CTU FEL Prague

Master thesis

EV vehicles

and charging demand

Bc. Jakub Zíka

May 20, 2025

CTU FEL

Contents

Introduction

The invention of the automobile in the late 19th century revolutionized human mobility, enabling unprecedented freedom to traverse long distances. However, this breakthrough hinged not only on the internal combustion engine (ICE) itself but also on the parallel development of a critical support system: gasoline stations. Just as early motorists relied on scattered gas stations to power their journeys, the rise of ICE vehicles necessitated a standardized, accessible network of refueling infrastructure to sustain their adoption. This relationship between vehicles and their energy infrastructure became a cornerstone of modern transportation, shaping urban planning, economic systems, and global energy policies.

Today, as societies pivot toward sustainability, electric vehicles are heralding a similar paradigm shift. Yet their widespread adoption faces a challenge mirroring the early days of automobiles: the need for reliable and efficient charging infrastructure. While electric vehicles eliminate tailpipe emissions, their practicality depends on overcoming "range anxiety" and ensuring charging availability aligns with user behavior—issues that gas stations largely resolved for ICE vehicles over a century of iteration. Predicting EV charger usage, therefore, is not merely a technical exercise but a good step in designing infrastructure that mirrors the ubiquity and convenience of gas stations. And helps smoothen transition.

Electromobility and Climate Change 1

This chapter explores the motivations behind the transition to electric mobility, examining the environmental causes driving this shift, the evolution of electric vehicles, regulatory frameworks accelerating adoption, and the critical infrastructure challenges that must be addressed to enable widespread electrification of transportation. One of the identified challenges, which is demand for chargers will be introduced. It will be the main topic of this thesis and our focus on predicting it based on existing data for new potential charger locations to aid the decision making process.

1.1 Climate change

Climate change represents one of the most pressing global challenges of our time, with transportation being a significant contributor to greenhouse gas emissions. The burning of fossil fuels is the primary driver of anthropogenic climate change, with transportation accounting for approximately 24% of direct CO_2 emissions from fuel combustion globally [1].

The environmental impact of ICE vehicles extends beyond carbon dioxide emissions. These vehicles also produce other polutants. Which have negative health outcomes like increased mortality (about 3% of mortality in Europe [2]), asthma attacks, higher cardiovascular hospital admissions, higher repiratory hospital admissions [3].

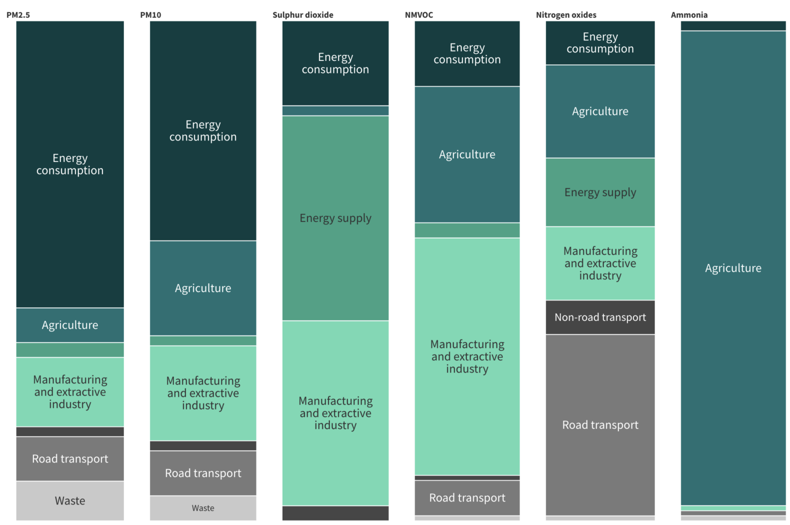


Figure 1.1: Share of EU emissions in 2020. PM2.5 = fine particulate matter; PM10 = particulate matter; Energy consumption = residential, commercial and institutional energy consumption; NMVOC = non-methane volatile organic compounds. Source [4]

While the scientific consensus on anthropogenic climate change is robust, it is worth acknowledging that the transition to electric mobility is not without controversy. The manufacturing process of electric vehicle has higher negative enviromental impact compared to internal combustion engine vehicle due to battery manufacturing. But they have a chance to mitigate this over their lifetime use if the share of generated energy for their use comes from sustainable sources. Combined with remanufacturing and recycling of old batteries. [5]

[5]: Xia et al. (2022), 'A Review of the Life Cycle Assessment of Electric Vehicles'

1.2 Electric vehicles

The history of electric vehicles (EVs) is marked by a parallel development alongside internal combustion engine vehicles, rather than being a purely modern innovation. In fact, electric vehicles were among the first automobiles developed in the late 19th century, with inventors like Gustav Trouvé creating electric cars as early as 1881 [6]. During the early automotive era, electric vehicles competed directly with steam and gasoline powered vehicles. And even had dominance in early 20th century (An electric vehicle taxi existed for a short period of time in 1987 London [7]).

[6]: Wakefield (1998), *History of the Electric Automobile: Hybrid Electric Vehicles*

[7]: (2012), *The Surprisingly Old Story Of London's First Ever Electric Taxi*

However, the limitations of early battery technology—particularly in terms of energy density, range, and recharging infrastructure—combined with the discovery of abundant petroleum reserves and the introduction of the electric starter for gasoline engines, led to the dominance of ICE vehicles throughout most of the 20th century . And use of electric engines in transportation remained in city trams and subways.

The modern resurgence of electric vehicles began in the late 1990s and early 2000s, driven by advances in lithium-ion battery technology, growing environmental concerns. The introduction of hybrid vehicles like the Toyota Prius served as a transitional technology, familiarizing consumers with electric drivetrains while alleviating range anxiety through the backup of a gasoline engine. The launch of the Tesla Roadster in 2008 demonstrated that electric vehicles could offer performance comparable to or exceeding that of high-end sports cars, challenging perceptions that EVs were inherently limited in capability [8].

[8]: (), *The Electric Vehicle Future Was Promised Decades Ago, What Happened?*

Today's electric vehicles have largely overcome many of the historical limitations that hindered their adoption. Modern EVs offer ranges exceeding 300-400 kilometers on a single charge, with high-end models approaching 600 kilometers. Fast-charging infrastructure has expanded significantly, enabling long-distance travel with reasonable charging stops. The total cost of ownership for EVs has become increasingly competitive with ICE vehicles due to lower operating and maintenance costs, despite higher initial purchase prices. A gap that continues to narrow as battery costs decline and economies of scale improve (see Figure 1.2).

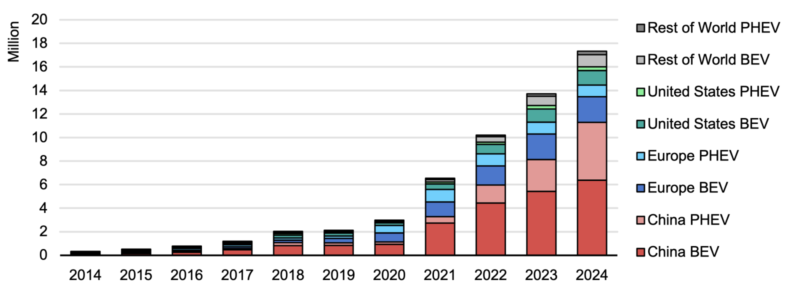
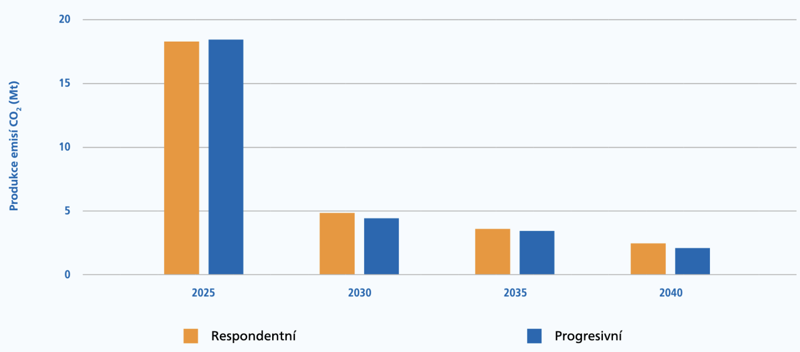


Figure 1.2: Global electric car sales, 2014-2024. Source: IEA analysis based on country submissions, ACEA, EAFO, EV Volumes and Marklines [9]

The environmental benefits of electric vehicles are substantial, particularly when powered by low-carbon electricity sources. Even when accounting for the current global electricity mix, which still includes significant fossil fuel generation, EVs typically produce lower lifecycle greenhouse gas emissions than comparable ICE vehicles. As electricity grids continue to decarbonize, this advantage will only increase (see Figure 1.3). Additionally, the shift of emissions from millions of individual tailpipes to centralized power plants offers significant air quality benefits in urban areas and creates opportunities for more efficient pollution control.



	2025	2030	2035	2040
Respondentní	18,3 Mt	4,8 Mt	3,6 Mt	2,4 Mt
Progresivní	18,5 Mt	4,4 Mt	3,4 Mt	2,1 Mt

Figure 1.3: CO_2 emission from electricity production prediction for Czechia for selected years [10]

However, the transition to electric mobility faces many of the same infrastructure challenges that the early automobile industry encountered. Just as the widespread adoption of ICE vehicles required the development of a comprehensive network of gas stations, repair facilities, and roads, the EV revolution depends on the deployment of charging infrastructure, grid upgrades, and maintenance expertise. These parallels suggest that while the challenges are significant, they are not unprecedented and can be overcome through coordinated investment and policy support.

1.3 EU mandate

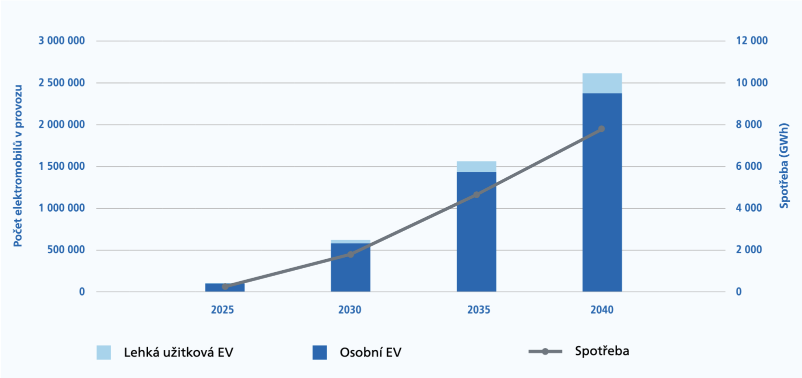
The European Union has established regulatory framework to accelerate the transition to electric mobility as part of its broader climate strategy. The cornerstone of this approach is Regulation (EU) 2019/631 [11], which sets CO_2 emission performance standards for new passenger cars and light commercial vehicles. This regulation is part of "Fit for 55" package proposed in July 2021 and subsequently adopted, which aims to reduce net greenhouse gas emissions by at least 55% by 2030 compared to 1990 levels [12][13].

The most transformative element of this regulatory framework is the mandate that effectively prohibits the sale of new internal combustion engine vehicles in the EU from 2035 onward. Specifically, the regulation requires a 100% reduction in CO_2 emissions from new cars and vans by 2035 compared to 2021 levels, which in practice means that only zero-emission vehicles—battery electric or hydrogen fuel cell—can be

[11]: (2019), *Regulation (EU) 2019/631 of the European Parliament and of the Council of 17 April 2019 Setting CO2 Emission Performance Standards for New Passenger Cars and for New Light Commercial Vehicles, and Repealing Regulations (EC) No 443/2009 and (EU) No 510/2011 (Recast) (Text with EEA Relevance.)*

[12]: (), *Časová osa – Zelená dohoda pro Evropu a balíček „Fit for 55“*

[13]: (), *Co je Fit for 55*



	2025	2030	2035	2040
Lehká užitková EV	3 905	41 984	128 239	239 036
Osobní EV	95 009	576 487	1 430 746	2 373 627
Spotřeba	219 GWh	1 773 GWh	4 641 GWh	7 788 GWh

Figure 1.4: Prediction on number of EVs in Czechia taking into account the (EU) 2019/631 regulation

sold as new vehicles after this date. This represents a clear and unambiguous signal to the automotive industry, infrastructure developers, and consumers about the direction of transportation policy in Europe.

The EU's approach includes intermediate targets to ensure a gradual transition: a 55% reduction in car emissions and a 50% reduction in van emissions by 2030 compared to 2021 levels. These targets are accompanied by incentive mechanisms for zero- and low-emission vehicles, penalties for manufacturers that exceed fleet-wide emission targets, and provisions for reviewing the effectiveness of the regulation.

This regulatory certainty has already catalyzed significant investment in electric vehicle production and charging infrastructure across Europe. Major automotive manufacturers have announced accelerated timelines for electrifying their fleets, with many planning to phase out ICE vehicle production well before the 2035 deadline. The mandate has also spurred innovation in battery technology, charging solutions, and vehicle design as companies compete to position themselves advantageously in the emerging electric mobility ecosystem.

By establishing a definitive end date for new ICE vehicle sales, the EU has moved beyond incremental improvements to fossil fuel efficiency and committed to a fundamental technological transition in personal transportation.

1.4 Electric Vehicle Chargers

Due to high urban density in Prague. Especially in the city center which mostly consist of apartment buildings. Street parking is the dominant type of parking here. Those parking locations do not have easy access to electric vehicle charging and they require construction as well as having the power source somewhere (underground). Opposed to single family housing where high number of parking spots is located on the houses private land. Where there is higher accessibility to private electric charger. Where the simplest solution can be an electrical outlet.

Modern EV charging infrastructure can be categorized by power output, which directly affects charging speed [14]:

[14]: (), *Regulation - 2023/1804 - EN - EUR-Lex*

- **Category 1 (AC):**
  - **Slow AC charging:** Utilizing standard household outlets (P < 7.4 kW). While inadequate as a primary charging solution for most users, they serve as emergency options or for overnight charging in residential settings.
  - **Medium-speed AC charging:** Operating at 7.4-22 kW, these chargers can fully replenish most EV batteries in 4-8 hours.
  - **Fast AC charging:** Operating at > 22 kW.
- **Category 2 (DC):**
  - **Slow DC charging:** Less than 50 kW.
  - **Fast DC charging:** 50-150 kW.
  - **Ultra-fast DC charging:** 150-350 kW and above.

DC fast charging stations can provide an 80% charge in 20-40 minutes for compatible vehicles. They are strategically deployed along major travel corridors and in urban centers to enable long-distance travel and quick top-ups for those without home charging access.

The installation costs vary based on charger type. A significant challenge is managing grid load, particularly during peak demand periods such as after-work hours. This has led to the development of smart charging systems with dynamic pricing. Chargers can also be installed in public lighting lamps as a cost-effective solution.

The technical challenges of charging infrastructure placement extend beyond the charging equipment itself to include grid integration considerations. High-power charging stations can place significant demands on local distribution networks, potentially requiring costly grid upgrades or reinforcement. Strategic placement that aligns with existing grid capacity can substantially reduce deployment costs and timelines.

Data-driven approaches to infrastructure planning incorporate multiple data sources including traffic patterns, demographics, points of interest, existing infrastructure utilization, grid capacity, and temporal mobility patterns. By integrating these diverse datasets, planners can develop models that predict charging demand with high temporal and spatial resolution, enabling more efficient infrastructure deployment that maximizes utilization while minimizing costs.

