Associate Professorship of Travel Behavior
TUM School of Engineering and Design
Technical University of Munich

Exploration of Traffic Area Segmentation on Aerial Imagery to Address the Parking Data Requirements of Travel Demand Models

A thesis presented in part fulfilment of the requirements of the Degree of Master of Science in Transportation Systems at the Department of Civil, Geo and Environmental Engineering, Technical University of Munich (TUM).

Author	Annie Thomas
Supervisor	Matthias Langer
	Associate professorship of Travel Behavior, TUM
External Supervisor	Jens Hellekes
	Remote Sensing Technology Institute, Deutsches Zentrum für Luft- und Raumfahrt (DLR)
Date of submission	22.09.2024

Abstract

Travel Demand Models (TDMs) are essential tools for predicting travel behaviour, helping decision-makers to evaluate transportation policies by simulating how people travel, where they go, and the modes and routes they choose. Although extensively studied, large-scale TDMs often lack a model for parking, or have a limited consideration of parking due to the absence of comprehensive parking data, despite its significant impact on travel behavior and traffic flow. This thesis addresses this gap by identifying the necessary data for integrating parking into TDMs to explore how aerial image datasets can fulfill these needs by serving as an additional input data source.

The methodology includes a thorough literature review and interviews with transport modelers to identify the data demand, along with discussions with data providers in major German speaking cities to assess data supply, with the goal of identifying key parking data gaps. Using Berlin as a case study, the study attempts to extract the identified parking data from a traffic area segmentation dataset based on aerial imagery, developed by DLR, through geospatial analyses. The results reveal that modelers require more detailed parking data, including parking location, type, capacity, cost, occupancy, search time, and egress distance. Among these requirements, the aerial image dataset proved particularly useful for providing detailed information on parking location, capacity, and type of parking based on access. Through geospatial analysis, the study estimates 1.3 million parking spaces in Berlin, 55% of which are on-street and 45% off-street, marking the first statewide estimate of off-street parking capacity. Ultimately, this study contributes to making remote sensing data more accessible to a broader community, demonstrating its potential benefits for applications in the transportation domain.

Acknowledgements

This research was conducted as part of my master’s thesis for the Master of Science in Transportation Systems at the Technical University of Munich (TUM), in collaboration with the Remote Sensing Technology Institute (IMF) of the German Aerospace Centre (DLR).

Firstly, I would like to express my heartfelt gratitude to my supervisor, Matthias Langer, for his valuable guidance, patience, and for constantly keeping me on track to stay focussed on the objectives. I am deeply indebted to Jens Hellekes, my external supervisor at DLR, for helping me think outside the box and for co-shaping the direction of this thesis. Jens, your mentorship has made this journey very rewarding and also an enjoyable experience. I truly enjoyed our brainstorming sessions, and without your support, I wouldn't have accomplished half of what I have done. I would also like to sincerely extend my thanks to Dr. Ana Tsui Moreno Chou for her insights which helped refine this study, and my colleagues at DLR, Corentin Henry and Felix Michael Rauch, whose timely assistance ensured that I was never stuck when facing challenges with the dataset. I am also thankful to all the experts I interviewed for this thesis, whose knowledge and perspectives added significant depth to this work.

Life, much like this thesis journey, is full of ups and downs and I am grateful to my friends and family for making the journey meaningful. Special appreciation to my buddy and classmate, Najath Basheer, for standing by me through thick and thin, revolving in the same orbit for the past few years, sharing all the same challenges and situations. To Anu Sree Krishna, thank you for your constant support and for proofreading my thesis. Finally, I am grateful to my family, especially my brother, Akash Thomas, for encouraging me to take on bigger challenges and inspiring me to step out of my comfort zone to pursue my master’s in Germany. Thank you for being by my side through all my achievements and I couldn’t have done it without you.

[This page is intentionally left blank]

Master's Thesis Outline

of Annie Thomas

Date of Issue: 2024-03-22

Date of Submission: 2024-09-22

<u>Topic:</u> **Exploration of traffic area segmentation on aerial imagery to address the parking data requirements of travel demand models.**

Parking is an important aspect of urban infrastructure when it comes to city planning and traffic management. The integration of parking in transport models is essential as the choice of parking has been proven to influence activity scheduling decisions of different activities. However, recent studies have shown that travel demand models (TDMs), especially large-scale ones, often lack a model for parking or have a limited consideration of parking. The absence of comprehensive parking data is a significant hurdle to precisely model the parking needs of a city. To obtain this parking data, several approaches have been adopted ranging from manual surveys to mobile mapping of parking areas. People have utilized remote sensing data to map parking spaces, but none of them could link this data to the travel demand models. Remote sensing imageries are becoming more and more available to the public and the algorithms for processing aerial images are getting advanced.

This thesis explores the potential of traffic area segmentation on aerial imagery to serve as an additional input data source for TDMs focusing on the modeling of parking. Traffic area segmentation in this context refers to the automatic partitioning of captured aerial images into distinct segments of traffic area classes such as roads or parking areas. To do so, TIAS (Traffic Infrastructure and Surrounding), a novel aerial image dataset with traffic area annotations developed by DLR, is used. To assess the potential of this dataset, first, identification of the knowledge gaps regarding parking data within the current state of practice from a modeling perspective is necessary. The data requirements for the incorporation of parking applications in TDMs are identified through an in-depth literature review, expert surveys, and interviews. Once the data requirements are known, the already processed TIAS dataset is analyzed and evaluated through geospatial analysis to address possible requirements. To extract this information, various data reduction techniques are used to assess their ability to complement existing data sources. With the processed datasets provided by DLR in the form of simplified shapefiles, the identified parking requirements such as the capacity or type of parking which are presently unknown can be calculated by simple QGIS operations. For example, parking capacity can be inferred from the total area of parking, and the type of parking can be detected by observing the neighboring traffic area classes in the dataset. This data extraction serves as a first step to make the parking data implementation-ready for relevant models in the future.

The research questions arising from the discussion above follow as:

What are the data requirements necessary for integrating parking into travel demand models or to improve current practices?

Which requirements from the modeling side are inadequately addressed currently, and to what extent can this be addressed by aerial image datasets?

The goal of this thesis is to identify the potential of aerial image datasets to address the data needs of travel demand models focusing on the modeling of parking. My contribution to the scientific community will be identifying the parking data needs of relevant models to explore the applicability of aerial imagery to fulfill these needs. By the conclusion of this thesis, this knowledge will be applied to an unexplored dataset to assess its feasibility to serve as an additional input data source for TDMs. This knowledge can be further applied to other aerial image datasets and also to areas that are structurally similar. The thesis also highlights the fact that some existing datasets lack spatial coverage, a limitation that can be addressed by leveraging the advantages offered by remote sensing. The thesis concludes by outlining the limitations imposed by aerial image datasets and proposes recommendations for future work.

The student will present intermediate results to the mentors Matthias Langer (Research Associate, Professorship of Travel Behavior, Technical University of Munich) and Jens Hellekes (external supervisor from DLR) in the fifth, tenth, 15th, and 20th weeks.

The student must hold a 20-minute presentation with a subsequent discussion at most two months after the submission of the thesis. The presentation will be considered in the final grade in cases where the thesis cannot be evaluated.

Prof. Dr.-Ing. Rolf Moeckel | Student's signature

Table of Contents

List of Figures

List of Tables

List of Abbreviations

API	Application Programming Interface
IoU	Intersection over Union
MATSim	Multi-Agent Transport Simulation
MiD	Mobilität in Deutschland
OSM	OpenStreetMap
PSR	Park Search Route
PST	Park Search Traffic
PyPML	Python Parking Monitoring Library
QGIS	Quantum Geographic Information System
SP	Stated Preference
SPS	Smart Parking Systems
SUMO	Simulation of Urban MObility
TAPAS	Travel and Activity Patterns Simulation
TAZ	Traffic Analysis Zone
TDM	Travel Demand Model
TIAS	Traffic Infrastructure and Surrounding
TraCI	Traffic Control Interface

[This page is intentionally left blank]

1. Introduction

1.1. Research motivation

Parking is an essential aspect of car travel, influencing not only individual travel behavior but also overall traffic conditions, making it a subject of extensive research. On average, vehicles remain parked for 90% of its total time they are in the city, leading to highly inefficient use of urban land allocated for parking (RAC Foundation, 2011). Studies have also shown that, on average, around 30% of city center traffic can result from drivers cruising for parking (Shoup, 1997). Additionally, parking availability have proven to significantly influence people's activity scheduling decisions, affecting how people plan and carry out their daily routines. Given the substantial impact of parking on travel behavior and traffic flow, integrating parking in travel demand models is essential for more accurate representations of urban mobility. Travel demand models are tools used to predict travel behavior and patterns, helping decision-makers evaluate transportation policies and projects by simulating how people travel, where they go, and the modes and routes they choose based on various factors. Therefore, integrating parking aspects into these models help authorities make informed decisions about parking policies, such as raising parking fees to discourage car use and thereby reduce parking demand.

However, recent studies have shown that travel demand models (TDMs), especially large-scale ones, often lack a model for parking or have a limited consideration of parking (Waraich & Axhausen, 2012). This limitation is primarily due to the absence of comprehensive parking data, as pointed out by numerous studies (Christiansen et al., 2017; Gu et al., 2021; Schiller, 2004). Obtaining comprehensive parking data poses a significant hurdle, as current data sources often lack sufficient quality, consistency, spatial resolution, and completeness, such as missing information on private parking. To understand and address these current limitations, it is first necessary to understand the data used in parking models and identify which data requirements are insufficiently addressed. This thesis aims to understand the data required by modelers to effectively integrate parking aspects into those models. Once these requirements are identified, this thesis explores the potential of remote sensing as a solution to fill these gaps. Remote sensing, especially high-resolution aerial imagery, is a relatively unexplored data source for transport applications. It offers distinct advantages over usual data collection methods in terms of spatial resolution and coverage, including the ability to observe private areas that are otherwise inaccessible. As a case study, this thesis evaluates a novel aerial image dataset developed by DLR to assess its capability to fulfill the requirements that are insufficiently addressed from the parking modeling side.

1.2. Objectives and scopes

As mentioned in the previous section, the absence of comprehensive parking data is a significant hurdle to precisely model the parking needs of a city. The main objective of this thesis is to understand how parking is currently modelled in TDMs, with a focus on understanding the demand and supply of parking data to explore the potential of remote sensing, particularly aerial image datasets to serve as an additional input data source for TDMs. Based on these objectives, the following research questions are formulated to guide the study:

1) What are the data requirements necessary for integrating parking into travel demand models or to improve current practices?

2) Which requirements from the modeling side are inadequately addressed currently, and to what extent can this be addressed by aerial image datasets?

The scope of this thesis is focused on supporting the integration of parking data into TDMs and this also extends to two of the most evolved agent-based parking models explained in the following chapter. While understanding the current state of implementation-ready parking data, the research is confined to major cities in Germany along with Zurich, as Zurich has well established parking data. Although various user groups may require this data, the scope of the 'demand' in this research is specifically limited to the modeling community. Additionally, the study explores the potential of remote sensing to address the data gaps, with a specific focus on the process of traffic area segmentation in aerial imagery. Traffic area segmentation involves the automatic partitioning of aerial images into distinct traffic area classes, such as roads and parking areas.

1.3. Research approach and expected contributions

The thesis is structured into two parts: identifying the demand and supply, thereby identifying gaps in parking data utilized in TDMs, and extracting the required data from an aerial image dataset to address these gaps. The first part, focused on identifying demand, involves an in-depth literature review and expert interviews with modelers. To specify the exact data requirements, the parking data is categorized based on the type and level of detail required by modelers. Later, to assess the supply of data, a web-based inquiry followed by interviews with representatives from data-providing companies and municipal authorities in major German cities is conducted. This investigation is done to gain insights into the availability of currently used parking data. By matching the demand and supply, the data gaps are identified. Subsequently, geospatial analysis is performed on a novel aerial image dataset to evaluate the extent to which they can fulfil these gaps.

The main contribution to the scientific community will be identifying the parking data needs of relevant models and assessing whether the available supply data can meet these demands. This contribution is compounded by bridging the gap between the fields of transportation and remote sensing. Several approaches have been adopted to extract parking information, ranging from manual surveys to mobile mapping of parking areas. People have utilized remote sensing data to map parking spaces, but none could link this data to the travel demand models. As remote sensing imagery becomes more accessible and aerial image processing algorithms continue to advance, this research contributes by exploring how these technologies can be utilized to improve the performance of TDMs in modeling parking. Moreover, if the required parking features can be successfully extracted for one city and the data extraction framework is well defined, this approach can be easily transferred to other structurally similar cities with aerial images, unlike other methods that require creating new datasets and custom extraction processes due to varying data collection techniques. Ultimately, this study contributes to making remote sensing data more accessible to a broader community by demonstrating its potential benefits for applications in another domain.

1.4. Structure of the thesis

The thesis tackles the research questions across the following chapters, structured as shown in figure 1. The following chapter is literature review, which explores various aspects of parking, including its influence on travel behaviour, current state of practice in modelling parking and parking data collection methods. Following this, chapter 3 describes the methodology for identifying parking data demand and supply. Details of how the necessary data is gathered through literature reviews, expert interviews, and inquiries with data providers is also presented. The results of this analysis are then discussed in chapter 4, highlighting key findings related to parking data demand and supply, as well as identified gaps. Next, chapter 5 provides information on the potential data that can be used to fill the identified gaps. Data description of the aerial imagery dataset utilized in the study is also included. The subsequent chapter 5 outlines the methodology of extracting parking data from the aerial image dataset, through geospatial analysis. The results of data extraction are then presented in chapter 6. Finally, the thesis concludes by synthesizing the findings, addressing the research questions, and discussing the contributions of the study, along with describing the limitations of the work and finally, recommendations for future research and application.

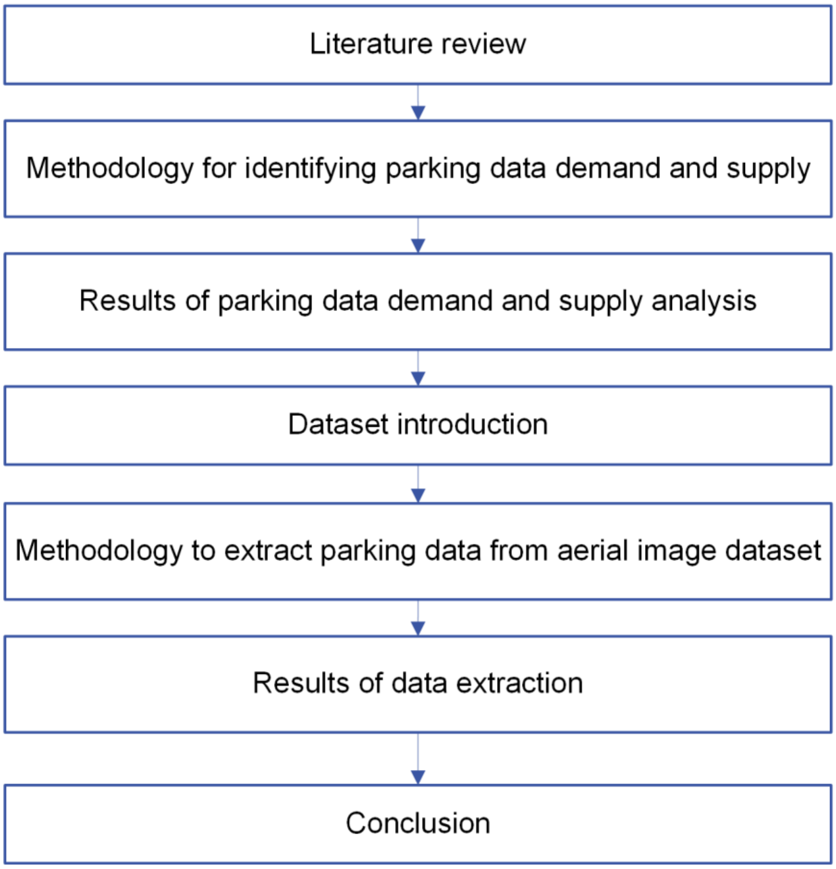


Figure 1. Overview of thesis structure

2. Literature review

This section reviews various aspects of parking from a modelling perspective, beginning with its impact on travel behaviour. Later the state-of-the-art in modelling parking within macroscopic and microscopic demand models, and explicit parking models are discussed. Additionally, this review discusses the changing face of parking data collection methods and the demand for such data in those models to identify the insufficiently addressed parking data in existing literature.

2.1. The impact of parking availability on traffic flow and travel behavior

Parking is arguably the most crucial element of car travel as every car trip starts and ends with walking to and from a parked vehicle and every single car needs to be parked somewhere when not in use (Tchervenkov, 2022, p. 21). However, while parking is essential, it also contributes to significant externalities. Recent reports on external costs in the transport industry have shown that EU-wide carbon dioxide emissions from road transport have risen by 21% since 1990, with private motorized road transport (cars) contributing the largest share (European Commission, Directorate-General for Mobility and Transport, 2022). Additionally, studies conducted around the world suggest that, on average, around 30% of city centres traffic could be due to parking search traffic (Shoup, 1997). Despite these externalities, the convenience of car for traveling from one place to another makes it the most popular mode, compelling cities to provide minimum parking requirements as part of urban planning regulations.

Many studies (Young, 1990;Young, 1990 Axhausen & Polak, 1991; Guo, 2013; Christiansen et al., 2017; Guo, 2013; Tchervenkov, 2022) over the years have shown that parking availability plays a significant role in shaping the mobility behavior and commuting mode choice. For instance, in activities where individuals have flexibility in choosing their destination (e.g., shopping), the availability of parking affects their decision to use a car as their mode of travel, with destinations offering parking being preferred. People are even willing to shift their departure time or activity itself, based on the parking supply constraints. Thus proper understanding and management of parking within a city can play a major role in developing more efficient transport policies aimed at managing traffic and reducing car use and related externalities toward a more sustainable future (Tchervenkov, 2022). (Tchervenkov, 2022)This is supported by several authors including Shoup, 1997) who emphasizes that parking pricing or parking restrictions are generally an effective travel management tool. Similarly, Waraich, 2016) mentioned that traffic-related policies can be designed around parking such as raising prices for parking at certain times of the day, or reducing parking supply in an area to impact travel demand. However, an important question remains, pointed out by Tchervenkov, 2022 - to what extent do the characteristics of parking supply shape mobility behaviour and how can we better integrate these insights into current transport simulation frameworks? Given the significant role that parking availability plays in influencing people's behaviour, it is necessary to further investigate how parking demand and supply are integrated within transport models.

2.2. State of practice: modeling parking in transport models

Transport models are essential for guiding authorities in developing informed policies and decisions by simulating possible traffic scenarios. However, travel demand simulation models, especially large scale ones, often lack a model for parking (Waraich & Axhausen, 2012, p. 4). To address this gap, several transport models have been developed over the years to explicitly integrate parking, aiming to capture the complex spatial and temporal interactions between travellers on real-world transportation networks (Tchervenkov, 2022). The following subsections discuss how parking is currently modelled in various transport models, including large-scale travel demand models where efforts have been made to incorporate parking into existing frameworks to understand its influence on overall traffic simulations. Additionally, explicit parking models, developed over the years to capture parking dynamics in high spatial and temporal detail, are also discussed.

2.2.1. Modeling parking within macroscopic transport models

While macroscopic transport models are recognised as valuable tools for strategic transport planning, parking is rarely explicitly considered in such models (Lubrich, 2023, p. 1). Given that parking aspects significantly influence commuting behaviour and traffic patterns, it is increasingly important to integrate parking elements into conventional transport model-setup. Recent research on parking in macroscopic transport models has focused on incorporating parking search and parking choice behaviour into the assignment step of the four-step modelling approach. Parking choice is the decision process of selecting a parking space from a given set of parking spaces located close to the agent’s destination (Waraich & Axhausen, 2012, p. 4). Gu et al., 2021 suggested an approach to include parking in macroscopic models through a macro-micro approach. The microscopic model simulated parking search behaviour including cruising for parking, considering both on-street and off-street parking with limited capacity. This model is then extended to a macroscopic framework, which allows to capture the integrated road-parking system dynamics to model parking and the proposed approach was applied to online parking pricing optimization. However, (Gu et al., 2021) have reported lack of parking data to effectively model parking in these models and to address this, Lubrich, 2023 proposed an extended macroscopic transport model, built on top of the software VISUM (PTV AG, 2020). Parking data was incorporated into the model by accessing data via an Application Programming Interface (API) by INRIX as an exogeneous data source to describe the parking infrastructure as part of the network model (Lubrich, 2023). The aspects of parking supply and demand were represented within the model setting, whereas parking choice is calculated via an optimisation of a park-search route (PSR). This model allows for explicit simulation of the parking search process, including elements of driving - looking for and choosing a parking spot, as well as walking to the activity location. The conceptual approach is applied to an existing macroscopic transport model for Cologne, Germany, and the results improved the spatial representation of parking patterns, allowing for detailed analysis of park search traffic (PST). Other approaches for analyzing parking within conventional transport models includes the work of Schiller, 2004 and Bagloee, Asadi, & Richardson, 2012. The common approach of all these models is to refine the model elements of network supply and travel demand to enhance the modelling techniques to consider the parking component within the traffic system as pointed out by Lubrich, 2023. The above-mentioned studies represent some of the recent research efforts to incorporate parking within macroscopic transport models. Further research is needed in this area to better understand how parking dynamics can be integrated into these models to open up more potential applications.

2.2.2. Modeling parking within microscopic transport models

Over the past few decades, agent-based transportation models have been developed in contrast to four-step models, notably to address the need to model complex interactions between individuals (Tchervenkov, 2022, p. 23). To capture such complex interactions associated with parking, various microscopic models have attempted to incorporate parking into their frameworks. This section of the study focuses on the most prominent and widely recognized microscopic models that have incorporated parking within their framework.

As mentioned by Rodríguez et al., 2022, there is a degree of heterogeneity among parking models, as these models were designed for different purposes. A common methodology for all these models stems from the approach introduced by Thompson & Richardson, 1998, who developed a parking search model that accounted for drivers' experience-based knowledge. The whole parking search process was divided into stages and utility function is assigned to factors such as travel time and cost. Subsequently, other models emerged, each focusing on specific aspects, such as influence of economic parameters on parking behavior, the parking search process, or the egress distance (the walking distance from the parking spot to the destination). Between 2008 and 2012, several explicit parking models emerged, each with different areas of focus, which are discussed in the upcoming section.

Later, Waraich & Axhausen, 2012 proposed a simple parking choice model implemented into the agent-based model MATSim that includes the simulation of parking choice and parking occupancy for the city of Zurich. MATSim (Horni, Nagel, & Axhausen, 2016) is a microscopic, agent-based, extendable simulation framework that populate the region of interest with synthetic agents that execute daily travel plans on a transportation network. Based on the results of previous iterations, agents can modify their plans to increase their scoring based on certain criteria, until an equilibrium is reached. The MATSim framework has already been utilized for a wide range of transportation applications, including parking. The simulation, by default, does not consider parking infrastructure or supply constraints (Waraich, 2016, p. 89). It means that the default model assumes infinite parking capacity such that an agent that arrives at their activity destination can directly park in the nearest link. This elimination of an important parking attribute might lead to the inaccurate estimation of parking phenomenon such as parking search time and egress distance, resulting in an unrealistic representation of travel behavior. The parking search model built on top MATSim is developed to address this modeling gap. In the proposed parking choice model, the agent can choose a parking space out of the four choices distinguished by parking regulations - public, private, preferred and reserved parking. Even though the replanning of the agent’s activities takes into consideration of parking occupancy, walking distance and, price preferences into account, the model omits parking search process, which means it cannot account for the resulting traffic and congestion. This basic model has applications in traffic policy design, performance-based parking pricing, etc. Waraich, Dobler, & Axhausen, 2012 later proposed an algorithm which was able to evaluate parking strategies to avoid overestimation of parking search time by describing how individual valuation of parking search components (e.g. search time, cost and walk time) affect the parking choice of agents.

