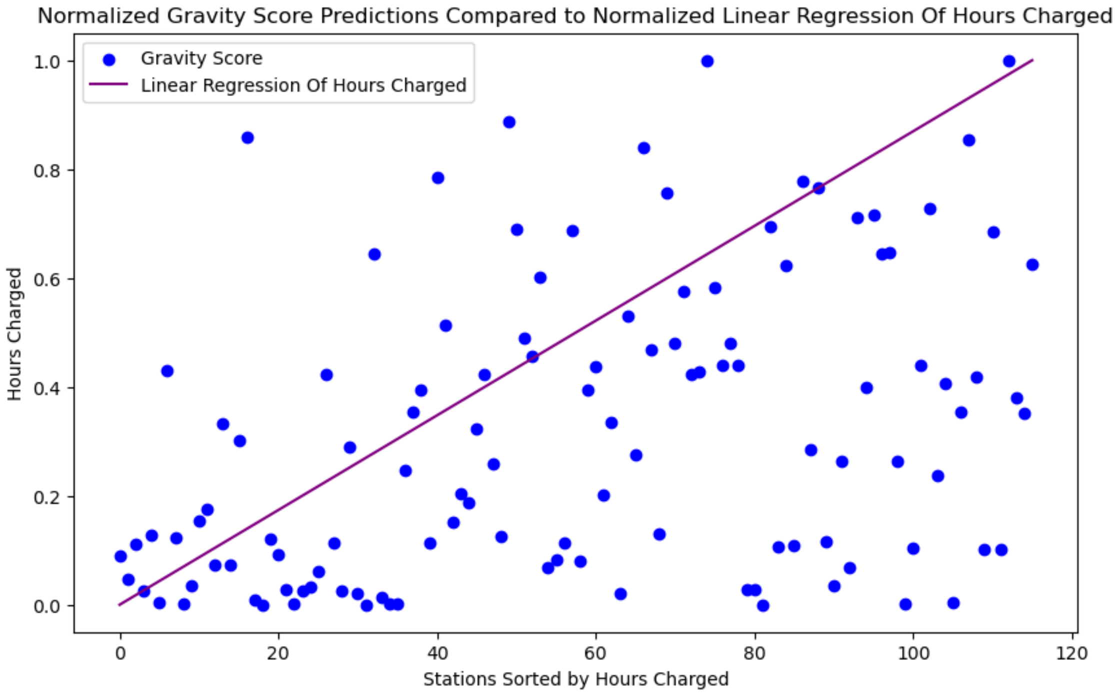


Figure 8.17: Optimized model predictions for motorway stations.

is that motorways are generally used for longer travels which requires charging the vehicle at some point. When charging at a point which is not the final destination, the stop is typically short, and the amenities which are relevant to the station are located in the immediate vicinity. Additionally, in the case where two stations are mirrored across the motorway, the score of the location should not be affected by the amenities on the other side of the motorway. The distance threshold was therefore reduced from 1000 to 250 meters.

A new run of evaluations was performed to examine the impact of the applied modifications. The results are described in Fig. 8.17 and Tab. 8.8.

	Real	**Real Lin**	**Norm Real**	**Norm Lin**
MSE	1100.5482	399.5075	0.0475	0.1312
RMSE	33.1745	19.9877	0.2179	0.3622
NRMSE	0.1873	0.3594	-	-

Table 8.8: Optimized model evaluation for motorway stations.

Comparing the performance of the optimized model to the baseline model of the motorway stations, the optimized model actually performs a bit worse across all evaluations. At first these results where suprising, which resulted in having to investigate the underlying reason. To do so, a number of stations were picked out and examined. The reason is partially demonstrated by the weights assigned to each amenity. The weights are provided in Tab. 8.9.

Starting off by comparing the weights of the amenities included in both models, the weighing schemes are somewhat similar. The most noticeable deviation is for the motorway amenity. The optimized model assigns a significantly higher weight to

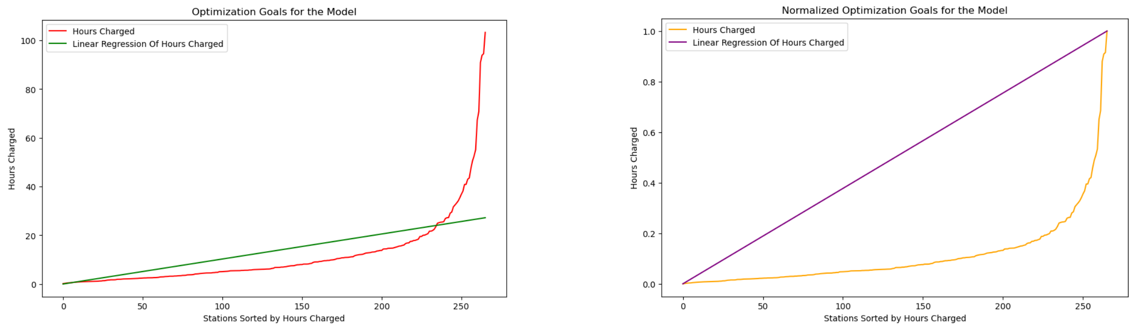
Category	Baseline Motorway	Optimized Motorway
Traffic_weight	0	0
Apartments_weight	64.075824	-
Cinemas_weight	0	-
Clinics_weight	44.899511	-
Colleges_weight	106.741930	-
CommunityCentres_weight	62.024438	-
ConferenceCentres_weight	7.152690	-
Dentists_weight	22.476864	-
EventsVenues_weight	73.866253	-
FastFoods_weight	148.578647	99.209856
Hospitals_weight	11.825605	-
Hotels_weight	0	4.934691
LeisureFitnessCentres_weight	36.728505	-
LeisureStadiums_weight	14.492850	-
MotorwayLinks_weight	38.928719	35.860793
Motorways_weight	2.575250	99.111507
Parkings_weight	77.258026	-
PlacesOfWorship_weight	25.584316	-
Restaurants_weight	48.262772	66.901433
ShopMalls_weight	0.389973	-
ShopSupermarkets_weight	8.780407	44.843685
SportsCentres_weight	102.766357	-
SportsHalls_weight	21.836938	-
Stadiums_weight	6.594788	-
SubstationMinorDistributions_weight	58.224335	-
Theatres_weight	0	-
Universities_weight	90.531584	-

Table 8.9: Weights for Baseline and Optimized Motorway Models

the motorway amenity, thus relying more heavily on the amount of vehicles passing by the area. The baseline model assigns a very low weight to the motorway amenity, implying that the number of passers by is not relevant for the model. The baseline model has the luxury of being able to compensate by adjusting the weights of a larger set of amenities. Cases where a type of amenity is only present at a single station, say a station with a high usage pattern, the weight of that amenity may be excessively inflated while maybe not having any impact in a real life scenario. Thus only amenities which are present at multiple stations need to be balanced to fit the optimization objective. As a result, the baseline model will have the advantage of being able to approximate the extreme cases better, while the optimized model is limited to a more generalized solution.

8.3.2 Evaluation of the Urban Model

Similarly to the motorway model, a new baseline model was trained however, on the stations located in urban areas. The new set of stations result in a new distribution, why the optimization goals again were tailored to the new distribution and are shown in Fig. 8.18.



(a) Distribution and optimization goals. (b) Normalized optimization goals.

Figure 8.18: Distribution of data and optimization goals for the optimized urban model.

The predictions for the normalized linear objective and accuracy for each optimization objective of the urban model are shown in Fig. 8.19 and Tab. 8.10.

	Real	Real Lin	Norm Real	Norm Lin
MSE	210.2199	59.8570	0.0257	0.0803
RMSE	14.4990	7.7367	0.1604	0.2834
NRMSE	0.1406	0.2844	-	-

Table 8.10: Baseline model evaluation for urban stations.

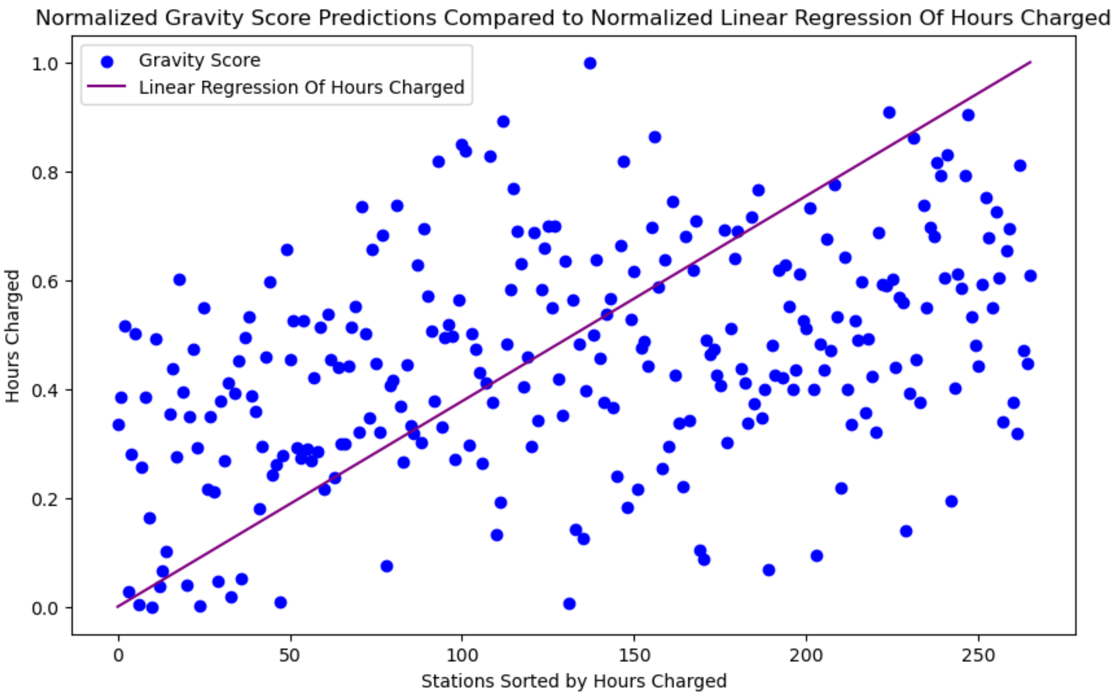


Figure 8.19: Baseline model predictions for urban stations.

An optimization of the urban model was performed in a similar manner to the motorway model. The first modification was to remove the motorway and motorway links, as these were determined not relevant in the case of urban stations. Secondly, the parking amenity was removed because charging necessitates a parking spot, why placing a station at a location without the possibilty for parking is infeasible. A re-evalution of the distance threshold for the urban model resulted in maintaining the original distance threshold of 1000 meters. The reasoning behind was that charging the car in a city setting will often relate to the final destination. Additionally, stops are typically longer as errands are run when visiting the city. In summary, the only modification was recuding the number of amenities to 24. The results of the optimized urban model are shown in Fig. 8.20 and Tab. 8.11.

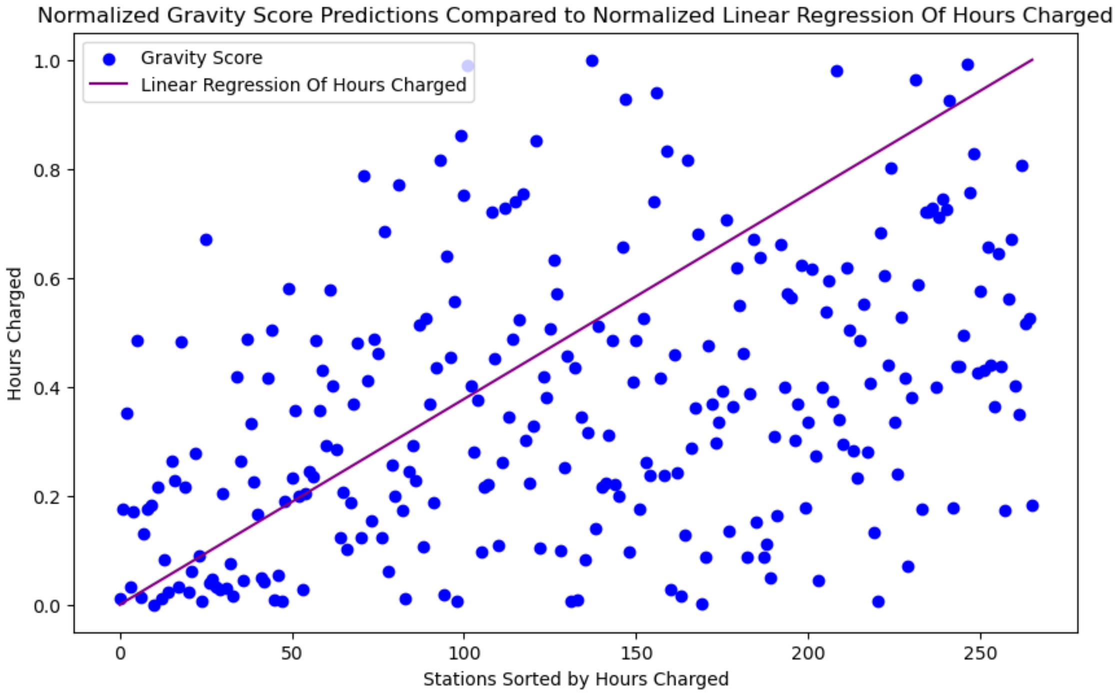


Figure 8.20: Optimized model predictions for urban stations.