1.5 Goals of the thesis

This thesis explores the challenge of predicting electric vehicle charging demand in Prague through data analysis and machine learning approaches. Working with data provided by PREdistribuce, this research investigates whether spatial and temporal features can effectively predict charging patterns. The specific objectives addressed in this thesis are:

- To examine the current state of electromobility and climate change, establishing the context for the growing importance of charging infrastructure planning
- To integrate publicly available spatial data sources, including Basic Settlement Units (ZSJ) demographics and OpenStreetMap points of interest, with charging data to create a comprehensive feature set

- To develop a neural network model with latent profiles that attempts to capture underlying patterns in charging behavior while maintaining interpretability
- To evaluate the model's performance against baseline approaches (average model, linear regression, XGBoost) using various error metrics
- To analyze the learned latent profiles to determine if meaningful charging patterns could be extracted despite the model's overall performance limitations
- To identify data limitations and suggest potential improvements for future research in charging demand prediction

This research represents an initial exploration into charging demand prediction in the context of Prague, acknowledging the challenges of limited data availability and quality. While the predictive performance did not exceed baseline models, the methodological approach and analysis of limitations might provide insights for future research directions in this field.

Related research 2

This chapter reviews literature on electric vehicle (EV) charging demand prediction and related methodologies, providing context for our research on charging patterns in Prague.

We examine four main research areas: data-driven approaches to understanding charging behavior, simulation-based methodologies for predicting demand, stochastic modeling approaches capturing probabilistic behavior, and research specific to the Czech context. The review highlights spatial and temporal dimensions of charging demand, data requirements of different approaches, and trade-offs between model complexity and accuracy.

This analysis identifies gaps in existing research regarding Prague's urban environment and informs our methodological approach.

2.0.1 Understanding EV Charger Use by Data

To understand why certain chargers are being utilized the way they are. Research utilizes traditional and Bayesian statistics. As the person that plans charging connector (CP) it is good to have insight into what influences charging demand. That is, why a certain charger is utilized. And what factors contribute to it. So far, we dont care about expansion of the infrastrcture. But could provide insights that allow to place new chargers more strategically.

[15] gathered counts of types of point of interest (POI)[1] near every charger of interest [2] from Open Street maps. Then for each charger computed its utilization. Which is its average daily power consumption. Then used linear regression to test which of the category of point of interest (POI) contributed to the consumption. The study had some statisficaly significant results. They also trained neural network model for capturing non linear relationships. User can use the model to select any point where a charging connector (CP) might be placed and see its estimated utilization and evaluate worthiness of placement. However it does not work with other chargers in the area and does take into account charger density. The paper has identified that certain categories of point of interest (POI)[3] are correlated with charger demand.

[15]: Hecht et al. (2024), 'Global Electric Vehicle Charging Station Site Evaluation and Placement Based on Large-Scale Empirical Data from Germany'

1: types like shops, sport areas, schools, offices

2: Search radius of 2000m around each charger with linear decrease in importance in relation to distance

3: POI data obtained from OpenStreetMaps

[16] uses log-Gaussian Cox process. Which is a statistical model that can handle dependence between points on a map (EV chargers). It has identified that workplace population and traffic flow are positively related to demand ofcharging connector (CP) while commerce is in a negative relation. [4].

[16]: Dong et al. (2019), 'Electric Vehicle Charging Point Placement Optimisation by Exploiting Spatial Statistics and Maximal Coverage Location Models'

4: <workplace population, traffic flow, commerce description here>

Simulation vs Model

Model is concerned with representing a system of interest. Purpose of it is so it matches its real system in some behaviour while being simpler than reality. There are many ways how to construct models.

From simplest small replicas (small car models to use in wind tunnel) to complicated mathematical and software ones.

Simulation is then a use of model to try to replicate some real system of interest to gain insights how a real system might behave .

An example can be a toy car and a ramp which would be our model. Then pushing the car off the ramp to see how far it is able to land. This is much less expensive than pushing real cars of larger ramps.

[17]

2.0.2 Simulations and EV Charger Use

Trafic models are used to estimate charging connector (CP) demand. They are a mathematical model of real-world trafic. And they can estimate charging demand by introducting EV vehicle agents into the simultion. Those agents navigate replica of real road network. While also being able to emulate traffic jams, alternative path finding. Negative of these models is that they are harder to develop, fine tune to match real situation. And performance costly. But once developed and validated with real-world. They can be highly utilized.

Micro,Macro,Meno-Scopic

Traffic models can be separated into three groups according to [18][19]([20] also writes about splitting simmulation models into 3 groups with a more theoretical view):

- **Microscopic** - Simulation of individual agents/vehicles. Can be as detailed as emulating accelerating, lane changing, turning. Higher fidelity isuUsable for example when designing intersection with light signals optimizing for traffic or pedestrian flow. But the model can also have less detailed fidelity for vehicles. Which simplifies the simmulation and enables larger area of interest. Like cities or even countries.
- **Mesoscopic** - Provide less fidelity than microscopic. improve
- **Macroscopic** - Work with aggregate information. Work more on analytical model and proven mathematcial relationships between traffic flow. improve

Matsim [21] is Agent based model (ABM) . It's OpenSource software licensed under GPLv2, implemented in Java. And started in 2004 in Zurich. It falls into category of microscopic models due to it being able to simluate individual agents. The model simmulates simplifed version of vehicle movement and traffic which is needed to correctyl simulate entire city including congestions. The simulation usually concerns only one day. To create such a model the following are needed:

[21]: Community (2025), *MATSim*

Slight foreshadowing but this approach was not taken in this thesis however it was considered.

- **Road/communication Network** - Spatial data for area of interest reagarding: roads, intersections, tram or train tracks etc. Obtained from technical data published by goverments or third party map providers. (Mainly OpenStreetMaps)

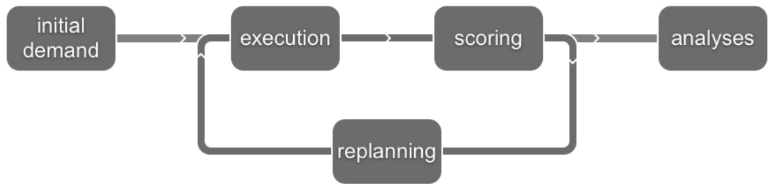


Figure 2.1: Stages of matsim simulation. [22]

- **Population** - Agents with their day plan, where are they going to live and commute to work. Those are generated with help of expert knowledge and travel diaries survey.
- **Facilities** - Locations where activities can be performed. Obtained from third parties like OpenStreetMaps.

The agents plan assigns unique daily schedule for each agent in the population. See 2.1 This estabilished the initial demand. During the execution step, Matsim simulates agents commute and tries to satisft the agents schedule. The agent might fail to be on schedule due to road or public transport congestion from other agents. After finishing the simulation the agents actual daily schedule is scored. Penalizing arriving at work late or being stuck for too long in traffic giving agents lover score. While agents who got where they wanted qucikly and without trafic jams get better score. During replanning agent schedule can be modified so it either adjusts the agents schedule, like less time spent home or taking a different route to work. When the simulation is sufficiently optimized the results of it can be then used for various analysis as the output of the modle is detailed log of agent activities.

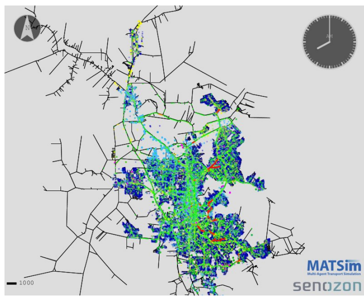


Figure 2.2: Matsim, Joinville example. Model intended to help the city encumbered by high traffic volumes. `https://matsim.org/gallery/joinville/`

To create the agents schedule, day intentions need to be provided to the model. Those can be either manually crafted. With the increased availabiliy of census data, data driven approach can be taken. This trades away the possibility to experiment with different kinds of populations and policies[5] but simplifies the process of creation where only data is needed and not expert knowledge on human behavior. [23] models multiple probability densities for what activity will an agent do, how long it will take and what activities are available in the area. This utilizies several datasets. New schedules can be generated from the model by sampling from the distributions.

Finaly, moving to use of Matsim for ev scenarios. [24] studies the impact of EVs on the electric id and electricty production in Croatia. By having a simmulation of power production network for the whole country as well. They first have a simmulation regarding the current state and how that impacts the electrical grid. This sets the baseline and also allows to correctly calibrate the model. Then they increase the ev adoption and see what is the hourly energy consumption of EVs and their impact on the grid.

5: like exmaining effect of different school time start on traffic

[23]: Drchal et al. (2019), 'Data-Driven Activity Scheduler for Agent-Based Mobility Models'

[24]: Novosel et al. (2015), 'Agent Based Modelling and Energy Planning – Utilization of MATSim for Transport Energy Demand Modelling'

Stochastic models provide a way for modeling charging behavior of electric vehicles by capturing randomness and uncertainty in travel patterns. They work with more simplified representation fo reality compared traffic models[6]. These models typically employ probability distributions to characterize variables such as departure times, travel distances, parking durations, and charging decisions. By sampling from these distributions using techniques like Monte Carlo simulation, they generate synthetic power demand profiles.

6: Of course trafic models have large part of stochasticity inside themselves

[25] presents an example of stochastic modeling applied to EV charging behavior. Their approach uses a non-parametric copula function to model the dependence structure between six variables: departure time, number of journeys, and total distance traveled across two consecutive days. For this they utilize real-world GPS data collected from electric vehicles. Their model simulates complete journey schedules for individual vehicles and implements a probabilistic charging decision model at each destination, conditioned on the state of charge, parking time, and journey number. The approach captures the variability in charging behavior by incorporating factors such as battery characteristics and probabilistic charging point availability.

[25]: Brady et al. (2016), 'Modelling Charging Profiles of Electric Vehicles Based on Real-World Electric Vehicle Charging Data'

[26] presents a probabilistic model for EV charging patterns in residential networks. The authors employ a stochastic approach using Markov chains to model transitions between driving, parked, and charging states. Their model accounts for various factors such as vehicle type, battery capacity, state of charge, driving habits, and charging preferences. By simulating daily activity patterns across different seasons and day types, they estimate charging needs for various EV penetration levels (34%, 50%, 65%). Results show distinct charging patterns between weekdays and weekends, with weekend peaks occurring later in evenings. This estimation of charging load patterns provides valuable insights for distribution network planning and management.

[26]: Ul-Haq et al. (2018), 'Probabilistic Modeling of Electric Vehicle Charging Pattern in a Residential Distribution Network'

[28] utilizes graphical model. It breaks down the problem into modelling several probability distributions. They posessed charging sessions data with anonymized car ids. With the car ids they reconstructed drivers charging behaviour. They constructed vector encoding the drivers charging history and then aplied hierachcical clustering to group drivers into groups based on their behavior. They separated the chargers into 6 segments based on the charging power and its location (single family house, multi unit dwelling, workplace, fast charger). Then modelled the groups probability distribution of charging at the segments (see the graphical model at Figure 2.3). Then by sampling from the distribution they were able to reconstruct the power demand profile. And importanlty, see the effect by changing the amount of drivers in the identified driver groups. Being beneificial to policy makers.

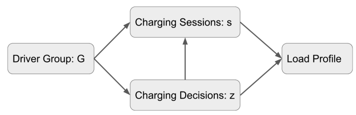


Figure 2.3: Graphical model of modelling EV driver charging behaviour [27]

[28]: Powell et al. (2022), 'Scalable Probabilistic Estimates of Electric Vehicle Charging given Observed Driver Behavior'

2.1 Relevant Research for Czechia

[29] introduces a standalone model for electric vehicle charging demand based on large-scale travel survey data from the Czech Republic. The model serves as a comprehensive input for subsequent charging infrastructure optimization problems. It develops a graph-based representation of the road network where traffic data is transformed from edge-defined to vertex-defined format. This mathematical approach allows for the introduction of a point-defined (vertex-defined) charging demand model where each road intersection (vertex) has an associated demand value.

[29]: Pekárek (2017), 'A Model of Charging Service Demand for the Czech Republic'

The model utilizes real traffic survey data to determine the spatial distribution of charging demand, calculating how demand at specific points is influenced by traffic patterns in surrounding areas. This is accomplished through a distance diminishing function that propagates demand through the network while simulating driver behavior. It applies this model to the entire Czech Republic road network (comprising over 30,000 road segments) and identifies several limitations, particularly related to data availability in municipal areas and potential calculation inaccuracies in regions with high-density road networks.

[30] focuses on optimization of charger placement in Czechia but on a private land.

[30]: Neumann (2021), 'Optimální Rozmístění Nabíjecích Bodů pro Firemní Flotily'

[31] focuses on detemrining the best placement of hydrgoen refill station. By using GIS multi-criteria decision making. By taking into account criteria such as: proximity to junctions, flood prone areas, flat terrain, safety distance from selected area types and high population area. They did not posess data about the existing hydrogen

[31]: Elomiya et al. (2024), 'An Advanced Spatial Decision Model for Strategic Placement of Off-Site Hydrogen Refueling Stations in Urban Areas'

2.2 Discussion

The literature review reveals several distinct approaches to modeling and predicting EV charging demand, each with its own strengths and limitations.

Simulation-based approaches, particularly traffic models and agent-based models like MATSim, offer high flexibility and the ability to model complex interactions between vehicles, infrastructure, and human behavior. These models are valuable in data-scarce environments where empirical charging data is limited or non-existent. The work by [24] demonstrates how such simulations can be extended beyond charging demand prediction to assess impacts on electrical grids. However, these approaches require significant computational resources, expert knowledge for calibration, and extensive input data about road networks, population characteristics, and travel patterns.

In contrast, stochastic models like those presented by [25] and [26] offer a more data-driven approach while still capturing the inherent variability in charging behavior. These models typically employ probability distributions to characterize key variables such as departure times, travel distances, and charging decisions. The graphical model approach by [28] is particularly noteworthy for its ability to identify distinct driver groups and model their charging behaviors separately, enabling more nuanced policy analysis.