Later, Bischoff & Nagel, 2017 introduced a parking search logic into the MATSim using within-day re-planning methodology and applied the model to Berlin. Compared to a standard simulation without parking search, results suggest that the parking search times sums up to 20% of the overall traffic in residential area (Bischoff & Nagel, 2017). Tchervenkov, 2022 builds upon these advancements of parking in MATSim through empirical and simulation studies by integrating large-scale survey and GPS tracking data in Zurich. Tchervenkov uses the approach of Bischoff & Nagel, 2017 to propose an improved agent-based transport simulation framework capable of capturing the influence of parking location, availability and price differentiation, applied to Zurich. His results showed that on-street parking is observed to be preferred than parking garages with egress distance having the most significant impact on parking choice in a garage, followed by parking search time and parking costs. Rybczak et al., 2024, took it to a further step by introducing a methodology to model parking cost and residential parking zones into a transport model base case in MATSim for the region of Leipzig in Germany. The results of Rybczak et al., 2024 suggested that parking pricing has a larger impact on the share of motorized transport than longer access and egress walks to the activity and raising parking costs could prove to be an effective tool to reduce motorized transport. However, like many other parking models developed within the MATSim framework, this model also couldn’t account for parking scarcity. The model made a compromise by assuming that the parking capacity is unlimited outside the residential parking zones so that agents can simply park on the nearest destination link (Rybczak et al., 2024). The agent-based parking models built on top of MATSim have the advantage of combining it with the traffic simulation to show it’s impacts on the traffic and congestion changes. However, they lack the detailing of simulation as well as the temporal and spatial resolution offered by the explicit parking models described in section 2.2.3.

Some of the other microscopic models where attempts have been made to incorporate parking into their simulation framework, includes SUMO, AIMSUN, VISSIM and TAPAS. SUMO, developed by DLR is an open-source general-purpose microscopic traffic simulator that allows to simulate how a given traffic demand which consists of single vehicles moves through a given road network (“SUMO documentation,” 2018). Erdmann et al., 2018 introduced a parking management framework within SUMO that focuses on large scale parking management optimizations. This framework utilizes general-purpose Python Parking Monitoring Library (PyPML), a tool integrated within the SUMO framework that uses Traffic Control Interface (TraCI) to gather and aggregate the parking monitoring information. Rodríguez et al., 2022 introduced a model called DYNAPARK, built on top of Aimsun simulation, using python programming language. The objective of the model applied to a central zone of the city of Santander, Spain, is to simulate dynamic pricing policies and their influence on parking demand and mobility behavior. The model integrated on-street parking into a traditional microsimulation model, and the simulation results showed that the impact of on-street parking must be considered when implementing a traffic simulation model, as it led to an approximately 50% increase in traffic density. Similarly, Weeraddana, Edussooriya, & Abeysooriya, 2020 used PTV VISSIM, a microsimulation software, to simulate the impact of roadside parking on traffic flow characteristics, focusing on how parking manoeuvres reduce lane capacity and affect vehicle speeds under Sri Lankan heterogeneous traffic conditions. Their study found that each 100 parking manoeuvres per hour resulted in a 7% reduction in lane capacity, highlighting the significant influence of on-street parking on traffic congestion in urban areas. The microscopic activity-based travel demand model TAPAS, developed by DLR is currently making efforts in integrating parking aspects in modelling. Mar´ıa L´opez D´ıaz, 2020 describes a simulation scenario that includes the implementation of a fee-based parking zone in the city centre of Berlin using TAPAS. The aforementioned studies provide only a brief overview of the most notable research on this topic, as there are numerous other studies that also focus on this area.

2.2.3. Modeling parking within explicit parking models

With the advancements in the computational power, spatially explicit models with high temporal resolutions have been developed with a focus on parking, to improve the performance and scope of agent-based models. These so-called agent-based parking models can potentially simulate in detail the change by a policy measure both from the agent's perspective (e.g. change in search time, walk distance, cost) or from the overall systems perspective (e.g. change in travel mode, parking revenue or traffic counts) (Waraich et al., 2012). In this section, some of such most evolved agent-based parking models emerged in recent years are discussed.

Benenson, Martens, & Birfir, 2008 developed PARKAGENT, a spatially explicit, agent-based model as an ArcGIS extension designed to model the parking choice of agents based on the strategies derived from the behavioural surveys for Basel and Tel Aviv. The model is applied to Tel Avis's residential neighbourhoods with a severe parking shortage, where they analysed the impact of additional parking spaces on parking behaviour. The agent enters the simulation with a knowledge of the parking capacity as it drives towards the destination, and if the agent cannot find a parking in 10 minutes, it parks in the nearest paid off-street parking. On the contrary SUSTAPARK by Dieussaert et al., 2009 includes a detailed traffic model for cities, applied in Leuven, Belgium that simulates the influence of park search into the overall traffic of a city. The model is based on cellular automata and the agents follows rules defined by surveys on parking behavior. In the simulation, the agents search for a parking space close to the destination and the search is based on a disutility function that increases with increase in search time. Similar to PARKAGENT, the agents park car off-street as the disutility surpasses their threshold value. While other explicit parking models, such as SimPark (Vuurstaek et al., 2018), PARKSIM (Young & Thompson, 1987), PARKANALYST (Levy, Martens, & Benenson, 2013), do exist, PARKAGENT and SUSTAPARK are the most frequently mentioned models by researchers.

Bischoff & Nagel, 2017 noted that while SUSTAPARK could model the behavioural changes in terms of parking, other choice dimensions such as departure times or mode choice are not considered. Waraich et al., 2012 points out that one limitation of the two models is their exclusion of the potential influence that parking supply shortages can have on travel time, mode choice, or destination selection. This exclusion limits these model's capability of providing feedback to the traffic simulation, thereby making them incapable of simulating parking policies. Waraich et al., 2012 also highlights that agent based parking model proposed by Waraich & Axhausen, 2012 (mentioned in section 2.2.2) has advantages over these explicit models as they are built on the top of an existing agent-based traffic simulation (MATSim). This integration allows the simulation to reflect the impact of parking searches on overall traffic flow. While the traffic simulation in MATSim offers far more features than SUSTAPARK, the modelling on the roads is less detailed compared to both PARKAGENT and SUSTAPARK (Waraich et al., 2012).

2.3. State of practice: parking data collection methods

The following section discusses the trends in parking data utilized in transport models as well as the demand and current practices related to this data.

2.3.1. Parking data utilized in transport models

Over the years, the parking data utilized in transport models has undergone significant transformations. Bonsall, 1991 discusses different methods of collecting data on parking. In his paper, he also explained the role of advancing technologies in addressing the parking data gaps in modelling. Axhausen & Polak, 1991 used stated preferences survey data to model the response of traveller's choice of parking type and location to show their impacts on parking policy measures. Bonsall & Palmer, 2004 used simulated data to model parking behaviour to show that the purpose of a trip significantly influences various aspects of parking choice. As public data on parking spaces became increasingly available, researchers began to combine different datasets, for example manual surveys, GPS data and building cadastre which contains type of land uses or their buildings and location. The technological advancements in navigation devices and smartphones which can collect comparatively precise location data made it possible to collect better parking data, especially parking search patterns. Tchervenkov, 2022 integrated large-scale survey and GPS tracking data in Zurich for the modelling of parking to capture both spatial and temporal variations. Lubrich, 2023 incorporated parking data based on Smart Parking Systems (SPS) accessed via API as an exogeneous data source improving the accuracy and temporal resolution of parking models in macroscopic transport models.

In addition to the data currently used in parking models, researchers have identified other data sources with potential for application in transport models. Recent advancements have continued to exploit a variety of data including remote sensing data. Henry et al., 2021 demonstrated the potential of deep learning techniques combined with OpenStreetMap (OSM) data to accurately estimate parking spaces using aerial imagery. Similarly, Hellekes et al., 2023 discussed the potential of traffic area segmentation on remote sensing data to create parking constraints in travel demand models. The studies described here provide only a brief overview of the parking data sources utilized in various modelling approaches but highlights the most recent developments in this field. As technologies continues to advance, the methods and sources of parking data will likely evolve further.

2.3.2. Parking data demand in transport models

As discussed in the beginning, studies have shown that transport models have long struggled with accurately representing parking due to a significant lack of reliable parking data. A central challenge for empirical studies of parking and car use, is access to a sufficient amount of valid and reliable data (Christiansen et al., 2017, p. 199). This challenge is widely recognized in the literatures. Gu et al., 2021 and Schiller, 2004, argue that despite the importance of integrating parking elements into wider transport models, integrated models have not found way into every day's practice of transport models, due to, among others, lacking data sources about parking. Erdmann et al., 2018 confirms this by stating that large scale parking management studies are more complicated to achieve primarily due to the lack of reliable aggregated data on parking. Nurul Habib, Morency, & Trépanier, 2012 also emphasizes that data scarcity is the primary barrier in integrating parking behavior in activity-based travel demand modeling. Guo, 2013 notes that parking data are typically difficult and costly to acquire and even when acquired, they are seldom made publicly available. As a result, simulated data naturally become an alternative, provided that the simulations are able to mimic the underlying behaviours of complex parking aspects such as parking search.

In response to these data limitations, some researchers used assumptions for the missing data, which could potentially misrepresent travel behavior. For instance, Waraich & Axhausen, 2012, in their agent-based parking choice model for the city of Zurich, assumed that the parking supply outside of the city was unlimited as the supply data outside the city was only sparsely available. For the same reasons, the pricing models for paid parking were simplified, applying a flat rate to all garage parking, which does not reflect the complexities of parking cost dynamics. The model also struggled to accurately model private areas, as data on private off-street parking was lacking. A similar approach was adopted by Bischoff & Nagel, 2017 in their efforts to integrate parking search within MATSim. Their simulation required data on the number of parking spots on each link, but for links without this information, a direct on-street parking spot was assumed. The challenges in modelling private off-street parking due to data limitations have also been pointed out by Benenson et al., 2008 for their model PARKAGENT and Dieussaert et al., 2009 for the model SUSTAPARK. For the latter, when no data was available for the number of private parkings, estimates were made based on car ownership rates and certain assumptions. Lubrich, 2023, who advanced the use of parking data by utilizing API data, highlighted that API data only provide parameters on public parking facilities, thus leaving private facilities out of scope.

In addition to mentioning the general lack of parking data, several researchers have pointed out specific gaps in the parking data type required for their applications. To specify the exact data requirements, it is necessary to categorize the parking data. Based on the type of parking data required by modelers, this thesis categorizes the parking data into seven types: parking location, type of parking (based on access), parking capacity, parking cost, parking occupancy, parking search time, and egress distance. While some other parameters are mentioned in the literature, they often can be related to these seven attributes, as some can be inferred from others. For instance, parking duration can be derived from parking occupancy data, and egress time can be inferred from egress distance. Table 1.1 provides a summary of the specific parking data demands identified by various modelers, differentiated by the models and authors. Not all models discussed in the previous section are included, as some only mention a general lack of parking data without specifying the exact type of data required by them. Additionally, for some models, while they have some data for a particular parking attribute, the data may be incomplete, which is also reflected in the table. For example, VISUM has some data on parking capacity, but it lacks data on private areas and hence listed as a required data (X) in the table. The specific data required for each model, corresponding to each parking attribute, is listed in Appendix A.

(Table is presented in the next page)

Table 1.Parking data requirements for transport models from literature review

Model	**Author**	**Parking location**	**Type of parking**	**Parking capacity**	**Parking cost**	**Parking occu-pancy**	**Parking search time**	**Egress distance**
VISUM	Lubrich, 2023			X		X		
	Gu et al., 2021						X	
MATSim	Waraich & Ax-hausen, 2012		X	X	X	X		X
	Bischoff & Nagel, 2017			X				
	Bischoff et al., 2019			X		X		X
	Rybczak et al., 2024		X		X			
SUMO	Erdmann et al., 2018			X				
SUS-TAPARK	Dieussaert et al., 2009			X				

Although the literature highlights the need for specific parking data, the exact level of detail required is not clearly stated. To gain these insights which are not stated in the literature, interviews with modelers were conducted. The following sections of this thesis will explore the specific data needs of modellers as well as the level of detail of data required, to understand the insufficiently addressed parking data in current transport simulation frameworks.

3. Methodology for identifying parking data demand and supply

The methodology outlined in this section focuses on identifying both the demand for parking data from the modelling perspective and the supply of parking data available in major German-speaking regions. To achieve this, interviews with modelers and data providers, as well as investigation into various sources, are the approaches adopted. However, before designing interview questions and structuring these discussions, it is essential to first establish a framework for classifying parking data and detailing the specific types and levels of data used by modellers and provided by data suppliers. By creating a clear structure, the specific types of data and their required granularity can be also pinpointed while trying to extract the data. This framework serves as a foundation for understanding the classifications and granularity of parking data, ensuring that both the demand and supply are comprehensively identified and addressed.

3.1. Framework for classifying parking data

As discussed in section 2.3.2, seven key attributes of parking were identified through a thorough literature review. These attributes are the most frequently mentioned in studies focusing on parking modelling, either as the type of input data required or as the output data generated by the models. These attributes will be referred to as 'dimensions' in this study. The seven dimensions are: parking location, type of parking, parking capacity, parking cost, parking occupancy, parking search time, and egress distance. While other parking data types, such as parking duration or egress time, exist, they can often be inferred by referencing one or more of these core dimensions or integrated as levels within them. Each of these dimensions is further subdivided into several categories, referred to as 'levels', based on the granularity of the data. The levels were identified by reviewing publicly available information, including scientific literature and model's software manuals, as well as carefully considering and defining additional levels based on potential parking scenarios. The following section outlines the definitions of each dimension and their specific levels.

1) Parking location:

Parking Location refers to the geographic area where individual parking slots or large parking lots are situated (note that a parking slot refers to a single space designated for one vehicle, while a parking lot refers to an area or facility that contains multiple parking slots). Four levels are identified for the dimension parking location, differentiated based on the spatial granularity. In city level, parking data is aggregated across the entire city and in zonal level parking data is aggregated based on TAZ, parking zones or zones self-defined by a specific spatial boundary. In link level, the parking areas are aggregated to roads and point level, the most granular level, provides exact coordinates or specific location of individual spaces or facilities.

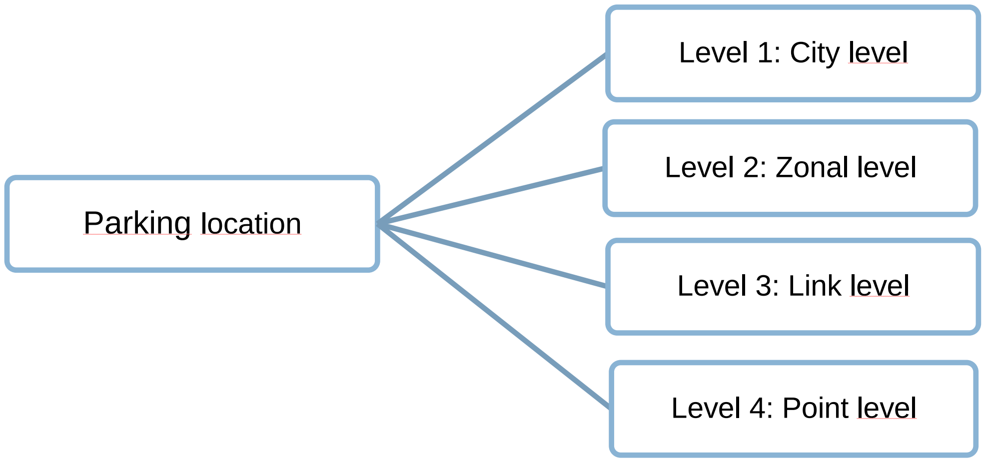


Figure 2. Levels of the dimension parking location

2) Type of parking:

Type of parking here refers to the classification of parking spaces based on access. Parking types are mainly classified here into:

a) On-street Parking

b) Off-street Parking

    i. Private garage

    ii. Private lots

    iii. Public garage

    iv. Public lots

This dimension is crucial because distinguishing between on-street, off-street, public, and private parking helps determine user preferences on parking types based on access, which in turn supports decision-making regarding the appropriate balance of on-street and off-street parking in a city. The associated levels are as follows:

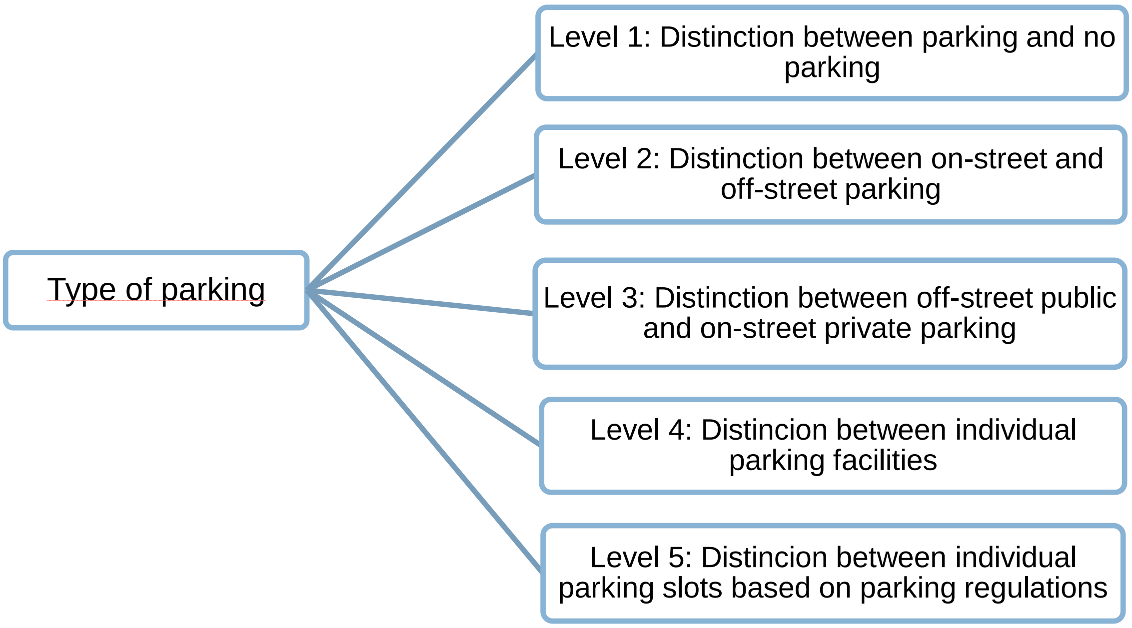


Figure 3. Levels of the dimension type of parking

3) Parking capacity:

Parking capacity refers to the total number of parking spaces provided in a specific area. It is important to distinguish parking capacity from parking availability: while capacity measures the overall infrastructure available for parking, availability refers to the number of unoccupied or vacant parking spaces at a given time. The levels of the dimension parking capacity are shown in figure 4.

4) Parking cost:

Parking cost refers to the fee associated with using parking facilities, which can vary based on the type of parking, location, and time of day. The subscription costs such as residential parking permit fee, employers' subscription cost are also included in the dimension of parking cost. The significance of this dimension lies in its influence on mode choice and its potential to contribute to illegal parking behaviours due to cost considerations. The associated levels are shown in figure 5.

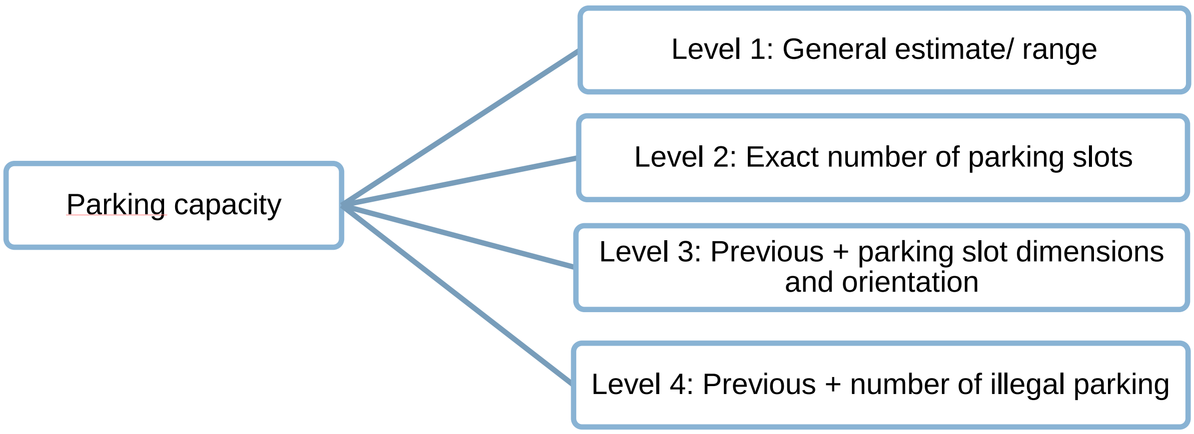


Figure 4. Levels of the dimension parking capacity

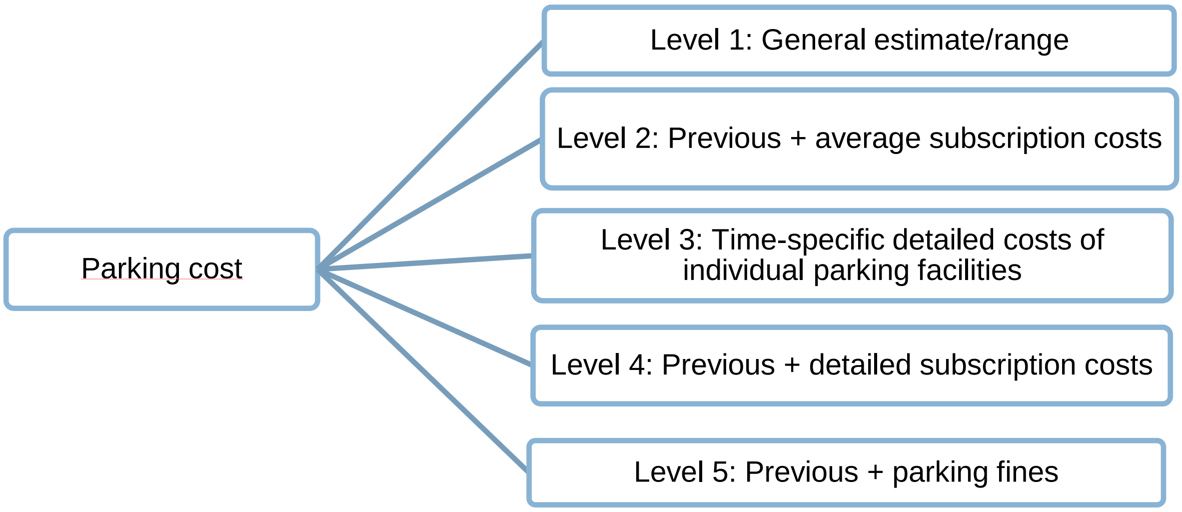


Figure 5.Levels of the dimension parking cost

5) Parking Occupancy:

Parking occupancy refers to the number of parking spaces that are occupied at a given time. The levels for the dimension parking occupancy are shown in figure 6.