Metric	Real	Lin	Norm Real (scaled)	Norm Lin (scaled)
MSE	217.1657	75.7455	0.0262	0.1059
RMSE	14.7365	8.7032	0.1620	0.3254
NRMSE	0.1429	0.3199	-	-

Table 8.11: Optimized model evaluation for urban stations.

The results of the optimized urban model show similarly to the motorway model, that the optimized model performs worse than the baseline model. A key component to the performance of the model is the removal of the parking amenities impact. By removing the parking amenity which was present at the vast majority of stations all predictions are pushed towards a lower value. Especially stations which rely heavily on the influence of parking amenities move towards a near zero value. This is further supported by the weights assigned for each amenity in the urban model, which are shown in Tab. 8.12.

Looking at the weights, starting with parking, a clear difference is observed. In the baseline model parking receives the highest weight of all amenities, while in the optimized model parking has been removed and thus cannot have an influence. The high weight is assigned to boost stations which do not a high number of amenities

Category	**Not Optimized**	**Optimized**
Traffic_weight	85.381931	163.853855
Apartments_weight	32.421863	50.229553
Cinemas_weight	38.263511	0.776341
Clinics_weight	105.353098	163.378273
Colleges_weight	27.784277	33.529536
CommunityCentres_weight	64.952296	6.014114
ConferenceCentres_weight	72.520957	219.133318
Dentists_weight	1.697793	0.971615
EventsVenues_weight	9.086568	0.362813
FastFoods_weight	95.488811	202.443266
Hospitals_weight	4.612098	22.661898
Hotels_weight	8.205093	2.008185
LeisureFitnessCentres_weight	53.385352	41.741974
LeisureStadiums_weight	1.355816	52.198716
MotorwayLinks_weight	20.171741	-
Motorways_weight	5.024778	-
Parkings_weight	116.464135	-
PlacesOfWorship_weight	25.344908	58.861586
Restaurants_weight	22.263102	88.695833
ShopMalls_weight	6.518590	6.634251
ShopSupermarkets_weight	55.238020	89.282856
SportsCentres_weight	15.461831	101.631898
SportsHalls_weight	1.087380	32.071118
Stadiums_weight	36.910138	51.305717
SubstationMinorDistributions_weight	91.555408	85.704374
Theatres_weight	1.590670	0
Universities_weight	27.886506	0

Table 8.12: Weights for Baseline and Optimized Urban models

in the vicinity, which is a problem given that parking is not optional but a necessity why it should not influence the score. A domino effect is caused by the removal of the parking amenity, where the weights of the remaining amenities are adjusted to compensate for the missing value associated with parking.

8.3.3 Hyperparameter Tuning of Genetic Algorithm

As for all machine learning models, the hyperparameters of the genetic algorithm can have a significant impact on the performance of the model. In the context of evolutionary algorithms this concept is referred to as parameter setting [48]. Parameter setting can be further divided into two categories, parameter tuning and parameter control [46]. Fig. 8.21 describes the global taxonomy of parameter setting in EAs. To search for the optimal hyperparameters, two methods were followed, parameter tuning and adaptive parameter control. Initially, parameter setting should have been performed earlier on in the process, however due to computational com-

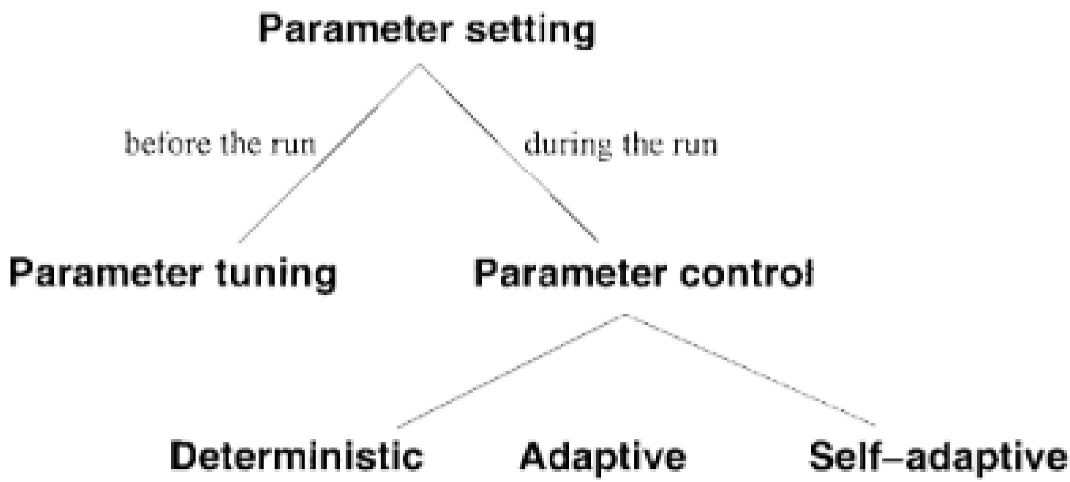


Figure 8.21: Global taxonomy of parameter setting in EAs [46].

plexity the parameter setting was performed after investigating the solution space complexity, and splitting the model.

For GAs the parameters are population size, number of generations, mutation rate, and crossover rate. Adaptive parameter control is a method where the parameters are adjusted during the optimization process, based on feedback from the optimization process. The adaptive control was based on the following rules:

- If the average fitness of the population increases, the crossover rate is increased by 5%. However, an upper bound is set at 90%.
- If the average fitness of the population increases, the mutation rate is decreased by 5%. However, a lower bound is set at 10%.
- If the average fitness of the population decreases, the crossover rate is decreased by 5%. However, a lower bound is set at 40%.
- If the average fitness of the population decreases, the mutation rate is increased by 5%. However, an upper bound is set at 40%.

To reduce the computational compelxity as much as possible, the parameter setting was conducted on the motorway model, due to a significant reduction in number of weights to optimize. The adaptive parameter control was initialized with the same starting parameters as all other models. The number of generations was static at 30, and the populations size similarly remained at 50. The results are shown in Fig. 8.22.

The best measured error was an RMSE of 0.3622, which is exactly the same error as for the optimized motorway model. As seen in the convergence happens at the same speed as is stable around the minimum value even through multiple increases

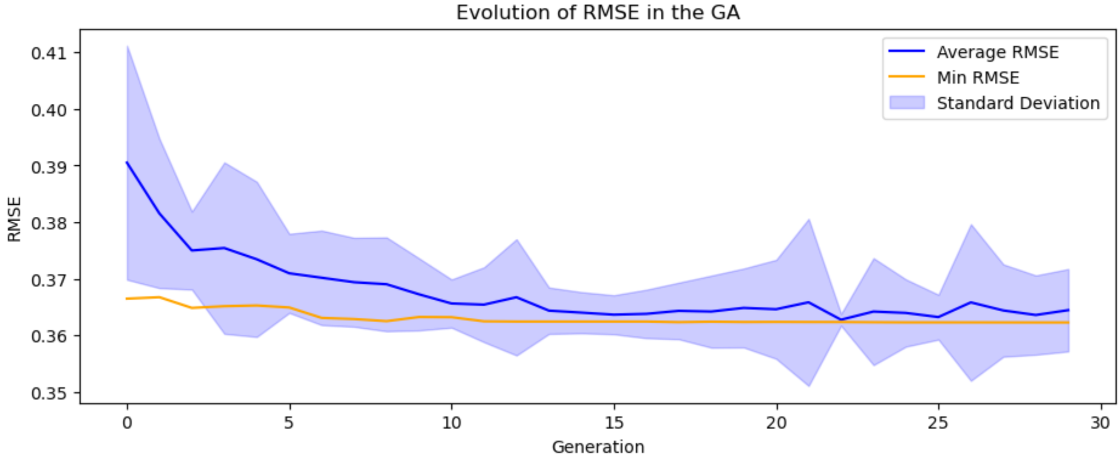


Figure 8.22: Adaptive parameter control of the genetic algorithm.

in mutation rates. While the model converges after relatively few generations, which might result in the adaptive control to not have a significant impact, the adaptibility of the hyperparameters was not succesfull in improving the model. Therefore another approach was taken to look for potential optimization of the hyperparameters, which was simply to perform a grid search on the parameter tuning side of the taxonomy. The grid search was performed on the optimized motorway model, and configured with the parameters:

- **Crossover rate:** $[0.3, 0.4, 0.5, 0.6, 0.7]$
- **Mutation rate:** $[0.1, 0.2, 0.3, 0.4, 0.5]$

The core difference is that the parameters are static, and set at the initialization of the model. As a result, the model will explore all the different tuning options before convergance. The results of the grid search showed that tuning of the parameters did not have a significant impact on the performance of the model. The error values were all within the range of of a 0.000x difference. For all error values the reader is referred to the appendix. The conclusion is therefore that the hyperparameters will not influence the model, and efforts towards improving the model should be focused on the quality of the data. This will be further discussed in Chapter 9.

8.4 Predicting Average Daily Hours Charged at New Locations

The final experiment was designed to investigate whether the scores calculated by the gravity based model(s) could be used to predict a more concrete value for the usage of new charging stations. The prediction is based on using all predetermined features, which mayb influence the usage of the station. Those include first and foremost the total gravity score for the locations, minimum capacity, maximum capacity, the number of plugs and the owner of the station. The prediction was performed using a linear regression model. A model was trained for each of the two categories. Given the limited number of stations in each category k-fold cross validation was used to asses the performance and generalization of the model, where $k = 10$ for the urban model and $k = 5$ for the motorway model. The results are shown in Tab. 8.13.

Dataset	Metric	Urban Mean ± Std	Motorway Mean ± Std
Baseline	R^2	0.5978 ± 0.2463	0.6161 ± 0.0906
	MSE	68.2816 ± 27.4380	364.6569 ± 202.3924
	RMSE	8.0818 ± 1.7223	18.1351 ± 5.9814

Table 8.13: Error metrics and R^2 for a number of configurations, where w/o denotes the leaving out of a feature.

The baseline model describes the case where all independent variables are used to predict the daily average hours charged. Both models have a R^2 value of approximately 0.6, meaning that 60% of the variance can be described by the combination of all independent variables. However, this value fluctuates significantly when repeating the experiment with different splits of the data, thus the model does not generalize well. This can be explained by either the complexity of the model, or the quality and quantity of data. Given the small sample size, and the gained knowledge of lackluster data quality, the latter is the most likely explanation. As a result, investigating the impact of the gravity feature was not further explored, as any obtained results would not have any statistical significance.

8.5 Heatmap Visualization of Gravity Scores

Visualizing the gravity scores provides a more intuitive understanding of how the gravity score is distributed across the country. To visualize the gravity scores, a simple application was developed to display the scores on an interactive map. The application was developed as a tool to identify geographical areas on a larger scale than having a short list of candidate locations for charging stations. The application has an interactive map where the user can select the area of interest, and the application will divide the area into a grid of 400x400 points. Gravity scores are calculated for each point in the grid, and the points are displayed on the map. The model used for calculating the gravity scores for the application is the model trained on the entire set of data. Unfortunately, there was not time to adjust such that stations along the motorway and in urban areas were calculated by seperate models. The following figures show a few examples of the visualization tool, where gravity scores are displayed in different areas and at varying granularity.