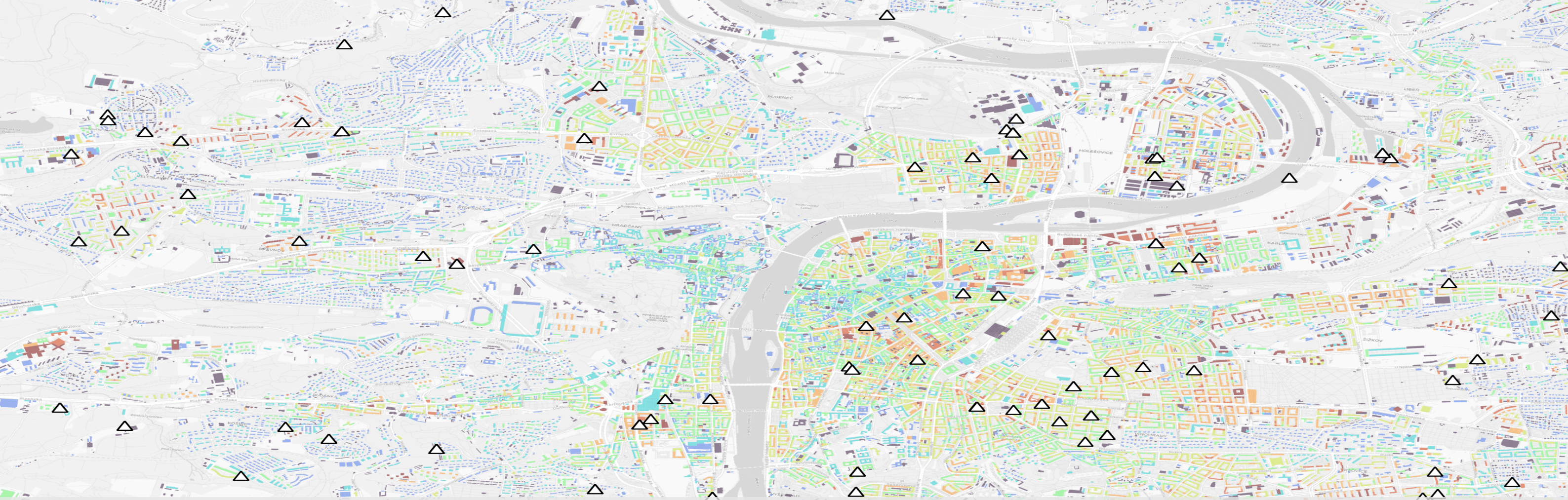
The Czech-specific research by [29] provides valuable insights into modeling charging demand in the Czech Republic using a graph-based representation of the road network. However, this approach focuses on the national level and does not address the unique urban characteristics of Prague.

Several patterns emerge from this literature review:

1. As data availability increases, research tends to shift from simulation-based approaches toward more data-driven statistical and machine learning methods. This progression reflects the natural evolution of research in emerging domains, where initial exploratory models based on expert knowledge gradually give way to empirical models as more data becomes available.

2. Stochastic models that capture the probabilistic nature of human decision-making processes have proven effective in modeling charging behavior. These models can represent the inherent uncertainty in factors such as arrival times, charging durations, and energy consumption patterns.
3. The spatial context of charging stations significantly influences their usage patterns, with factors such as proximity to workplaces, commercial areas, and transportation hubs playing important roles in determining charging demand.
4. Temporal patterns in charging behavior exhibit distinct variations across different days of the week and seasons, necessitating models that can capture these cyclical patterns.

The methodological approaches in the literature also reveal a clear trade-off between model complexity and data requirements. Simulation-based approaches like MATSim require extensive configuration and calibration but can operate with limited charging data, making them suitable for early-stage planning when empirical data is scarce. Conversely, data-driven approaches like those employed by [28] and [15] require substantial historical charging data but can potentially deliver more accurate predictions with less configuration effort.



3 Data

This chapter introduces the data sources and processing methods that form the foundation of our research on electric vehicle charging demand. We begin by classifying the types of data used in our analysis and explaining their relevance to the research problem. We then describe the charging infrastructure data, including chargers and charging sessions, and detail how we transform this raw data into the target variables for our predictive model. Finally, we present the spatial and contextual data sources that provide the features for our model, along with the transformations applied to prepare them for analysis.

The data landscape presented here directly supports the modeling approach described in Chapter **??**, where these processed datasets serve as inputs to our neural network with latent profiles.

Types of data

- **Spatial** (geospatial) data are those which have assigned position in real world and are invariant in some timeframe. Such data are locations of charging station, administrative boundaries, road network, buildings.
- **Temporal** data are characterized by their variation over time without specific geographical coordinates. In this thesis, these include charging session durations, energy consumption patterns throughout the day, historical charger utilization rates, and seasonal variations in charging demand.
- **Spatio-temporal** data incorporate both location and time elements, providing insights into how phenomena evolve across space and time. Examples relevant to this research include mobility patterns of people, real-time charger availability, and dynamic variations in APC across different city zones during different hours of the day.

Title image is a map of Prague with all chargers denoted as triangles in available datasets. The layer below displays all buildings in Prague with color being the number of floors

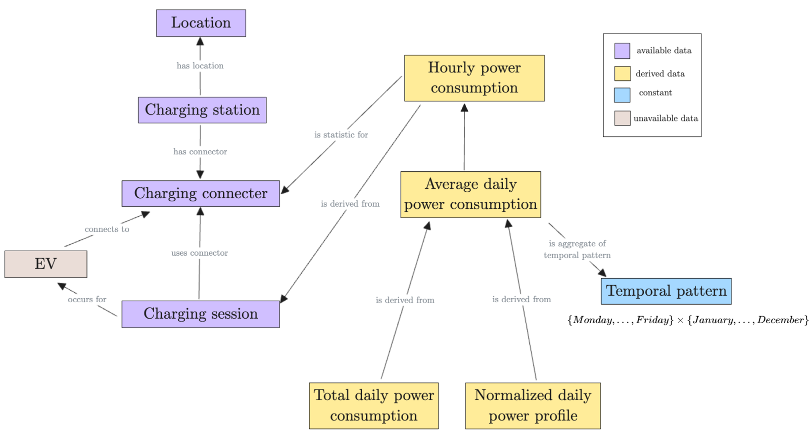


Figure 3.1: A charging ontology showing the relationships between entities in the EV charging ecosystem

3.1 EV Chargers and Charging Sessions

3.1.1 Description

To introduce our problem domain and establish a framework for connecting the various data elements, we present a simple charger ontology [1]. Inspired by the AURORAL EV-charger Ontology [32], this ontology helps clarify the relationships between entities in the EV charging ecosystem. The model aligns with the data obtained from PRE (Prague's electricity distribution company).

1: Ontology describes subjects of some system and the ways they are related to each other.

[32]: (), *Ontology Documentation Generated by WIDOCO*

Below is a description of the individual components of the charging ontology as illustrated in Figure 3.1.

- **charging station** (as visible in Figure 3.2) is equipment that connects an EV to a source of electricity to recharge them. A charging station typically consists of physical infrastructure including power conversion hardware, connectivity modules, authentication systems, and user interfaces. Charging stations vary in their power delivery capabilities, ranging from slow AC chargers (3.7-22 kW) commonly found in residential and workplace settings to fast DC chargers (50-350+ kW) deployed in public corridors and commercial hubs. Within our dataset, charging stations from PRE's network predominantly consist of public AC and DC installations distributed throughout Prague's urban and suburban areas.

  Formally, we define the set of all charging stations as $S = \{s_1, s_2, \ldots, s_m\}$, where each station $s \in S$ is associated with a location $l_s \in \mathbb{R}^2$ representing its geographic coordinates.

Figure 3.2: Picture of charging station. It has one connector on each of its sides. One of which has charging cable attached.

- **charging connector** (as visible in Figure 3.3) one or many are part of a CS. These physical interfaces allow for the actual connection between the vehicle and the charging infrastructure. Connectors follow different standards depending on region and charging speeds. Each connector type supports specific charging protocols and power levels. In our studied network, the majority of charging stations feature multiple connectors, enabling simultaneous charging of different vehicles and supporting various connector standards to accommodate the heterogeneous EV market.

Figure 3.3: View of 1 of the 2 charging connectors the CS has

For each station $s \in S$, we define the set of connectors as $C_s = \{c_1^s, c_2^s, \ldots, c_{n_s}^s\}$, where n_s represents the number of connectors at station s. Each connector $c \in C_s$ has a unique identifier $\mathrm{id}^{c,s}$ assigned by PRE.

- **charging session** occurs when an EV arrives at a CS and connects to a CP. This interaction initiates a session that is logged by the CS together with various parameters including connection time, disconnection time and total power consumed. The charging session captures both spatial, temporal patterns (duration, time of day, day of week) and energy consumption behaviors.

  For each connector $c \in C_s$ at station s, we define the set of charging sessions as $V_{c,s} = \{v_1^{c,s}, v_2^{c,s}, \ldots, v_{k_{c,s}}^{c,s}\}$, where each session $v_i^{c,s} = (t_{\text{start},i}^{c,s}, t_{\text{end},i}^{c,s}, p_i^{c,s})$ is characterized by its start time, end time, and energy consumed during the session.

- **Location** denotes the geographical position where the charger is installed. This spatial attribute is crucial to our analysis framework as it allows for correlation between charging demand and various features of the surrounding environment.

  The location function $l : S \rightarrow \mathbb{R}^2$ maps each station $s \in S$ to its geographic coordinates.

Power consumption assumption

Before we describe how charging session have been processed into hourly power consumption and average power consumption with temporal pattern, it's important to clarify our simplifying assumption regarding power consumption during charging sessions. In reality, power consumption can vary significantly during a charging session, typically following a non-linear pattern. However, for modeling purposes, we make the following assumption:

The vehicle, from connection time, is charged at the maximum power that both the electric vehicle can handle and the charging connector can provide. This leads to charging sessions being divided into two distinct phases: 1. A charging phase, during which power is actively delivered to the vehicle 2. An idle period, where the vehicle has been fully charged but remains connected, consuming no power[2]

This simplification is illustrated in Figure 3.4, which contrasts our rectangular approximation with a more realistic charging curve.

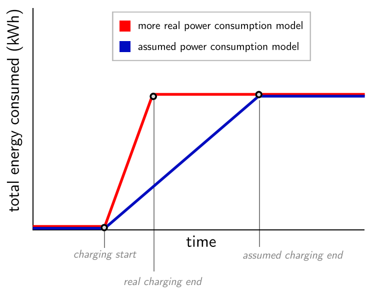


Figure 3.4: Chart comparing realistic power consumption vs our assumption.

2: Some charging station providers financially penalize this idle period, as another EV could be charging during this time. This practice can lead to improved availability of charging station.

3.1.2 Charging Sessions Dataset

PRE provided us with charging session data collected from **907** charging stations across Prague. We collected **385,404** individual charging sessions spanning **January 2022 - June 2024**. For each session, the following information is available:

- **Charger ID**: A unique identifier assigned by PRE to each charger and connector

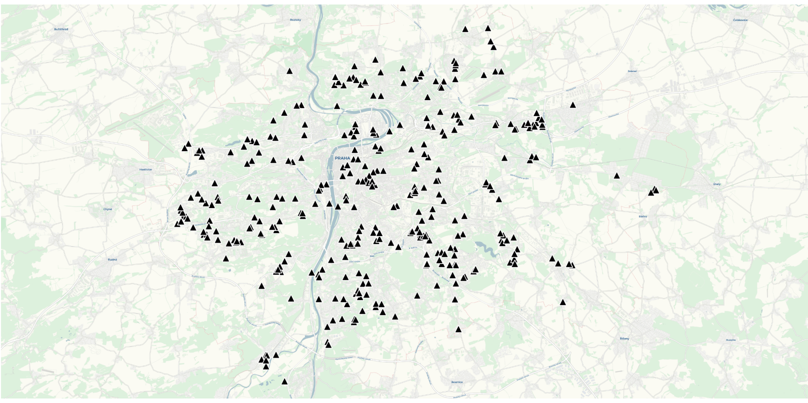


Figure 3.5: Map of all PRE charging station in Prague for which we have available charging session. See Figure 1 for larger image.

- **charging station location**: The physical address of the charging station
- **Connector type**: Categorized as AC, DC, or UFC
- **Session timestamps**: Start and end times of each charging session
- **Energy consumption**: Total power consumed during the session

We identified the precise location of chargers using supporting documentation from PRE.

3.1.3 Transformations

We are interested in obtaining the power demand of chargers with hourly granularity. This allows us to analyze temporal patterns in charging behavior and develop predictive models for any charging connector at any charging station.

We derive hourly power consumption $h_t^{s,c}$, where s specifies the charging station, c identifies the charging connector at the station, and t represents the hour. The allowed range of t is limited by the start of the first session until the end time of the last session.

To obtain $h_t^{s,c}$ from the sessions, the following transformation process is applied for each station-connector pair (s, c):

1. Compute the total active timespan of the charger. Create an empty hours list

$$h_t^{s,c} = 0, \forall t$$

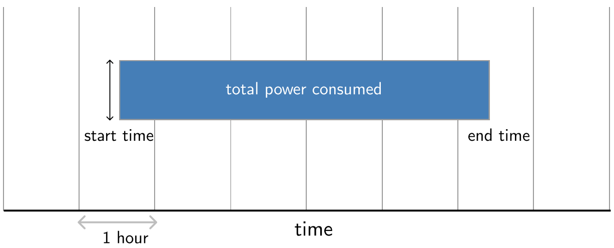


Figure 3.7: charging session overlaid with time axis, showing the temporal distribution of charging events.

2. Take each charging session k $v_k^{c,s}$ for the station-connector pair (s, c). Divide it into hourly chunks, and for each chunk redistribute the power consumed weighted by the fraction of an hour the chunk

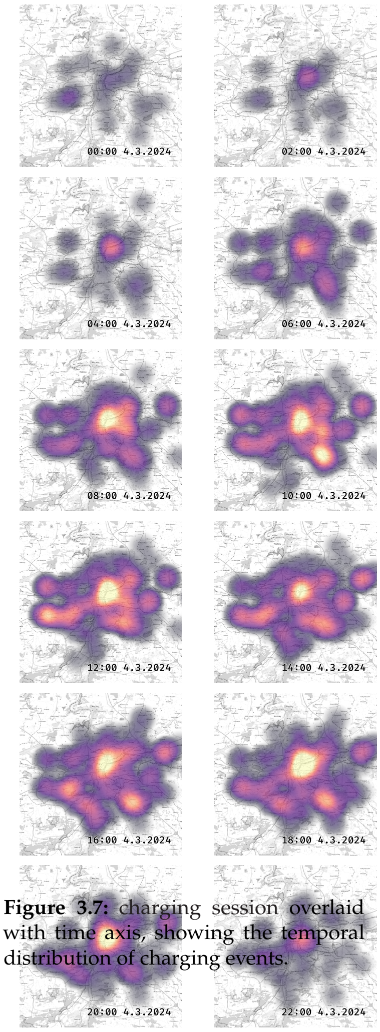


Figure 3.6: Heatmap images of current charging sessions for 4th of March (Monday) separated into 12 blocks starting at 0:00. Brightest yellow denotes 15 charging sessions happening at the given time block.

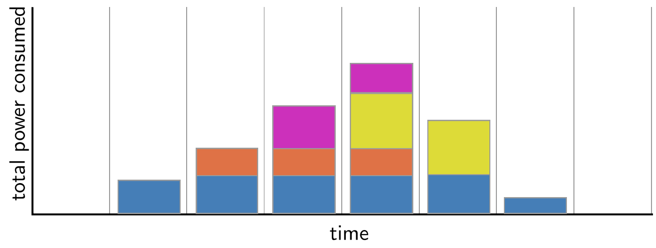


Figure 3.9: charging connector power consumption averaged over a temporal pattern $o \in O$, showing how consumption patterns emerge when aggregated across similar time periods.

occupies (this is necessary to correctly handle start and end hour chunks).

$$h_t^{s,c} = \sum_{i=1}^{|V^{s,c}|} \frac{\mu(T \cap [t^{c,s}_{\text{start}}; t^{c,s}_{\text{end}}])}{\mu([t^{c,s}_{\text{start}}; t^{c,s}_{\text{end}}])} * p_i^{c,s}, \forall t$$

where μ is a function to measure the length of an interval:

$$\mu((a;b)) \mapsto b - a$$

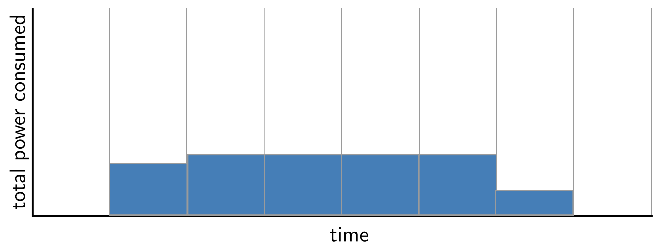


Figure 3.8: charging session cut into hourly chunks with assigned power consumption proportional to the fraction of the hour the session occupied.