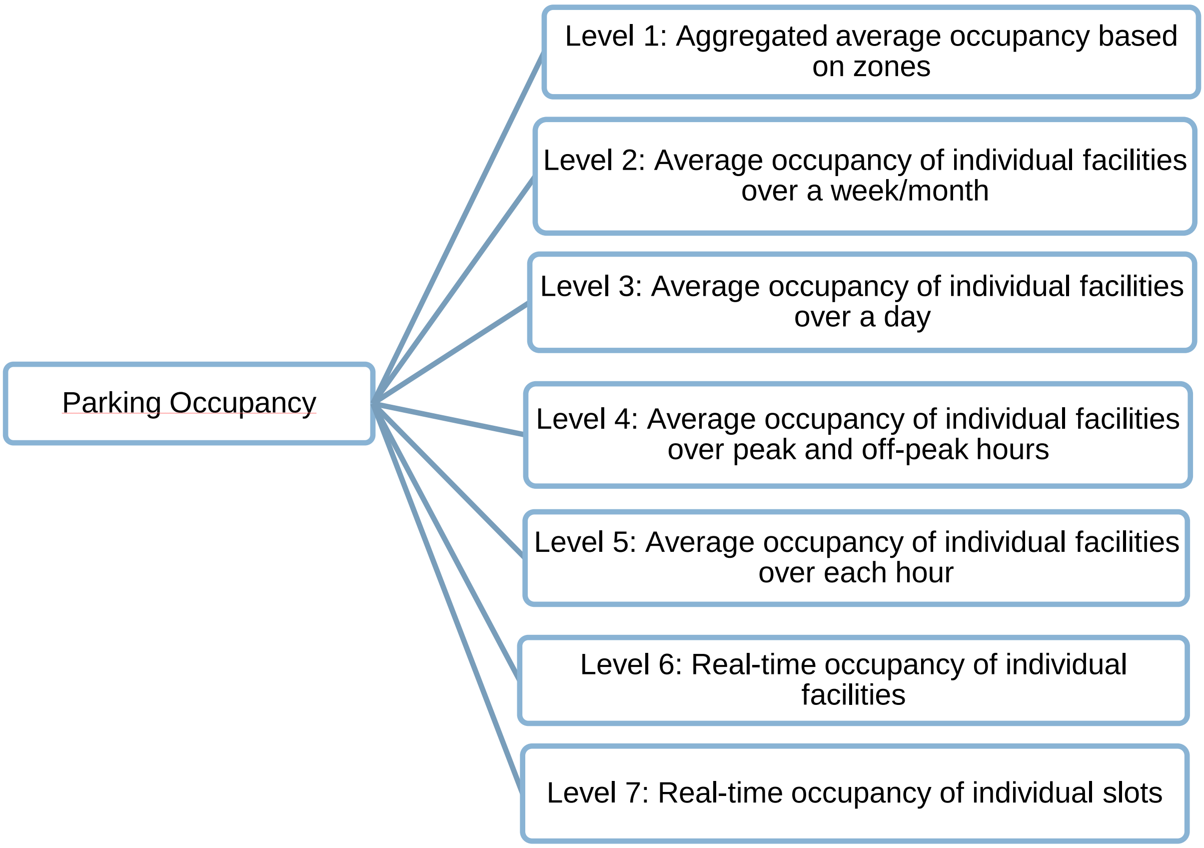


Figure 6. Levels of the dimension parking occupancy

6) Parking search time:

Parking search time refers to the amount of time a driver spends looking for an available parking spot. This dimension is crucial because parking searches has the potential to create congestion in a city thereby impacting the overall traffic flow. The associated levels are shown in figure 7.

7) Egress distance:

Egress distance refers to the distance a person travels on foot from their parked vehicle to their destination. This dimension is relevant because it influences user convenience, and the overall attractiveness of parking options as longer egress distances can discourage the use of certain parking facilities. The levels of the dimension are shown in figure 8.

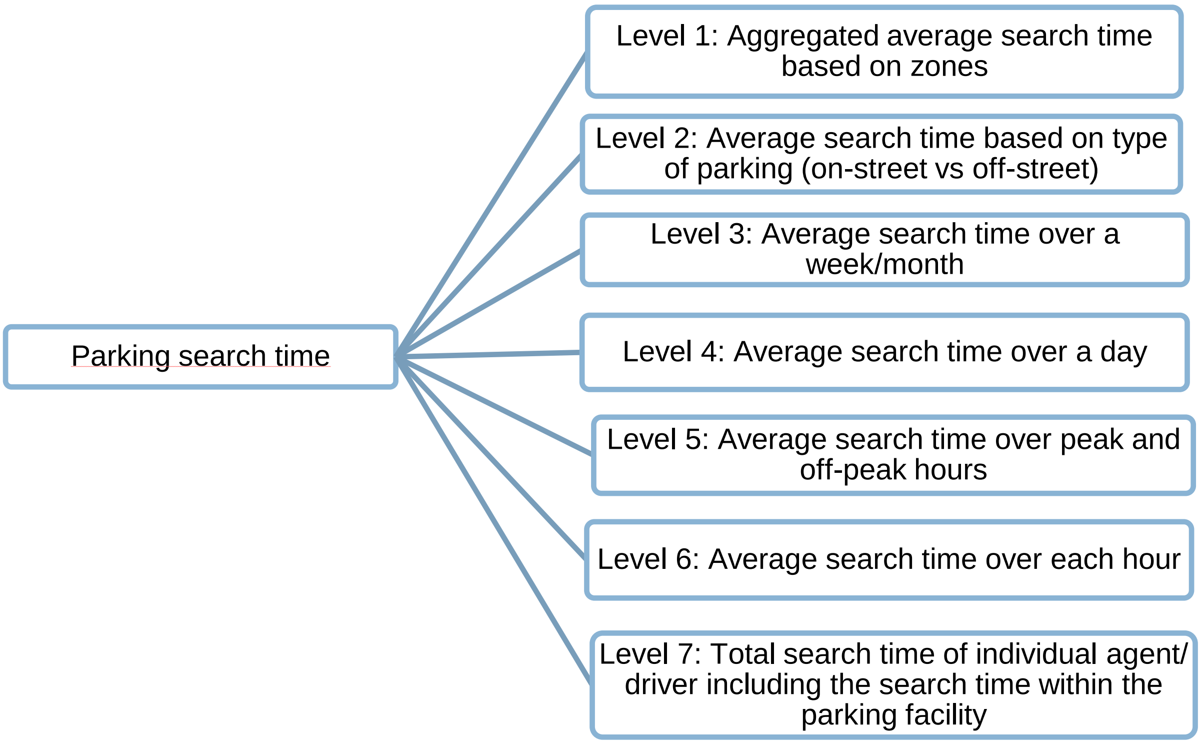


Figure 7. Levels of the dimension parking search time

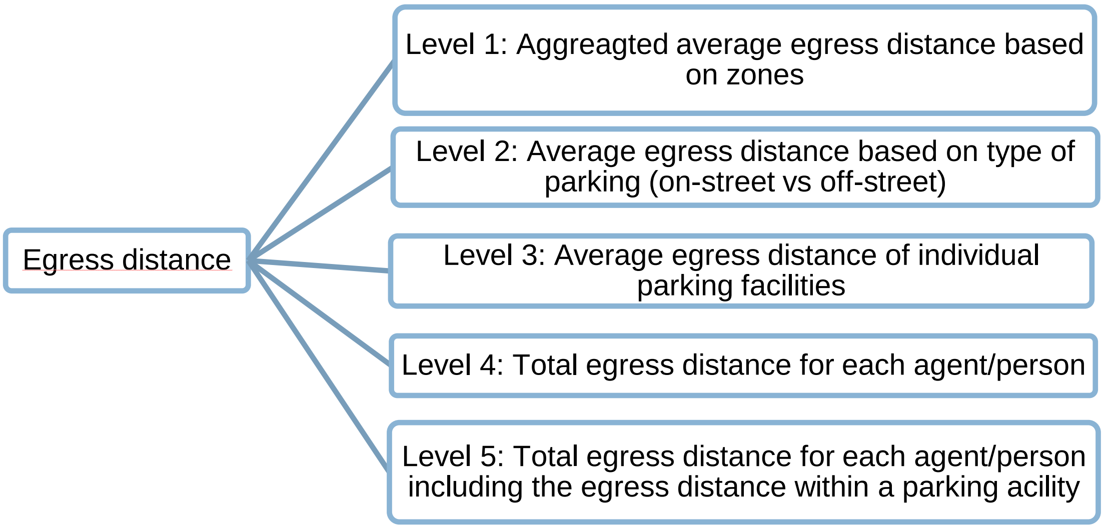


Figure 8. Levels of the dimension egress distance

Although these levels were defined based on insights from the literature, it is possible to further subdivide them based on spatial and temporal resolution. While nested levels were also considered, they add significant complexity, particularly when attempting to visualize the data, and formulating interview questions, making it more challenging to interpret.

3.2. Identifying modelling gaps

3.2.1. Visualization technique to map modelling gaps

As different models are studied here to identify the modelling gaps, visualizing both current modelling practices and the requirements of parking data helps in comparing the data needs of various transport models. Although there are several ways to present this information, the spider diagram, also known as radar chart, proves particularly effective in displaying multivariate data. Its structure allows displaying parking data dimensions along its arms/axes (also known as dimensions), with each arm divided further into levels representing the different data granularities. This approach provides a clear and simple method to compare the data needs across models and highlight the modelling gaps. For this reason, the spider diagram is the chosen visualization technique for identifying modelling gaps in this thesis.

The structure of a dummy spider diagram used to identify the modelling gaps is shown below.

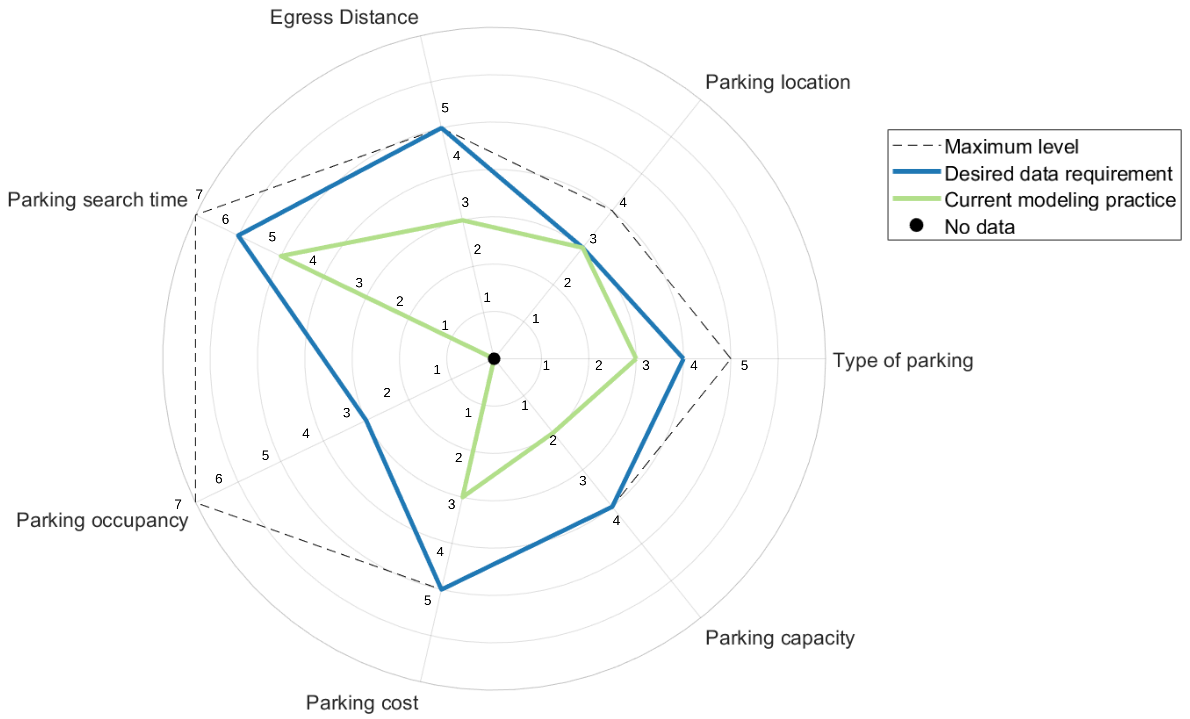


Figure 9. Structure of a dummy spider diagram (demand data)

The numbers 1-7 represent the respective levels for each dimension, while the centre point (black dot) indicates 'no data' is considered. The levelling scheme is arranged such that as the levels increase (from 1 to 7), the data becomes more detailed, with a corresponding rise in complexity and difficulty in data collection.

In the current practice (green line), 'no data' means the model neither requires (if input) nor generates data (if output) for that particular dimension, currently. In the desired data requirements, 'no data' indicates that the data is not needed (if input) or expected to be generated (if output), either now or in the future (depending on the model's scope).

Example of interpretation:

For ‘type of parking’, even though the maximum possible level is level 5, the modelling applications require data only till level 4 and the current modelling practice, due to limited data availability is level 3. Therefore, a modelling gap is observed in the dimension ‘type of parking’ in between levels 3 and 4.

The terminologies associated and their relevance are:

Maximum level: Unlike most spider diagrams, where all dimensions have the same number of levels, the dimensions in this diagram have varying numbers of levels. Since the classification of levels differ significantly, it’s difficult to standardize the number of levels across all dimensions. To address this, the visual design was adjusted by introducing the term "maximum level." This adjustment was necessary due to the limited capabilities of the software used to generate the image (MATLAB), which required the inclusion of this new term. Even though the number of levels are not same across all dimensions, the level of detail in data increases for all the dimensions with increasing level, such that the levels closer to ‘no data’ are coarser levels and the farthest levels are the most granular levels.

Current modeling practice: This term refers to the state of data that modelers currently use, which may not necessarily mean the actual availability of data but represents the current practice in modeling. The reasons why a modeler chooses a particular level as their current practice can be the lack of implementation-ready data, the model doesn’t have the capability to model with higher granular data as of now or the specific application requires only data till the specified level. However, when a dimension is the generated output of the simulation (for example occupancy, search time etc.), the level of current practice reflects level till which the output has been generated.

Desired data requirement: This term refers to the level of data that a model/modeler wish to have for their application, if the full capabilities of a model are utilized. The data requirements can correspond to the current applications or also the future applications. If a dimension is the generated output of the simulation, for instance parking search time, then the desired data requirement means the level of data needed for validating the results.

Modeling gap: This refers to the gap between the current modeling practices and the desired data requirements. The gaps are formed by the constraints contributed by the present modeling approaches due to various reasons. Closing this gap is motivated by the need to exploit the full capabilities of the model for more precise results, more insights and the develop new applications.

Experts representing the models considered in this thesis are consulted and asked to select a level for each dimension, for both ‘current modelling practices’ and the ‘desired data levels’. This is done to identify the modelling gaps in parking. At the end, each model will have a seperate spider diagram based on the results given by interviewees. Ultimately, the spider diagrams of all the models will be overlaid to find the most repetitive gaps.

3.2.2. Selection of models and interview candidates

Several models have incorporated aspects of parking into their frameworks, whether be travel demand models or explicit parking models. However, as discussed in the literature review, eight models have been identified through published research. Experts representing these models are invited to participate in online interviews to select the levels for current modelling practices and the desired data requirement for their present or future applications. It is important to note that the responses will reflect the specific applications and expertise of the experts being interviewed, and they may not be fully representative of the entire modelling community. This is because different application has different input and output constellations, which might imply different 'desired' versus 'current' data availabilities. Nevertheless, this approach offers valuable insights that are often not available in published literature. Only those models for which responses were successfully obtained will be presented in the results section as the literature on these models alone offers only limited information on this topic. The models considered are:

1) VISUM

2) MATSim

3) SUMO

4) AIMSUN

5) TAPAS

6) VISSIM

7) PARKAGENT

8) SUSTAPARK

For interview, the first approach is to contact the authors of the papers where specific parking modelling applications were conducted using the identified models through email. If the original authors were unreachable, alternative contacts were pursued, such as developers of the model or other individuals associated with the model/software.

3.2.3. Interview questionnaire design

An extensive questionnaire was developed to gain an in-depth understanding of the current state of parking modelling and to identify modelling gaps, particularly in terms of data. The main goal of the interview is to gather information that are not presented in the literature. Based on this, three objectives are formulated for the interview:

1) To understand the applications and use cases while modeling parking within the specific model.

2) To identify the current practices and requirements of the parking data utilized in the respective model to identify modelling gaps in parking data

3) To gather insights from modelers on their expectations, and concerns regarding the quality and availability of parking data provided by data suppliers.

The questionnaire contains 20 questions, mostly repeating about marking the levels for each dimension. The complete questionnaire can be found in Appendix B.

3.3. Methodology to identify supply of parking data

3.3.1. Visualization of parking supply data

A similar approach to the one used for identifying parking data demand is applied here for visualizing parking supply data. A spider diagram is used to represent the different dimensions and respective levels of supply data. However, the spider diagram in this case don't have a demand data line, and it only presents the current state of available data. The diagram distinguishes between data from open geoportals, which can be accessed and downloaded for scientific research, and the data provided by commercial companies.

The structure of a dummy spider diagram used to visualize the state of parking supply is shown below.

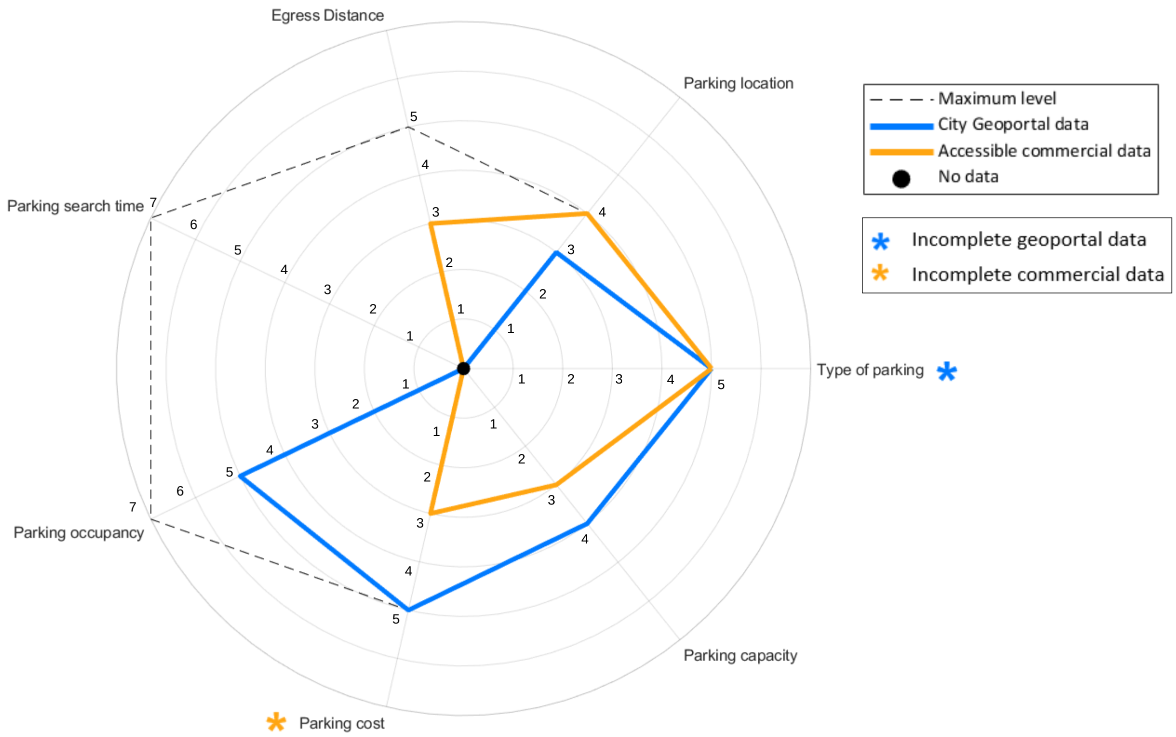


Figure 10. Structure of a dummy spider diagram (supply data)

In this diagram, the blue line represents usable data from city geoportals, while the orange line indicates accessible data from commercial providers. Similar to the demand supply diagram, the numbers 1-7 represent the respective levels for each dimension of parking data.

Since the data may not always be complete, asterisks are used to indicate missing or incomplete data: blue asterisks for incomplete geoportal data and orange asterisks for incomplete commercial data. The black dot in the centre of the diagram signifies that the data provider does not have any data available for that particular dimension.

Example of interpretation:

For the dimension ‘type of parking’, both the city geoportals and accessible commercial data has data up to the highest granular level of level 5. However, the data from city geoportals are incomplete. This can happen when they have only data on on-street parking, but no off-street parking. Hence the level 5 which included data or parking regulations for each slot exists for on-street parking, but not for off-street parking. Similarly, for the dimension ‘parking cost’, the city geoportals have the highest possible granular data. The accessible commercial data providers have data till level 3, but the data is incomplete. This can happen for example when they might have only data for parking garages, but not for on-street parking.

3.3.2. Investigation on parking data providers throughout German speaking areas

Through a web-based inquiry, several parking data providers, in addition to geoportals, were identified. These providers include companies that manage commercial parking data, as well as organizations offering public parking guidance systems. The focus is on contacting entities that provide publicly accessible or commercially available parking data relevant to the German-speaking regions. Reaching out to these providers allows for a better understanding of the current availability, quality, and accessibility of parking data. The following are the major parking data providers observed for the study area:

1) City geoportals
2) APCOA
3) Parkopedia
4) OSM
5) Parking guidance system Zurich
6) Tomtom

7) Inrix

8) Easypark + PARK NOW

9) Parkster

10) Paybyphone

While there may be additional data providers beyond those identified, these represent the primary sources discovered through this investigation. The data provided by these providers will be used to assess the current state of parking data availability across the selected cities.

3.3.3. Selection of cities and interviewee candidates

Five major cities in the German speaking areas are selected for studying parking data patterns: Munich, Berlin, Zurich, Hamburg, and Cologne. These are some of the major cities in German speaking regions, with well-established infrastructure, and availability of parking data. Focusing on larger cities allows for a more thorough examination of data supply, as they tend to have better access to resources and systems for data collection. The interview candidates were chosen based on their involvement with parking data management, either through city geoportals, commercial providers, or other relevant stakeholders.

Individuals associated with the data provided in the city geoportals and contact persons in the websites of commercial data providers are consulted and asked to select a level for each dimension. Hence each city considered in this thesis will have a separate spider diagram based on the results given by interviewees. In the end, the spider diagrams of all the cities will be compared.

3.3.4. Interview questionnaire design

An extensive questionnaire was developed to gain an in-depth understanding of the current state of parking data across the five cities. The main goal of the interview is to understand the current state of parking data and also to gather information on the quality and comprehensiveness of the data, that are not explicitly described in the websites. The interviews also addressed limitations in data collection and questions about the reliability of the data were also asked. The questionnaire contains 11 questions, mostly repeating about marking the levels for each dimension. The complete survey questionnaire can be found in Appendix C.