Fig. 8.23 displays a 400x400 grid of gravity scores across the entire country of Denmark. Given the relatively small granularity of the grid, the model generalizes urban areas as hotspots. Individual spots along larger roads such as motorways and access roads also light up as potential spots identified by the model.

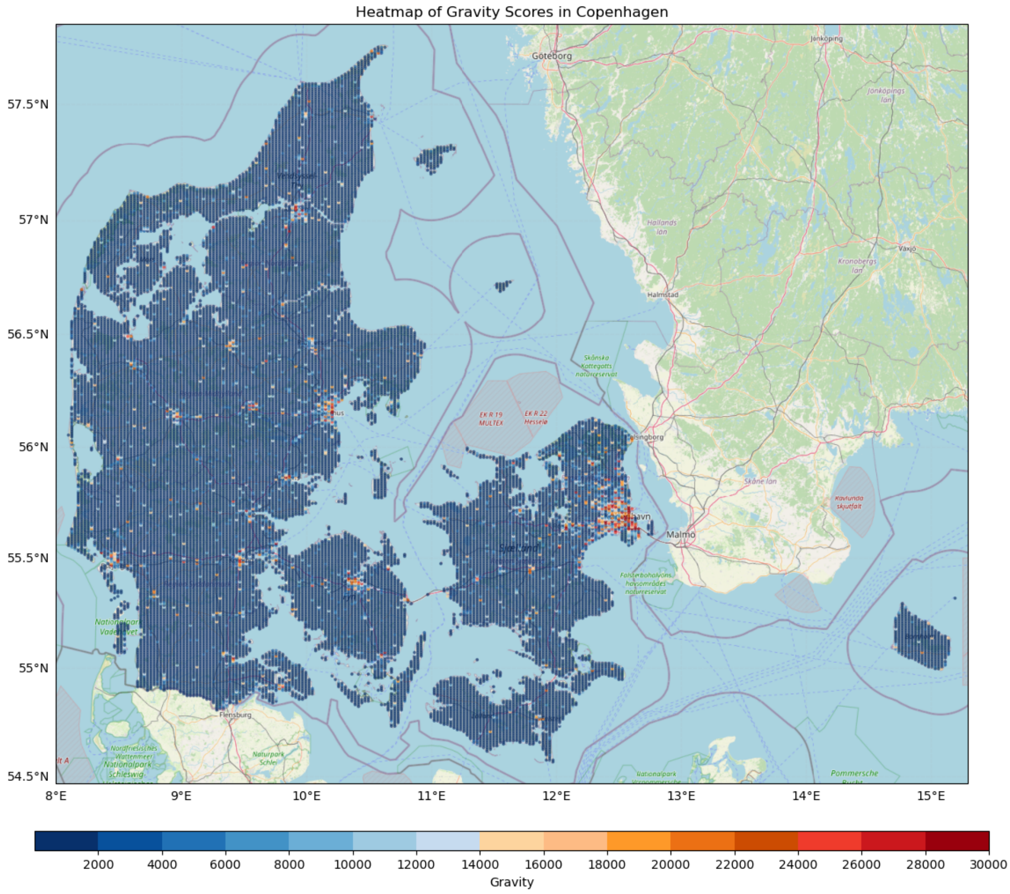


Figure 8.23: Heatmap of gravity scores across Denmark.

In order to inspect the relative distribution of gravity scores in more detail, the boundaries of the grid is adjusted to focus on a smaller area. Fig. 8.24 displays the gravity scores in and around the capital of Denmark, Copenhagen. The granularity of the grid is still 400x400, however the grid covers a significantly smaller area, which allows for a detailed inspection of the gravity scores in the area. Note that for the more detailed grids, the transparency of the points is increased to emphasize the map underneath. The colors should therefore not be interpreted 1:1 with the colors in the figure over the entirety of Denmark.

Within the inner city of Copenhagen, the gravity scores are generally high and indistinguishable from each other. However, this is caused by a mix of granularity and that points within the inner city will have higher scores compared to the rest of the area. To avoid this, it is possible to further increase granularity to only look at the inner city, and define the color range based on the updated range of scores which is reflected in Fig. 8.25. Additionally, the model avoids placing stations in infeasible places such as parks, lakes and above ground central railways.

Overall, the model is able to assign high scores to popular areas, and low scores to locations where deploying a charging station is infeasible.

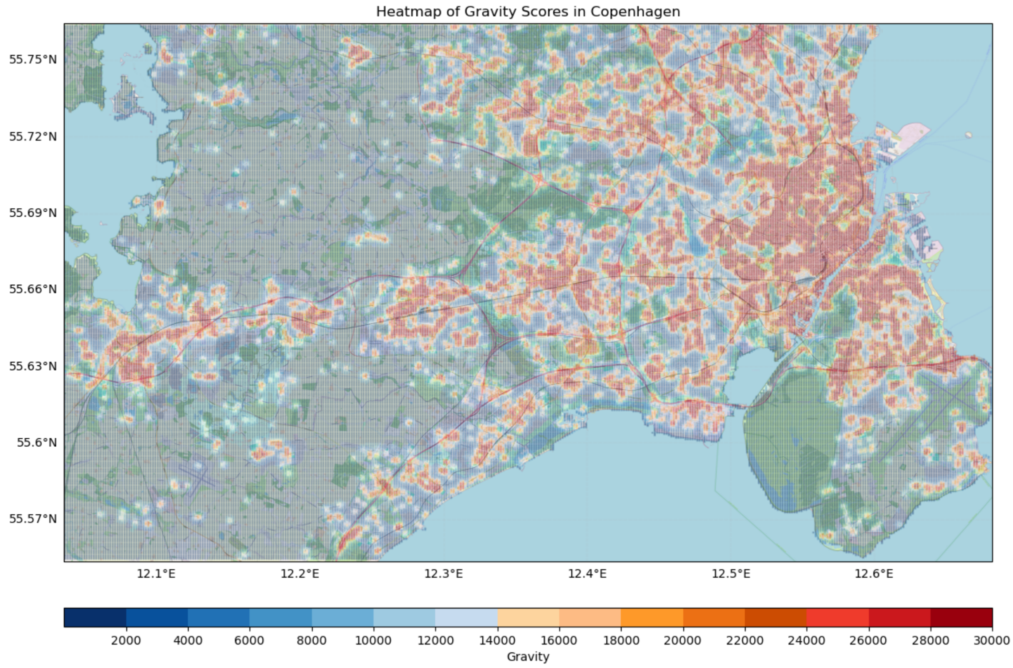


Figure 8.24: Heatmap of gravity scores in and around Copenhagen.

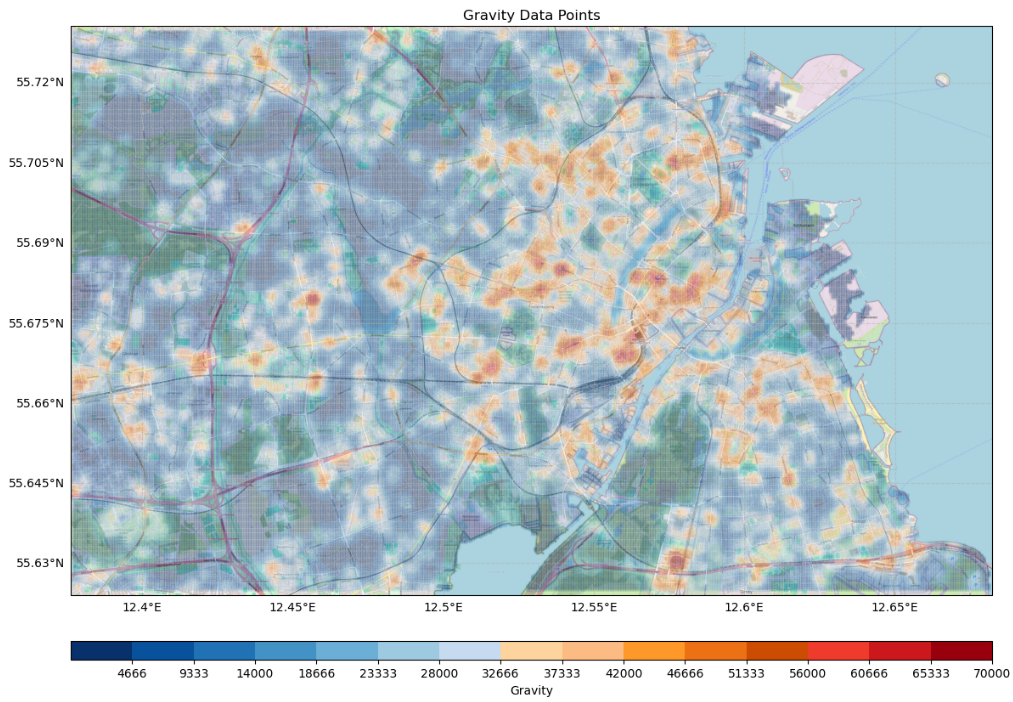


Figure 8.25: Increased granularity and update color range for gravity scores in the inner city of Copenhagen.

The application can be accessed on this github repository. To run the application, navigate to the folder named MapApp and follow the instructions provided in the README file.

8.6 Hybrid Forecasting Model for Station Usage Prediction

The work in this thesis has focused on the development of a gravity based model, and has been evaluated on the ability to predict usage of charging stations particularily at new locations. While the gravity based model has shown the ability to rate the potential of a location, the model is not very succesfull at predicting the usage of a station. To improve prediction accuracy the integration with other models was considered. As mentioned in the acknowledgements, the work in this thesis has been in collaboration with both Norlys and another thesis which has similarly been conducted in collaboration with Norlys. The other thesis has focused on forecasting the charging load of already existing charging stations in Denmark. The ambition has throughout the process been to integrate the findings of the two theses, with the aim of finding a solution which is able to predict the usage of stations at new locations. Forecasting models are analytical tools which use historical data to predict future trends. The models use statistical methods to analyze patterns in the historical information. Common forecasting models include Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing (ETS) and several machine learning models like neaural networks. Specifically, Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks and Gated Recurrent Unit (GRU) have been applied in recent years to time series forecasting [49].

The integration between the two theses has been performed by combining the gravity based model developed in this thesis with a hybrid model consisting of a GRU network and a Graph Neural Network (GNN) primarily developed in the other thesis. Thus, the exact functionality of the hybrid model is not described in this thesis, however a brief description is provided along with the efforts made to integrate the two models.

In order to facilitate the integration of gravity scores into a forecasting setting, a hybrid model was created with the purpose of predicting usage by using both spatial and temporal information. As mentioned, the model consists of two main components, a GNN and a GRU. The GNN is responsible for creating a graph representation of the charging stations, where stations are represented as nodes in the graph and edges are the euclidean distance between the stations. The GNN then uses the graph including all static features associated with each node, to create spatial embeddings for each node. The list of static features includes the gravity score, the contribution of each amenity, owner of the station, the minimum and maximum capacity of the station, the number of each plug type at the station and finally the connectioon to the four nearest neighboring nodes. The node embeddings are described by a 16 dimensional vector, which are then concatenated to the timeseries data for each timestamp. As a result each timestamp for a given location will have a number of dynamic features, and the embeddings describing the spatial information as static features. The timeseries data is fed into the GRU network, which then uses a combination of these features to predict the usage of the station.

To predict the use of a station placed at a new location, it is not possible to use historical data since no data is available at the given location. Therefore, three station have been randomly selected to be used as test cases. To use these stations as test cases, the timeseries data for the three stations has been extracted from the dataset. Opposingly, the static features will typically be available for a new stations, or the model can be run with different configurations to see how the model performs with different static features. Nevertheless, the three stations are included when generating the GNN, with the assumption that the static features are available. Note that in the cooperating thesis, 8 charging station where removed due to inconsistencies in the timeseries data for those stations. Therefore the total number of stations is reduced to 392 from the original 400. Thus, the embeddings generated by the GNN is determined by the relation between all 392 stations. The embeddings generated by the GNN are then mapped to the timeseries data of all stations including the three test stations. The GRU network is then trained on the data of the "pre-existing" 389 stations.