3. Group the hourly power consumption into days to create daily hourly power consumption

$$H_d^{c,s} = \begin{bmatrix} h_{d_1}^{c,s} \\ \vdots \\ h_{d_{24}}^{c,s} \end{bmatrix}$$

where d is a day and d_i denotes the i-th hour range of day d

We are also interested in the aggregate behavior of charging points and stations. To analyze this, we compute averages over specific temporal patterns, such as days of the week or months of the year. For a temporal pattern $O \in \{\text{Monday}, \dots, \text{Sunday}\} \times \{\text{January}, \dots, \text{December}\}$, we calculate the average power consumption with temporal pattern.

Using either daily hourly power consumption or average power consumption with temporal pattern, we can derive two important metrics: 1. Total daily power consumption ($|H_d^{c,s}|_1$) - the sum of power consumed over a 24-hour period 2. Normalized daily power consumption ($\frac{H_d^{c,s}}{|H_d^{c,s}|_1}$) - the hourly distribution of power consumption as a proportion of the total

To summarize, the data derived from charging session that will be used further in this thesis include:

- **hourly power consumption** - The power consumption of a specific charging point during a specific hour. This is the most granular level of consumption data and serves as the foundation for all other derived metrics.

- **daily hourly power consumption** - A 24-element vector representing the hourly power consumption of a charging point over a specific day. This captures the daily charging pattern for individual days.
- **average power consumption with temporal pattern** - The average power consumption pattern for a charging point across a specific temporal pattern (e.g., all Mondays, all weekdays in January). This metric smooths out day-to-day variations to reveal consistent temporal patterns.
- **total daily power consumption** - The total power consumed by a charging point over a 24-hour period. This metric indicates the overall charging demand without considering its temporal distribution.
- **normalized daily power consumption** - The normalized distribution of power consumption across a 24-hour period. This metric captures the shape of the charging demand curve independent of its magnitude.

3.2 Basic settlement unit (ZSJ)

The spatial context of charging stations influences their usage patterns. As an attempt to capture this context we incorporate data from census based on Basic Settlement Units (ZSJ), which provide demographic and urban characteristic information at a fine-grained spatial resolution.

Basic Settlement Units (ZSJ) are territorial elements defined by the Czech Statistical Office for statistical and administrative purposes. They represent parts of municipalities with distinct spatial, technical, and urban planning characteristics or groupings of residential or recreational buildings. ZSJ units denote city districts, small villages, or settlements that would otherwise be joined to their belonging municipality [33].

[33]: (), *Základní sídelní jednotka*

Originally created as basic presentation units for census data, ZSJ units now serve as spatial reference units for various analyses. Currently, there are approximately 23,000 ZSJ units in Czechia, with 953 located within Prague [34]. The Czech Statistical Office collects and maintains this data, which we obtained for our research [35].

[34]: (), *Mapa základních sídelních jednotek*

[35]: (), *Český statistický úřad*

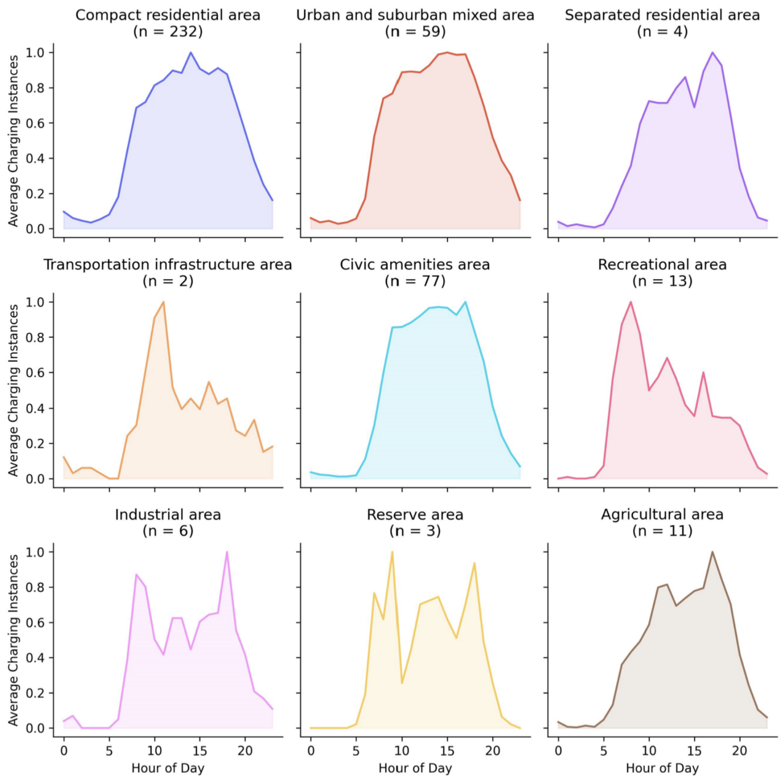


Figure 3.10: Average charger demand dependent on ZSJ type

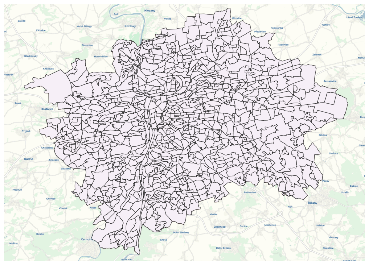


Figure 3.11: All basic settlement unit (základní sídelní jednotka) boundaries in Prague, showing the spatial segmentation used for demographic and urban characteristic analysis.

The ZSJ dataset provides contextual information that might explain variations in charging demand across different locations. By linking charging stations to their containing ZSJ units, we can incorporate demographic and urban characteristics into our predictive model. The effect of ZSJ type on the average power consumption of chargers can be seen in Figure 3.10.

3.2.1 Description

The ZSJ dataset contains the following data for each area:

- **Id**: A unique identifier assigned to each ZSJ unit. This code allows for unambiguous identification of each basic settlement unit within the national registry system.
- **Name**: The official name (název) of the basic settlement unit, representing the commonly used designation for that specific area or settlement.
- **Character**: Classification indicating the functional and urban character of the ZSJ, such as residential, industrial, mixed-use, or recreational area. See visualization in Figure 2.
- **Area**: The total surface area of the ZSJ in square meters (výměra), which can be derived from geometry data but is provided as a pre-calculated attribute for convenience.
- **Number of addresses**: Count of valid addresses (počet adres) within the ZSJ boundaries, indicating the density of addressable locations. See visualization in Figure 4.

- **Population**: Number of permanent residents (počet obyvatel) recorded within the ZSJ, typically based on census data or continuous population registry. See visualization in Figure 3.
- **Geometry**: The spatial representation of the ZSJ boundaries as a polygon in the S-JTSK coordinate system, enabling GIS analysis and visualization of the territorial unit.

Coordinate reference system - WGS84 and S-JTSK coordinate system

To be able to measure locations on earth as coordinates, a mathematical model of the earth is necessary. The most well known is WGS84 (EPSG:4326) [36] which is used by the Global Positioning System (GPS). This model assumes the earth is an ellipsoid and uses ellipsoidal coordinates to locate any point on the earth's surface. This ensures it can be used worldwide and is therefore useful for navigation. However, this approach causes issues such as continental drift, which would render the work of public offices like the Czech Geodetic and Cadastral Office more difficult due to the need to recalculate the position of objects of interest as they shift a few centimeters each year.

For this and historical purposes, the S-JTSK [37] regional coordinate system is still employed by many public Czech offices. This system can be used only in the region of Czechia and Slovakia. It is anchored to local monuments, thereby mitigating the issue of continental drift. It also provides a local Euclidean approximation, allowing for the calculation of distances between points using ordinary Euclidean distance, albeit with some loss of precision.

[36]: GmbH (https://www.klokantech.com/) (), *WGS 84 - WGS84 - World Geodetic System 1984, Used in GPS - EPSG*

[37]: (), *ČÚZK: Geoportál*

3.2.2 Transformations

The primary transformation performed on the ZSJ data is the conversion of absolute counts to density measures. This is achieved by dividing the quantitative field values (population, number of addresses) of each ZSJ by the area of its geometry polygon. This normalization allows for more meaningful comparisons between ZSJ units of different sizes and better reflects the intensity of human activity in each area.

3.3 People Mobility

We hypothesize that human mobility patterns could influence EV charging demand. As it may be more probable, that the charging demand may be due to inter-municipality peoples commute. Our model incorporates mobility data derived from mobile phone positioning information. Slice of the data can be seen at Figure 5.

3.3.1 Description

The mobility data was sourced from the Prague Institute of Planning and Development (IPR). This dataset uses anonymized mobile phone connections to cellular towers to track movement patterns.

The data is aggregated into origin-destination matrices for privacy preservation. A person's origin is defined as the location where the person (their mobile device) spent the night and morning hours. The destination is where the person spent the majority of daytime hours.

The dataset is structured as a matrix where rows represent origin areas (where people commute from) and columns represent destination areas (where people commute to).

3.3.2 Transformations

We used data from March 2022 for our analysis. The original dataset included 1,443 municipalities.

For each municipality within Prague, we computed two metrics: 1. The number of people commuting into the municipality from other areas within Prague 2. The number of people commuting into the municipality from outside Prague's boundaries

This transformation reduced the dimensionality of the data while preserving some of the information about commuting patterns that might influence charging demand.

It would be beneficial to also work with the distance from the municipalities. Because if people commute from longer distance with EVs their demand to charge could be larger.

3.4 Open Street Map

point of interest (POI) are points on a map with relevance to the domain of study. These can include buildings, shops, parking spots, landmarks, and other features that might influence charging behavior.

Previous research [15][16] has identified statistically significant relationships between POIs and charging demand using linear regression. In particular, [15] extracted POI data from Open Street Map (OSM), which we also utilize in our research.

Open Street Map [38] is a crowdsourced project aimed at constructing, maintaining, and openly providing map data. The data are available through map applications for end users for purposes like navigation, or can be exported for computational analytics.

[15]: Hecht et al. (2024), 'Global Electric Vehicle Charging Station Site Evaluation and Placement Based on Large-Scale Empirical Data from Germany'

[16]: Dong et al. (2019), 'Electric Vehicle Charging Point Placement Optimisation by Exploiting Spatial Statistics and Maximal Coverage Location Models'

[38]: OpenStreetMap contributors (2017), *Planet Dump Retrieved from https://planet.osm.org*

3.4.1 Description

OpenStreetMap data represents geographic features through a tagging system. The fundamental data model consists of three element types: nodes, ways, and relations. Each element can contain any number of tags in the form of key-value pairs.

For this research, we extract specific POI types relevant to charging behavior. The OSM data structure allows for detailed categorization through its tagging schema. Common tags include:

- **amenity**: Services and facilities (restaurants, parking, fuel stations)
- **shop**: Retail establishments (supermarket, mall, convenience)
- **leisure**: Recreational facilities (park, sports centre, fitness centre)
- **tourism**: Tourist attractions (hotel, museum, attraction)

- **building**: Building types (residential, commercial, industrial)
- **landuse**: Land usage patterns (retail, residential, industrial)
- **road**: Highways, walkways, bikelanes, and other transportation infrastructure

3.4.2 Transformations - Amenities

To extract a comprehensive set of points of interest, we used the software library OSMOX [39], which is capable of extracting data from OSM data. The extracted features correspond to the categories defined in the OSM wiki [3], ranging from public amenities and transportation hubs to accommodations, shops, and tourist attractions.

[39]: Peetz (2025), *MorbZ/OsmPoisPbf*

3: `https\protect\protect\leavevmode@ifvmode\kern+.2222em\relax//wiki.openstreetmap.org/wiki/Map_feature`

3.5 Spatial data transformations/feature engineering

Figure 3.12: Heatmap showing the density of Points of Interest in Prague, highlighting areas with high concentrations of amenities and services.

Spatial data types

- **Point** - A single location in space, represented by coordinates (x,y). In our research, charging stations and points of interest are represented as points.
- **Line/Multiline** - A set of connected points forming a path or multiple paths. Roads, rivers, and other linear features are represented in this format.
- **Polygon/Multipolygon** - A closed area defined by a boundary. ZSJ units, building footprints, and administrative boundaries are represented as polygons.

In this section, we describe the methods used to link spatial data to charging points. These transformations are essential for creating the feature vectors used in our predictive model, as detailed in Chapter **??**.

Point in polygon - Given a point and a set of mutually disjoint areas with features, this method assigns the features of the area containing the point to that point. For example, we assign ZSJ demographic characteristics to charging stations based on the ZSJ polygon in which they are located. This method may be unreliable when the point is near a boundary with other polygons/areas. A more precise solution would be spatial interpolation[4].

4: `https://r-spatial.org/book/12-Interpolation.html`

Nearest neighbors by radius - Given a point k and a set of points $\mathcal{P}$, this method returns the subset of $\mathcal{P}$ whose distance from k is less than a threshold K. It also stores the distance of each point to k. The actual distance function depends on the coordinate system used.

$$\text{Distance} : \mathcal{P} \times \mathcal{P} \longrightarrow \mathbb{R}^{+}$$

Following [15], we set K to 2000 meters. The choice of distance function depends on the coordinate system of the points.

[15]: Hecht et al. (2024), 'Global Electric Vehicle Charging Station Site Evaluation and Placement Based on Large-Scale Empirical Data from Germany'

Nearest neighbors with importance [15] - Similar to nearest neighbors by radius, but instead of using raw distance, this method

[15]: Hecht et al. (2024), 'Global Electric Vehicle Charging Station Site Evaluation and Placement Based on Large-Scale Empirical Data from Germany'

computes an importance factor that decreases linearly with distance:

$$Importance_K(a, b) = \frac{\max(K - Distance(a, b), 0)}{K}$$

This approach assigns higher weights to closer points of interest and zero weight to points beyond the threshold distance K. We use this method to incorporate the influence of nearby amenities and services on charging demand.

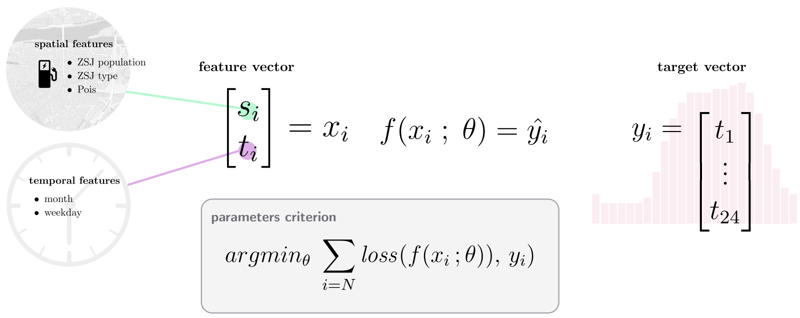
Normalization by area - Since polygons can have varying areas and some features are expressed in absolute numbers, normalizing by the area of the polygon provides density per unit area (usually per km^2). For a polygon $a \in A = \{a_1, \ldots, a_m\}$ and a function that assigns some feature $F \subseteq \mathbb{R}$ to the polygon: $feature : A \to F$, the density is computed as:

$$\text{Density}(a, f) = \frac{feature(a)}{Area(a)}$$

This transformation is particularly important for demographic features like population and address counts, as it allows for meaningful comparison between areas of different sizes.