4. Results of parking data demand and supply analysis

4.1. Assessment of parking data demand for transport models

The following results build upon the methodology outlined in the previous section, presenting insights gained from interviews with transport modelers. Despite efforts to reach a broader set of models mentioned in the methodology, responses were obtained from representatives of five models: VISUM, MATSim, SUMO, TAPAS, and PARKAGENT. The interviewees are either developers of these models or authors of research papers that applied these models in parking applications. A full list of the interviewees and their designations can be found in Appendix D, while Appendix G provides a table of all dimensions and levels. By referring to Appendix G, the levels of the dimensions can be easily cross-referenced to better interpret the findings presented.

Spider diagrams are presented for each model to illustrate both the modelling requirements and current practices related to parking data, enabling a clear comparison across the different transport models. It is important to note that the ‘desired data levels’ and ‘current modelling practices’ described here are specific to the modelling approaches shared by the interviewees. Different modelling approaches with varying data input/output structures could results in different desired and current modelling practices. Additionally, the availability of parking data is contingent upon the specific city in which each model is applied, meaning that the data discussed here may differ depending on the cities studied.

4.1.1. VISUM

VISUM is a widely used macroscopic travel demand model, particularly for strategic and tactical applications in transport modelling. While conventional models like VISUM are popular in transport planning, the integration of parking within the model framework remains a relatively underexplored topic. However, parking aspects have been modelled in VISUM to evaluate parking management strategies (Lubrich, 2023). Although VISUM doesn’t have an explicit parking feature by default, the system can be adapted by using proxies, such as connectors, to incorporate parking elements into the model.

Results using spider diagram

The dimensions - parking location, type of parking, parking capacity and parking cost are required as inputs for the modelling, while occupancy and egress distance are a mix of input and output. Parking search time is a model output and data on the outputs are required for validation.

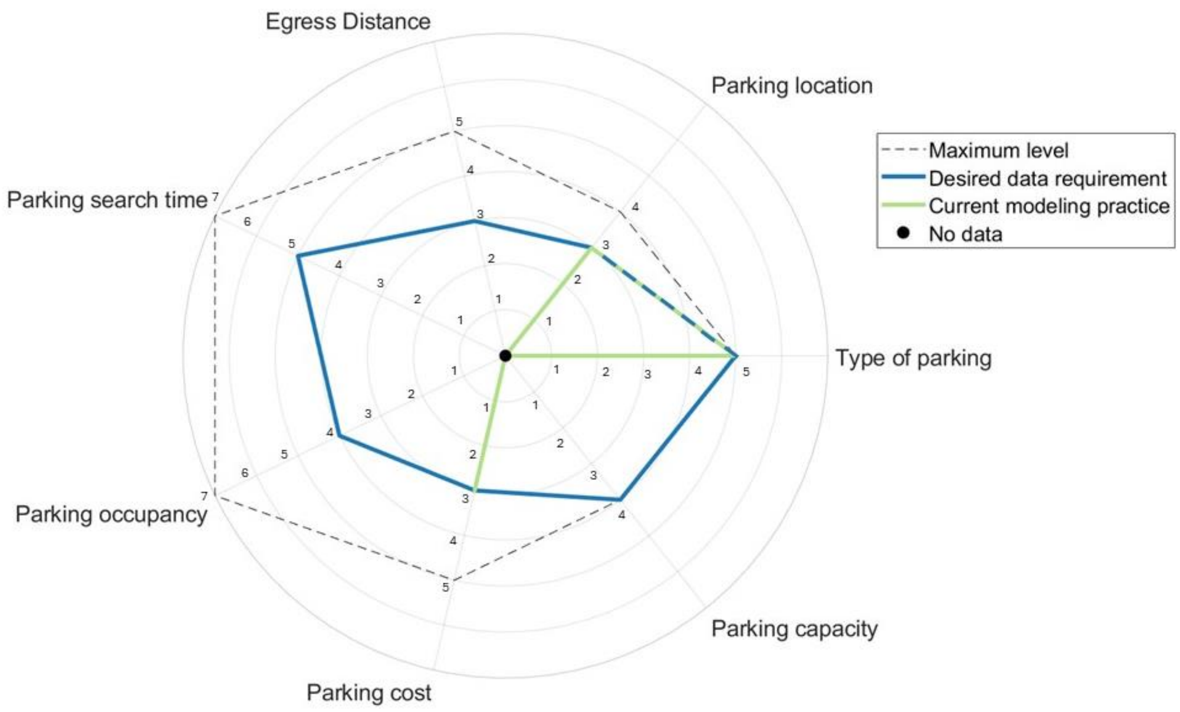


Figure 11. Results as spider diagram for the model VISUM

For the dimension ‘parking location’, the current modelling practice and the desired data level are both at level 3 (link level) as it meets the needs of the modelling application. For ‘type of parking’, the model application of Interviewee 1 which focused on evaluation of parking management strategies, utilized data at level 5 (distinction between individual slots based on parking regulations) making it the current practice. For evaluating parking management strategies, it is essential to know the regulations for each parking slot making level 5 the desired data level as well. For the same application, ‘parking capacity’ data wasn’t required making the current modelling practice level 0 (no data). However, the interviewee received data up to level 2 (exact numbers) for the study area in Cologne. Future use cases need data at level 4 (exact number, parking slot dimensions and orientation, and number of illegal parking) as data on illegal parking is necessary for assessing management strategies. Similarly, although 'parking cost' data up to level 4 was received, only selected data (PM peak hour cost) were utilized, setting both the requirements and current practices at level 3 (time specific detailed cost of individual parking facilities).

‘Parking occupancy’ data were not used in the model resulting in the current practice being at level 0 for this model set-up. The reason was that the input data provider (a commercial parking API service) did not provide any explicit occupancy information. Instead, some proxy input data on parking demand was used via probabilities to find a free parking spot per link. In addition, the model calculations include the parking-route and parking-location choice behaviour of each traveller searching for parking, which could be interpreted as output data on parking occupancy. However, the desired data level is level 4 (average occupancy of individual facilities over peak and off-peak hours), at least to validate the parking location choice being calculated by the model and the gap is due to limited data availability. ‘Parking search time’ is an output of the VISUM parking model, generated up to level 5 (average search time over peak and off-peak hours). However, no data on search time were obtained for validation purposes, resulting in the current practice being at level 0. The dimension of ‘egress distance’ is a mix of model input and output in VISUM. While the parking model in VISUM could calculate egress distance based on some model calculations up to level 3 (average egress distance of individual parking facilities), the absence of validation data places the current practice at level 0. It is noted that the ‘desired data level’ and ‘current modelling practices’, as noted above, are only valid for the model approach proposed by Interviewee 1. There are also other model approaches, that have other data input/output constellations, which might imply different ‘desired data level’ and ‘current modelling practices.

Gaps and Interpretation

While there are no gaps for parking location and type of parking data, there is a significant gap between levels 0 and 4 of ‘parking capacity’, despite this data not being required for the model application. This gap is partly attributed to lack of data (data gap is between levels 2 and 4). ‘Parking occupancy’ has a significant gap between levels 0 and 4, as no usable data was obtained for the Cologne study area. Significant gaps are also found in the validation data for search time, and egress distance.

Despite lacking an explicit parking feature, VISUM can incorporate parking elements using proxies like connectors. VISUM can generate satisfactory outputs for parking search time and egress distance but lacks the necessary validation data, leading to potential inaccuracies in reflecting real-world conditions. These gaps underscore the necessity for more comprehensive and granular parking data for the successful modelling of parking evaluation strategies using VISUM. The model’s ability to model parking is robust where data is available, making it essential to improve data quality and availability to enhance VISUM's precision and performance in parking modelling.

4.1.2. MATSim

As mentioned in section 2.2.2, the MATSim simulation, by default, does not consider parking infrastructure or supply constraints. The default setup is agents just drive to their destination link and park their car without any parking constraints or search process. However, various parking extensions have been developed and used within the MATSim community to enhance this basic functionality. For the most evolved studies on MATSim parking, an agent arrives a link and have to search for a parking place unlike the default one. Although MATSim as a transport model can be used for strategic, tactical and operational applications, the applications for parking modelling mainly lies within the tactical-operational levels. Examples of applications in modelling parking include designing traffic policies by means of reduction of parking supply or adjusting parking costs.

Results using spider diagram

As MATSim operates at the link level, both the desired and current levels for the ‘parking location’ are at level 3. This is because in both the default and evolved simulations in MATSim, the agent arrives at a link and searches for a parking space. For the dimension ‘type of parking’, the current modelling practice is at level 1 (distinction between parking and no parking). The reasons for staying at level 1 are the lack of comprehensive data and the limited computational power of the model. However, the desired level for future applications is level 4 (distinction between individual parking facilities). As already mentioned, MATSim by default doesn’t consider ‘parking capacity’. However, for studies focusing on cities with parking capacity data such as Zurich (Waraich & Axhausen, 2012), data at level 2 (exact numbers) has been used, making it the current modelling practice. However, for further applications, data on the parking slot dimensions and orientation as well as the number of illegal parking is required (level 4). For cities without parking capacity data such as Leipzig, parking is modelled under the assumption that there is infinite capacity in the link [Interviewee 2].

Figure 12. Results as spider diagram for the model MATSim

For the dimension ‘parking cost’, the current practice is level 3 (time specific detailed cost of individual parking facilities) as the modelling applications of Interviewee 2 require only data till this level. However, future applications require level 5 which includes additional information on subscription costs and parking fines. It is important to note that cost is a significant attribute in MATSim, as the model incorporates marginal utility of money, meaning that each monetary term has differently assigned score based on the agent’s income. Consequently, parking costs are not merely an attribute for parking; rather they reflect the behavior of individuals corresponding to price fluctuations in general [Interviewee 2].

The dimensions ‘parking occupancy’, ‘search time’ and ‘egress distance’ are outputs of MATSim simulations and Tchervenkov, 2022 utilized and generated these dimensions at higher granularities. For the dimension ‘parking occupancy’, data has been generated till level 6 (real-time occupancy of individual facilities) making it the required level for validation of the results. Parking occupancy data is usually computed from the activity patterns of agents in MATSim as these patterns gives a rough idea about how long the slot is occupied. The data on activity location and duration is often obtained from GPS or mobile phone data. In order to get more accurate values on occupancy derived from parking duration, access-egress time should be also added in the calculations, which is not currently considered in MATSim. Tchervenkov, 2022 used the overall parking occupancy over a typical day considering a 30-hour day (until 6:00 in the following morning) making the current validation data level at level 5. Similarly, 'parking search time' can be generated up to level 6 (average parking search time over each hour excluding time spent searching within the facility) placing the requirement of validation data at level 6. Maurer et al., 2023 registered the parking occupancy and parking search time on a weekday morning by following vehicles in cycles, generating level 5 (average parking search time over peak and off-peak hours) for validation data making it the current modeling practice. For 'egress distance', output can be generated up to level 4 (egress distance of each agent excluding distance inside the facility) when both the activity and parking location are known. Validation data can also be obtained till level 4 as even with GPS data, parking location within the garages is hard to obtain to achieve level 5.

Gaps and interpretations

While there is no gap in the dimension 'parking location', a significant gap exists in the dimension 'type of parking' between levels 1 and 4, primarily due to data limitations and limited computational power of the model. Similarly, the gap between levels 2 and 4 for 'parking capacity' and levels 3 and 5 of 'parking cost' is partly caused by insufficient data. For the dimensions 'parking occupancy' and 'search time', there is a one-level validation data gap between levels 5 and 6, caused by the difficulties in collecting data at higher levels. For egress distance, there is no gap, as level 4 is both the desired and currently utilized levels.

In conclusion, while MATSim is a general-purpose agent-based simulation tool widely recognized for its successful applications in large scale transport modelling, its capacity to accurately model parking is limited by the absence of capacity constraints in the model and the limited availability of granular parking data. Additionally, as MATSim is primarily focused on the links rather than individual parking slots, it does not capture the behaviours within large parking facilities. However, the current applications of MATSim in parking modelling do not require this level of detail. Despite these limitations, the model has still been effectively used by researchers, particularly with parking extensions that exploit its full potential to capture the complexity of parking dynamics. Addressing the challenges of insufficient data (especially type of parking, parking capacity and parking cost), limited computational power and the absence of a behavioural model to simulate sudden changes in agent's behaviour could enhance MATSim's ability to better reflect parking scenarios [Interviewee 2].

4.1.3. SUMO

SUMO, being a general-purpose microsimulation traffic model, is capable of generating high detailed parking simulations. The applications in modelling parking in SUMO falls under tactical-operational model. Each agent in the simulation drives to a specific parking spot and if it's occupied, it starts to search for other spots. The applications in modelling parking within SUMO includes observing parking search times, the traffic caused by parking, resulting congestion in the network, and the amount of time wasted during these searches. This capability enables a rough understanding of how parking dynamics affect the urban traffic flow. One of the main advantages of SUMO over other models is its flexibility to simulate parking at different levels of granularity, as explained below.

Results as spider diagram

Figure 13. Results as spider diagram for the model SUMO

For the dimension 'parking location', the required level and current modelling practice are at level 4 (point level) as SUMO primarily relies on OSM as its main data source, which provides this information. However, the quality and heterogeneity of the data available in OSM is limited, highlighting the need for an additional, more accurate data source. While SUMO can also work with level 3 (link level), city level and zonal level are generally too coarse for detailed simulations. For 'type of parking', the current modelling practice is at level 4 (distinction between individual parking facilities) while the requirement is at level 5 (distinction of individual slots based on parking regulations). Depending on the application, SUMO can also work with lower levels of this dimension. For instance, all on-street parking can be aggregated into a single large parking facility, based on the specific modelling needs. However, using lower levels can affect the preciseness of the results, depending on the applications.

For 'parking capacity' the current practice is at level 3 (exact numbers and parking slot dimensions and orientation), whereas the requirement is at level 4, which includes additional information on the number of illegal parking. SUMO includes data on parking slot dimensions and orientation for visualization, and this also helps in limiting parking for certain vehicles based on slot dimensions. If data on 'illegal parking', is available, SUMO can utilize it to restrict specific user groups to certain parking areas based on regulations, although there is currently no implementation-ready data on this. Nevertheless, depending on the application and project needs, SUMO can work with lower levels of 'parking capacity' due to its flexibility to work with different granularities. Regarding the dimension 'parking cost', SUMO does not currently use this information, which is a limitation of the model. Hence, both the current modelling practice and the near-future requirement remain at level 0 [Interviewee 3]. However, it is important to note that if the future applications require parking cost data, the model can handle the most granular data. While SUMO does not directly integrate parking costs, the model recently added features to restrict certain user groups (traffic participants) from specific parking areas, as parking costs might be too high for those groups. Additionally, SUMO supports some abstract 'preference' metric that can be attached to each parking area with a configurable weighting factor that trades this value against factors like time and occupancy [Interviewee 3]. Though this metric could be adapted to account for parking costs, it has not yet been used for this purpose.

'Parking occupancy' is an output of the SUMO simulation and hence data on it is required as an input for the validation. SUMO can generate parking occupancy up to level 7 (real time occupancy of individual parking slots), making it the desired data level for validation. However, as per Interviewee 3, the data available so far is only till level 5 (average occupancy of individual facilities over each hour). Similarly, 'parking search time' is also a simulation output, which SUMO can generate up to level 7 (total search time of individual agents including the search time within the parking facility). SUMO has not modelled parking search times inside the parking facilities as these details are not important for its general applications. While the validation data requirement for parking search time is at level 7, no such data is currently available. Even though search time references (e.g. a constant value of 10 minutes) exist in the literature, search times vary significantly with the local conditions, making this topic challenging to approach data wise [Interviewee 3]. Similarly, 'egress distance' is also a simulation output which SUMO can generate up to level 5 (total egress distance for each agent including the egress distance within a parking facility). However, since most of the parking areas in SUMO are modelled as monolithic facilities, agents do not have to walk within the facility on extra roads or sidewalks. As a result, egress distances inside parking facilities have not been modelled so far. Since no data is available on egress distances for the validation, current practice remains at level 0. It is important to note that the 'desired data level' and 'current modelling practices' mentioned above, are specific to the model approach proposed by Interviewee 3.

Gaps and interpretations

While there are no obvious gaps for the dimension 'parking location', the quality of data obtained from OSM is often inconsistent and lacks homogeneity, which limit its reliability. For 'type of parking' there is a small gap between levels 4 and 5, due to lack of data and limited computational power of the model. Similarly, the gap between levels 3 and 4 in the dimension 'parking capacity' is attributed to the absence of data on illegal parking. 'Parking cost' is currently not included in SUMO's parking choice model, and there are no plans for its integration soon. On the other hand, 'parking occupancy', 'search time', and 'egress distance' can be simulated in high detail. However, there are significant gaps in the validation data for these outputs, mainly due to the scarcity of available data.

SUMO can be a valuable tool for simulating parking dynamics in high detail, even when the data availability is limited, due to its ability to work with different levels of data granularity. However, excluding parking costs as a model attribute in SUMO can lead to the miscalculations in parking demand. Additionally, this omission limits the model's ability to assess the effectiveness of parking policies such as dynamic pricing and parking permits. Furthermore, while aggregating all parking areas into a single facility may be suitable for certain applications, this approach can overlook the specific effects of on-street parking on traffic flow, parking search times, and congestion caused by these searches. Nonetheless, the interview results shows that microsimulation traffic models like SUMO are capable of doing much more in modelling parking than what they are currently doing now, if they have the desired data. This underscores the necessity of more granular data in terms of parking regulations, capacity, occupancy, search time and egress distance.

4.1.4. TAPAS

TAPAS is an agent-based microsimulation model that models the behavioral changes of agents in response to both spontaneous event changes and planned long-term changes such as policy measures including speed limit reductions, parking space reductions, and changes in parking costs. Although, parking is not explicitly modelled in TAPAS, it is included as a component of the egress time associated with each TAZ (Traffic Analysis Zone). For example, instead of decreasing parking capacity as a policy, the egress distance can be increased, making the location less attractive. TAPAS also closely works with SUMO, performing the first three modelling steps (trip generation, trip distribution, and mode choice) before handing over the fourth step, trip assignment, to SUMO. TAPAS simulations provide total travel times from the pre-computed distance and travel time matrices. These travel times are fed into SUMO, resulting in updated travel time matrices that are iteratively refined to equilibrium. The model applications for parking in TAPAS fall under the strategic-tactical levels.

Results using spider diagram

TAPAS doesn’t output any data on parking and all the dimensions specified here are utilized as inputs for the model. However, certain information such as parking occupancy or egress distance can be inferred from the model outputs. The model output includes the list of trips made, start and end locations with the TAZ information, mode of transport, travel time, and information about the activity performed and the agent itself.

As TAPAS focuses broadly on TAZs, the current practice for the dimension ‘parking location’ is at level 1 (zonal level). However, in future, TAPAS would like to utilize more granular parking location data at level 4 (point level). Currently, TAPAS distinguishes between four categories of parking (CAT-0 to CAT-3) at the TAZ level, with the primary difference between these categories being the price per hour. However, based on the levelling scheme for the dimension ‘type of parking’, the current practice can be aggregated to level 1 (distinction between parking and no parking), as no distinctions are made here based on the access to parking (on-street vs off-street).

Figure 14. Results as spider diagram for the model TAPAS

When an agent is bound to its vehicle and is aware that there is no parking availability in a specific location/TAZ, the model assumes an infinitely high egress time, effectively preventing the agent from choosing that location. For agents not bound to a particular mode of transport, since location choice precedes mode choice, the agent would end up choosing a different mode of transport. While it is possible to model an infinitely large egress time when there is no parking availability in the destination, in reality, individuals may find parking somewhere even if it's illegal [Interviewee 4]. Alternatively, TAPAS uses a dummy value, such as a travel time of -1, to indicate that driving between two zones is not possible with the chosen mode, ensuring the journey won't be made. However, the desired level is level 5 (distinction of individual slots based on parking regulations) as the additional attributes to each parking slot could help in future applications.

‘Parking capacity’ is currently not considered in TAPAS as it falls outside the model's current scope. Instead, the model assumes that if an agent can enter a TAZ, parking is available for the agent. In the future, there is an interest in incorporating parking capacity as a model component for experimental purposes to see if it is calibratable or if it’s possible to implement different behaviors such as preferences of agents when given multiple options [Interviewee 4]. For the near future, level 1 (general estimate) would be the required level as exact numbers could complicate the model by agents potentially developing excess strategies demanding higher computational power [Interviewee 4]. ‘Parking cost’ as already discussed, is a configurable attribute of each TAZ, with four parametrizable attributes (CAT 0 - CAT 3) with the model being sensitive to these costs. This keeps the current practice at level 1 (general estimate/range) as the cost values utilized in the model are not the detailed costs with information on parking subscriptions. As incorporating detailed subscription costs requires more information such as information on who holds parking subscriptions, the desired level for the future applications is limited to level 3 (time specific detailed cost of individual parking facilities) that excludes the subscription costs.

At present, TAPAS doesn’t consider ‘parking occupancy’ data (level 0) as the focus of the model is on behavioral changes rather than detailed planning [Interviewee 4]. But parking occupancy estimate could be derived by simply checking how many agents are at an activity in a TAZ and used their car to get there. In the future, there is a desire to incorporate occupancy data at a basic level (level 1 - aggregated average occupancy for TAZs). ‘Parking search time’ is not directly modelled currently but incorporated as a constant value inside an egress matrix (level 1- aggregated average search time for TAZ’s /parking zones). As parking search time varies a lot with the local conditions of the agent’s location, the value for each TAZ is estimated either based on some empirical data or expert knowledge known by experience. Surprisingly, the requirement for search time is also at level 1 (aggregated average search time for TAZ) as there are concerns regarding the potential model sophistications associated with incorporating complex data [Interviewee 4]. In addition, the detailed simulation of search time is handed over to SUMO as part of the TAPAS-SUMO coupling. Similarly, for ‘egress distance’, the current practice is at level 1 (aggregated average egress distance for TAZ’s /parking zones) as the egress distance value is a constant inside an egress matrix in TAPAS. In future TAPAS would need level 5 (total egress distance for each agent including the egress distance within a parking facility) as parking itself is modelled in TAPAS as a component of the egress time. It should be also considered that the ‘desired data level’ and ‘current modelling practices' as noted above, are only valid for the model approach proposed by Interviewee 4.

Gaps and interpretations

Unlike other models where parking occupancy, search time and egress distance are mostly outputs of the simulation, all the dimensions specified here are used as input data in TAPAS for the modelling purpose. Even though TAPAS in general is very detailed on the activity-level, the model is pretty blind inside the zones, thus lacking detailed insights into activities within each TAZ [Interviewee 4]. For the dimension 'parking location', there is a gap between level 2 and 4 while 'type of parking' shows a significant gap between level 0 and 5. For 'parking capacity' there is a small gap between level 0 and 1 and for 'parking cost' there is a gap between level 1 and 3. Although 'parking occupancy' is not considered in the model, there is a small gap due to the need for incorporating occupancy at a basic level in the model. There is no gap in parking search time, and this is to avoid the potential complexity arising with incorporating more detailed data, as explained by the interviewee. There is a substantial gap in 'egress distance' due to its high significance in the model applications, but poor data availability.