Understanding how the hybrid model is constructed, the next step was to evaluate how the model performs on stations at locations the model has never been introduced to, and to examine the impact of spatial embeddings on the performance of the model. To explore the impact of adding the knowledge of the gravity scores to formulate the spatial embeddings, the performance of the hybrid model was evaluated on all three stations with three different settings. The settings are described in the list:

- **Baseline:** The baseline setting uses the timeseries data of the new station to perform the predictions. This is exactly what is normally done in timeseries forecasting. Thus, the difference is that the spatial embeddings and the training split of the timeseries has never been introduced to the GRU network, as the model is only trained on the "pre-existing" stations. The baseline setting is used to evaluate the obtainable accuracy of forecasting on unknown stations.

- **Constant significant lags:** This setting is used to evaluate the impact of the spatial embeddings on the performance of the model, when temporal elements are mitigated. Therefore target variable is set to a constant value equal to the daily mean across the entire timeseries for the station in all rows.

- **Updated significant lags:** The final setting evaluates the impact of performing iterative multi-step ahead forecasting. Given that there is no historical data the batch size must be set to 1, such that all predictions are based on updated previous values. All rows are initialized as in the constant significant lags setting, but for each prediction the target variable is updated with the predicted value, before the next prediction is made. A graphical representation of the third setting is shown in Fig. 8.26.

From the figure it is explained how the previous predictions are used to update the future predictions, and thereby creating a rolling window of predictions. A description of the three stations used as test cases is provided in the list:

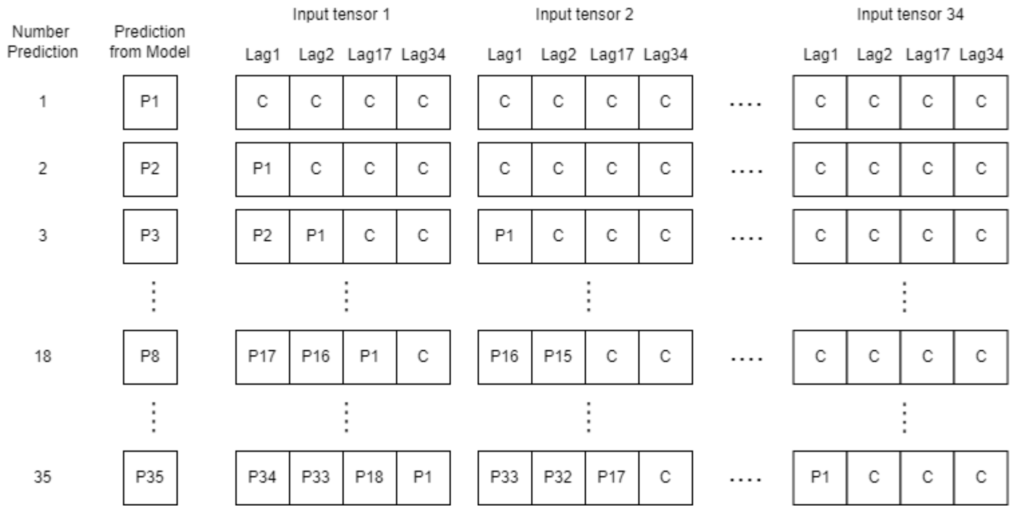


Figure 8.26: Graphical representation of Updated significant lags setting.

- **Test 1:** The first test station is placed in a urban environment of a small city called Oksby. The station is owned by Clever, and has a total of 4 plugs. The minimum capacity is 11 kw and maximum capacity of 150 kw. The station has a gravity score obtained for this location is: 18259.2.
- **Test 2:** The second test station is placed near one of the most populated motorways in Denmark, Køge Bugt Motorvejen. The station is owned by Tesla and has a total of 26 plugs. The minimum and maximum capacity are both 150 kw. The gravity score at this location is: 14627.7.
- **Test 3:** The last test station is placed in a larger urban area, Holstebro. The station is owned by Clever and has 6 plugs. The minimum and maximum capacity are both 300 kw. The gravity score at this location is 53736.9.

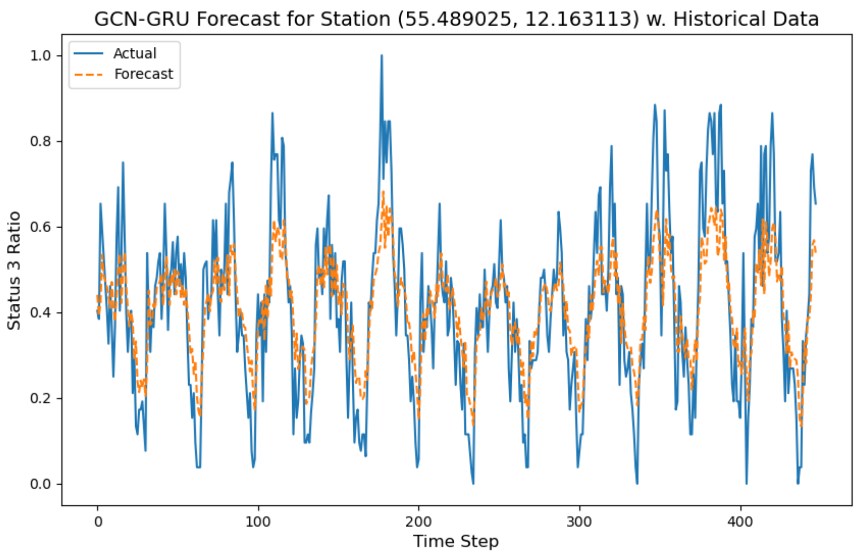


Figure 8.27: Test 2 - Baseline setting.

The results for each of three stations on each of the three settings are shown in Tab. 8.14. From the errors it is observed that the baseline model performs significantly better than the settings where historical data is not available. The predictions for test case 2 with the baseline setting is shown in Fig. 8.27 and the updated significant lags setting in 8.28.

From the figures the importance of historical data is evident. The baseline setting is confident in predicting much more extreme values, since these can be observed from the historical data, while in the absence of historical data, the predictions are more conservative and do not deviate much from the mean. This is reflected in the error values. However, looking at the cosine similarity it is observed that approximately 70% to 90% of the trend can be captured by static features alone. This indicates that by knowing the static features of a station including the gravity score, only the gain of predictions need to be approximated in order to obtain predictions which are close to the actual values. Unfortunately, how to approximate the gain has not been investigated in this thesis.

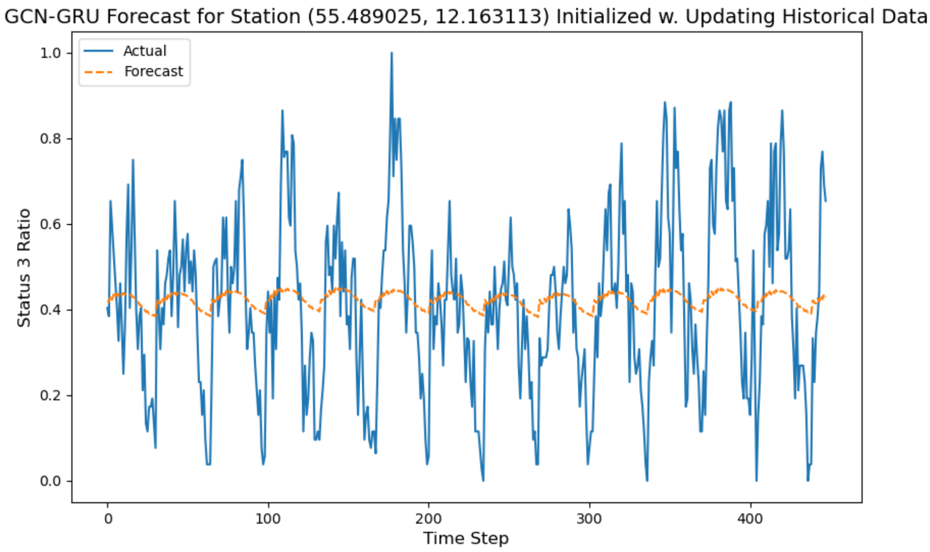


Figure 8.28: Test 2 - Updated significant lags setting.

Test	**Setting**	**MSE**	**RMSE**	**Cosine Similarity**
Test1	Baseline	0.0465	0.2156	0.8788
	Constant significant lags	0.1032	0.3212	0.7165
	Updated significant lags	0.1078	0.3283	0.7155
Test2	Baseline	0.0154	0.1243	0.9637
	Constant significant lags	0.0368	0.1917	0.9088
	Updated significant lags	0.0369	0.1921	0.9085
Test3	Baseline	0.0042	0.0646	0.8688
	Constant significant lags	0.0062	0.0790	0.7991
	Updated significant lags	0.0094	0.0972	0.7916

Table 8.14: Model performance metrics across different tests and settings.

Chapter 9

Discussion

9.1 Implementation of the Model

This section will discuss the implementation of the model, and the potential improvements that could be made. The discussion will include the distance decay function, the baseline weighting of amenities, the distinction between similar amenities, and the competitiveness in facility location.

9.1.1 Distance Decay Function

Early in the project it became evident that this was a problem of facility location. At the heart of facility location problems is the determination of demand in and around the area of inspection. Given that the request of siting charging stations spanned across the entire country of Denmark, it was necessary to distinguish both between inter-area demand on a micro level and intra-area demand on a macro level. The micro demand determination is based on the proximity of amenities to the charging station. Given formulating the proposed solution as a gravity model, the distance between a candidate location and amenities is a key factor in determining the demand. Thus an appropriate distance decay function is necesarry to capture the difference in demand between amenities that are close to the candidate location and those that are further away. During this project, a simple distance decay function of $1/d^2$ was chosen to model the inverse relationship between distance and demand, however no empirical evidence was provided to support this choice. Effort was made to find literature looking into the correlation between distance from charging stations to amenities or other impactful features, however only studies stating the distance to which features where impactful were discovered, and not the difference between distances. In addition, no data was available to conduct an analysis of the correlation, why an assumption was made that the distance decay function would be an inverse power function, which is a common choice in both gravity models and facility location problems. Therefore, analyzing the impact of distance remains a potential area of improvement for the model. A possible solution to combatting this limitation could be to include crowd insights data, which could provide information on the number of people moving between locations, and measure the decay as distance increases.

9.1.2 Baseline Weighting of Amenities

Currently, the baseline weighting of parameters is based on a statistically derived decision. Instead, the weights could be determined by a professional decision, specifically from the existing strategic department in Norlys. This approach would allow the baseline model to reflect the current strategic decisions rather than a decision based on statistical assumptions. Thereby, the results would reflect whether the proposed solution is an improvement to the existing strategy.

9.1.3 Distinction Between Similar Amenities

The current model is not able to distinct between similar amenities located in different areas. As a result, locations in the proximity of a large number of amenities will always score highly. Being able to distinct between similar amenities would enhance the interpretation of demand on the macro basis. This could be achieved with the inclusion of crowd insights data, which could provide information on the people mass at different parts in the country. This could potentially allow the model to penalize location with a high number of amenities, but a low number of consumers. Another approach could be to regularize the total amount of amenities that has influence on the gravity score, however this would require to define a treshold for the number of amenities considered, while also necessitating a definition of which amenities to include and which to neglect.

9.1.4 Competitiveness in Facility Location

Competitiveness between charging stations is not considered in the existing model despite being a common inclusion in facility location problems. Competitiveness is a two sided coin, as it can both be an attraction and a deterrent. In essence, the presence of a competitor can be an attraction and act as an indication of demand. Flip the coin and the presence of a competitor results in shared demand amplifying the return of investment. The inclusion of competitiveness would require data on the number of charging stations in the proximity of the candidate location. Without a complete database of public charging stations across Denmark, this would lead to an additional uncertain factor in the model. In contrast, acknowledging the presence of significant competitors and surrendering demand in order to leverage uncontested demand is a strategic decision that could be beneficial in the long run.

9.2 Discussing the Obtained Results

Obtained results from the previous chapter are interpreted and discussed in this section. The discussion will include reviewing the results of the model and provide an analysis of why the model has performed as it has. In addition, the discussion will include an interpretation of the results in comparison to those obtained in the existing litterature, and outline the core differences between the two. The discussion of the results follows the structure of the experiments and results presented in the previous chapter.