These spatial data transformations form the foundation of our feature engineering process, enabling us to capture the complex relationships between charging demand and the surrounding urban environment. The resulting features are used as inputs to our predictive model, as described in Chapter **??**.

Research and implementation 4



In this chapter, we formulate the research problem of estimating electric vehicle charging demand and present our approach to solving it. We begin with a problem statement, followed by a detailed description of our feature engineering process. We then introduce our neural network architecture with latent profiles, which is designed to capture both temporal and spatial patterns in charging behavior. Finally, we explain our dataset splitting strategy, training procedure, and the baseline models used for comparative evaluation.

4.1 Problem statement

Our primary objective is to estimate the average power consumption with temporal pattern (APC) of electric vehicle charging stations based on a set of spatial and temporal features. Through this research, we aim to determine whether our data-driven approach is viable for predicting charging demand at new locations.

Formally, we define our model f_θ as a function that maps from a feature space to a target space:

$$f_\theta : X \to Y$$

Where $X \subset \mathbb{R}^M$ represents the set of all feature vectors (with M dimensions), and $Y \subset \mathbb{R}^{24}$ represents the set of all target vectors (average hourly power consumption over a day). The available data are split into *training*, *test*, and *validation* sets, with the training set used to minimize the empirical risk. The details of this splitting strategy are explained in Section 4.4.

The model function f has trainable parameters θ, which we optimize to minimize the following empirical risk:

$$loss_{total}(f, \theta) = \frac{1}{P} \sum_{(x,y)\in\mathcal{T}} \alpha \cdot a_x + \beta \cdot b_x$$

Title image provides a high level simplified overview of our problem

Where α and β are weighting coefficients for the two components of our loss function:

$$a_x = loss_{power}(\left\|f(x_i;\theta)\right\|_1, \left\|y_i\right\|_1)$$
$$b_x = loss_{norm}(\frac{f(x_i;\theta)}{\left\|f(x_i;\theta)\right\|_1}, \frac{y_i}{\left\|y_i\right\|_1})$$

The first component, a_x, measures the error in predicting the total daily power consumption (TDPC) using the $L1$ norm of the predicted and actual consumption vectors. The second component, b_x, measures the error in predicting the normalized daily power consumption (NDPC) by comparing the normalized predicted and actual consumption patterns. This dual-objective loss function aligns with our neural network architecture, which separately models the total power consumption and the normalized consumption pattern.

Mean square error (MSE) has been chose as loss functions for $\text{loss}_{\text{power}}$ and $\text{loss}_{\text{norm}}$. This loss function penalizes large difference in predictions more than smaller ones.

$$MSE(y, y') = (y - y')^2$$

4.2 Model features and feature engineering

Our feature vector consists of both **temporal** and **spatial** components. Ideally, features regarding the charger's capabilities (particularly maximum power output) would also be included, as these would significantly influence average consumption. However, due to limitations in our current dataset, these features are not available. This limitation motivates our interest in estimating the normalized daily power consumption (NDPC), which may be less dependent on the charger's maximum power output.

The features in our model fall into two categories based on data type: categorical and numerical. We process these feature types as follows:

Categorical features are transformed using one-hot encoding. This technique converts a categorical feature with n possible values into n binary features, where each binary feature corresponds to one possible value. For each observation, exactly one of these binary features has the value 1, indicating the presence of that categorical value, while all others are 0.

In our model, categorical features such as day of the week, month, and location characteristics are one-hot encoded before being fed into the neural network. This ensures that the model can effectively learn from these categorical variables without imposing arbitrary numerical relationships.

For example, if we have a categorical feature "day of week" with 7 possible values (Monday through Sunday), one-hot encoding transforms this into 7 binary features: "is_Monday", "is_Tuesday", etc. If an observation occurs on Wednesday, then the "is_Wednesday" feature would be 1, while all other day features would be 0. This transformation allows the model to properly handle categorical variables without imposing an arbitrary ordinal relationship between category values.

Numerical features are standardized by subtracting the mean and dividing by the standard deviation. This normalization ensures all features are on a comparable scale, preventing features with larger magnitudes from dominating the learning process and helping achieve faster convergence during training.

The feature vector x_i for each sample is constructed as follows:

$$x_i = \begin{bmatrix} s_i^T \\ t_i^T \end{bmatrix}$$

Where s_i represents the spatial features and t_i represents the temporal features.

The **spatial features** vector s_i is defined as:

$$s_i = \begin{bmatrix} s_i^1 \\ \vdots \\ s_i^R \end{bmatrix}$$

Table 4.1 provides a detailed overview of the spatial features used in our model.

Table 4.1: Overview of spatial features used in the feature vector. The additional processing is described in Section 3.5

Index	Name	Type	Value from	Additional processing
s_1	ZSJ population	numeric	charger in ZSJ polygon	normalization by the polygon area
$s_{2:10}$	ZSJ type	categorical	charger in ZSJ polygon	one-hot encoding
s_{11}	ZSJ number of addresses	numeric	charger in ZSJ polygon	normalization by the polygon area
s_{12}	Number of people commuting into the district from inside Prague	numeric	charger in the district polygon	normalization by the polygon area
s_{13}	Number of people commuting into the district from outside of Prague	numeric	charger in the district polygon	normalization by the polygon area
$s_{14:162}$	Points of Interest	numeric	number of PoIs by euclidean distance	importance calculation (value of single PoI is 1 if its distance from charger is 0, 0 if it is of distance 2km or further)

The **temporal features** vector t_i is defined as:

$$t_i = \begin{bmatrix} t_i^1 \\ \vdots \\ t_i^P \end{bmatrix}$$

Table 4.2 provides a detailed overview of the temporal features used in our model.

Table 4.2: Overview of temporal features used in the feature vector.

Index	Name	Type	Value from	Additional processing
$t_{1:7}$	day of the week	categorical	average power consumption with temporal pattern (APC)	one-hot encoding
$t_{8:19}$	month	categorical	average power consumption with temporal pattern (APC)	one-hot encoding

4.3 Architecture of the Latent Neural Network

Our machine learning problem formulation allows for various potential solutions, particularly within the class of neural networks. Before introducing our specific architecture, let's briefly review the fundamentals of neural networks.

Neural networks are computational models consisting of layers of interconnected nodes or "neurons" that process information. A typical neural network contains an input layer that receives data, one or more hidden layers that perform computations, and an output layer that produces the final result. Each connection between neurons has an associated weight that is adjusted during the training process. Information flows through the network via activation functions, which introduce non-linearity and allow the network to learn complex patterns. The training process involves feeding the network with labeled examples and using optimization algorithms, typically variants of gradient descent, to minimize a loss function by adjusting the weights. Backpropagation is the primary algorithm used to calculate gradients and update weights efficiently.

For our specific problem, we propose a neural network with latent profiles, as illustrated in Figure 4.1. The key innovation in our architecture is the construction of a latent profile matrix $R \in \mathbb{R}^{24 \times K}$, where K is the number of latent profiles. The network learns these profiles and predicts how they should be combined to generate the final consumption pattern for a given location and time period.

The network utilizes the following layers:

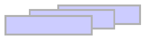
Trainable layers:

- **Fully connected (Linear transformation)**
  $\text{Linear}^n_m(x) = Wx + b$
  $\text{Linear}^n_m : \mathbb{R}^n \longrightarrow \mathbb{R}^m \ , \ W \in \mathbb{R}^{m\times n} \ , \ b \in \mathbb{R}^m$
  W, b are learnable parameters

  This is a standard linear transformation that maps input vectors to output vectors through a weight matrix and bias vector.

- **Latent vectors (Embedding)**
  $\text{LatentVec}_K = R$
  $\text{LatentVec}_K : \emptyset \longrightarrow \mathbb{R}^{24\times K}$
  R is learnable

  This layer represents our latent profiles matrix, which contains K different 24-hour consumption patterns that the network learns during training.

Non-parametric operations:

- **Softplus (Smooth activation)**
  $f(x) = \ln(1 + e^x)$
  $f : \mathbb{R} \longrightarrow \mathbb{R}^+$

  Softplus is a smooth approximation of the ReLU activation function. It ensures that the output is always positive, which is appropriate for our case since power consumption cannot be negative.

- **Normalization**
  $\text{Norm}(x) = \frac{x}{\|x\|_2}$
  $\text{Norm} : \mathbb{R}^n \longrightarrow \{y \in \mathbb{R}^n : \|y\|_2 = 1\}$
  input dimension matches output dimension

  This operation normalizes vectors to have unit L2 norm. We use it to normalize our latent profiles and to compute the normalized daily power consumption pattern.

- **Leaky-relu**
  $$\text{LReLU}(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha x & \text{if } x \leq 0 \end{cases}$$
  $\text{LReLU} : \mathbb{R} \longrightarrow \mathbb{R} \ , \ \alpha$ is a hyperparameter

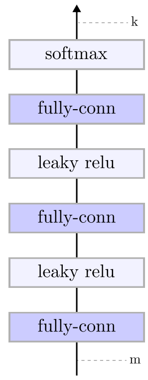
  Leaky ReLU is an extension of the standard ReLU activation function that allows a small gradient when the unit is not active. In this work, there is no clear motivation for its use over tanh or standard ReLU.

These layers are combined to form three functional modules, each with a specific purpose in our neural network architecture:

Network modules:

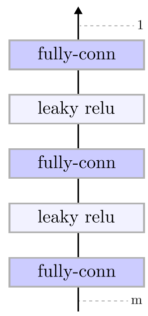
- **f module (Latent profile probabilities)**
  $f : \mathbb{R}^d \longrightarrow \mathbb{R}^K$
  $f = \text{Softmax} \circ \text{Linear}_K \circ \text{LeakyReLU} \circ \text{Linear}_{64} \circ \text{LeakyReLU} \circ \text{Linear}_h$
  Where d is feature size, h is hidden size, and K is latent profiles count
  Outputs normalized weights for latent profiles

  The purpose of this module is to predict the contribution of individual latent profiles to the resulting normalized consumption pattern. In other words, this module estimates the daily rhythm of the charger without considering the actual total power. The input is a feature vector, which is transformed by two linear layers with LeakyReLU activations. The output is a vector of K values transformed by softmax to ensure the sum equals 1, representing the mixing weights for the latent profiles.



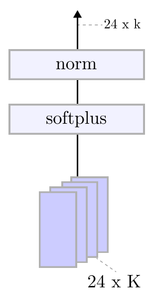
- **g module (Total power)**
  $g : \mathbb{R}^d \longrightarrow \mathbb{R}$
  $g = \text{Linear}_1 \circ \text{LeakyReLU} \circ \text{Linear}_{32} \circ \text{LeakyReLU} \circ \text{Linear}_{h_g}$
  Where d is feature size and h_g is hidden size for g module

  This module predicts the total power consumption. It consists of three linear layers joined with LeakyReLU activations. Its input is a feature vector, and it outputs a single scalar. This scalar is multiplied with the combined output of the h module to obtain the final prediction, effectively scaling the normalized consumption pattern to the appropriate magnitude.



- **h module (Latent profiles)**
  $h(\mathbf{x}) = f(\mathbf{x}) \cdot R^T$
  $h : \mathbb{R}^d \longrightarrow \mathbb{R}^{24}$
  Where $R \in \mathbb{R}^{24 \times K}$ is the normalized latent profiles matrix
  24 is the time granularity (hours in a day), K is latent profiles count

  The h module provides



The outputs of the h and f modules are combined as follows:

$$\text{Combine}(R, p) = \sum_{i=1}^{K} p_i \, R_i$$
$$\text{Combine} : \mathbb{R}^{24\times K} \times \mathbb{R}^{K} \longrightarrow \mathbb{R}^{24}$$

This combined output is then multiplied by the scalar from the g module to produce the final prediction of hourly power consumption.

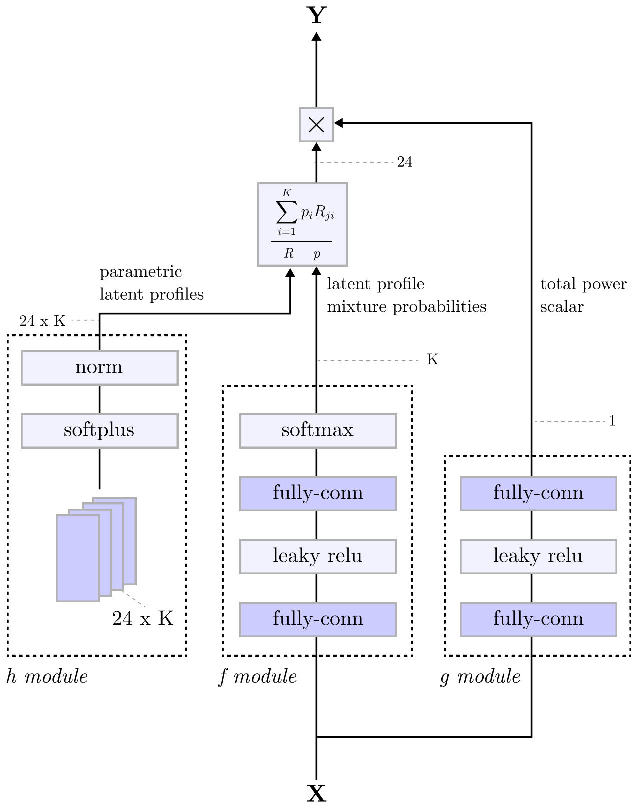


Figure 4.1: Latent neural network architecture. Light blue rectangles denote NN layers without trainable parameters, while blue denotes layers with parameters learned by SGD. Notation borrowed from Fleuret's book "Little Book of Deep Learning"

The model is implemented in Python using the PyTorch library [40].

[40]: Paszke et al. (2019), *PyTorch: An Imperative Style, High-Performance Deep Learning Library*

4.4 Dataset splitting

Our dataset splitting strategy is designed to evaluate the model's ability to generalize to new locations, which is crucial for our use case of predicting demand for new charging stations.

We split our data into three sets: training, test, and validation. However, we cannot simply randomly assign samples to these sets because of the hierarchical nature of our data. Specifically, one physical location may have multiple charging points (CPs) within a charging station (CS), and each CP may have multiple average power consumption (APC) measurements for various temporal patterns (TP). There might be high correlation between measurements from the same location, which would lead to data leakage if we split randomly.

To address this issue, we split the data based on location. Features for each unique location can only be present in one of the sets. This ensures that when we evaluate our model on the test or validation set, it is truly being tested on locations it has never seen during training.

The three datasets serve distinct purposes in our research:

- **Train dataset** (72% of the whole dataset) is used for training the model using stochastic gradient descent for empirical risk minimization.

$$\mathit{train}: \ \mathcal{T} = \{(x_i, y_i) \in X \times Y | \ i = 1, \ldots, P\}$$

- **Test dataset** (21% of the whole dataset) is used for inspecting the model's performance on unseen data and tuning hyperparameters. It provides an estimate of the true model risk.

$$\mathit{test}: \ \mathcal{S} = \{(x_i, y_i) \in X \times Y | \ i = 1, \ldots, R\}$$

- **Validation dataset** (7% of the whole dataset) is used to obtain the final model risk assessment on a model with already trained parameters and chosen hyperparameters.

$$\mathit{validation}: \ \mathcal{V} = \{(x_i, y_i) \in X \times Y | \ i = 1, \ldots, O\}$$

4.5 Training procedure

To train our neural network model, we utilize the standard PyTorch training procedure with mini-batch stochastic gradient descent (SGD) optimization. The model is trained iteratively over multiple epochs, with early stopping implemented to prevent overfitting.