TAPAS selectively focuses on certain type of parking data, and while some dimensions are currently considered, both the current practice and requirements remain at lower abstract levels. Nonetheless, the efforts of TAPAS in advanced behavioral modeling could potentially address the challenges faced by other models in accurately simulating parking behavior. Although TAPAS currently relies on commonly available data sources such as MiD (Mobilität in Deutschland), which lack the necessary detailed information, there is ongoing interest in incorporating more detailed data for experimental purposes to see if it is calibratable or to explore the impact of this additional information on agent's behavior [Interviewee 4]. This underscores the need for more homogeneous and detailed parking data for TAPAS in terms of parking location, type of parking, capacity, cost and egress distance, to more effectively model the parking phenomenon.

4.1.5. PARKAGENT

PARKAGENT is an agent-based, spatially explicit model developed as an ArcGIS extension to simulate parking behaviour in a city. Unlike traditional models, PARKAGENT can simulate the behaviour of each agent in high detail capturing the complex parking dynamics occurring between agents. There are several studies involving PARKAGENT and one of such studies was conducted in Tel Aviv, where detailed parking data was collected through field surveys, providing accurate, high detailed data on parking attributes. PARKAGENT's applications include modelling parking search behaviour in city centres, assessing the impact of additional parking supply in residential areas with parking shortages, examining how occupancy rates influence parking searches etc., Overall, the applications of modelling in PARKAGENT lies within the tactical-operational levels.

Results as spider diagram

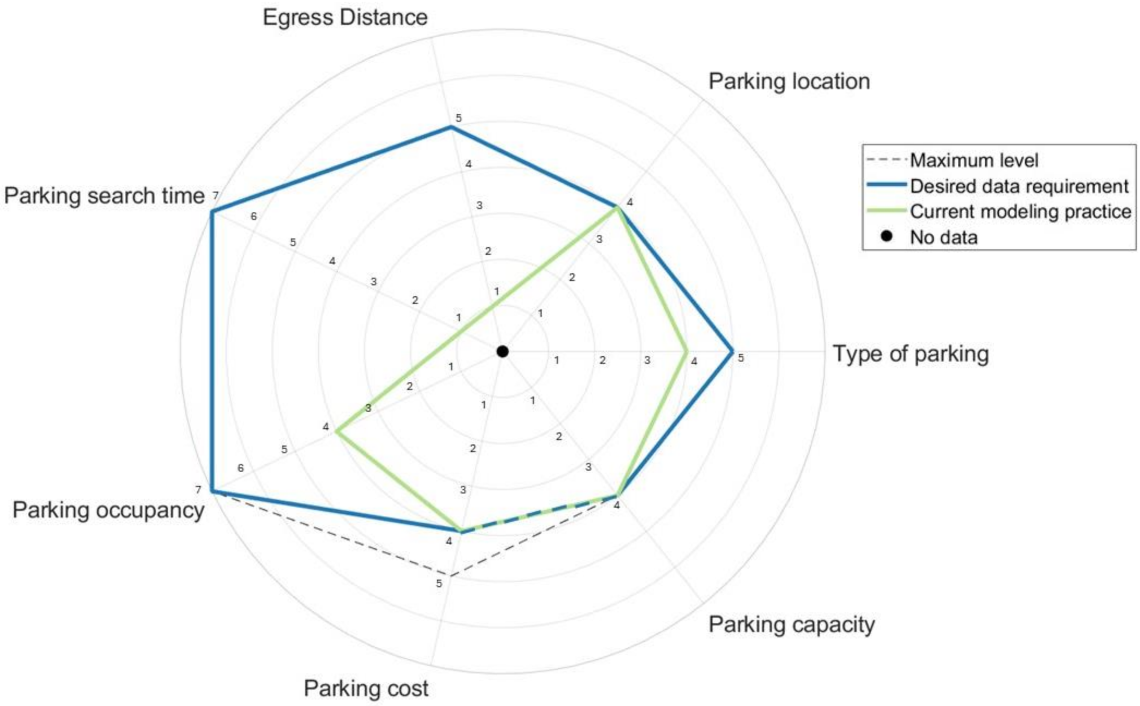


Figure 15. Results as spider diagram for the model PARKAGENT

Similar to other agent-based models like SUMO and MATSim, PARKAGENT also requires data on parking attributes at its highest granularity, as the model is capable of utilizing highly detailed data [Interviewee 5]. For the dimension 'parking location', both the required level and current modelling practice are at level 4 (point level), as PARKAGENT is a spatially explicit model built on high-resolution urban GIS, with layers containing information on every element of the traffic infrastructure. For 'type of parking', the current modelling practice is at level 4 (distinction between individual parking facilities) as the model was able to incorporate data on all types of individual parking facilities, such as curb parking (on-street), multistorey and public lots (off-street parking). Even though the model's application only required data till level 4, the required level is 5 (distinction of individual slots based on parking regulations), as PARKAGENT's applications can utilize and incorporate data at this higher level. For 'parking capacity', both the current practice and requirement are at level 4 (exact numbers, parking slot dimensions and orientation and illegal parking). The field survey performed in the study area collected detailed data on the exact number of parking slots including the number of illegal parking's. The dimensions 'parking cost' serves as both an input and output in the model. While parking cost data was required as an input to model other attributes, the model also generated parking cost values across different driver groups, as one of the key model outputs. Data on parking costs along with information on the residents who hold parking subscriptions, which corresponds to level 4 were utilized in the model. However, interviewee 5 had information on the parking fines, such as when fines are issued and the likelihood of being fined. These components were not integrated into drivers' parking decision process or used in calculating the attractiveness of parking spots. Given this, both the current modelling practice and requirement levels can be considered as level 4.

'Parking occupancy' is also a mix of input and output in the model. Parking occupancy data on peak and off-peak hours in the study area were collected through extensive field surveys on every working day during 2 consecutive weeks making the current practice at level 4 (average occupancy of individual facilities over peak and off-peak hours). Additionally, overnight parking occupancy was recorded once, between 23:00 and 4:00 h. The requirements for the validation data on parking occupancy are at level 7 (real time occupancy of individual slots), as the model generated data up to this level. 'Parking search time' is another key output of the model, with data generated up to level 7 (total search time of individual agents including the search time within the parking facility). Within the model, a parking lot has its internal structure, hence search times inside big parking facilities can be generated. However, despite this capability, it has not been applied to the use cases in the studies. In the modelling approach, based on observations from study area, a threshold of 10 minutes was set for parking search time after which drivers park at the nearest paid parking lot. This threshold can be considered as the available data for parking search time, placing the current modelling practice at level 1 (aggregated average search time for TAZ's /parking zones). Similarly egress distance is a model output, where data can be generated up to level 5 (total egress distance for each agent including the egress distance within a parking facility). Like the threshold value used for parking search time, a comparison between the location of each parked car recorded during the survey and the home address of the owner (determined by matching car plate numbers with statistical data) showed that most drivers park within 350 meters of their residence. This makes the current practice for validation at level 1 (aggregated average egress distance for TAZ's /parking zones).

Gaps and interpretations

There are no substantial data gaps in PARKAGENT across its dimensions, as Interviewee 5 was able to collect all necessary data through extensive field surveys. The small gap observed between levels 4 and 5 in the 'type of parking' dimension exists because the specific modelling application did not require the highest level of data detail, even though the model is capable of handling it. Similarly, the gap between levels 4 and 7 in 'parking occupancy' is because, occupancy data serves as both an input and output, and only data up to level 4 was required for the application, meaning it does not represent a 'data gap'. Although the model estimates 'parking search time' and 'egress distance' at high resolution, gaps remain in the validation data. According to the interviewee, parking search time is highly localized and cannot be effectively represented by average values (as in the levels of the dimension) due to its heterogeneous nature. Interviewee 5 expressed that, instead of average values that doesn't characterize anything, local distribution of search time is required. Similarly, the interviewee had the opinion that egress distance is not some knowledge by itself in the model, as it automatically comes out of the model based on the known destination and parking location.

PARKAGENT is a powerful microscopic parking simulation tool capable of modelling parking behaviour with high detailing. It not only generated high resolution output data, but also utilized accurate input data specific to its applications. In comparison to other models discussed, PARKAGENT incorporated more comprehensive data, particularly concerning parking capacity and occupancy. Additionally, it is one of the few models that considers the internal structure of parking facilities, simulating parking dynamics within these facilities. It is important to note that the model's study area is in Israel, and the data collection methods or availability discussed might not directly apply to German-speaking cities.

4.2. Most significant modelling gaps

The modeling gaps for all five models are identified by comparing desired data requirements and current modelling practices. To reveal the most frequently recurring gaps, the gaps identified across these models have been overlaid as shown in figure 16. The varying thickness of the lines represents the frequency of occurrence of these gaps, such that, the thicker lines indicate more recurring gaps. It should be noted that the gaps are not always due to a lack of data; they may also stem from the constraints caused by the present modelling approaches, which could arise for various reasons. Addressing these gaps is crucial for exploiting the full potential of the models, allowing for more accurate results, deeper insights, and the possibility of developing of new applications.

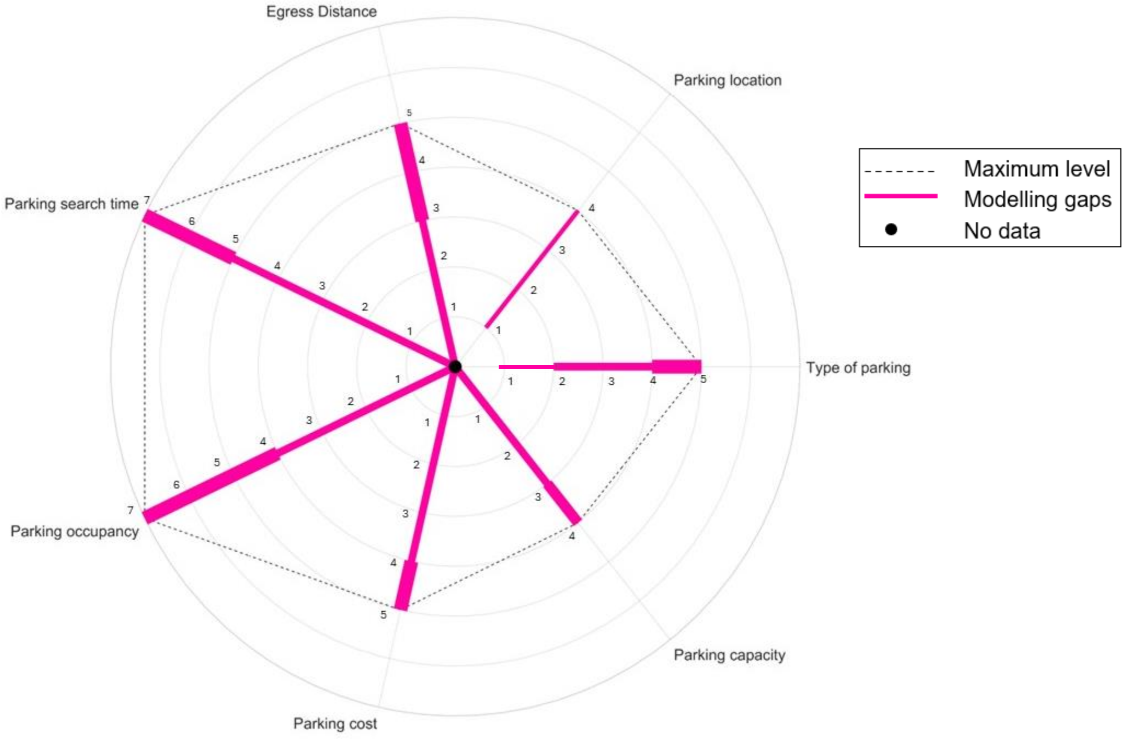


Figure 16. Overlaid spider diagram highlighting data gaps in transport models

From the figure, it can be seen that at least one model shows a gap between almost all the levels in each dimension. The most significant gaps are observed in the following dimensions: type of parking (levels 4 to 5), parking capacity (levels 3 to 4), parking cost (levels 4 to 5), parking occupancy (levels 4 to 7), parking search time (levels 5 to 7) and egress distance (levels 3 to 5). To be more specific, the most demanded data to improve these models for their respective applications are as follows:

Table 2. Most demanded parking data for each dimension

Dimension	**Most demanded data**
Type of parking	Information on the parking regulations applicable to individual parking slots
Parking capacity	Number of illegal parking
Parking cost	Information on parking fines
Parking occupancy	Average occupancy of individual facilities over peak and off-peak hours, over each hour, real time occupancy of individual facilities and individual parking slots
Parking search time	Average search time over peak and off-peak hours, over each hour, total search time of individual drivers including the search time within the parking facility.
Egres distance	Total egress distance for each driver, total egress distance for each driver including the egress distance within a parking facility.

While majority of these levels in the dimensions are justified and necessary for improving the current modelling practices, some interviewees felt that certain levels may be unnecessary or redundant. Interviewee 5 pointed out that certain information, such as ‘egress distance’, can often be inferred or automatically generated because the parking location and destination of agents are already known. Additionally, while average parking search time is often demanded as a key data requirement, estimating average search times in city centres can be problematic. As Interviewee 5 emphasized, averages do not characterize the complexity of parking behaviour, and the resolution of search time require another reference. For example, a local distribution of search times can account for the highly localized and heterogeneous nature of parking.

Regarding 'type of parking', there were different viewpoints. Interviewee 5 emphasized that the most relevant distinction for drivers is whether they need to spend additional time inside a parking facility, such as a multistorey parking garage, which is often avoided due to the time it takes to enter and exit. Thus, distinguishing between curb parking, open space, and multistorey parking is considered more relevant [Interviewee 5]. On the other hand, Interviewee 3 suggested adding another level to differentiate between on-street parking with markings and on-street parking without markings. Unmarked parking areas pose a significant challenge in modelling due to the lack of detailed data on how vehicles utilize such spaces. For instance, in some cases, vehicles park at the side of the roads without obstructing traffic, while in other cases, parking spaces also function as driving lanes when not in use [Interviewee 3]. The latter situation is particularly important, as inadequate management and understanding of these spaces can reduce road capacity and contribute to congestion due to lane blockages. Also, Interviewee 4 expressed that 'parking occupancy' estimate could be derived by simply checking how many agents are at an activity in a TAZ and used their car to get there.

Additionally, Interviewee 5 pointed out that the relevance of certain levels may vary depending on the region modelled. For instance, the relevance of illegal parking data can vary by cities, as this depends on factors such as the size of fines and the intensity of enforcement. These factors, defined by the likelihood of being caught, determine whether specific data, like the number of illegal parking, is necessary. Thus, the decision to include such data should be based on whether it is relevant to the specific conditions and region being modelled [Interviewee 5]. However, these suggestions put forward by the interviewees could not be implemented by adjusting the levels, as doing so would impact the results of other modelers.

4.3. Conclusion on interview results for parking demand data

In comparing the five parking models, each exhibits different strengths and limitations in terms of the level of detailing in parking simulation and granularity of required input data. MATSim has advantages with its extensive literature, multiple use cases, and successful applications in various cities. However, despite MATSim being a powerful transport model with large scale applications, its ability to model parking is limited by the absence of capacity constraints in its default framework as the model does not have individual parking slots, but links. SUMO, while capable of generating high-detail simulations of parking aspects, is limited by the exclusion of parking costs, limiting its ability to model how changes in parking prices impact mode choice and agent behaviour. Nevertheless, SUMO's flexibility allows it to work effectively with varying levels of data granularity, offering advantages in certain modelling applications. For a model offering detailed simulations with less detailed data, SUMO proves useful.

PARKAGENT, in contrast, is one of the most advanced agent-based parking models, offering highly detailed simulations that are unmatched by other microsimulation models. While MATSim and SUMO claim to simulate parking at a microscopic resolution, accounting for all behavioural changes of agents, neither considers microscopic details like the potential collision between cars in parking lots or the possibility of a person to open the door and get out of the car [Interviewee 5]. However, PARKAEGENT has to make trade-offs between modelling detail and performance, particularly when it comes to linking parking outputs with broader traffic simulations, where MATSim and SUMO have an edge. Although TAPAS, another agent-based microsimulation model does not independently connect parking outputs to other traffic parameters such as travel times, its coupling with SUMO allows for the integration with travel time, mode choice, and destination choice. Finally, while VISUM, a macroscopic model, integrates parking through proxies and is useful for broader transport planning, lacks the granularity and precision in parking-specific outputs compared to the other models. However, this should not be viewed as a limitation, as VISUM is a macroscopic transport model and the data requirements and level of detail in outputs are dependent on the specific application.

Regarding data utilization, agent-based microsimulation modes such as MATSim, SUMO and PARKAGENT demands data at its highest possible resolution as they have the opinion that more the data, the better is the outcome. However, Interviewee 4 has the opinion that while more data can theoretically improve model precision, it also brings significant challenges in terms of data collection, computational requirements, and model explainability. For example, more data might lead to the development of more strategies in the simulations which may cause the agents to become stuck - an undesirable outcome, pointed out by the interviewee. Additionally, Interviewee 4 doubted whether such improvements would significantly enhance model accuracy or fall outside the intended scope. TAPAS, which focuses on advanced behavioural modelling, operates using data at abstract levels and simulate output at lower detailing levels, to balance data complexity and performance. On the other hand, studies using PARKAGENT and MATSim (especially, studies of MATSim in Zurich) rely on accurate data collected through extensive filed surveys, unlike other models that rely on less accurate data sources such as OSM or MiD.

A common limitation across all models is the challenge of balancing computational power with data availability. Additionally, there is a lack of a robust behavioural model that can simulate how individuals react to sudden changes such as changes in costs and space availability. While agent-based microsimulation models are capable of large-scale simulations, they often face computational difficulties, especially when applied to larger cities. For example, parking search modelling has rarely been attempted in large cities like Berlin; instead, most scenarios have been tested in smaller cities [Interviewee 3]. Interviewee 2 further emphasized that, complexity also occurs when new regions are added in the model or the size of TAZs is reduced, resulting in larger and intricate OD (Origin - Destination) matrices.

In conclusion, the interview findings indicates that transportation models have the potential to integrate more detailed and high-resolution information than they currently use. As widely known, models built with accurate, comprehensive data tend to yield more precise and reliable results compared to those based on assumptions or generalized data. Even though all gaps identified here are not solely caused by the absence of comprehensive parking data, the interviews underscored that there is a clear need for more complete and consistent data for parking modelling. For drawing more conclusions on the availability of data and its impacts on modeling, it is essential to examine the supply of parking data, which will be the focus of the next section.

4.4. Assessment of parking data supply

As part of the investigation into the current availability of parking data, interviews were conducted with representatives from the city geoportals of Berlin, Hamburg, Munich, and Zurich. Despite reaching out to other interview candidates, several either did not respond or declined to provide the requested information. Consequently, the state and comprehensiveness of such paid comprehensive data is unknown and thus not included in this section. This limitation highlights a significant barrier to understanding the comprehensiveness of parking data, as proprietary datasets remain behind paywalls, limiting their transparency and accessibility for scientific research.

The following section presents the findings gathered from interviews with representatives of cities (geoportal data), and web inquiry for publicly accessible data from commercial data providers for these cities. It is important to note that ‘accessible commercial data’ refers to data from providers like ‘Parkopedia’ or ‘APCOA’ where the data can be viewed, but it is unsure whether it can be downloaded and used by the public for research purposes. The information on the interview candidates can be found in Appendix D, while a comprehensive overview of data availability for each city, based on the data providers, can be found in Appendix E.

4.4.1. Munich

For Munich, the results shows that the city geoportal provides information across all dimensions except for ‘parking search time’ and ‘egress distance’. However, it’s important to note that the geoportal data has limitations: for ‘type of parking’, data is available only for on-street parking, and for ‘parking cost’, the information is limited to parking garages and other off-street parking. Munich has 'parking occupancy' data for 20 parking garages monitored through the 'Parkleit system'. These garages are publicly accessible and located within the inner city. However, there is no 'parking occupancy' data available for on-street parking and for the available ones, the current state of data is level 5 (average occupancy of individual facilities over each hour). Real-time occupancy tracking using a mobility data platform is also on their agenda for future implementation. On the other hand, the accessible commercial websites do not provide parking occupancy or search time. Although they claim to offer information on private parking, the data is far from comprehensive. However, the platforms can quickly display the egress distance from parking lots when a destination is specified.

(spider diagram is presented in the next page)

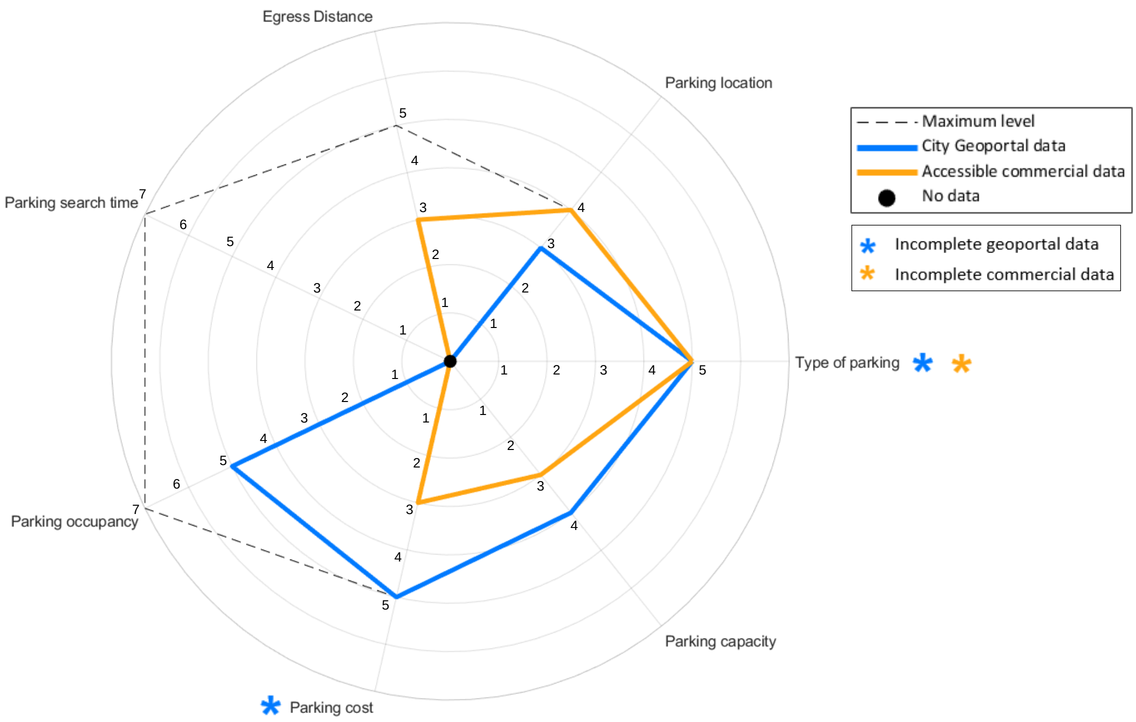


Figure 17. Parking data availability in Munich

4.4.2. Hamburg

For Hamburg, the spider diagram reveals that the city geoportal doesn't provide data for parking occupancy, search time and egress distance. While 'type of parking', data is available, the data is limited to only public and on-street parking. On the other hand, the accessible commercial data providers have almost all information for Hamburg except 'parking search time'. However, they don't have data on private parking and the 'parking occupancy' data is limited to certain garages.