9.2.1 Patronization of Charging Facilities

From the results presented in chapter 8 it was shown that the average use of a station was not directly correlated with the gravity score. Tendencies to overevaluate the gravity score of stations with low usage and underevaluate stations with high usage were evident. This can possibly be explained by Tab. 9.1 which displays two cases of stations placed at similar locations with similar gravity scores, but with a significant variation in usage. The first case describes four stations located at a motorway exit in Aabenraa. Within a distance of 200 meters, four different charging stations are located, each owned by a different company. As mentioned, the usage of each station varies significantly, which can only possibly be explained by the variables that differ. These are the kW capacities of each station, the number of plugs and the owner of the station. Three of the stations consist purely of fast chargers, leaving only two inconsistencies. The first inconsistency is the owner, which can be described by the Circle K and IONITY stations. They have an equal number of outlets, however the IONITY station averages double the usage of the Circle K station. Thus, demand is not only generated by geographical descrepancy but is also a question of EV owners patronizing or staying loyal to a specific provider.

Lat	Lon	GS	Use	# plugs	Min Cap	Max Cap	Owner
Case 1							
55.065	9.367	1016.88	0.892	2	43.0 kW	50.0 kW	E.ON
55.065	9.366	1037.79	11.172	6	60.0 kW	300.0 kW	Circle K
55.065	9.366	1022.76	22.401	6	350.0 kW	350.0 kW	IONITY
55.068	9.360	1037.51	74.794	16	150.0 kW	150.0 kW	Tesla
Case 2							
55.489	12.162	2210.76	5.019	6	60.0 kW	300.0 kW	Circle K
55.489	12.161	2210.94	139.303	16	300.0 kW	300.0 kW	Clever
55.489	12.163	2187.14	177.976	26	150.0 kW	150.0 kW	Tesla

Table 9.1: Two cases of stations placed at the same location with similar gravity scores, but with different owners and plug capacities.

The second inconsistency is the number of plugs. Given that the current metric for evaluating the location is based on the average hours charged, stations with a lower number of plugs will not be able to achieve equal numbers to stations with a higher number of plugs, despite an equal assessment of the locations. This leads to the realization that the optimization objectives of the model are not a realistic representation of the actual distribution. Hence, to be able to model the actual distribution it would be necesarry to either accuractely include the impact of inconsistencies between stations, or to cluster the stations into groups of similar stations and optimize towards the distribution of defined groups. However, given the already limited data points, the results obtained from the corresponding model would be even more uncertain than the results currently displayed.

When determining the optimization objectives, it was considered how the gravity score of a location would be able to give the best indication of the locations potential. Two ways of modelling this where considered, the first being representing the use of the station as a ratio of how much the station is used. The second being multiplying the ratio by the number of plugs at the station, to get an indication of the hours charged. Considering only the ratio would result in a loss of information, as there would be no indication of the potential demand covered by the station at a certain location. This could result in stations charging a combined 5 hours a day to be equal to a station charging 50 hours a day. Rather, multiplying the ratio by the number of plugs would give an indication of the hours charged at a location, which indirectly reflects the number of plugs, and thus partially removing an uncertainty. This would also lead to clustering stations with fewer plugs towards the lower end of the ranking, and vice versa for stations with more plugs. Keeping the problem statement in mind, the objective was to evaluate how profitable a station would be at a given location, and the best way of defining the optimization objective given the data available was the give an estimate of the hours charged at a location.

9.2.2 Data Limitations

Another key factor in the obtained results lies within the data limitations. First and foremost the data provided by Norlys did not contain a clear definition of the target output. Therefore it was necessary to construct a definition of what should be interpreted as the truth from the available data.

After going through large portions of the literature, and deciding upon modelling the solution as a problem of gravitation it became a primary focus to find the most accurate data available to facilitate the solution. Much effort was put into locating and acquiring the most up to date and accurate data available. Despite said efforts, the obtained data still had limitations. Particularily data from OSM was limited, as it is community driven and thereby dependent on contributors to map out amenities. As a result, the score of certain locations are underevaluated due to a lack of registered amenities in the area. This is shown definitively in the motorway model, as the number of amenities in the area is limited and the lack of registered amenities therefore have large impact on the gravity score.

9.3 Pros and Cons of the Gravity Model

The advantage of using the proposed method is the granularity in which the model is able to estimate demand. Compared to a clustering approach where demand is captured in an area, the gravity model is able to indicate micro demand capture within the area of the cluster. Given the formulation of the problem described by Norlys, it was necesarry to be able to distinct between demand on a micro level, as the decision of which candidate locations that are presented to Norlys are determined by another instance. The negative sides of the gravity model is the data requirements. The model is extremely data hungry and requires data from a wide range of sources, while also relying heavily on the consistency and reliability. As mentioned earlier, gaps in the data will lead to the incapability of capturing distinction between similar locations in terms of amenity proximity, but unregistered in one location. Therefore, investment in data collection from quality sources should be prioritized the accuracy and effectiveness of the model.

9.4 Comparison to Existing Literature

Given the new nature of the proposed model and the limited amount of existing research on the topic, it is difficult to compare the results of the model to existing methods applied in the field. Rather than comparing results, a discussion of the modelling approach is compared to the methods presented in the literature.

Most of the literature presented models demand as a clustering problem, where demand is captured in an area. In [8] the authors captured demand by forming clusters using a fuzzy k-means approach. The centroids are determined by minimizing the distance between destination points within the cluster and the clsuter centroid. Charging stations are therefore proposed to be deployed at the cluster centroids. However, it occurs that the cluster centroid are not always suitable for the deployment of charging stations, and a solution to this problem is not presented. In contrast, the model proposed in this thesis is able to estimate locations on a finer granularity, where feasible locations can be evaluated, however on the basis of a different formulation of demand.

Zeb et al. [9] studied the cost associated with deploying charging stations, and the impact on the grid. These aspects are interesting aspects that could be included as an additional layer to the proposed model. This will allow Norlys to evaluate the potential of the locations in terms of cost and grid impact, and thereby make an informed decision on the return of investment associated to each location.

Rodrigues et al. [10] provides an interesting approach to demand forecasting. A combination of socio-demographic data and sales figures are used to predict demand at specific areas, and a distance decay function is used to distribute the demand to the surrounding areas. The original gravity model is then used to capture the spatial distribution of demand, and the supply is estimated based on the captured demand at each point. This approach is a simplistic concept of formulating demand, and requires significantly less data than the proposed model. However, the variables used to predict demand are evaluated on the base of historic data, and the correlation is only around 35%. Interestingly, a lot of the approaches taken in this study are similar to the method developed in this thesis. Note, that even though the paper is officially accepted by IEEE Xplore, it has no citations, and is only a case study in the municipality of Porto.

Chapter 10

Conclusion

The overall goal of this thesis was to explore the strategic siting of public electric vehicle charging stations. The project was completed in collaboration with one of the leading danish energy companies, Norlys. Exploring the existing literature, it was found that the current methods for siting electrical vehicle charging stations did not evaluate locations on a granular level that satisfied the project requirements. Therefore, a novel approach was developed, using a modified gravity-based model to evaluate the potential of a new location with a gravity score. The score reflected the attractiveness of a location based on multiple factors, including geospatial data, traffic density, and socio-demographic data.

The core of the model was the gravity equation, which was used to calculate the attractiveness of a location by determining the interaction between the location and the surrounding amenities. A weighted approach was used to determine the importance of each amenity. The optimal set of weights was found using an optimization approach using a genetic algorithm. The results showed that the model had trouble capturing the actual distribution of observed data, which was determined to be partially due to limitations in the data used in the project.

Nevertheless, the geographical features were used in a hybrid forecasting model, with the aim of investigating the impact of spatial awareness in forecasting models. These results showed that the spatial features could capture some of the underlying patterns, however struggled to produce accurate prediction without the inclusion of historical data.

Concluding upon the research goals set in the problem statement, this thesis has achieved the following: The existing literature has been thoroughly investigated to identify the current methods for siting electric vehicle charging stations. The literature review showed that the current methods in the existing literature have had partial success in solving the problem, and no method has been determined to be the optimal way of solving the problem. Thus, through inspiration from the existing literature a novel method has been developed to solve the problem. The chosen method included using spatial awareness to evaluate the attractiveness of a given location, by using a gravity-based model. To be able to evaluate the model, a set of criteria was defined which reflected the actual use of charging stations. The results were presented by visual representation of gravity maps, which showed attractiveness of locations. This was achieved by presenting plots of gravity maps and by developing an interactive web application. Finally, the method was discussed in relation to other methods in the field, however concrete comparisons were not possible due to the novelty of the method.

Bibliography

[1] *Elbiler - Danmarks Naturfredningsforening.* https://www.dn.dk/vi-arbejder-for/klima/elbiler/. (Visited on 05/10/2024).

[2] R. van den Hoed et al. "Data Analysis on the Public Charge Infrastructure in the City of Amsterdam". In: *2013 World Electric Vehicle Symposium and Exhibition (EVS27)*. Nov. 2013, pp. 1–10. DOI: 10.1109/EVS.2013.6915009. (Visited on 10/05/2023).

[3] *Behovet for ladestandere til elbiler eksploderer.* https://www.dtu.dk/newsarchive/2022/12/behovet-for-ladestandere-eksploderer. (Visited on 06/03/2024).

[4] *Massiv CO2-reduktion og ambition om 1 mio. grønne biler i 2030.* https://www.regeringen.dk/nyheder/2020/massiv-co2-reduktion-og-ambition-om-1-mio-groenne-biler-i-2030/. (Visited on 05/10/2024).

[5] *NYT: Elbiler udgjorde 36 pct. af de nye biler i 2023.* https://www.dst.dk/da/Statistik/nyheder-analyser-publ/nyt/NytHtml?cid=46493. (Visited on 05/10/2024).

[6] *Udvikling i Ladeinfrastruktur Og Bestand Af Elog Plug-in-Hybridbiler.* May 2024. (Visited on 05/13/2024).

[7] Christian Østergaard. *Norlys satser stort på elbiler: Vil opføre 300.000 ladestandere.* https://borsen.dk/nyheder/baeredygtig/kaempeaftale-saetter-dansk-software-i-300000-ladestandere. (Visited on 06/03/2024).

[8] N. Andrenacci, R. Ragona, and G. Valenti. "A Demand-Side Approach to the Optimal Deployment of Electric Vehicle Charging Stations in Metropolitan Areas". In: *Applied Energy* 182 (Nov. 2016), pp. 39–46. ISSN: 0306-2619. DOI: 10.1016/j.apenergy.2016.07.137. (Visited on 05/30/2024).

[9] Muhammad Zulqarnain Zeb et al. "Optimal Placement of Electric Vehicle Charging Stations in the Active Distribution Network". In: *IEEE Access* 8 (2020), pp. 68124–68134. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2984127. (Visited on 01/22/2024).

[10] Gonçalo Rodrigues et al. "Electric Charging Demand Forecast and Capture for Infrastructure Placement Using Gravity Modelling: A Case Study". In: *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. Sept. 2023, pp. 2112–2117. DOI: 10.1109/ITSC57777.2023.10421918. (Visited on 05/30/2024).

[11] Dmytro Karamshuk et al. “Geo-Spotting: Mining Online Location-Based Services for Optimal Retail Store Placement”. In: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD ’13. New York, NY, USA: Association for Computing Machinery, Aug. 2013, pp. 793–801. ISBN: 978-1-4503-2174-7. DOI: 10.1145/2487575.2487616. (Visited on 05/28/2024).

[12] Mevlut Uyan. “GIS-based Solar Farms Site Selection Using Analytic Hierarchy Process (AHP) in Karapinar Region, Konya/Turkey”. In: *Renewable and Sustainable Energy Reviews* 28 (Dec. 2013), pp. 11–17. ISSN: 1364-0321. DOI: 10.1016/j.rser.2013.07.042. (Visited on 05/29/2024).