4.6 Other models for quantitative comparison

To evaluate the effectiveness of our latent neural network approach, we compare it quantitatively with other machine learning models. The results from this comparison can either motivate further research with our model or indicate that simpler approaches may be sufficient.

One challenge in this comparison arises from our custom loss function, which combines two L1 losses. The baseline models are trained to minimize standard L1 loss, which might seem to create an unfair comparison. However, we provide more detailed analysis of this issue in Chapter 5.

The models we use for comparison are:

- **Average** - The simplest baseline, which takes the average over all feature vectors in the dataset and uses that value for all predictions. Both the train and test datasets $\mathcal{T}$ and $\mathcal{S}$ are utilized for this "model." It does not take into account the input features and uses a single value for all predictions.
  The implementation is straightforward:

$$f_{\mathcal{D}}(z) = \frac{1}{|\mathcal{D}|} \sum_{(x,\_) \in \mathcal{D}} x$$
$$f : \mathbb{R}^m \to \mathbb{R}^n$$
$$n, m \in \mathbb{N}, \mathcal{D} \in (X \times Y)^p, p \in \mathbb{N}$$

- **Linear regression** - A standard linear model that learns a linear mapping from input features to output predictions.
  We use the implementation from the Python library Scikit-Learn [41]:

$$f(z) = \alpha + \beta z$$
$$f : \mathbb{R}^m \to \mathbb{R}^n$$
$$\alpha \in \mathbb{R}^n, \beta \in \mathbb{R}^{n \times m}$$

[41]: Pedregosa et al. (2011), 'Scikit-Learn: Machine Learning in Python'

- **XGBoost** - A scalable end-to-end tree boosting system [42].

[42]: Chen et al. (2016), 'XGBoost: A Scalable Tree Boosting System'

The performance of these models is compared with our latent neural network based on mean absolute error and mean square error for three metrics: average power consumption with temporal pattern (hourly consumption), total daily power consumption (total daily consumption), and normalized daily power consumption (normalized daily consumption pattern).

Results 5

In this chapter, we present and analyze the results of our prediction model. First, a quantitative comparison with other models is provided using various error metrics to evaluate the model's performance. Next, we analyze the latent profiles learned by our model and examine how they are utilized in predictions. We then assess the model's ability to predict total power consumption. Finally, we summarize the key findings from our experiments and their implications.

5.1 Quantitative comparison with other models

Based on the defined loss functions in Chapter 4, we present the loss of the model. The losses represent error on prediction for an hour, calculated by dividing the loss by 24, instead of using the 24 values representing a full day.

The losses can be seen in Table 5.1.

Table 5.1: Table containing losses for several metrics. The last difference row provides percentage comparison between latent profiles NN model and train average.

Model name	MAE	MSE	MAE norm profile	MSE normal profile	MSE total power	MAE total power	MSE mixture loss
Latent profiles NN	3.6717	1661.9267	0.0021	0.0003	3.5574×10^5	64.0718	3.4302
Latent profiles NN (no data)	3.7540	1667.8263	0.0020	0.0003	3.5677×10^5	66.3788	3.4138
Train average	3.8964	1664.8744	0.0021	0.0003	3.5561×10^5	69.6754	3.4125
Test average model	3.8964	1664.8744	0.0021	0.0003	3.5561×10^5	69.6754	3.4125
Linear regression	4.2535	1823.4076	0.0060	0.1130	4.0771×10^5	79.0246	1130.0005
XGBoost	4.0460	2430.1903	0.0022	0.0004	3.9855×10^5	68.6162	4.1041
Difference %	+6.1204%	+0.1774%	−1.7618%	−0.5702%	−0.0377%	+8.7459%	−0.5150%

From inspection of the table, it can be seen that the model did not provide better results. We hypothesize that the model derived most of its performance from learning to predict the average. We conclude that the current features we used for predicting power consumption did not help the model perform better.

5.2 Latent profiles analysis

Although our model's prediction quality was not improved, the model learned to utilize the latent profiles. In this section, we examine how the model learned to utilize these profiles. To reiterate, we hypothesized that charging demand can be modeled as a mixture of K charging profiles, and meaning could be derived from both the learned profiles and the prediction of probabilities per location.

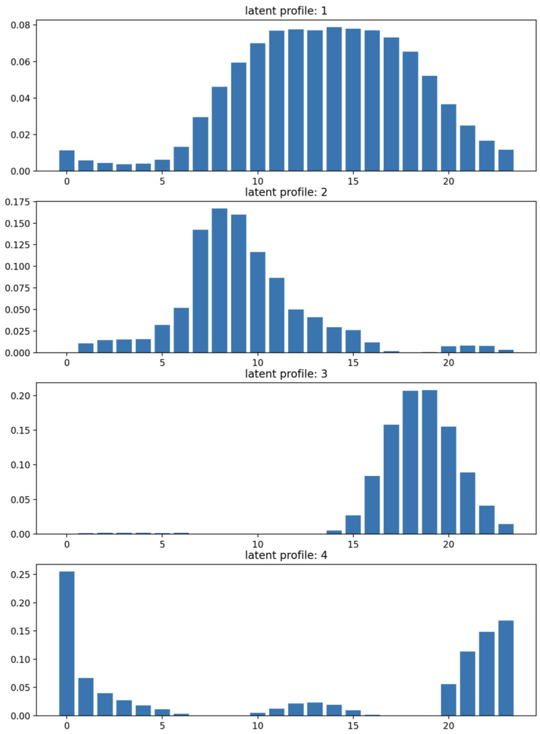


Figure 5.1: Latent profiles learned by our model

We found that setting $K = 4$ was the lowest possible number at which the train and validation loss could not be improved further.

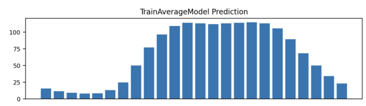


Figure 5.2: Train dataset $\mathcal{T}$ labels average

The learned profiles are visible in Figure 5.1.

The first profile closely corresponds to the average over the train dataset $\mathcal{T}$ labels (see Figure 5.2). While also resembling the average power consumption of the ZSJ unit of "compact residential area" (see Figure 3.10). A large part of the predictions for data from the test dataset utilize this profile (see Figure 5.4). We interpret this as our model falling back to simply giving the average of "compact residential area" chargers as its prediction[1].

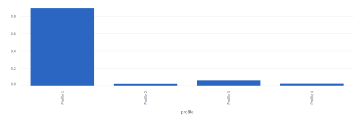


Figure 5.3: Average predicted latent profile probabilities by our model on the train dataset.

5.3 Power consumption

The second part of what our model predicts is the total power, of .This was already hypothesized in the beginning section, due to the absence of data about charger maximum power output. The comparison of true values versus predicted values can be seen in Figure 5.5. This may be due to the absence on the chargers

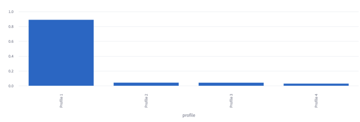


Figure 5.4: Average predicted latent profile probabilities by our model on the test dataset.

1: This could be mitigated by stratified sampling

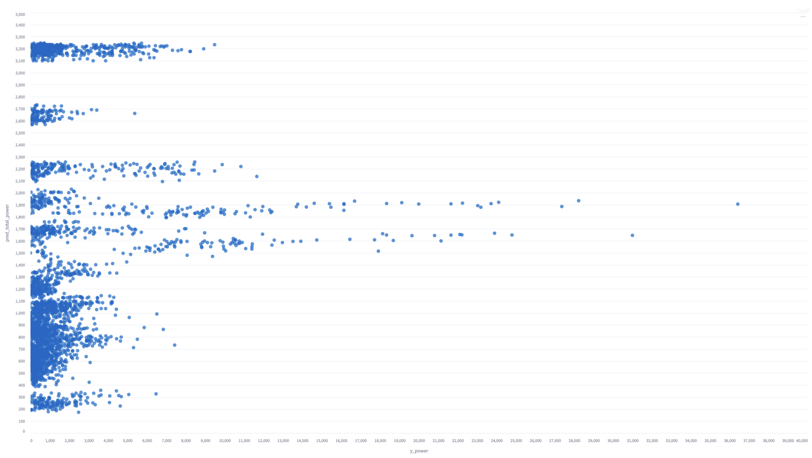


Figure 5.5: Comparison of true total power (x axis) and our prediction (y axis)

5.4 Conclusion

The takeaway from our findings is that the data used for the model were not sufficient and would require gathering a larger dataset of features. Potential additional data sources will be discussed in the conclusion. This could also be attributed to the fact that the model did not manage to fit the training dataset well.

Conclusion 6

In this thesis, we tried to understand the problem of charging demand. For this we have gathered data we tried to utilze existing data to predict the demand. We also tried for the model to have interpretable results.

6.1 Practical implications

Due to the low results the model could not be used in its current state to help planners with the charger placement problem.

We attribute the poor results mainly to the lack of data we provided to the model. This stems from the large variance in data quality and the difficulties we encountered when obtaining the data. And more naive use of spatial processing in regarding of processing the polygons.

6.2 Future steps

- focus on studying what factors affect the charging behaviour before building a complex model.
- address data bias
- incorporate more data (spatial). Like financial. Or identify from what kind of areas people are driving towards the charger. Increase data quantity and quality. This can be used for our model as well as for future models. And potentially more urban prague based models incorporating spatial features.
- data quality and centralized storage of the gathered spatial data. Like store in duckdb or postgress with postgis extension. Allowing for easier data manipulations and extraction of features. With stronger aspect on charger data.. So investing time into data pipelines.
- consider more simulated approach - large randomness due to stochastic nature of the data, maybe the idea of trying to predict the statistic is not a good one. A simulation framework, with model learning the connection if spatial and temporal features to the prob (That is try to predict some value of a random variable instead of deterministic feature) distribution of key charging aspects (start time, total lenght of stay, power consumed). With a stronger aspect on trying to identify if the connection even is there. From the learned prob distributions via monte carlo simmulation a charging demand could be obtained. With the addition of uncertainty. With the hope that the model could incorporate uncertainty for areas whose behaviour might simply be too stochastic. The stochastic model could deliver more fine grained data as well as provide predictions for the data our model tried to predict as well. This flexibility would lead to more potential use cases and could capture the inherent stochasticity. Also it is hard to compare the baseline right now due to tho model having to come up with the aggregate statistic for some temporal pattern that is still highly stochastic.

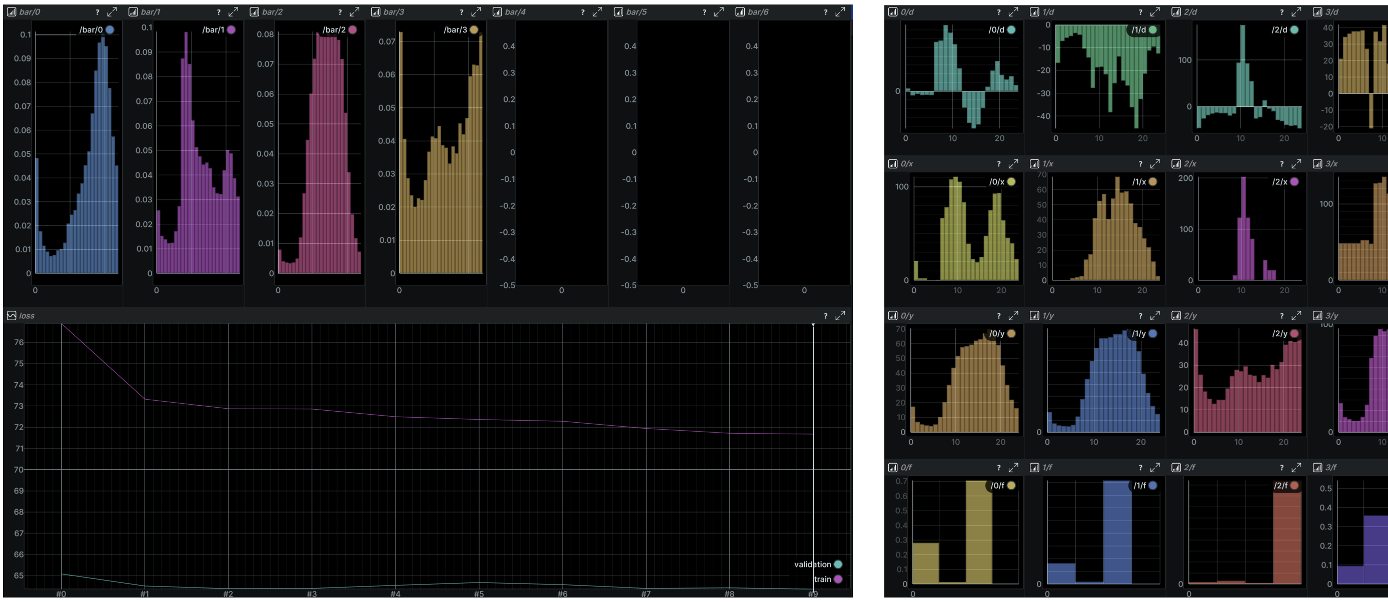
Appendix

Supporting software

0.1 Training dashboard

To offer visibility into the training process. A GUI dashboard using Python library Rerun [43]. The dashboard[43]: allows real time inspection of the training process. A plot of training and validation loss. And most importantly, presents the latent learned latent profiles. Also offers a way to see prediction results for random slice out of validation dataset.

(), *Rerun*



Rerun visualization dashboard.

Figure 1

1 Tool for visualisation of prediction results

To be able to inspect the prediction results in spatial context a tool was also developed with use of Streamlit[44][44]: The dashboard is in a form of a web application. It consists of two screens. First one allows inspection of overall model losses as visible in 5.1. In bar charts and tabular formats. Where the bar chart shows all the types of losses relatively to each other.

The second screen displays table of all charging location together with day of the week and month. The individual rows of the table are selectable. When selection haapens location of the charger in map is shown and prediction from the model together with other models used for comparison is computed and its results are shown in a bar chart graph. Together with losses agains the original label y value.