(spider diagram is presented in the next page)

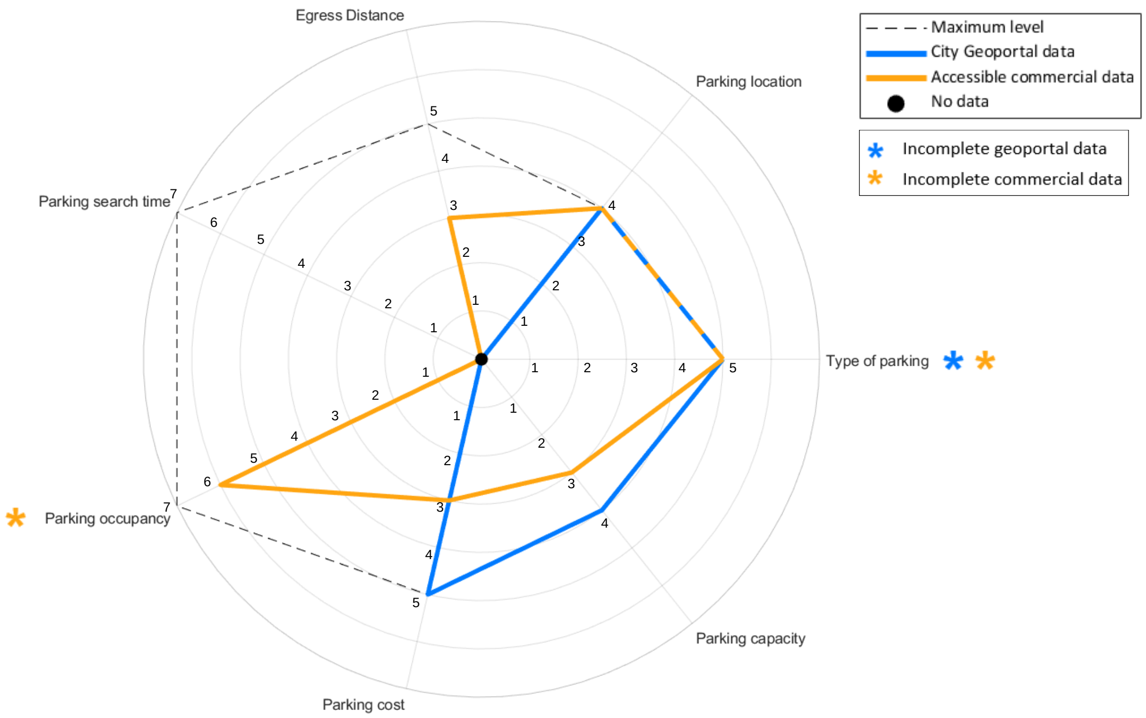


Figure 18. Parking data availability in Hamburg

4.4.3. Berlin

Berlin's geoportal offers relatively comprehensive data compared to other cities. The spider diagram shows that the geoportal provides data across most dimensions, except for 'egress distance'. For 'parking location' the city has mapped public on-street parking spaces with centimetre-precision with coordinates of each parking slot. The dataset also includes additional characteristics of each parking slot such as prices, restrictions, and geometry, and is considered highly reliable, with around 98-99% accuracy [Interviewee 6]. For 'type of parking', the data is limited to public on-street parking, with no data on private parking, particularly underground parking [Interviewee 6]. Also, the dataset does not include public off-street parking, as Berlin does not have publicly operated garages, and private parking data remains incomplete, largely due to the reluctance of private operators to share their information with the administration. For larger facilities, such as shopping malls, the administration has data on the number of parking spaces through building permits and other administrative processes, but this dataset is still far from comprehensive [Interviewee 6]. Research has been conducted on the number of illegally parked vehicles, but this is not done regularly as monitoring illegal parking on a consistent basis would require the regular use of scan cars, which is not feasible under the current legal framework.

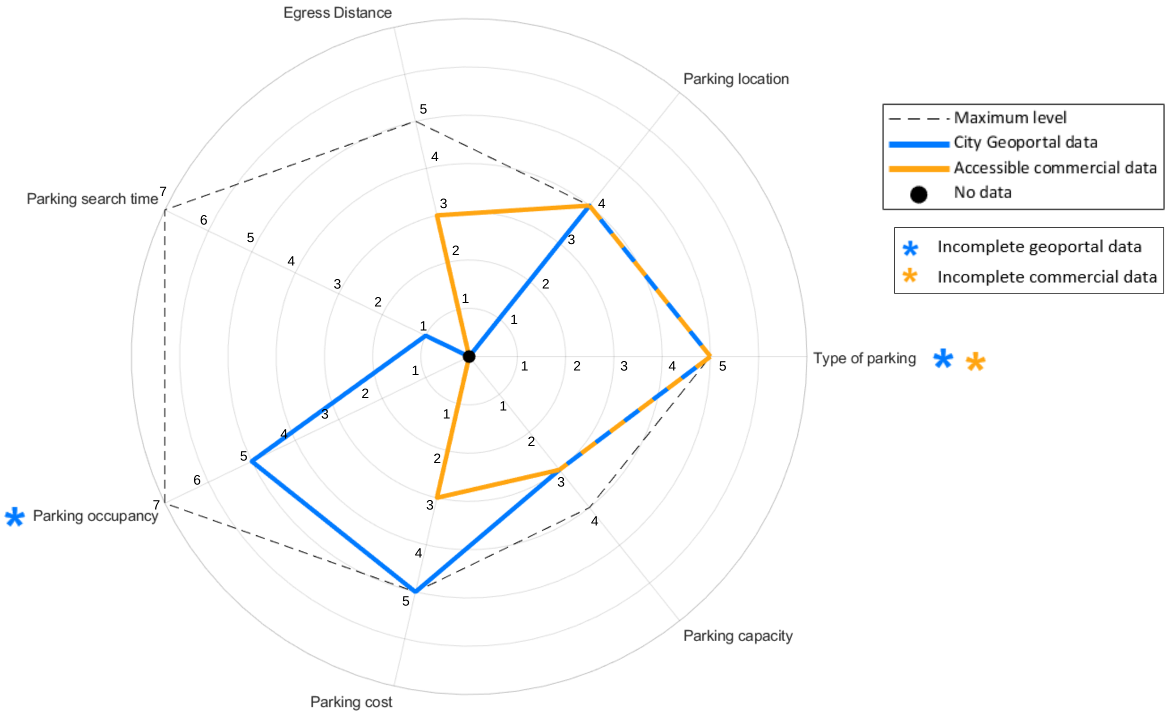


Figure 19. Parking data availability in Berlin

For 'parking cost', Berlin's dataset is quite comprehensive. The regulatory office reports detailed data on parking fines, and the city has exact figures for parking costs, including payments made through apps and parking machines. Subscription costs are uniform for all residents, which makes it easier to collect data. For 'parking occupancy', there has been an experiment conducted on the occupancy of every parking lot in a special area, with data collected over the course of one year, covering approximately 20 km² area. However, this dataset is limited in scope and does not provide full city coverage. Real-time parking occupancy data would require the use of scan cars or detectors at every parking lot, which, according to Interviewee 6, would be highly cost-intensive and may not offer sufficient benefits to justify the investment. For 'parking search time', while some examinations have been conducted within a parking zone, the results are basic and not yet applicable to different areas of the city. The available data has only been measured or calculated at an average level. According to Interviewee 6, obtaining more precise data would require real-time vehicle data, which is technically not feasible at present, as only newer cars transmit traffic data to providers. Therefore, parking search time can only be estimated based on a limited set of input data [Interviewee 6]. For 'egress distance', there is currently no available data. According to Interviewee 6, gathering accurate information on egress distances would require conducting interviews, but these often do not reflect reality and can be subjectively distorted. Additionally, precise monitoring of individuals would raise data protection concerns, making it challenging to collect reliable data for this dimension.

The interviewee also noted that one of the main limitations in fully integrating comprehensive parking data into their dataset is the lack of knowledge about the link between 'parking data' and 'mobility behaviour'. For example, the effects of a longer parking search distance or higher parking fees on the modal shift or commuting behaviour is unknown. This gap in knowledge also affects the integration of 'egress distance', as there has been little research into how mobility behaviour relates to parking in this context. When asked about future plans to improve the parking data in their dataset, the interviewee explained that there are considerations to merge 'parking occupancy' data with 'traffic data' in order to gain better insights into the duration of parking searches. However, the interviewee also noted that this integration poses significant technical challenges, and it remains uncertain whether it can be implemented.

The accessible commercial data, like that of other cities, lacks information on 'parking occupancy' and 'parking search time'. However, commercial platforms can quickly display egress distance from parking lots when a destination is specified, although, like other cities, they do not provide detailed data on private parking.

4.4.4. Cologne

For Cologne, data collection faced some limitations, as an interview with the relevant person from the city portal could not be arranged despite efforts. The available information from the city's website is provided as CSV or JSON files with basic details, and the platform lacks a map-based interface like other cities. Additionally, the data is somewhat outdated, dating back to November 2020. The accessible geoportal data offers limited information on 'parking location', 'type of parking', and 'parking capacity', with the 'capacity' data only covering public parking. The available commercial data is similar to that of other cities, with no information on private parking but detailed data on 'parking location', 'type of parking', 'parking capacity', 'parking cost', and 'egress distance'. However, they don’t provide data on 'parking occupancy' or 'parking search time'.

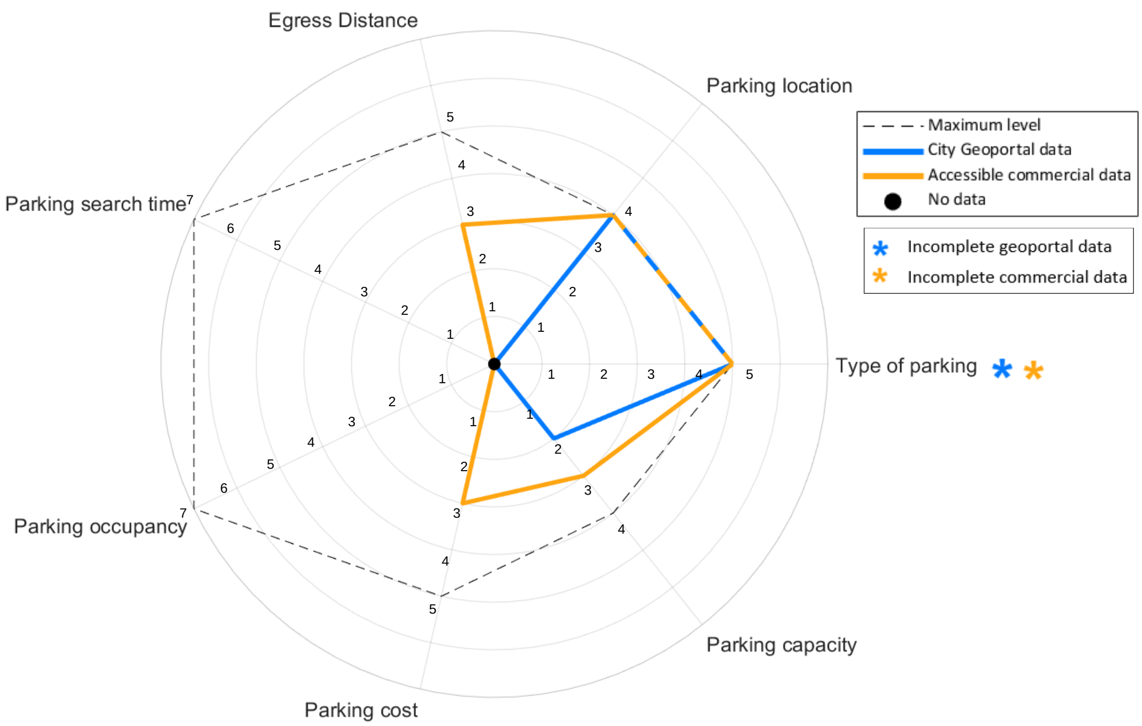


Figure 20. Parking data availablity in Cologne

4.4.5. Zurich

Zurich has relatively comprehensive parking data compared to other cities. Data for the geoportal is manually collected every two years through inspections, and the dataset is updated at the end of each year (Stadt Zürich). Additionally, publicly accessible parking spaces in multi-city car parks are offered in a separate dataset as open data (PLS Zurich).

The geoportal offers detailed data on 'parking location', 'type of parking', 'parking capacity', and 'parking cost'. However, according to Interviewee 9, there is no data on 'parking occupancy', 'parking search time', or 'egress distance', as this information is not considered as necessary for traffic management and planning purposes. For 'type of parking', the data is limited to publicly accessible street parking. Interviewee 9 noted that the reason for not reaching level 5 (including information on parking regulations) is that traffic regulations have not yet been digitized, on-street recording of signage has not yet taken place, and the evaluation and digitization of this data are still pending. Regarding 'parking capacity', the city has exact figures, but it does not track the number of illegally parked vehicles. Interviewee 9 explained that while the police issue fines, they do not conduct statistical surveys due to legal restrictions. As for 'parking occupancy', Interviewee 9 emphasized that the city does not require this data, and there are still technical challenges to addressing how such information could be managed and collected. Finally, Interviewee 9 mentioned that there are no plans to expand the parking dataset in the near future.

The accessible commercial data, similar to that of other cities, lacks information on 'parking search time'. However, Zurich has excellent data on the occupancy of parking garages through a public-private partnership of parking garage operators Parkleitsystem AG (PLS Zurich). This dataset includes real-time updates on the operating status of parking garages (e.g., malfunctions, closed), with information updated every minute. Additionally, commercial platforms can quickly display egress distance from parking lots when a destination is specified, though, like in other cities, they do not provide comprehensive data on private parking.

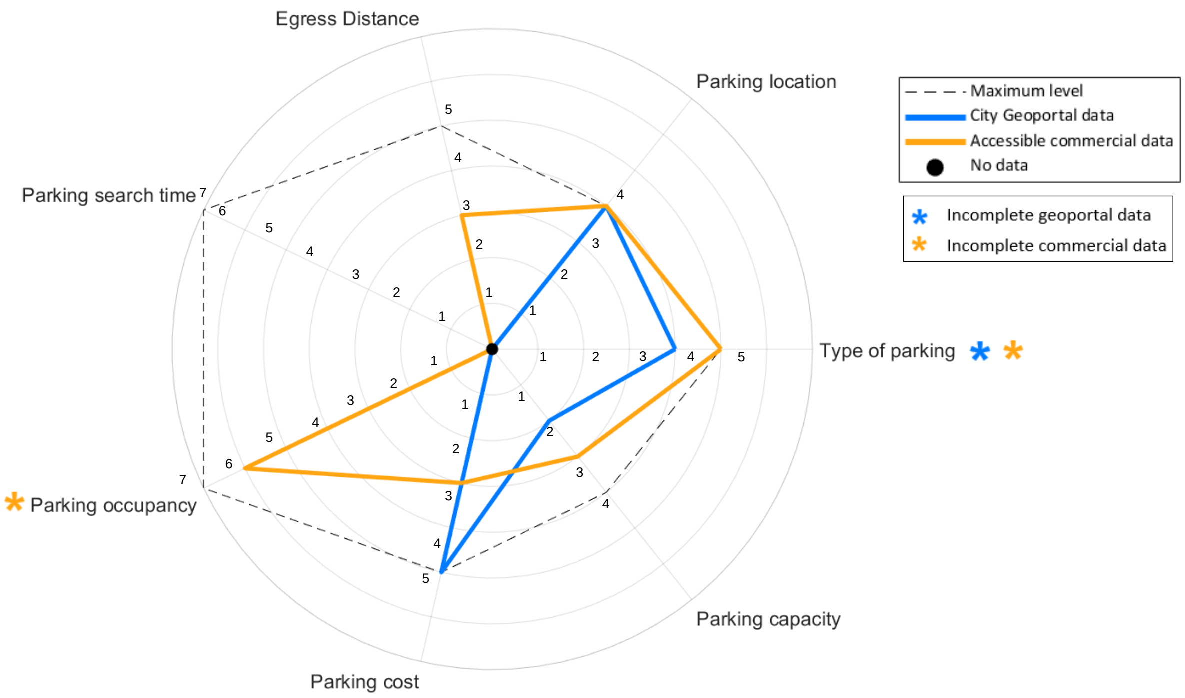


Figure 21. Parking data availability in Zurich

4.5. Comparative analysis of cities

When comparing the parking data across the five cities, Berlin and Zurich stand out as having the most comprehensive datasets. The geoportals of both cities have reliable data for public on-street parking, with data collection in Berlin still ongoing, even for parking search times. The dataset in Berlin claims an accuracy of at least 95 percent, with the most recent update provided in July 2023 (Senatsverwaltung für Mobilität, Verkehr, Klimaschutz und Umwelt, 2023). Munich also has relatively good parking data, with plans to implement a mobility data platform that can offer data such as real-time occupancy of monitored parking garages. Likewise, Hamburg is actively working to expand its parking data to enhance both coverage and accuracy. In contrast, data for Cologne was less accessible, as no interview could be conducted with the relevant representatives, and the city's geoportal only offers limited information on parking, leaving the city's dataset comparatively underexplored. However, neither geoportal has data on private parking, and while commercial providers claim to offer such data, it remains far from comprehensive. Importantly, paid commercial services may hold more extensive datasets, but due to their inaccessibility for this study, their actual coverage remains unknown.

Overall, the pattern of data provided by the geoportals and accessible commercial platforms are similar across these cities, the difference mainly reflected in the data provided by geoportals in parking occupancy. However, comparing cities based on the levels identified in the spider diagrams proves difficult as these levels do not capture the full scope of parking data across cities. For example, while Berlin, Munich, and Zurich appear to have similar parking occupancy data from the spider diagram, Berlin's experimental data covers only a small area (20 km²), Munich's data is limited to 20 parking garages, and Zurich's data includes all parking garages. This also highlights the inconsistency in data collection across cities and the challenges in forming a cohesive understanding of the current state of parking data.

The results from the interviews and web-based inquiry indicates that the cities considered in this study are actively working to further improve their datasets, using a wide range of advanced tools and technologies. It is important to note, however, that larger cities like Berlin and Zurich tend to have more comprehensive datasets due to their greater resources, while smaller cities may struggle with fewer financial resources and less guidance. In conclusion, while most cities appear to already have good data, there is still a clear need to enhance data comprehensiveness and sharing, especially regarding private parking and parking occupancy. Nevertheless, the continued improvement of parking datasets, especially in larger cities with more resources, suggests that the overall quality of parking data will keep improving in the coming years.

4.6. Comparing demand and supply of parking data

The spider diagrams for the five models and five cities are overlaid to identify the maximum parking data demand levels (shown in magenta) and the maximum supply data levels (shown in blue) in the figure below. It should be noted that comparing the desired data level and current modelling practice in the spider diagrams of models highlights the modelling gaps which do not necessarily reflect data gaps. However, the gaps formed by lack of data can be observed when comparing parking data demand with available supply data.

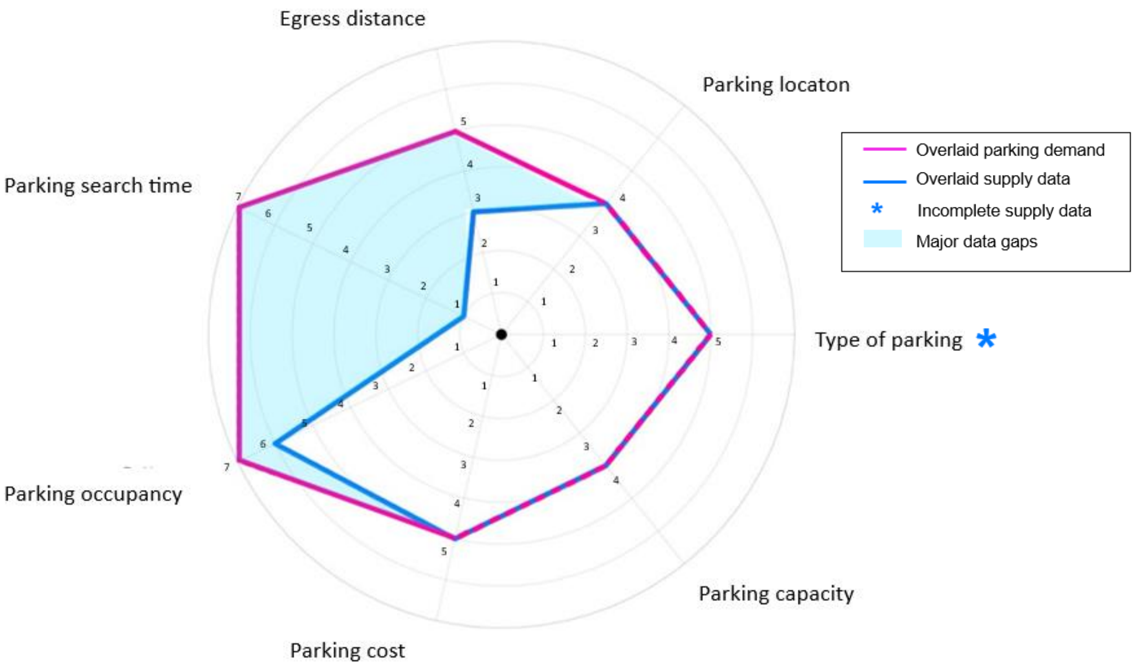


Figure 22. Overlay diagram of parking data demand and supply

As illustrated in the figure above, gaps are evident in the dimensions - parking occupancy, search time, and egress distance, indicating that no city currently provides sufficient data for these dimensions to meet the needs of modelers. What cannot be seen directly in the spider diagram is that data gaps also exist across all dimensions for private parking data (as marked by the blue asterisk). This is applicable to all dimensions as the supply data primarily covers public on-street parking, while comprehensive data on off-street private parking is unavailable from any source. Therefore, the data gaps identified in this study by comparing demand and supply include gaps in parking location, capacity and cost information for private parking as well as overall gaps in parking occupancy, search time and egress distance including data for private parking. This is just an overview and for anyone aiming to model parking for a city using a specific model, comparing the respective spider diagrams can provide valuable insights.

The spider diagram in the end, is just a tool for visualizing the information and only the main gaps are identified through the diagram. Beyond these visualized data gaps, there are deeper problems related to the heterogeneity, quality, and accessibility of parking data reported by the interviewee's (modelers). A common request from the modelers is to make the data open source, at least for smaller geographical regions, to allow experimentation with models to understand its capability in using more granular data. Interviewee 4 pointed out the high cost of commercial data, noting that many commercial companies possess valuable 'parking search time' data but restrict access to public research institutions, preferring to sell to other private companies instead. The interviewee suggested that even aggregated datasets for specific zones or cities, rather than the entire country, could be made available to the research community, without compromising the company's profitability. To add on this, Interviewee 3 pointed out that providing more data to researchers could potentially benefit the businesses as well, as more detailed modelling could increase demand for more specialized data, such as occupancy information. Interviewee 3 also expressed the need for location data, rather than more complex data like occupancy or search time, as commonly used sources like OSM suffer from issues of quality and inconsistency. Additionally, Interviewee 2 also mentioned that data availability remains a significant issue and noted the importance of assessing the actual quality of the data. While providers may claim the data is reliable, it would be beneficial for the modellers to know how good the data truly is, as the reality often differs once it is applied in practice.