[13] Rob van Haaren and Vasilis Fthenakis. “GIS-based Wind Farm Site Selection Using Spatial Multi-Criteria Analysis (SMCA): Evaluating the Case for New York State”. In: *Renewable and Sustainable Energy Reviews* 15.7 (Sept. 2011), pp. 3332–3340. ISSN: 1364-0321. DOI: 10.1016/j.rser.2011.04.010. (Visited on 05/29/2024).

[14] Younes Noorollahi, Hossein Yousefi, and Mohammad Mohammadi. “Multi-Criteria Decision Support System for Wind Farm Site Selection Using GIS”. In: *Sustainable Energy Technologies and Assessments* 13 (Feb. 2016), pp. 38–50. ISSN: 2213-1388. DOI: 10.1016/j.seta.2015.11.007. (Visited on 05/29/2024).

[15] B. P. Mbilinyi et al. “GIS-based Decision Support System for Identifying Potential Sites for Rainwater Harvesting”. In: *Physics and Chemistry of the Earth, Parts A/B/C*. Mainstreaming Integrated Water Resources Management in the Development Process 32.15 (Jan. 2007), pp. 1074–1081. ISSN: 1474-7065. DOI: 10.1016/j.pce.2007.07.014. (Visited on 05/28/2024).

[16] Mehmet Erbaş et al. “Optimal Siting of Electric Vehicle Charging Stations: A GIS-based Fuzzy Multi-Criteria Decision Analysis”. In: *Energy* 163 (Nov. 2018), pp. 1017–1031. ISSN: 0360-5442. DOI: 10.1016/j.energy.2018.08.140. (Visited on 05/29/2024).

[17] Nitin Liladhar Rane et al. “An Integrated GIS, MIF, and TOPSIS Approach for Appraising Electric Vehicle Charging Station Suitability Zones in Mumbai, India”. In: *Sustainable Cities and Society* 97 (Oct. 2023), p. 104717. ISSN: 2210-6707. DOI: 10.1016/j.scs.2023.104717. (Visited on 05/29/2024).

[18] *Trafiktællinger – Nøgletal - Datasæt*. https://www.opendata.dk/vejdirektoratet/taellinger-nogletal-mastra#resource-open_data_noegletal_view.json. (Visited on 02/09/2024).

[19] *PROJ — PROJ 9.4.1 Documentation*. https://proj.org/en/9.4/index.html. (Visited on 06/03/2024).

[20] Jon Louis Bentley. “Multidimensional Binary Search Trees Used for Associative Searching”. In: *Communications of the ACM* 18.9 (Sept. 1975), pp. 509–517. ISSN: 0001-0782, 1557-7317. DOI: 10.1145/361002.361007. (Visited on 05/30/2024).

[21] Magnus Tang and Bastian Aaron Kramer. *Thesis_Magnus_Bastian/RnD_Article_Siting_Charging_Stations_for_Electric_Vehicles_Data_ at Main · Charging-Stations-Placement/Thesis_Magnus_Bastian.* https://github.com/Charging-Stations-Placement/Thesis_Magnus_Bastian/blob/main/RnD_Article_Siting_Charging_Stations_fo (Visited on 05/31/2024).

[22] *Bestand Af Motorkøretøjer - Statistikbanken.* https://www.statistikbanken.dk/statbank5a/Graphics/MapAnalyser.asp?Maintable=BIL54&l (Visited on 02/19/2024).

[23] *OpenStreetMap*. https://www.openstreetmap.org/. (Visited on 02/05/2024).

[24] *Geofabrik Download Server*. https://download.geofabrik.de/. (Visited on 02/05/2024).

[25] *Map Features - OpenStreetMap Wiki.* https://wiki.openstreetmap.org/wiki/Map_features#Geological. (Visited on 05/31/2024).

[26] Csaba Csiszár et al. "Urban Public Charging Station Locating Method for Electric Vehicles Based on Land Use Approach". In: *Journal of Transport Geography* 74 (Jan. 2019), pp. 173–180. ISSN: 0966-6923. DOI: 10.1016/j.jtrangeo.2018.11.016. (Visited on 05/31/2024).

[27] Sarah Hoffmann. *Osmium: Python Bindings for Libosmium, the Data Processing Library for OSM Data.* (Visited on 05/31/2024).

[28] *GADM*. https://gadm.org/index.html. (Visited on 05/31/2024).

[29] R. Wolfson. *Essential University Physics*. 3rd ed. Vol. 1. Pearson Education, 2016. ISBN: 978-1-292-10266-5.

[30] Peter A. G. van Bergeijk and Steven Brakman. *The Gravity Model in International Trade: Advances and Applications*. Cambridge University Press, June 2010. ISBN: 978-1-139-48828-0.

[31] Tammy Drezner. "Gravity Models in Competitive Facility Location". In: *Contributions to Location Analysis: In Honor of Zvi Drezner's 75th Birthday*. Ed. by H. A. Eiselt and Vladimir Marianov. Cham: Springer International Publishing, 2019, pp. 253–275. ISBN: 978-3-030-19111-5. DOI: 10.1007/978-3-030-19111-5_11. (Visited on 05/28/2024).

[32] Altaf Q. H. Badar. *Evolutionary Optimization Algorithms*. Boca Raton: CRC Press, Oct. 2021. ISBN: 978-1-00-320647-7. DOI: 10.1201/9781003206477.

[33] Dipankar Dasgupta and Zbigniew Michalewicz, eds. *Evolutionary Algorithms in Engineering Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1997. ISBN: 978-3-642-08282-5 978-3-662-03423-1. DOI: 10.1007/978-3-662-03423-1. (Visited on 05/17/2024).

[34] Melanie Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, Mar. 1998. ISBN: 978-0-262-63185-3.

[35] John H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT Press, Apr. 1992. ISBN: 978-0-262-58111-0.

[36] Thomas Bartz-Beielstein et al. “Evolutionary Algorithms”. In: *WIREs Data Mining and Knowledge Discovery* 4.3 (2014), pp. 178–195. ISSN: 1942-4795. DOI: 10.1002/widm.1124. (Visited on 05/20/2024).

[37] Larry J. Eshelman and J. David Schaffer. “Real-Coded Genetic Algorithms and Interval-Schemata”. In: *Foundations of Genetic Algorithms*. Ed. by L. DARRELL Whitley. Vol. 2. Foundations of Genetic Algorithms. Elsevier, Jan. 1993, pp. 187–202. DOI: 10.1016/B978-0-08-094832-4.50018-0. (Visited on 05/17/2024).

[38] Abdullah Konak, David W. Coit, and Alice E. Smith. “Multi-Objective Optimization Using Genetic Algorithms: A Tutorial”. In: *Reliability Engineering & System Safety*. Special Issue - Genetic Algorithms and Reliability 91.9 (Sept. 2006), pp. 992–1007. ISSN: 0951-8320. DOI: 10.1016/j.ress.2005.11.018. (Visited on 06/03/2024).

[39] Kalyanmoy Deb. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons, July 2001. ISBN: 978-0-471-87339-6.

[40] K. Deb et al. “A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II”. In: *IEEE Transactions on Evolutionary Computation* 6.2 (Apr. 2002), pp. 182–197. ISSN: 1941-0026. DOI: 10.1109/4235.996017. (Visited on 06/03/2024).

[41] Eckart Zitzler, Marco Laumanns, and Lothar Thiele. “SPEA2: Improving the Strength Pareto Evolutionary Algorithm”. In: ETH Zurich, May 2001, 21 p. DOI: 10.3929/ETHZ-A-004284029. (Visited on 06/03/2024).

[42] *Distance on a Sphere: The Haversine Formula*. https://community.esri.com/t5/coordinate-reference-systems-blog/distance-on-a-sphere-the-haversine-formula/ba-p/902128. Oct. 2017. (Visited on 06/03/2024).

[43] Wolfgang Torge, Jürgen Müller, and Roland Pail. “Geodesy”. In: *Geodesy*. De Gruyter Oldenbourg, Apr. 2023. ISBN: 978-3-11-072330-4. DOI: 10.1515/9783110723304. (Visited on 04/15/2024).

[44] *Telia Crowd Insights*. https://business.teliacompany.com/crowd-insights.

[45] Félix-Antoine Fortin et al. “DEAP: Evolutionary Algorithms Made Easy”. In: *Journal of Machine Learning Research* 13.70 (2012), pp. 2171–2175.

[46] John Grefenstette. “Optimization of Control Parameters for Genetic Algorithms”. In: *IEEE Transactions on Systems, Man, and Cybernetics* 16.1 (Jan. 1986), pp. 122–128. ISSN: 0018-9472. DOI: 10.1109/TSMC.1986.289288. (Visited on 05/23/2024).

[47] Chibuzor N. Obiora et al. “Forecasting Hourly Solar Radiation Using Artificial Intelligence Techniques”. In: *IEEE Canadian Journal of Electrical and Computer Engineering* 44.4 (2021), pp. 497–508. ISSN: 2694-1783. DOI: 10.1109/ICJECE.2021.3093369. (Visited on 05/25/2024).

[48] Ahmad Hassanat et al. “Choosing Mutation and Crossover Ratios for Genetic Algorithms—A Review with a New Dynamic Approach”. In: *Information* 10.12 (Dec. 2019), p. 390. ISSN: 2078-2489. DOI: 10.3390/info10120390. (Visited on 05/26/2024).

[49] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. “Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions”. In: *International Journal of Forecasting* 37.1 (Jan. 2021), pp. 388–427. ISSN: 0169-2070. DOI: 10.1016/j.ijforecast.2020.06.008. (Visited on 05/26/2024).

Chapter 11

Appendix

Datasets

Longitude	Latitude	Name
12.539886223809525	55.69419510476191	København
12.4902969	55.65709462	København
12.48951872	55.65691124	København
12.48896554	55.65665136	København
12.505433957142856	55.65372124285714	København
12.5158021090196	55.684151641568654	Frederiksberg
12.455062942857142	55.76001798571429	Gladsaxe
12.527414644444445	55.689253533333336	Frederiksberg
12.516059377419356	55.68450386129032	Frederiksberg
12.503659544	55.68573019599999	Frederiksberg
12.527979621052632	55.694895163157895	Frederiksberg
12.35717306	55.71892854	Ballerup
12.35812506	55.7124946	Ballerup
12.358342766666668	55.71358095	Ballerup
12.35817108	55.71467428	Ballerup
12.35819334	55.71575722	Ballerup
12.3582144	55.71684722	Ballerup
12.35823766	55.71793034	Ballerup
12.36115946	55.71856078	Ballerup

Table 11.1: Example of OSM Amenity Data

OMRÅDE	Per_Change_2023_2024
København	39.798850574712645
Frederiksberg	39.70080552359033
Dragør	35.24904214559387
Tårnby	23.013321084060635
Albertslund	43.8058748403576
Ballerup	49.98074701578744
Brøndby	19.9698568198945
Gentofte	33.190307925290256
Gladsaxe	44.43748041366343
Glostrup	38.34661354581673
Herlev	47.04830053667263
Hvidovre	45.36031589338598
Høje-Taastrup	48.90178201409034
Ishøj	41.05960264900662
Lyngby-Taarbæk	41.36690647482014
Rødovre	40.98582039162728
Vallensbæk	44.27480916030535
Allerød	39.49967083607637
Egedal	43.71609403254973

Table 11.2: Example of Commune Data

Longitude	Latitude	AADT_normalized
9.406551000153188	56.44727501833787	0.05240059807287629
9.404161587955736	56.44835196178049	0.005198526968656551
9.709261732806684	55.73051380358308	0.05251135230922582
8.664532819121744	56.09492252444594	0.06056872300365489
9.12949984166756	55.49315436804926	0.07523673718019713
9.76089489806664	55.85120419864989	0.007226713921807509
9.174611543558436	56.58875315680864	0.006056872300365489
9.404084694877229	55.51422084329878	0.006582954923025805
9.399116918662916	55.51312289784229	0.0066867870196035
10.094158038383515	56.31137082432293	0.004665522206224388
12.558601075368337	55.6611136834091	0.034832207331930445
11.28706716800421	55.50537603756709	0.03850094141100897
11.289141785680924	55.50128356408648	0.03484605161147414
11.292332919425448	55.494069194848265	0.03274864326060472
9.989185496585764	56.05910504861096	0.006762930557093809
8.995675226399317	56.10427302660906	0.005496178978845941
11.29130721189084	55.49549056822348	0.032990918152619335
8.590139575963201	56.373412091996045	0.02573651567172444
8.615421967819278	56.36126519822712	0.04014841067670838
8.76905568566716	56.5186207844191	0.003592590541588216
8.810902328137754	56.53552788731374	0.0009068003101118618
8.66719437139599	56.38964414189227	0.0008791117510244767