(2021), *Streamlit • A Faster Way to Build and Share Data Apps*

model	MAE	MAE normalized profile	MAE total power	MSE	MSE normalized profile	MSE total power
ChargingProfileModel	86.3404	0.0447	1553.9763	33729.441	0.0057	7748441.2767
LinearRegression	107.9335	0.0962	2181.7242	41590.8896	2.9124	10658337.9148
TrainAverageModel	87.9659	0.0435	1629.9461	33973.4273	0.0057	7853195.2634
ValidationAverage Model	87.9659	0.0435	1629.9461	33973.4273	0.0057	7853195.2634
XGBoost	92.4132	0.0456	1639.3703	42282.8242	0.0065	7993035.6818

Figure 2: Data dashboard website showing bar chart with relative losses of the models

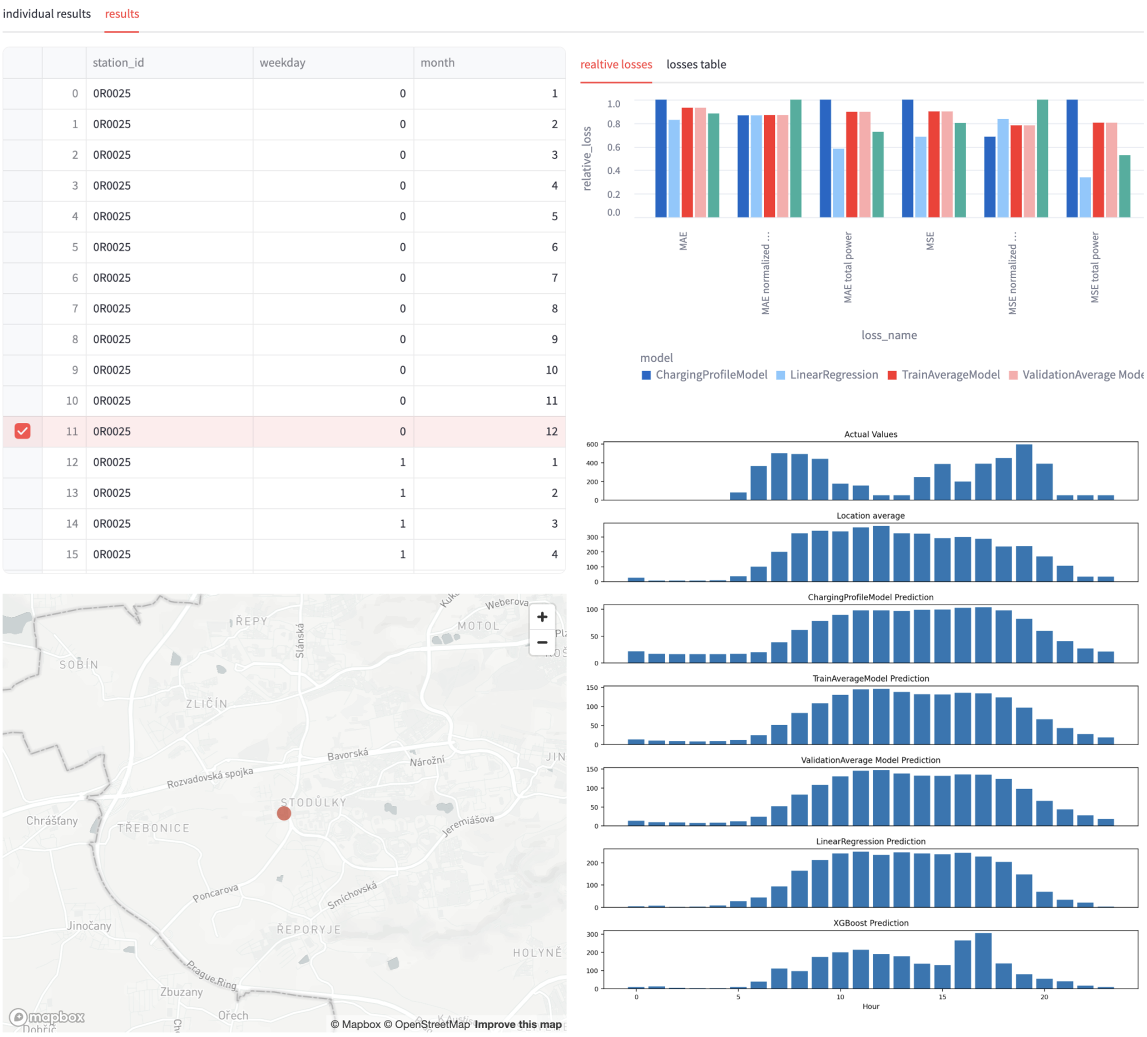


Figure 3: Data dashboard website showing table with data from validation data set with one selected row. Of which the prediction is computed from the main model and the other models for comparison.

Large figures

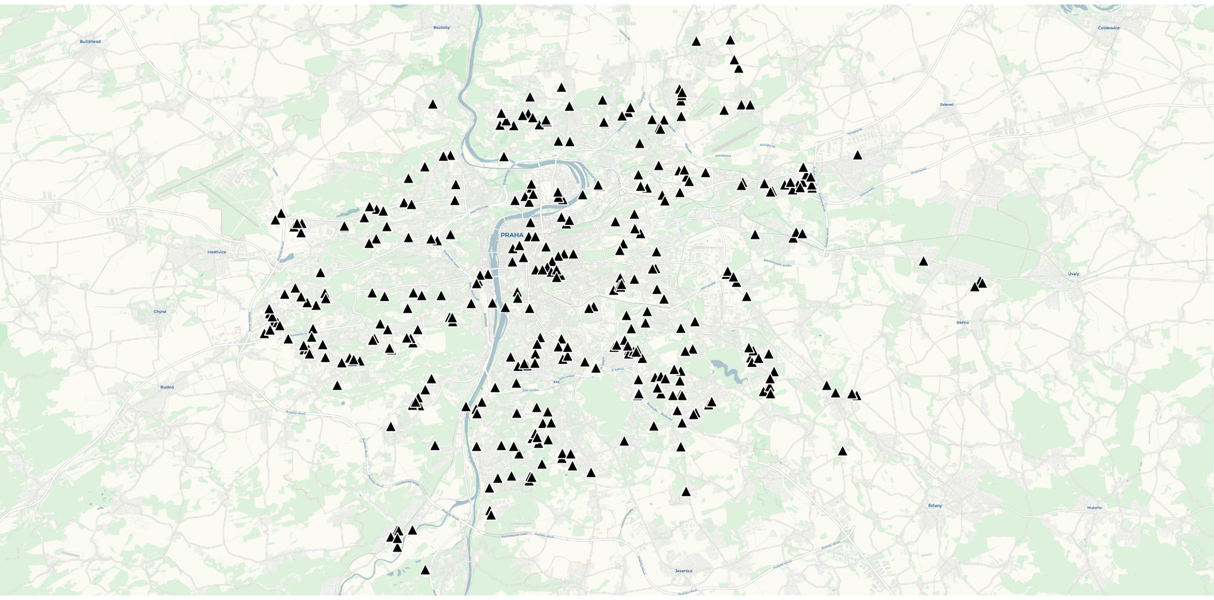


Figure 1

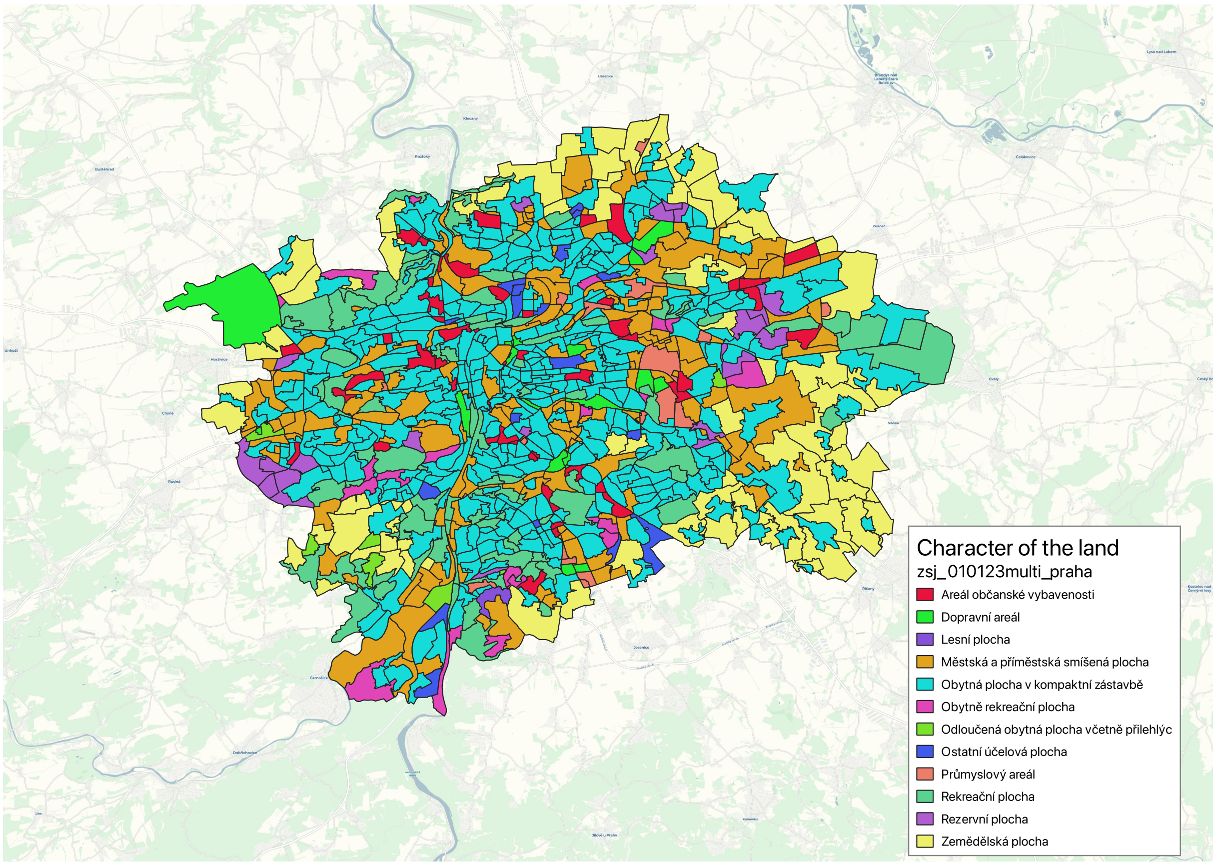


Figure 2

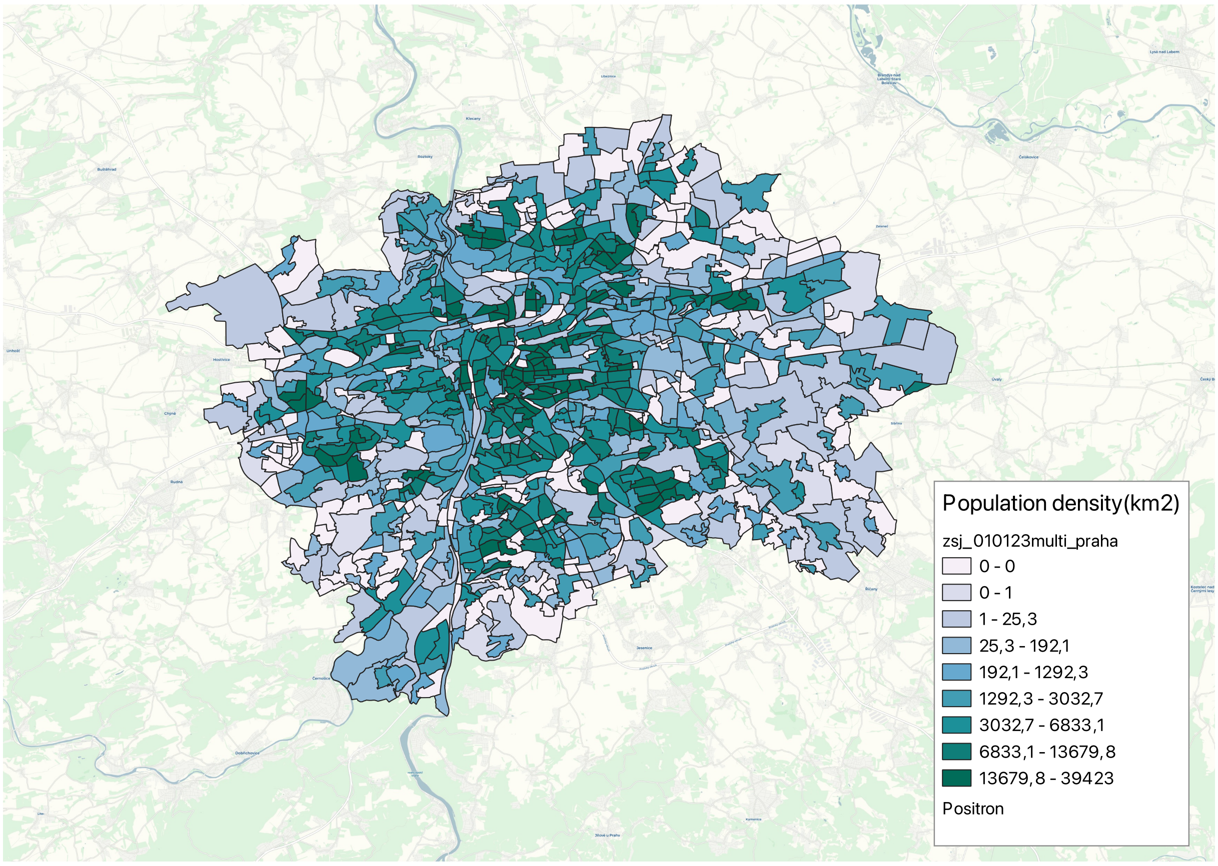

Population density(km2)
zsj_010123multi_praha
0 - 0
0 - 1
1 - 25,3
25,3 - 192,1
192,1 - 1292,3
1292,3 - 3032,7
3032,7 - 6833,1
6833,1 - 13679,8
13679,8 - 39423
Positron