The data suppliers have different concern from their side. A major issue raised by the data suppliers is the lack of detailed knowledge about the relationship between parking data and mobility behaviour. For instance, understanding the effects of longer parking search times or higher parking fees on commuting patterns or modal shifts remains unclear [Interviewee 6]. Interviewee 9 mentioned that their datasets do not include information on parking occupancy, search time, or egress distance because this data is not necessary for them for traffic management or urban planning purposes. Additionally, Interviewee 6 highlighted another challenge in parking data collection: ensuring user privacy, particularly when gathering data on parking occupancy. Real-time data collection, while beneficial, can raise privacy concerns. If it involves precise monitoring of individuals, it would be problematic from a data protection perspective [Interviewee 6]. Furthermore, implementing technologies like scan cars or real-time detectors for parking occupancy and search time is often costly and legally challenging [interviewee 6, 9].

Ultimately, the fragmented nature of parking data collection across cities, with varying assumptions, technologies, and update frequencies, creates significant challenges for cross-city comparisons. The results of this study clearly reveal the pressing need for a unified, more complete data source - one that not only covers public and private parking, but also adheres to standardized data collection methods, which are essential for effectively transferring modelling practices between cities.

5. Dataset introduction: addressing gaps with novel aerial image dataset

In the previous chapter, we identified the modelling gaps that arise mainly due to the lack of comprehensive parking data in transportation models. This chapter outlines a relatively unexplored data source that has the potential to fill the identified data gaps.

5.1. Potential data sources to fill the gaps

To address the identified data gaps, it is important to first understand the different data collection methods used to capture parking data at varying levels of detail. As discussed in section 2.3, common methods include road inspections, surveys, GPS and mobile phone data, mobile mapping using LiDAR etc. While these methods can provide useful data, they are often costly, time-consuming, and limited in scope. Additionally, none of these methods effectively capture off-street parking data on private ground. The absence of parking data in private areas has led many models to assume infinite capacity in those areas, which undermines the ability of such models to accurately reflect transportation phenomenon. Currently, no comprehensive dataset exists in Germany that includes off-street parking data for a larger area such as Berlin, making it impossible to determine whether these parking capacity assumptions result in underestimation or overestimation. Apart from the incompleteness and inconsistencies of data, heterogeneity of data across different regions is another issue. Varying assumptions, errors, and formats across different data collection methods complicate the transferability of models between cities. Therefore, there is a growing need for a more complete and standardized data that can address all these limitations.

As mentioned in section 2.3, researchers recently have pointed out the potential of remote sensing data to create parking constraints in transportation models. Remote sensing, particularly aerial imagery provides high-resolution spatial data across large areas, including private properties, offering data that are otherwise inaccessible by traditional methods. Aerial images can cover large areas quickly which gives consistent dataset during the time of flight. Furthermore, georeferencing of aerial images is much more precise than using imagery from moving cars (Hellekes et al., 2023) .Once captured, these datasets are often made publicly available, making them a valuable resource for long-term, transferable studies. As remote sensing imageries are becoming more and more available to the public and the algorithms for processing aerial images are getting advanced, there is a growing need to explore how this data can be utilized in other domains that require spatial data, such as transportation.

5.2. Traffic area segmentation on aerial imagery

One of the most effective ways to utilize aerial imagery for parking data extraction is through a process called 'traffic area segmentation'. This involves identifying and segmenting different traffic areas such as roads, parking areas, bikeways, and footways from high-resolution aerial images. Automated segmentation models allow the large-scale extraction of these features from aerial images, including areas on private property. However, this method comes with challenges, such as occlusions from buildings and trees, shadowing and lighting variations, and its inability to capture temporal changes in traffic conditions. Despite these limitations, the advantages it offers such as broad coverage and access to otherwise inaccessible areas make it a valuable data source that is worth exploring.

A notable example of traffic area segmentation performed on aerial imagery is the TIAS (Traffic Infrastructure and Surrounding) dataset, developed by DLR. TIAS is a novel dataset consisting of high-resolution aerial images with labels of traffic areas. This dataset accurately reflects traffic areas seen from an aerial perspective by providing detailed, fine-grained labels of relevant features enabling the reconstruction of traffic networks for motorized vehicles, bicycles, pedestrians, and rail traffic.

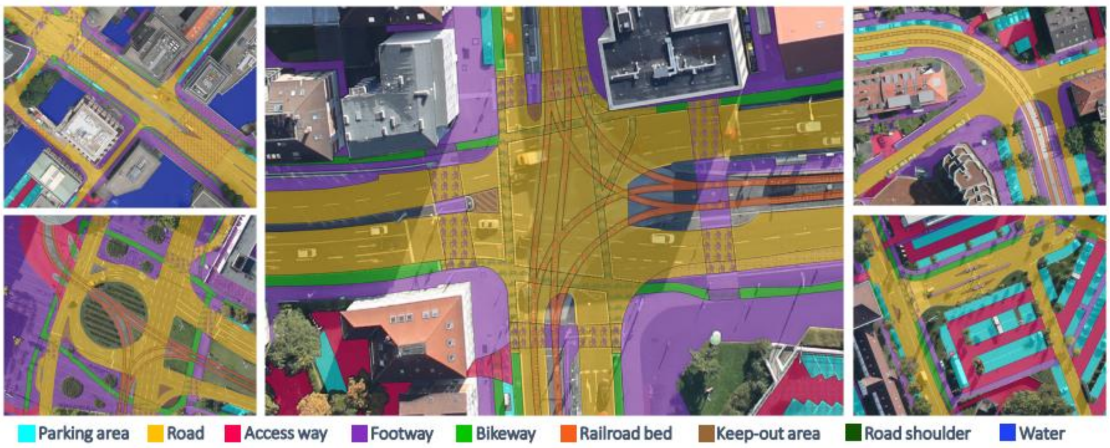


Figure 23. Traffic area classes in TIAS dataset (source: GeoDPA DLR, 2024)

The segmentation method applied in TIAS is semantic segmentation. Semantic segmentation is a computer vision technique used to classify each pixel in an image into a specific category or class. Unlike object detection, which identifies objects with bounding boxes, semantic segmentation provides pixel-level precision, meaning every pixel in the image is assigned a label corresponding to a predefined class.

For simple analysis, TIAS provides traffic area polygons in shapefile format, including parking areas, making it a valuable resource for addressing the previously identified gaps. The parking area polygons can be extracted and used for various analyses, such as determining parking capacities and type of parking based on access through geospatial filtering. This thesis explores the potential of TIAS dataset to serve as an additional data source for transportation models. The objective of the analysis is to create a statewide inventory for parking areas. For this, the analysis is conducted as a case study of Berlin, chosen due to the availability of complementary datasets that help to evaluate the performance of TIAS, and to compare and enrich the extracted data.

5.3. Data description: TIAS dataset

The input data received is a shapefile containing predictions of traffic areas from 2022, for Berlin and its surrounding areas (see figure 24). The shapefile includes polygons classified into three categories: road, access way and parking area.

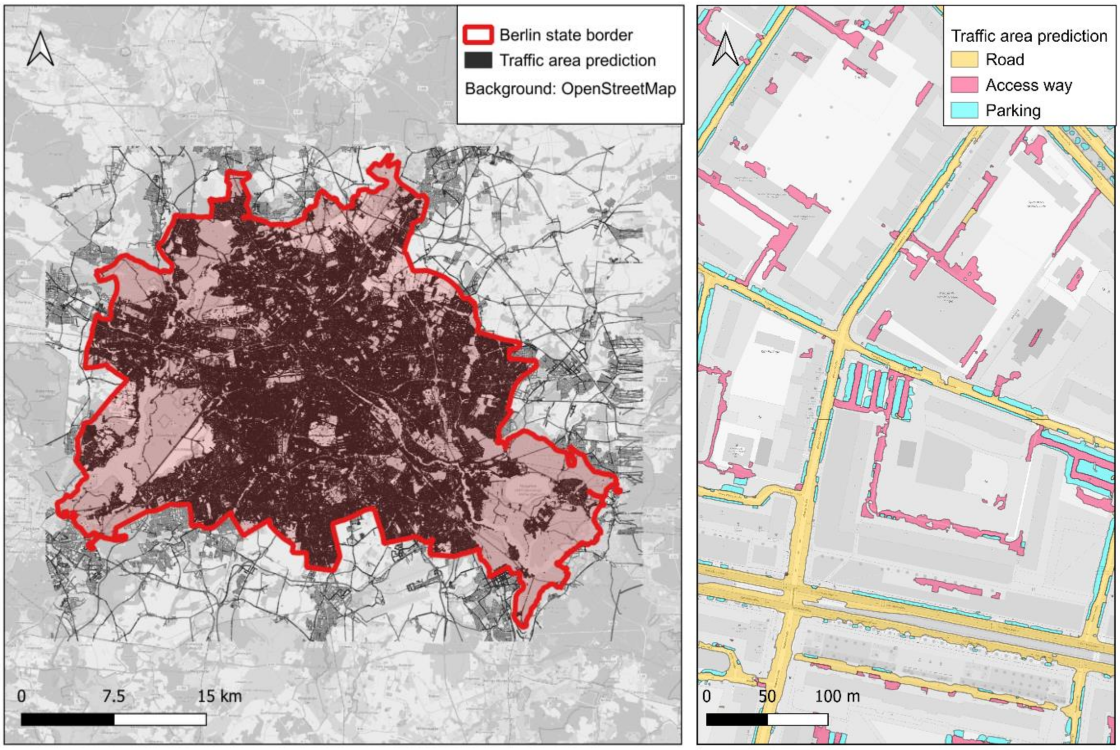
The definitions of the associated classes are as follows:

Road: Any public or private area where motorized vehicles may drive without special permission but only stop temporarily.

Access way: A special type of road that ensures access to parking areas, parking buildings, and other places without through traffic, e.g. industrial areas.

Parking area: Any public or private area dedicated to parking, where a motorized vehicle may stop for an indefinite or defined period.

The validation metrics for TIAS model predictions for each of these classes are outlined in Appendix G. As the focus of this thesis is on parking areas, only the polygons representing parking areas are extracted and used for the analysis. Although the predictions were received for a broader area extending beyond the boundaries of Berlin, the analysis is restricted to the area within Berlin's boundaries. The shapefile polygons are georeferenced and contain only basic classification information, specifying whether a given polygon represents a road, access way, or parking area. No further attributes, such as parking capacities or access information are included in the dataset.



a) Map showing extend of traffic area prediction around Berlin

b) Zoomed-in map showing the 3 distinct classes in traffic area prediction

Figure 24. Map showing the spatial extend of predictions recieved and the respective classes (own illustration)

6. Methodology to extract parking data from aerial image dataset

6.1. Identifying extractable data from the aerial image dataset

Among the seven identified parking dimensions, three - namely parking location, type of parking and parking capacity can be extracted directly from the TIAS dataset. The dataset allows for the extraction of multiple levels of these dimensions owing levels can be extracted. For parking location, point level, link level, zonal level and city level can be extracted. However, as point level and link level are the most demanded and used levels by the modelers, this analysis focuses on extracting information at those levels of the dimension parking location. Similarly for parking capacity, both exact numbers of parking slots and the number of illegal parking can be extracted, and these are included in the analysis. Other dimensions, such as parking occupancy, search time, and egress distance, require temporal data that the TIAS dataset does not provide, as it captures only static imagery. Additionally, parking cost cannot be extracted from this dataset and requires other sources or methods for its determination. All the preprocessing and geospatial filtering for data extraction in this study are performed in QGIS.

This chapter is further divided into two sections: the methodology for extracting the type of parking based on access and the methodology for extracting information on parking location and capacity.

6.2. Extraction of type of parking based on access

As discussed in section 3.1, this study classifies parking mainly into two categories based on access: on-street and off-street parking. The definitions of these categories are as follows:

On street parking: These are parking spaces that are located directly on the streets, usually along the curb. Examples include the parking parallel to the road such as the parking bays along roads as well as cars parked in footways. Essentially, all parking found along a city street is classified as an on-street parking.

Off-Street Parking: These are in designated areas away from the main road, such as parking lots, garages, or private driveways and they are usually separated from the street infrastructure. The semi-public parking which are located on private ground, but accessible for general public such as parking lots near a shopping area/ touristic spot also falls under this category. To determine if a parking area is categorized as on-street or off-street, it is necessary to identify whether the parking is located on public or private ground, distinguished based on land ownership. The property dataset is exported as WFS layer in QGIS, and on-street parking is identified by overlaying the parking predictions onto the public property layer in QGIS, categorizing all predictions within these areas as on-street parking. A similar process is applied for identifying off-street parking, using the private property layer for the overlay.

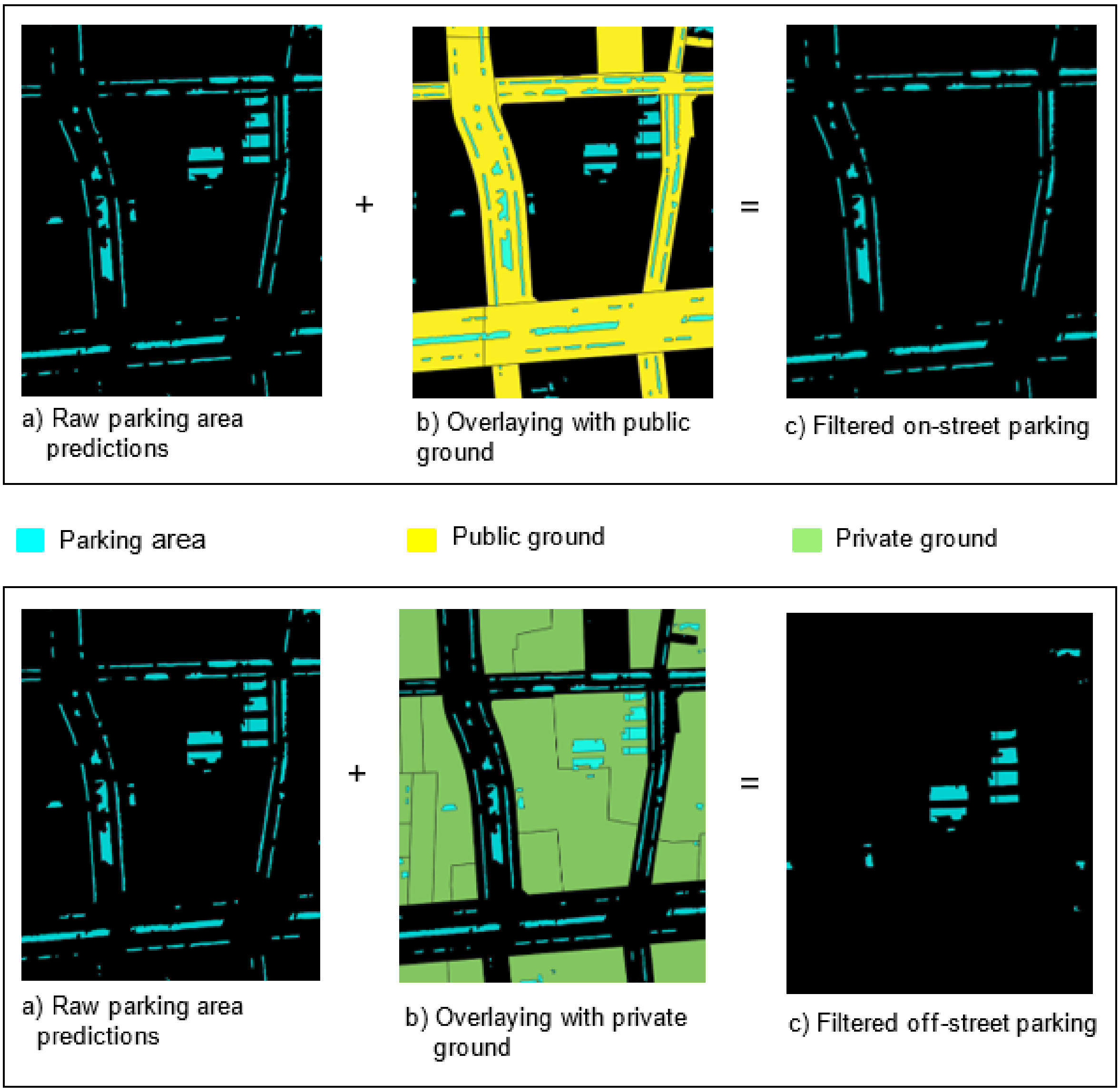


Figure 25. Method for classifying parking area predictions into on-street and off-street parking (own illustration)

The dataset used for land ownership classification (public/private) in this study is sourced from the (Berlin Geoportal [FIS Broker]) (Amtliches Liegenschaftskatasterinformationssystem) and will henceforth be referred to as the cadastral dataset. The cadastral dataset consists of 24 classes based on land use type (see figure 26). Out of these, 3 classes – 'road traffic' (Straßenverkehr), 'plaza' (Platz) and 'path' (Weg) are considered as public ground as they are accessible to the general public. The remaining 21 classes including 'industrial and commercial area' (Industrie Und Gewerbeflaeche) and 'residential area' (Wohnbauflaeche) are considered as private ground as they are not generally accessible by the public. The respective classes are exported as separate layers of public and private ground in QGIS to allow the parking area predictions to be overlaid for the analysis.

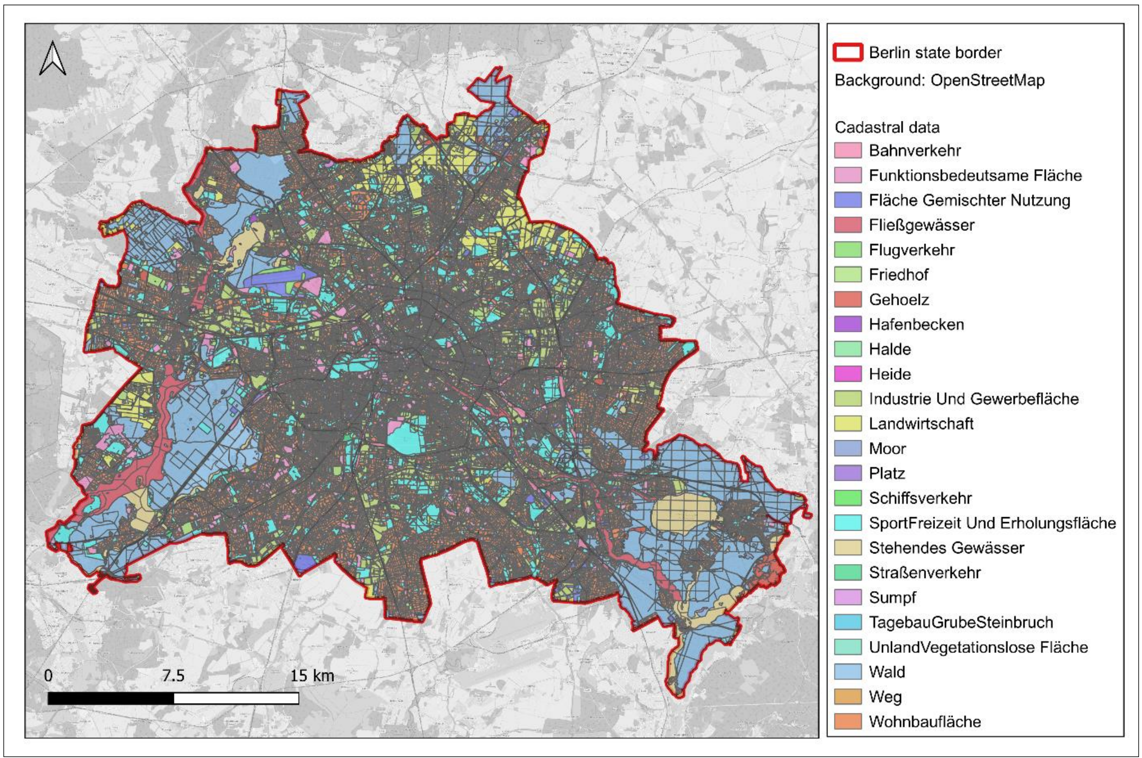


Figure 26. Cadastral dataset with land ownership classification in Berlin (own illustration)

6.3. Extraction of information on parking capacity and location from predictions

The methodology used here for estimating parking capacity builds on the approach suggested by Hellekes et al., 2023. The parking capacity, or number of parking slots, can be easily estimated from the area of each parking polygon (which represents a parking area that may contain multiple parking slots), by dividing it with the area of an individual parking slot.

However, the prediction polygons are irregular and dispersed, largely due to the limitations of semantic segmentation methods. Only objects visible in the aerial imagery are labelled, and assumptions about hidden objects are generally avoided. As a results, occlusions caused by elements like trees and buildings can lead to parts of the parking areas being missed, as pixels representing parking spaces under occlusions are not classified. Therefore, the polygons in the predictions do not always reflect the true dimensions of parking areas.

To assess the impacts of these limitations, it is necessary to compare the predictions against a reliable and verified reference dataset, with parking capacity as the primary comparison metric. This comparison helps determine whether the predictions overestimate or underestimate the parking areas. Additionally, the reference dataset will be utilized to develop certain metrics that is used to correct the inaccuracies in the predictions. The reference dataset, sourced from the FIS Broker (Straßenparkplätze innerhalb des Berliner S-Bahnringes), claims 95% accuracy regarding public street parking areas in Berlin (source: "Kartierung von Straßenparkplätzen innerhalb des S-Bahn-Rings," 2023). It should be noted that this dataset contains information only on public street parking within the Berlin S-Bahn ring and adjacent areas (see figure 27 to know the spatial extend of this dataset). It provides details on parking types following a different classification scheme, orientation, and capacity, among other attributes.

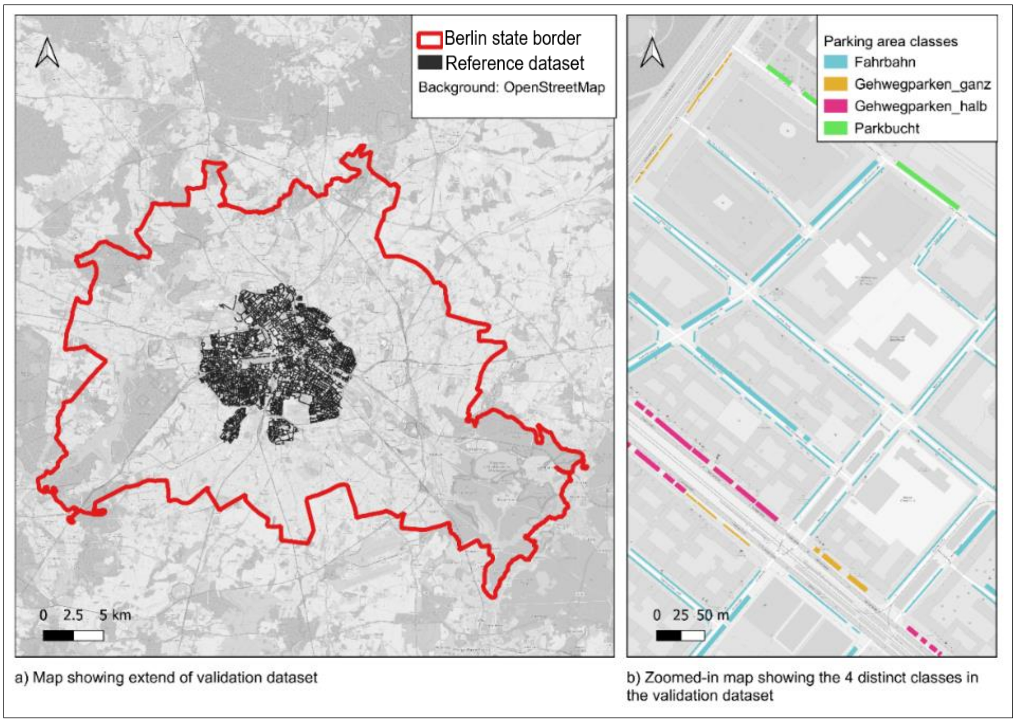


Figure 27. Spatial extend and classes in reference dataset (own illustration)

The flowchart below illustrates the approach used to extract parking capacity from the predictions utilizing the reference dataset. A detailed explanation of each step follows.