Table 11.3: Example of Traffic Data

Longitude	Latitude	Daily Average Use
10.648394	55.501504	0.0788680120621654
8.836123	56.781975	0.1568627450980382
10.182387	57.061677	0.2573282075925014
9.12968	55.13012	0.3690531177829088
11.342991	55.41403	0.4418658621489852
9.535372	56.739888	0.4830214830214828
9.911765	56.359149	0.49185509360563917
11.859938	55.592095	0.5313730030099546
9.99036	57.2485	0.6306515186645018
9.241929	56.964766	0.692388025063815
10.182387	57.061677	0.7772104896727763
8.805212	56.390391	0.8026283767339965
11.29599	55.48796	0.8462779442267772
10.161239	57.152389	0.884188626907072
9.36722	55.065492	0.8921705426356562
9.794386	56.391501	0.9061413673232861
8.298722	56.196524	0.9192501802451332
10.064056	56.466197	0.9604056011588589
8.859013	56.031235	0.9812469556746202
10.410471	54.886893	0.987683011852196
11.915559	55.00714	1.0258266956311848
9.910397	55.26112	1.0553963705826155
10.06316	56.264016	1.058319833371902
10.046712	56.326678	1.0654706430568488
8.221635	56.700175	1.089944649446494
11.790111	55.254194	1.1259842519685028
9.867087	56.255411	1.1615152219381777
9.76432	56.147603	1.176892799259087
11.871619	54.899064	1.2233054781801274
9.93721	57.2226	1.2390097177232757

Table 11.4: Example of Station Daily Average Use Data

Results

Github Repository

The code and data used in this project can be found at the following Github repository: Link. Here you will find all plots and results generated in this project.

Statistical Overview

Category	Sum	Min	Max	Mean	Median	Std Dev	Q1	Q3	IQR	TC
Status	5882.42	0.08	177.98	14.71	6.94	22.72	3.16	15.56	12.40	400
Apartments	3415.74	0.44	138.28	14.06	8.27	18.17	4.48	15.74	11.26	243
Charging Station	5402.05	0.08	177.98	15.75	6.94	24.17	3.24	16.47	13.24	343
Cinema	1517.37	0.44	93.89	13.31	8.20	15.67	4.69	14.69	10.00	114
Clinic	1424.99	0.53	94.53	13.97	8.18	16.51	5.50	15.45	9.95	102
College	1448.55	0.44	177.98	20.12	11.27	29.37	5.75	18.18	12.43	72
Community Centre	2340.00	0.08	177.98	15.00	8.18	22.36	4.65	15.39	10.73	156
Conference Centre	163.90	1.36	34.07	11.71	10.02	9.58	5.18	15.35	10.17	14
Dentist	1291.33	0.69	103.28	14.04	7.00	17.50	4.54	15.09	10.54	92
Events Venue	1589.34	0.16	103.28	12.82	8.44	14.90	4.68	14.68	10.00	124
Fast Food	5582.93	0.16	177.98	15.73	7.91	23.66	3.75	16.07	12.32	355
Hospital	793.55	0.99	55.16	11.34	7.59	10.99	4.45	14.64	10.19	70
Hotel	789.92	0.98	90.90	13.17	7.58	16.88	4.02	14.97	10.95	60
House	3407.26	0.37	138.28	12.21	6.99	16.27	3.57	14.07	10.50	279
Leisure Fitness Centre	2489.40	0.44	103.28	13.99	9.46	15.60	5.23	15.61	10.37	178
Leisure Stadium	574.45	1.03	90.90	12.22	9.40	14.37	4.91	13.38	8.47	47
Motorway Link	3295.18	0.63	177.98	20.47	7.39	31.09	3.22	20.08	16.87	161
Motorway	3288.78	0.16	177.98	20.43	7.53	30.93	3.22	20.65	17.43	161
Parking	5822.70	0.08	177.98	14.82	6.99	22.79	3.20	15.56	12.36	393
Place of Worship	3930.94	0.08	128.49	14.19	7.88	19.53	3.53	15.56	12.03	277
Power Substation	3460.48	0.08	138.28	16.02	8.68	21.35	4.53	17.61	13.08	216
Rest Area	178.96	0.26	52.57	7.78	2.56	14.27	1.35	5.41	4.06	23
Restaurant	3487.52	0.37	103.28	13.21	8.08	16.19	3.58	15.39	11.81	264
Services	71.13	1.24	29.32	5.47	2.99	7.44	2.44	5.27	2.84	13
Shop Mall	1708.59	0.44	103.28	17.43	9.23	21.95	5.28	18.15	12.88	98
Shop Supermarket	4039.03	0.08	128.49	13.20	7.48	18.12	3.53	14.73	11.20	306
Sports Centre	441.04	0.99	55.16	13.78	9.06	13.17	5.40	15.16	9.76	32
Sports Hall	746.66	1.48	125.19	15.24	8.87	21.81	5.58	15.85	10.27	49
Stadium	40.98	1.36	14.74	8.20	6.05	5.72	5.29	13.54	8.24	5
Substation Distribution	2191.24	0.08	177.98	19.92	9.46	29.44	3.88	22.22	18.34	110
Substation Minor Distribution	2179.70	0.53	177.98	18.63	10.41	27.04	5.58	19.50	13.92	117
Theatre	1409.31	0.44	125.19	13.82	8.25	15.94	4.58	17.57	12.99	102
Transformer Tower	17.29	4.43	7.11	5.76	5.75	1.34	5.09	6.43	1.34	3
University	577.11	0.44	55.16	14.43	10.17	13.49	4.86	18.26	13.39	40

Table 11.5: Statistics Overview 1000m

Category	Sum	Min	Max	Mean	Median	Std Dev	Q1	Q3	TC
Status	5882.42	0.08	177.98	14.71	6.94	22.72	3.16	15.56	400
Apartments	3201.93	0.44	138.28	14.42	8.20	18.83	4.45	15.97	222
Charging Station	5214.47	0.08	177.98	15.95	6.82	24.61	3.18	16.71	327
Cinema	1258.82	0.53	93.89	14.14	8.49	16.84	5.16	14.68	89
Clinic	1104.85	0.53	94.53	14.93	9.17	17.83	5.50	17.52	74
College	1194.94	0.44	177.98	20.96	9.68	32.40	5.44	18.38	57
Community Centre	1672.52	0.08	93.89	13.38	8.27	15.02	5.05	15.33	125
Conference Centre	86.36	1.36	27.21	9.60	5.63	8.45	2.89	13.15	9
Dentist	988.76	0.69	93.89	13.54	7.47	15.62	4.53	15.56	73
Events Venue	1219.95	0.16	103.28	14.19	8.92	16.80	4.81	15.06	86
Fast Food	5490.85	0.16	177.98	15.96	8.03	23.96	3.77	16.21	344
Hospital	637.01	0.99	55.16	11.18	7.88	10.79	4.35	14.74	57
Hotel	705.65	0.98	90.90	14.11	7.58	18.25	3.86	15.49	50
House	2781.93	0.37	93.89	11.64	6.99	13.58	3.87	14.38	239
Leisure Fitness Centre	2216.64	0.44	103.28	14.12	9.44	15.79	5.46	15.56	157
Leisure Stadium	276.67	1.06	43.40	11.07	7.47	10.43	4.55	13.61	25
Motorway Link	3116.01	0.63	177.98	21.64	7.44	32.57	3.19	21.87	144
Motorway	2945.54	0.88	177.98	22.83	7.99	33.75	3.59	25.37	129
Parking	5818.29	0.08	177.98	14.84	7.02	22.81	3.20	15.58	392
Place of Worship	2834.49	0.08	94.53	12.43	7.19	15.36	3.58	14.37	228
Power Substation	2884.28	0.08	138.28	16.48	9.40	20.97	5.32	18.87	175
Rest Area	175.31	0.88	52.57	8.77	2.89	15.09	1.80	6.41	20
Restaurant	3278.62	0.37	103.28	13.49	7.99	16.65	3.79	15.59	243
Services	71.13	1.24	29.32	5.47	2.99	7.44	2.44	5.27	13
Shop Mall	1483.13	1.13	103.28	18.54	8.78	23.69	5.49	18.40	80
Shop Supermarket	3826.46	0.08	128.49	13.57	7.61	18.74	3.45	14.89	282
Sports Centre	386.39	0.99	55.16	14.31	9.25	13.96	4.99	15.39	27
Sports Hall	629.28	1.64	125.19	18.51	10.13	25.45	5.90	18.22	34
Stadium	26.24	1.36	13.54	6.56	5.67	5.08	4.31	7.92	4
Substation Distribution	1366.35	0.08	128.49	19.52	9.69	24.76	4.61	25.29	70
Substation Minor Distribution	1785.58	0.53	138.28	16.38	10.40	20.11	5.58	18.24	109
Theatre	923.84	2.09	43.53	12.66	9.58	10.36	4.78	16.59	73
Transformer Tower	5.75	5.75	5.75	5.75	5.75	NaN	5.75	5.75	1
University	350.02	0.44	43.53	12.50	8.27	11.45	4.72	16.35	28

Table 11.6: Statistics Overview 500m

Category	**Sum**	**Min**	**Max**	**Mean**	**Median**	**Std Dev**	**Q1**	**Q3**	**IQR**	**Total Count**
Status	5882.42	0.08	177.98	14.71	6.94	22.72	3.16	15.56	12.40	400
Apartments	1383.84	0.96	90.90	13.06	7.59	14.17	4.54	15.97	11.43	106
Charging Station	4798.35	0.08	177.98	17.08	7.11	26.05	3.22	18.07	14.86	281
Cinema	307.67	1.07	67.26	12.82	7.02	15.23	4.97	14.75	9.78	24
Clinic	181.07	2.74	38.09	13.93	9.58	10.46	5.95	19.50	13.55	13
College	81.34	2.19	23.53	11.62	11.94	7.71	5.73	16.11	10.38	7
Community Centre	365.28	1.07	90.90	16.60	11.87	18.96	7.30	15.66	8.36	22
Conference Centre	14.50	1.36	13.15	7.25	7.25	8.34	4.30	10.20	5.90	2
Dentist	287.20	1.48	55.16	12.49	6.94	13.72	3.68	12.98	9.29	23
Events Venue	122.37	0.16	15.10	8.16	8.40	4.40	4.49	10.67	6.18	15
Fast Food	4452.39	0.26	177.98	18.55	9.16	26.56	4.67	19.51	14.84	240
Hospital	44.07	1.16	16.59	6.30	5.46	5.39	2.37	8.06	5.70	7
Hotel	145.75	2.09	43.53	11.21	5.98	11.40	4.11	14.93	10.82	13
House	910.50	0.37	93.89	10.84	6.50	13.37	3.49	13.24	9.76	84
Leisure Fitness Centre	617.26	2.11	103.28	13.42	7.93	19.32	5.67	13.56	7.88	46
Leisure Stadium	22.79	4.55	18.24	11.39	11.39	9.68	7.97	14.82	6.85	2
Motorway Link	1705.92	0.88	177.98	25.85	7.35	39.55	3.67	26.35	22.68	66
Motorway	745.31	0.88	125.19	16.20	6.33	24.82	2.67	17.71	15.04	46
Parking	5357.60	0.08	177.98	14.92	7.38	21.93	3.53	16.29	12.75	359
Place of Worship	553.99	0.08	90.90	13.19	8.44	15.25	5.09	16.14	11.05	42
Power Substation	1496.67	0.44	138.28	16.09	9.27	19.62	5.44	19.50	14.06	93
Rest Area	156.00	0.88	52.57	8.67	2.89	15.68	1.99	5.81	3.82	18
Restaurant	1659.28	0.37	94.53	14.81	8.28	18.64	4.13	17.58	13.46	112
Services	29.20	1.24	7.39	3.24	2.66	1.92	2.44	3.04	0.61	9
Shop Mall	698.29	1.16	103.28	22.53	11.04	29.46	5.43	15.56	10.13	31
Shop Supermarket	2281.20	0.08	128.49	14.26	8.33	18.48	4.50	15.56	11.06	160
Sports Centre	32.65	5.44	27.21	16.32	16.32	15.39	10.88	21.77	10.88	2
Sports Hall	11.81	3.82	7.99	5.90	5.90	2.95	4.86	6.95	2.09	2
Substation Distribution	16.69	0.44	9.07	4.17	3.59	3.65	2.17	5.59	3.42	4
Substation Minor Distribution	1432.25	0.53	138.28	17.90	10.41	22.56	5.62	20.05	14.43	80
Theatre	169.56	3.04	43.53	13.04	9.68	11.38	5.63	14.93	9.30	13
University	17.84	17.84	17.84	17.84	17.84	NaN	17.84	17.84	0.00	1