Figure 3

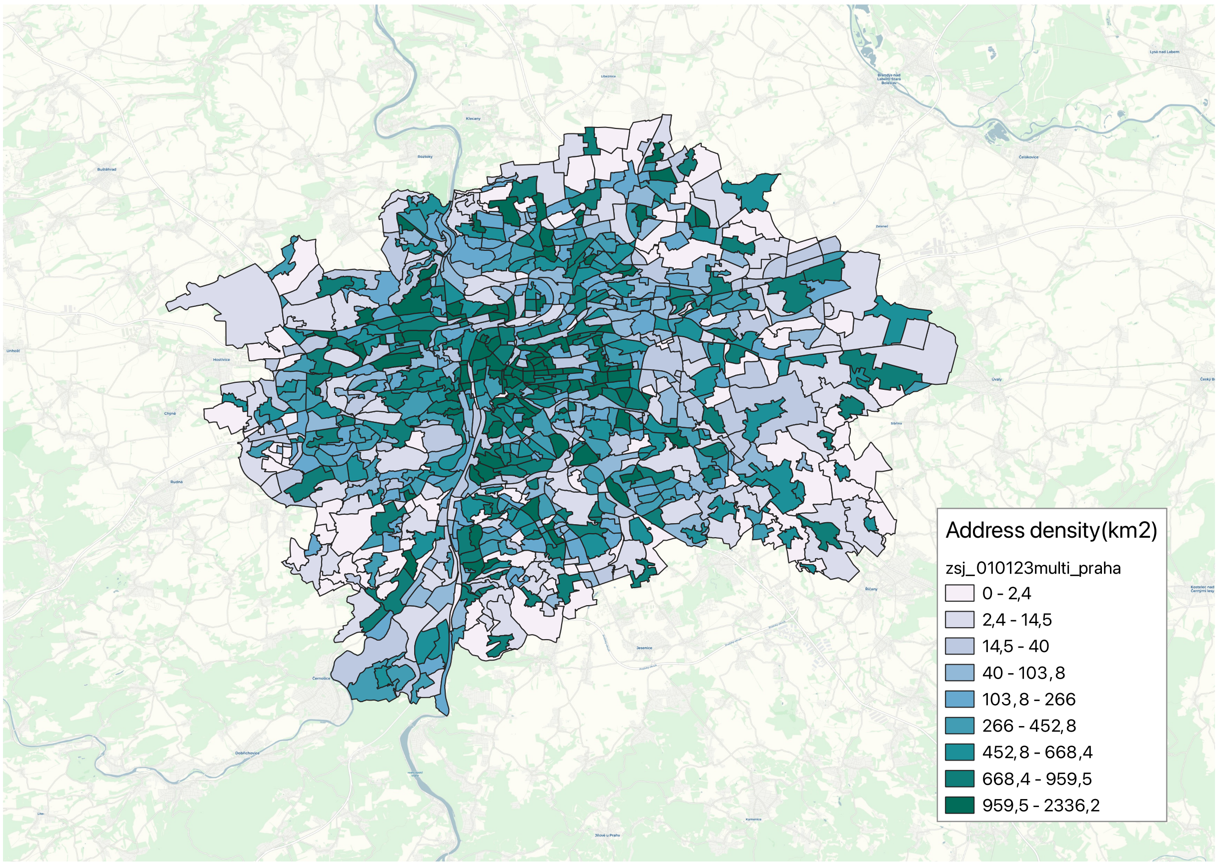


Figure 4

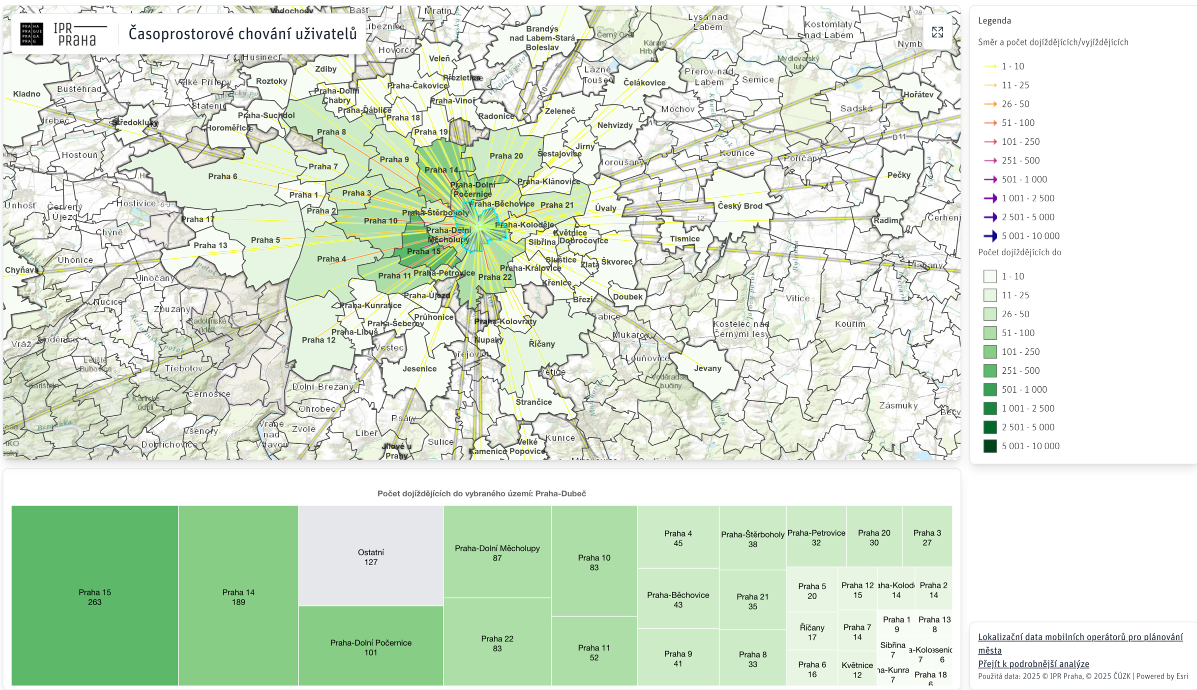


Figure 5

List of Figures

List of Tables

Bibliography

Here are the references in citation order.

[1] Roberta Quadrelli and Sierra Peterson. 'The Energy–Climate Challenge: Recent Trends in CO2 Emissions from Fuel Combustion'. In: *Energy Policy* 35.11 (Nov. 1, 2007), pp. 5938–5952. DOI: 10.1016/j.enpol.2007.07.001. (Visited on 05/16/2025) (cited on page 2).

[2] N. Künzli et al. 'Public-Health Impact of Outdoor and Traffic-Related Air Pollution: A European Assessment'. In: *The Lancet* 356.9232 (Sept. 2, 2000), pp. 795–801. DOI: 10.1016/S0140-6736(00)02653-2. (Visited on 05/16/2025) (cited on page 2).

[3] Nino Künzli et al. *Health Costs Due to Road Traffic-Related Air Pollution: An Impact Assessment Project of Austria, France and Switzerland: PM10 Population Exposure: Technical Report on Air Pollution*. July 1, 1999 (cited on page 2).

[4] *Contributions to EU-27 Emissions of BC, CO, NH_3, NMVOCs, NOx, Primary PM_10, Primary PM_2,5, SO_2 and CH_4 from the Main Source Sectors in 2020*. Nov. 24, 2022. URL: https://www.eea.europa.eu/en/analysis/maps-and-charts/contribution-to-eu-27-emissions (visited on 05/16/2025) (cited on page 2).

[5] Xiaoning Xia and Pengwei Li. 'A Review of the Life Cycle Assessment of Electric Vehicles: Considering the Influence of Batteries'. In: *Science of The Total Environment* 814 (Mar. 25, 2022), p. 152870. DOI: 10.1016/j.scitotenv.2021.152870. (Visited on 05/16/2025) (cited on page 3).

[6] E.H. Wakefield. *History of the Electric Automobile: Hybrid Electric Vehicles*. R.: Society of Automotive Engineers. Society of Automotive Engineers, 1998 (cited on page 3).

[7] *The Surprisingly Old Story Of London's First Ever Electric Taxi*. Science Museum Blog. July 9, 2012. URL: https://blog.sciencemuseum.org.uk/the-surprisingly-old-story-of-londons-first-ever-electric-taxi/ (visited on 05/16/2025) (cited on page 3).

[8] *The Electric Vehicle Future Was Promised Decades Ago, What Happened?* Museums Victoria. URL: https://museumsvictoria.com.au/article/the-promised-electric-vehicle-future/ (visited on 05/16/2025) (cited on page 3).

[9] International Energy Agency. *Global Electric Car Stock, 2013-2023*. Paris: IEA, 2024. URL: https://www.iea.org/data-and-statistics/charts/global-electric-car-stock-2013-2023 (cited on page 3).

[10] *ČEPS, a.s.* URL: https://www.ceps.cz/cs/uvod/novinka/ceps-vydava-maf-cz-2023-analyzuje-zdrojovou-primerenost-do-roku-2040 (visited on 05/16/2025) (cited on page 4).

[11] *Regulation (EU) 2019/631 of the European Parliament and of the Council of 17 April 2019 Setting CO2 Emission Performance Standards for New Passenger Cars and for New Light Commercial Vehicles, and Repealing Regulations (EC) No 443/2009 and (EU) No 510/2011 (Recast) (Text with EEA Relevance.)* Apr. 17, 2019. (Visited on 05/16/2025) (cited on page 4).

[12] *Časová osa – Zelená dohoda pro Evropu a balíček „Fit for 55"*. Consilium. URL: https://www.consilium.europa.eu/cs/policies/european-green-deal/timeline-european-green-deal-and-fit-for-55/ (visited on 05/16/2025) (cited on page 4).

[13] *Co je Fit for 55*. Fakta o klimatu. URL: https://faktaoklimatu.cz/infografiky/fit-for-55 (visited on 05/16/2025) (cited on page 4).

[14] *Regulation - 2023/1804 - EN - EUR-Lex*. URL: https://eur-lex.europa.eu/eli/reg/2023/1804/oj/eng (visited on 05/16/2025) (cited on page 5).

[15] Christopher Hecht et al. 'Global Electric Vehicle Charging Station Site Evaluation and Placement Based on Large-Scale Empirical Data from Germany'. In: *eTransportation* 22 (Dec. 1, 2024), p. 100358. DOI: 10.1016/j.etran.2024.100358. (Visited on 02/28/2025) (cited on pages 8, 13, 22, 23).

[16] Guanpeng Dong et al. 'Electric Vehicle Charging Point Placement Optimisation by Exploiting Spatial Statistics and Maximal Coverage Location Models'. In: *Transportation Research Part D: Transport and Environment* 67 (Feb. 2019), pp. 77–88. DOI: 10.1016/j.trd.2018.11.005. (Visited on 01/28/2024) (cited on pages 8, 22).

[17] Anu Maria. 'Introduction to Modeling and Simulation'. In: *Proceedings of the 29th Conference on Winter Simulation*. 1997, pp. 7–13 (cited on page 9).

[18] *Traffic Analysis Tools: Types of Traffic Analysis Tools - FHWA Operations*. URL: https://ops.fhwa.dot.gov/trafficanalysistools/type_tools.htm (visited on 04/13/2025) (cited on page 9).

[19] *Home | Austroads*. URL: https://austroads.gov.au/ (visited on 04/13/2025) (cited on page 9).

[20] Mirosław Lachowicz. 'Microscopic, Mesoscopic and Macroscopic Descriptions of Complex Systems'. In: *Probabilistic Engineering Mechanics*. Special Issue: Stochastic Methods in Mechanics — Status and Challenges 26.1 (Jan. 1, 2011), pp. 54–60. DOI: 10.1016/j.probengmech.2010.06.007. (Visited on 04/13/2025) (cited on page 9).

[21] Matsim Community. *MATSim*. MATSim. Mar. 31, 2025. URL: https://matsim.org/ (visited on 04/13/2025) (cited on page 9).

[22] Andreas Horni, Kai Nagel, and Kay Axhausen, eds. *Multi-Agent Transport Simulation Matsim*. London: Ubiquity Press, Aug. 2016, p. 618 (cited on page 10).

[23] Jan Drchal, Michal Čertický, and Michal Jakob. 'Data-Driven Activity Scheduler for Agent-Based Mobility Models'. In: *Transportation Research Part C: Emerging Technologies* 98 (Jan. 1, 2019), pp. 370–390. DOI: 10.1016/j.trc.2018.12.002. (Visited on 04/15/2025) (cited on page 10).

[24] T. Novosel et al. 'Agent Based Modelling and Energy Planning – Utilization of MATSim for Transport Energy Demand Modelling'. In: *Energy*. Sustainable Development of Energy, Water and Environment Systems 92 (Dec. 1, 2015), pp. 466–475. DOI: 10.1016/j.energy.2015.05.091. (Visited on 04/13/2025) (cited on pages 10, 12).

[25] John Brady and Margaret O'Mahony. 'Modelling Charging Profiles of Electric Vehicles Based on Real-World Electric Vehicle Charging Data'. In: *Sustainable Cities and Society* 26 (Oct. 2016), pp. 203–216. DOI: 10.1016/j.scs.2016.06.014. (Visited on 01/13/2024) (cited on pages 11, 12).

[26] Azhar Ul-Haq, Carlo Cecati, and Ehab El-Saadany. 'Probabilistic Modeling of Electric Vehicle Charging Pattern in a Residential Distribution Network'. In: *Electric Power Systems Research* 157 (Apr. 2018), pp. 126–133. DOI: 10.1016/j.epsr.2017.12.005. (Visited on 01/15/2024) (cited on pages 11, 12).

[27] Siobhan Powell et al. 'Charging Infrastructure Access and Operation to Reduce the Grid Impacts of Deep Electric Vehicle Adoption'. In: *Nature Energy* 7 (Sept. 2022), pp. 1–14. DOI: 10.1038/s41560-022-01105-7 (cited on page 11).

[28] Siobhan Powell, Gustavo Vianna Cezar, and Ram Rajagopal. 'Scalable Probabilistic Estimates of Electric Vehicle Charging given Observed Driver Behavior'. In: *Applied Energy* 309 (Mar. 2022), p. 118382. DOI: 10.1016/j.apenergy.2021.118382. (Visited on 12/29/2023) (cited on pages 11–13).

[29] Jan Pekárek. 'A Model of Charging Service Demand for the Czech Republic'. In: *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis* 65.5 (Oct. 2017), pp. 1741–1750. DOI: 10.11118/actaun201765051741. (Visited on 01/19/2024) (cited on pages 11, 12).

[30] Jan Neumann. 'Optimální Rozmístění Nabíjecích Bodů pro Firemní Flotily'. In: (June 2021). (Visited on 01/18/2024) (cited on page 12).

[31] Akram Elomiya et al. 'An Advanced Spatial Decision Model for Strategic Placement of Off-Site Hydrogen Refueling Stations in Urban Areas'. In: *eTransportation* 22 (Dec. 1, 2024), p. 100375. DOI: 10.1016/j.etran.2024.100375. (Visited on 02/09/2025) (cited on page 12).

[32] *Ontology Documentation Generated by WIDOCO*. URL: https://auroral.iot.linkeddata.es/def/electrical-vehicle/#https://auroral.iot.linkeddata.es/def/adapters#EnergyProperty (visited on 05/01/2025) (cited on page 15).

[33] *Základní sídelní jednotka*. RSO. URL: https://csu.gov.cz/rso/zsj_rso (visited on 05/04/2025) (cited on page 19).

[34] *Mapa základních sídelních jednotek*. Statistika a my. URL: https://statistikaamy.csu.gov.cz/mapa-zakladnich-sidelnich-jednotek (visited on 05/04/2025) (cited on page 19).

[35] *Český statistický úřad*. Statistika. URL: https://csu.gov.cz/domov (visited on 05/04/2025) (cited on page 19).

[36] Klokan Technologies GmbH (https://www.klokantech.com/). *WGS 84 - WGS84 - World Geodetic System 1984, Used in GPS - EPSG:4326*. URL: https://epsg.io (visited on 05/04/2025) (cited on page 21).

[37] *ČÚZK: Geoportál*. URL: https://geoportal.cuzk.cz/(S(wpj55enbhlagg2cfaowevz1x))/Default.aspx?mode=TextMeta&side=sit.trans&text=souradsystemy (visited on 05/04/2025) (cited on page 21).

[38] OpenStreetMap contributors. *Planet Dump Retrieved from https://planet.osm.org*. 2017. URL: https://www.openstreetmap.org (cited on page 22).

[39] Merten Peetz. *MorbZ/OsmPoisPbf*. Feb. 28, 2025. (Visited on 05/04/2025) (cited on page 23).

[40] Adam Paszke et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. 2019. URL: https://arxiv.org/abs/1912.01703 (cited on page 31).

[41] F. Pedregosa et al. 'Scikit-Learn: Machine Learning in Python'. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cited on page 33).

[42] Tianqi Chen and Carlos Guestrin. 'XGBoost: A Scalable Tree Boosting System'. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Kdd '16. ACM, Aug. 2016, pp. 785–794. DOI: 10.1145/2939672.2939785 (cited on page 33).

[43] *Rerun*. URL: https://rerun.io (visited on 05/11/2025) (cited on page 39).

[44] *Streamlit • A Faster Way to Build and Share Data Apps*. Jan. 14, 2021. URL: https://streamlit.io/ (visited on 04/17/2025) (cited on page 39).