Table 11.7: Statistics Overview 200m

Category	**Sum**	**Min**	**Max**	**Mean**	**Median**	**Std Dev**	**Q1**	**Q3**	**IQR**	**Total Count**
Status	5882.42	0.08	177.98	14.71	6.94	22.72	3.16	15.56	12.40	400
Apartments	841.68	1.03	55.16	12.75	7.67	11.84	4.80	15.45	10.65	66
Charging Station	4583.50	0.08	177.98	17.56	7.30	26.72	3.29	18.36	15.07	261
Cinema	58.56	3.29	15.56	9.76	9.87	5.85	5.04	14.87	9.83	6
Clinic	42.71	5.05	19.50	14.24	18.16	7.98	11.60	18.83	7.22	3
College	53.67	4.52	23.53	13.42	12.81	9.15	6.33	19.89	13.56	4
Community Centre	58.54	4.43	19.52	11.71	11.80	6.20	6.94	15.85	8.91	5
Conference Centre	13.15	13.15	13.15	13.15	13.15	NaN	13.15	13.15	0.00	1
Dentist	126.64	1.48	55.16	14.07	3.77	18.93	3.17	10.77	7.60	9
Events Venue	15.92	4.43	6.94	5.31	4.55	1.41	4.49	5.74	1.25	3
Fast Food	2896.89	0.91	177.98	18.93	9.25	27.02	5.08	20.65	15.57	153
Hospital	25.66	9.07	16.59	12.83	12.83	5.32	10.95	14.71	3.76	2
Hotel	91.21	3.04	43.53	15.20	10.68	15.34	4.44	18.36	13.92	6
House	422.13	0.37	47.37	10.05	8.20	9.50	4.49	11.80	7.31	42
Leisure Fitness Centre	157.64	5.08	19.50	9.85	7.64	5.01	6.04	11.77	5.73	16
Motorway Link	433.05	1.41	128.49	22.79	5.14	33.06	3.32	30.10	26.78	19
Motorway	236.24	1.41	91.59	16.87	4.93	25.88	2.34	22.15	19.81	14
Parking	3659.06	0.08	121.97	13.35	7.02	17.62	3.28	15.56	12.28	274
Place of Worship	198.68	0.08	90.90	14.19	7.06	22.95	4.34	12.50	8.16	14
Power Substation	635.52	0.53	52.53	14.12	10.41	12.10	5.75	19.95	14.20	45
Rest Area	97.58	0.88	52.57	6.51	2.56	12.88	2.04	4.54	2.49	15
Restaurant	727.60	0.37	94.53	13.23	9.25	15.39	4.14	16.21	12.08	55
Services	19.29	1.24	7.39	3.22	2.61	2.13	2.47	2.91	0.44	6
Shop Mall	214.38	1.16	90.90	17.87	10.96	23.78	8.27	15.56	7.29	12
Shop Supermarket	1430.69	0.08	90.90	13.01	9.15	14.08	4.13	15.23	11.10	110
Sports Centre	5.44	5.44	5.44	5.44	5.44	NaN	5.44	5.44	0.00	1
Substation Distribution	11.81	2.74	9.07	5.91	5.91	4.47	4.33	7.49	3.16	2
Substation Minor Distribution	631.70	0.53	52.53	15.79	10.73	12.92	5.71	24.00	17.29	40
Theatre	36.99	3.04	14.93	7.40	6.05	4.99	3.29	9.68	6.39	5

Table 11.8: Statistics Overview 100m

Category	**Sum**	**Min**	**Max**	**Mean**	**Median**	**Std Dev**	**Q1**	**Q3**	**IQR**	**Total Count**
Status	5882.42	0.08	177.98	14.71	6.94	22.72	3.16	15.56	12.40	400
Apartments	156.93	2.00	27.21	7.47	5.63	6.46	4.35	8.18	3.83	21
Charging Station	4438.57	0.08	177.98	18.34	7.90	27.43	3.45	19.49	16.04	242
Clinic	5.05	5.05	5.05	5.05	5.05	NaN	5.05	5.05	0.00	1
Community Centre	6.94	6.94	6.94	6.94	6.94	NaN	6.94	6.94	0.00	1
Dentist	3.59	3.59	3.59	3.59	3.59	NaN	3.59	3.59	0.00	1
Events Venue	6.94	6.94	6.94	6.94	6.94	NaN	6.94	6.94	0.00	1
Fast Food	1098.02	0.91	128.49	18.30	7.49	26.98	5.15	16.60	11.44	60
Hospital	25.66	9.07	16.59	12.83	12.83	5.32	10.95	14.71	3.76	2
Hotel	18.70	3.77	14.93	9.35	9.35	7.89	6.56	12.14	5.58	2
House	157.31	0.37	24.94	7.87	6.38	6.08	3.45	10.95	7.50	20
Leisure Fitness Centre	47.89	7.05	14.79	9.58	8.17	3.28	7.11	10.77	3.66	5
Motorway	84.19	2.26	50.25	21.05	15.84	22.50	3.92	32.96	29.04	4
Parking	1975.31	0.08	121.97	13.26	7.05	17.05	3.19	17.53	14.33	149
Place of Worship	51.84	0.08	22.40	10.37	10.95	8.39	5.46	12.94	7.48	5
Power Substation	406.41	4.14	52.53	19.35	14.74	12.92	8.87	25.61	16.74	21
Rest Area	22.20	0.88	7.30	3.17	2.16	2.52	1.67	4.26	2.58	7
Restaurant	113.66	0.37	24.25	8.12	6.09	6.82	3.32	9.72	6.40	14
Services	5.23	2.57	2.66	2.61	2.61	0.07	2.59	2.64	0.05	2
Shop Mall	10.77	10.77	10.77	10.77	10.77	NaN	10.77	10.77	0.00	1
Shop Supermarket	608.83	0.26	47.37	12.68	10.28	9.29	7.34	16.04	8.70	48
Substation Distribution	9.07	9.07	9.07	9.07	9.07	NaN	9.07	9.07	0.00	1
Substation Minor Distribution	423.56	4.14	52.53	20.17	14.74	13.87	8.87	27.21	18.33	21
Theatre	3.29	3.29	3.29	3.29	3.29	NaN	3.29	3.29	0.00	1

Table 11.9: Statistics Overview 50m

Genetic Algorithm Hyperparameters

Crossover	Mutation	Best RMSE
0.3	0.1	0.3634
0.3	0.2	0.3629
0.3	0.3	0.3629
0.3	0.4	0.3636
0.3	0.5	0.3636
0.4	0.1	0.3628
0.4	0.2	0.3627
0.4	0.3	0.3625
0.4	0.4	0.3627
0.4	0.5	0.3624
0.5	0.1	0.3628
0.5	0.2	0.3629
0.5	0.3	0.3625
0.5	0.4	0.3623
0.5	0.5	0.3627
0.6	0.1	0.3623
0.6	0.2	0.3622
0.6	0.3	0.3623
0.6	0.4	0.3628
0.6	0.5	0.3623
0.7	0.1	0.3624
0.7	0.2	0.3626
0.7	0.3	0.3628
0.7	0.4	0.3625
0.7	0.5	0.3623

Table 11.10: Best RMSE values for different Crossover and Mutation probabilities

Crossover	Mutation	Best RMSE
0.3	0.03	0.3628
0.3	0.04	0.3625
0.3	0.05	0.3628
0.3	0.06	0.3627
0.3	0.07	0.3631
0.4	0.03	0.3632
0.4	0.04	0.3624
0.4	0.05	0.3623
0.4	0.06	0.3628
0.4	0.07	0.3625
0.5	0.03	0.3645
0.5	0.04	0.3633
0.5	0.05	0.3624
0.5	0.06	0.3629
0.5	0.07	0.3623
0.6	0.03	0.3622
0.6	0.04	0.3622
0.6	0.05	0.3625
0.6	0.06	0.3622
0.6	0.07	0.3622
0.7	0.03	0.3622
0.7	0.04	0.3622
0.7	0.05	0.3623
0.7	0.06	0.3623
0.7	0.07	0.3622

Table 11.11: Best RMSE values for different Crossover and Mutation probabilities

Gravity Model Output

Lon	Lat	Gravity	Fast Food	Hotels	Motorway Links	Motor-ways	Restau-rants	Super-markets	Num. Station	Min Cap.	Max Cap.	Use	Owner
10.1612	57.1524	2120.27	0.0	0.0	1458.55	661.721	0.0	0.0	2	43.0	50.0	0.88	E.ON
9.36722	55.0655	1129.36	0.0	0.0	892.911	236.453	0.0	0.0	2	43.0	50.0	0.89	E.ON
9.76432	56.1476	2627.19	0.0	0.0	1259.32	1367.87	0.0	0.0	2	43.0	50.0	1.18	E.ON
9.93721	57.2226	608.926	0.0	0.0	0.0	608.926	0.0	0.0	2	43.0	50.0	1.24	E.ON
9.76695	56.1497	3011.12	0.0	0.0	1861.93	1149.19	0.0	0.0	2	43.0	50.0	1.28	E.ON
10.0659	56.1752	10067.5	0.0	0.0	5690.52	4376.95	0.0	0.0	2	43.0	50.0	1.41	E.ON
9.40002	55.5141	2882.92	0.0	0.0	2862.77	20.1489	0.0	0.0	2	43.0	50.0	1.45	E.ON
9.0001	56.038	42.575	0.0	0.0	0.0	42.575	0.0	0.0	2	43.0	50.0	1.56	E.ON
9.94186	57.222	837.534	0.0	0.0	0.0	837.534	0.0	0.0	2	43.0	50.0	1.6	E.ON
9.39401	55.512	3596.82	0.0	0.0	3586.08	10.7408	0.0	0.0	2	43.0	50.0	1.62	E.ON
9.51789	56.2008	4102.02	0.0	0.0	3642.61	459.409	0.0	0.0	3	60.0	300.0	1.86	Shell Recharge
9.32942	54.8262	1724.35	0.0	0.0	1508.42	215.926	0.0	0.0	2	43.0	50.0	1.99	E.ON
10.0654	56.1781	7811.53	0.0	0.0	3434.58	4376.95	0.0	0.0	2	43.0	50.0	2.16	E.ON
10.0242	57.1	1746.9	0.0	0.0	832.517	53.6076	0.0	860.774	2	300.0	300.0	2.18	Spirii
9.5784	56.1858	7072.42	0.0	0.0	6179.75	57.7295	0.0	834.936	8	600.0	600.0	2.23	Spirii
12.2314	55.5607	20076.7	9920.99	0.0	1931.22	8224.46	0.0	0.0	2	43.0	50.0	2.26	E.ON
11.6519	55.7006	192.106	0.0	0.0	0.0	192.106	0.0	0.0	2	43.0	50.0	2.3	E.ON
8.58928	55.5246	9.73065	0.0	0.0	0.0	9.73065	0.0	0.0	3	60.0	300.0	2.43	Shell Recharge
9.74208	56.8436	2823.21	0.0	0.0	1571.1	1252.11	0.0	0.0	2	43.0	50.0	2.44	E.ON
11.6567	55.7018	2163.89	0.0	0.0	623.285	1540.6	0.0	0.0	2	43.0	50.0	2.56	E.ON

Table 11.12: Example of Gravity Model form Optimized Motorway