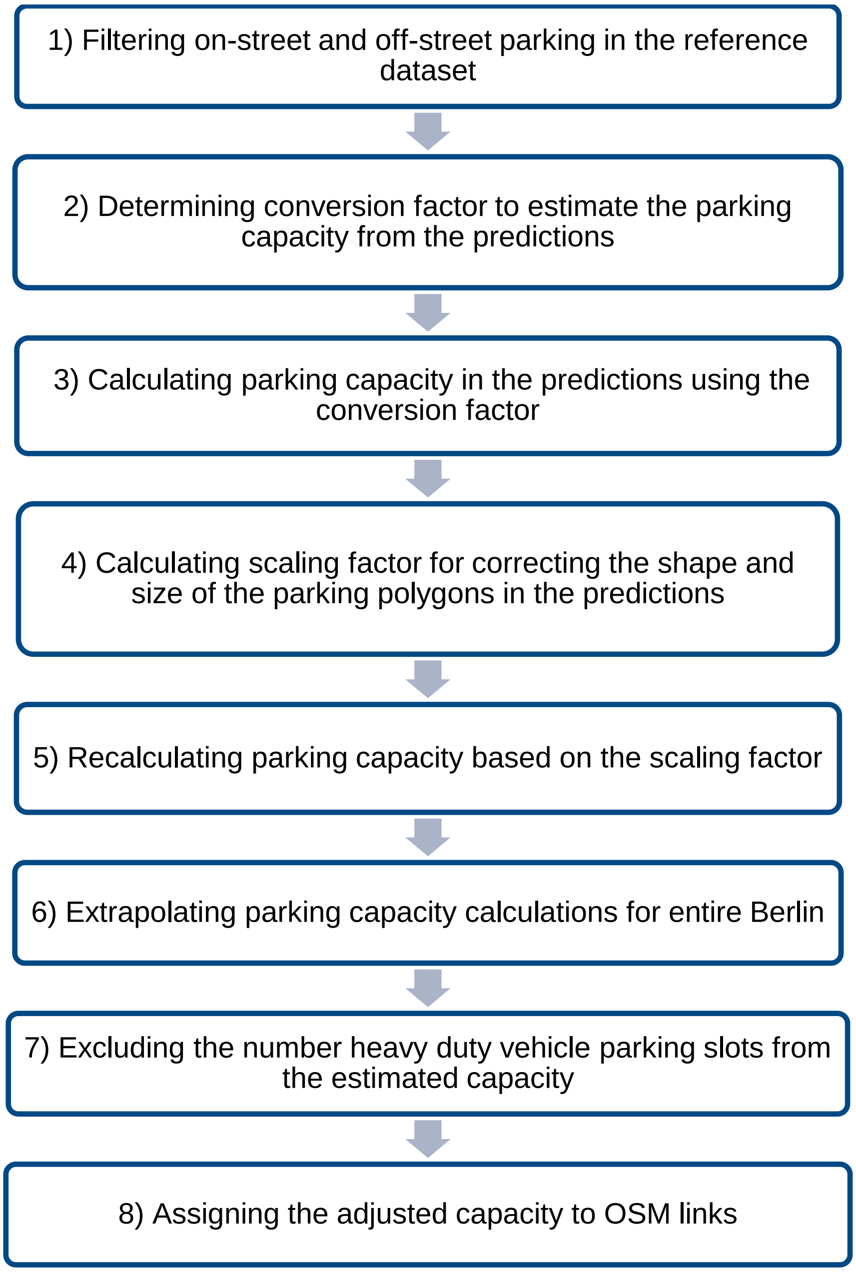


Figure 28. Flowchart showing the methodology for extracting parking capacity data

1) Filtering reference dataset

Based on access, four types of parking areas are identified in the reference dataset- parking bay (Parkbucht), half footway parking (Gehwegparken halb), full footway parking (Gehwegparken ganz), and road (Fahrbahn). While all the parking areas in the reference dataset seems to be on-street parking, a filtering approach to distinguish between on-street and off-street parking similar to the one applied to the predictions is necessary here as well. To ensure a fair comparison, the reference dataset is also exported as a WFS layer in QGIS and parking polygons are filtered using cadastral data.

2) Determining conversion factor for predicted parking areas

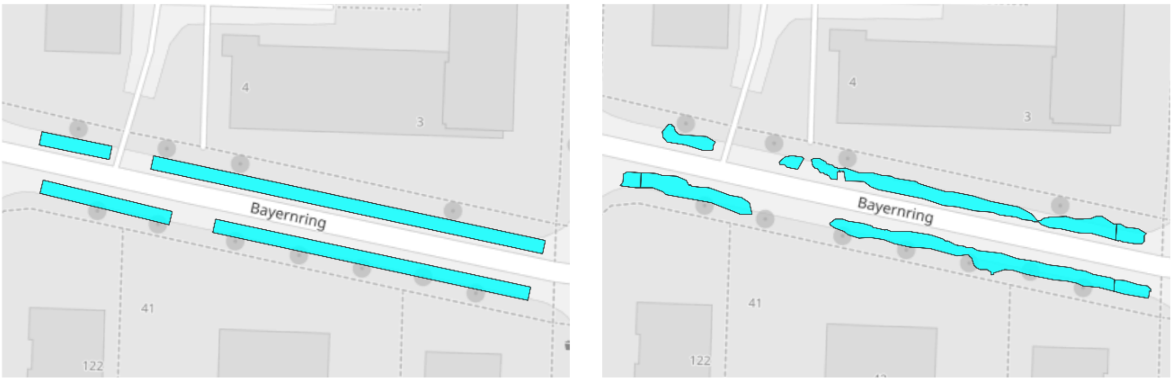
As previously mentioned, parking capacity is calculated by dividing the parking area (m2) in the prediction by the area of a single parking slot (m2). To avoid errors associated with aggregating the data (i.e., summing the total area of parking for the region of interest and then dividing by the conversion factor), each parking area polygon is individually divided by the divisor. This divisor representing the area of a single parking slot, is referred to as 'conversion factor'. Since there are three main types of parking orientations (perpendicular, parallel and diagonal), each with different dimensions, a weighted average of their areas is used to calculate the conversion factor.

3) Calculating parking capacity in the predictions using the conversion factor

The predictions are cropped to the boundaries of the reference dataset and each parking area polygon in the prediction is divided with the calculated conversion factor to obtain the number of parking slots within the area defined by the boundaries of the reference dataset.

4) Calculating scaling factor for correcting the shape and size of predictions

While the conversion factor is accurate for the right area, applying it directly to the predictions is unsuitable as the predicted parking areas do not accurately reflect the true shape or size (see figure 29). To correct this area loss in the predicted parking polygons caused by the limitations of segmentation model, a scaling factor is introduced.



a) Shape and size of a parking area polygon in the reference dataset

a) Shape and size of a parking area polygon in the prediction

Figure 29. Comparing shapes of parking polygons in reference dataset and prediction

$$\text{Scaling factor} = \frac{Total\ area\ of\ reference\ parking\ polygons}{Total\ area\ of\ predicted\ parking\ polygons}$$

The scaling factor is calculated by dividing the total area of reference parking polygons by the total area of predicted parking polygons, effectively calibrating the predictions to provide a rough estimate of how much area was missed by the segmentation model. Since the reference data is only available for on-street parking, the scaling factor can only be calculated for on-street parking. However, it is assumed that this factor can be also applied to off-street parking to measure how much off-street parking is realistically captured by the segmentation model. Therefore, the scaling factor is used to correct the underestimation in on-street parking and to provide a more realistic estimate for off-street parking predictions.

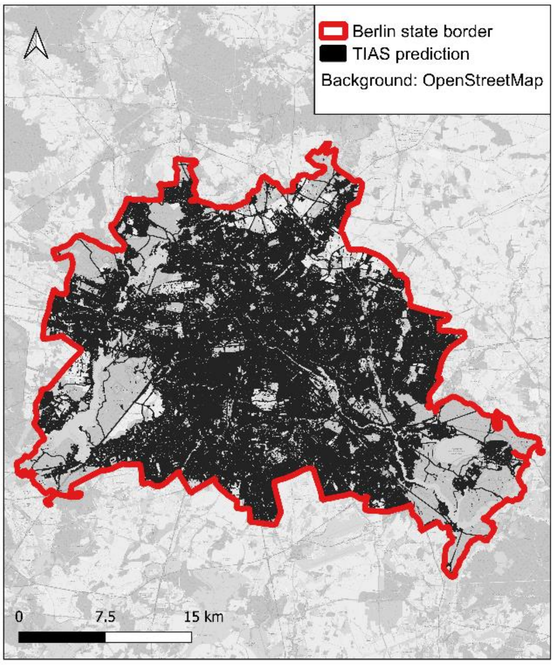
5) Recalculating parking capacity based on the scaling factor

Each parking area polygon in the prediction, both on-street and off-street, is multiplied by the scaling factor to adjust its size to reflect the true size of polygons. This is done to enlarge the size of individual polygons in the prediction to account for the true size of the polygons that was missed due to the challenges posed by the segmentation model and remote sensing approach. The adjusted area is then divided by the conversion factor (the weighted average area of an individual slot) to calculate the corrected parking capacity.

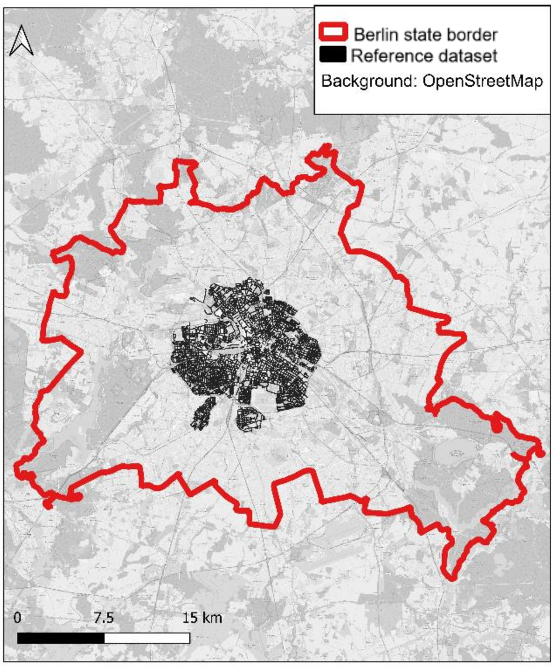
6) Extrapolating parking capacity calculations for entire Berlin

As shown in the figure 30, the reference dataset covers only the inner parts of Berlin, and both the scaling factor and conversion factor were determined based on this smaller area. It is important to note that the conversion factor may differ for areas outside the inner parts of Berlin due to variations in share of different parking orientations (perpendicular, parallel and diagonal). However, for the purposes of this analysis, it is assumed that the share of parking orientations is sufficiently similar across the city, allowing the same conversion factor to be applied throughout Berlin.

Although the scaling factor was derived from the inner part of Berlin, its purpose is to address some general limitations of the segmentation model. While this scaling factor may not fully apply to the outer areas of Berlin, it is used here to estimate parking capacity for the entire state. Although occlusion patterns may differ between urban and suburban regions due to variations in land use, building density, and tree cover, the overall parking infrastructure remains similar, making this assumption reasonable. Even though this is a coarse approach, it provides a reasonable estimate for the missing parts of parking polygons in the predictions. Based on these assumptions and by repeating the same steps, the prediction for the entire Berlin is converted into number of parking slots, using the same conversion and scaling factors.



a) Map showing parking area predictions cropped to the size of Berlin



b) Map showing the spatial extend of reference datset

Figure 30. Maps comparing the spatial extend of prediction and reference dataset (own illustration)

7) Excluding the number of heavy-duty vehicle parking slots from the estimated capacity

The segmentation model classifies all parking areas without differentiating between heavy-duty vehicles (HDVs) and light-duty vehicles (LDVs). However, since the models considered in this thesis are more concerned about car parking than HDV parking, it is necessary to differentiate between LDV and HDV parking.

DLR has a vehicle detection and classification dataset for the whole Berlin, which includes bounding boxes around vehicles and information about their orientation, as well as the vehicle classification into LDVs and HDVs (see figure 31). This dataset can be used to filter and separate parking polygons for LDVs and HDVs. The dataset also contains confidence score assigned to each detected vehicle (see figure 34) and a minor filtering is required to eliminate vehicle detections with lower confidence score. Once these detections are eliminated, the vehicle detection and classification dataset is overlaid onto the parking predictions to identify which parking spaces are primarily used by passenger cars (LDVs) and which are used by HDVs. However, there is a limitation: if a dedicated HDV parking space was unoccupied at the time the aerial image was captured, it may not be classified as HDV parking. This introduces some uncertainty into the method, but at the very least, it allows for the filtering of spaces currently being used by trucks. The vehicle detection data is captured around the same time as the segmentation predictions, making the comparison fair.

8) Assigning the adjusted capacity to OSM links

Once the LDV capacity is calculated, this data can be assigned to the corresponding OSM links in QGIS. By attributing the parking capacity to these links, it becomes possible to visualize and analyse the state-wide spatial distribution of parking spaces. This step is helpful for identifying the patterns in parking availability and how it is related to the land use types in urban settlement.

Figure 31. Example of vehicle classification in the dataset (HDVs and LDVs)

6.4. Estimation of illegal on-street parking

Illegal parking cannot be directly estimated from the traffic area segmentation dataset, as it is often difficult to determine whether a vehicle is parked legally or illegally, solely from the aerial view. However, the reference dataset (used in section 6.3) only maps on-street parking spaces where parking or stopping is legal, including areas with time restrictions. Therefore, parking areas present in the predictions but absent in the reference dataset can serve as an indicator of illegal parking. It is important to note that the aerial image dataset provides information only on the number of illegal parking visible at the time the aerial image was captured.

Although this method is limited to the inner parts of Berlin, where the reference dataset is available, it provides an estimate of the distribution of illegal parking across various parking zones within the city. By overlaying the reference dataset onto the predictions, the parking areas that appear only in the predictions can be categorized as illegal parking. Similar to how parking capacity is calculated using conversion factor and scaling factor, the number of illegal parking can be also estimated. While the data is not fully comprehensive due to the limited coverage of the reference dataset, and limitations of the segmentation model, it offers a reasonable approximation of illegal parking in the inner Berlin.

7. Results of data extraction from aerial image dataset

This section presents the results of the geospatial calculations discussed in the previous chapter. All the maps used in this section are own illustrations generated in QGIS. A summary of the key findings can be found in the concluding section of this chapter.

Visualization of data

Due to the small size and dispersed nature of parking polygons across the state of Berlin, creating a comprehensive visualization of the parking data presented in this section for the whole of Berlin is challenging. Therefore, Zone 92, a newly established parking zone located in the Tempelhof-Schöneberg district, is selected for visualization (see figure 32). This zone, established in September 2023, is observed and covered more comprehensively in the 'reference dataset' than other zones. Additionally, its location, neither too close to the city centre nor near the state boundary makes it a balanced and representative area for analysing the parking distribution.

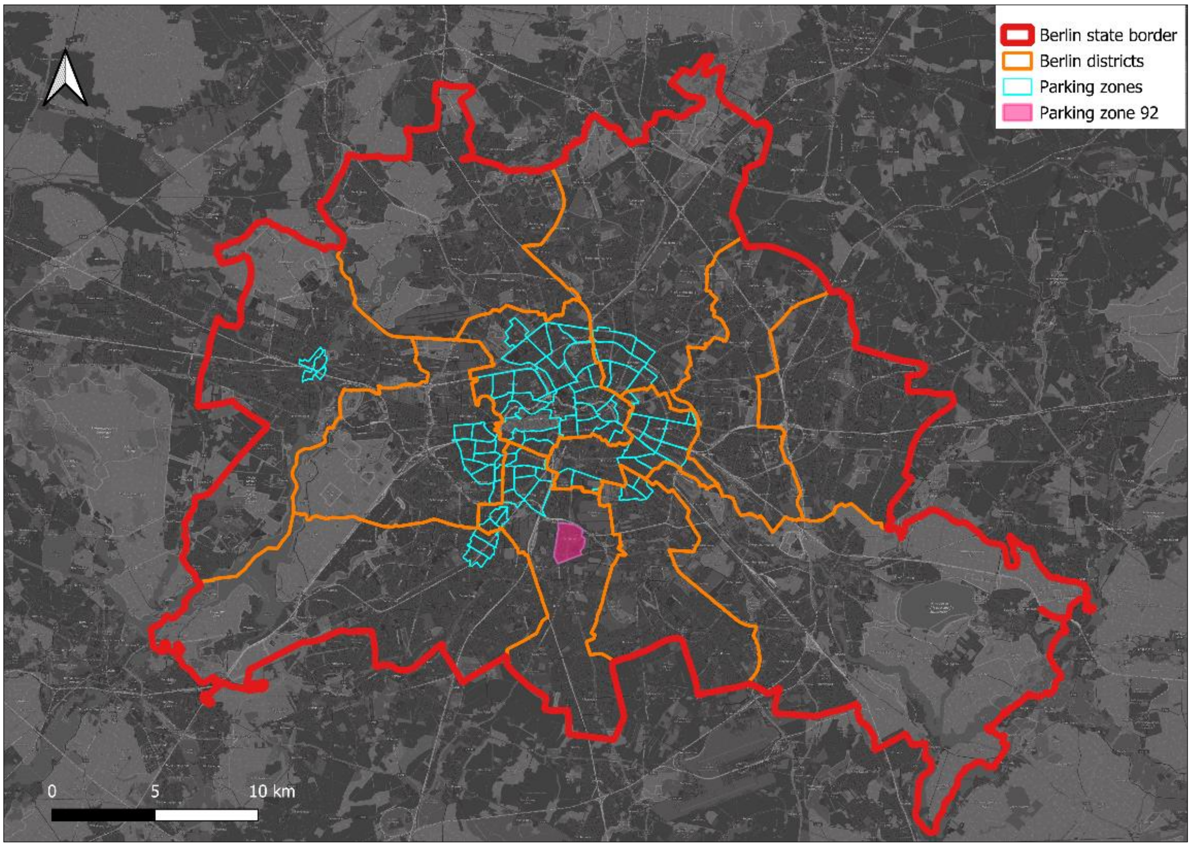
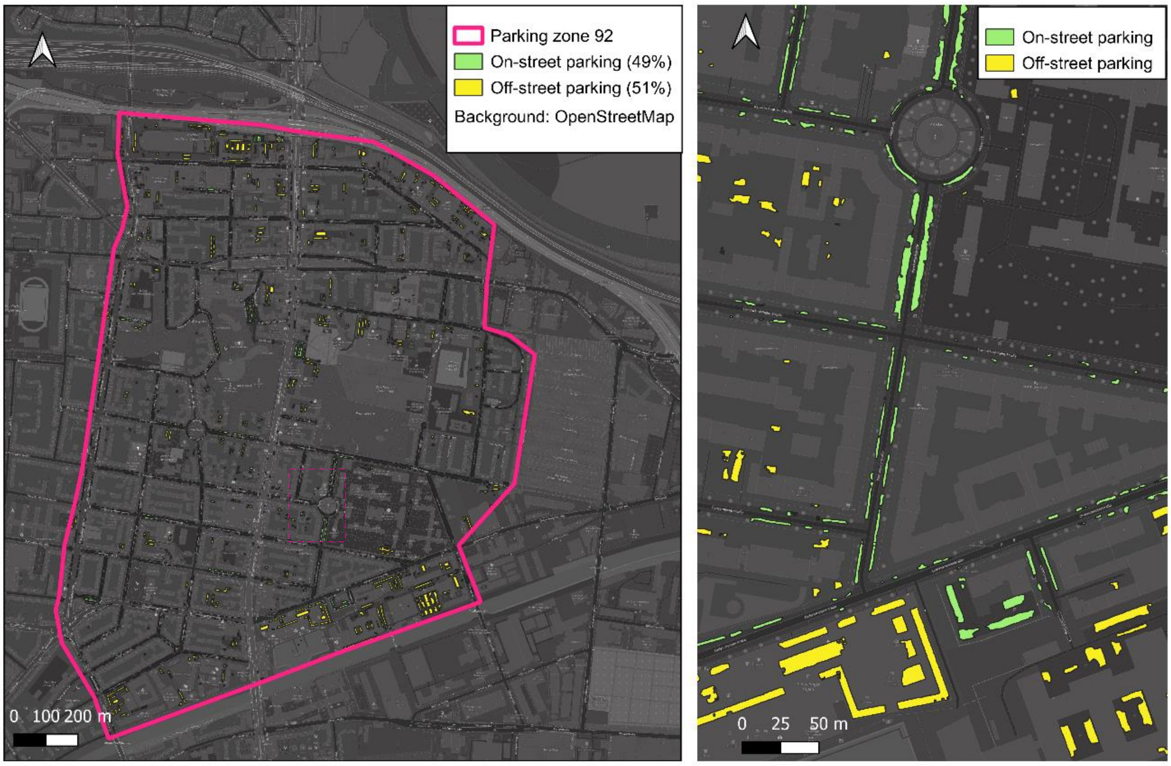


Figure 32. Map of Berlin state boundary, districts and parking zones with highlighted parking zone 92

7.1. Distribution of on street and off-street parking in Berlin

By overlaying parking predictions onto the public property layer in QGIS, on-street parking is identified. Similarly, overlaying parking predictions onto private property layer provides off-street parking. It is observed that 55.45% of the total parking area from the prediction in Berlin is on-street parking, while 44.54% is off-street parking, when considering parking area (m^2) as the comparison metric. The exact number of LDVs and HDVs along with the distribution of on street and off-street parking in Berlin across various districts is also presented in the upcoming sections. Within the representative parking zone used for visualization (zone 92), the distribution of predicted parking areas is relatively balanced with 49% of the identified parking areas classified as on-street parking and 51% as off-street parking. As there are no other reliable and comprehensive data sources on off-street parking, comparing the accuracy of these on-street vs off-street parking shares is not possible. However, the classified on-street parking will be compared and assessed with a reference dataset in the next section, with parking capacity as the comparison metric.



a) Map showing on-street and off-street parking area predictions in parking zone 92

b) Zoomed-in map showing on-street and off-street parking area predictions in parking zone 92

Figure 33. Distribution of on-street and off-street parking in zone 92

7.2. Parking capacity distribution in Berlin

1) Filtering reference dataset

A filtering approach to distinguish between on street and off-street parking, similar to the one applied to the predictions is done with the reference dataset. As expected, the filtering process confirmed that almost all the parking spaces in the reference dataset were on-street, with only a small percentage (0.26 %) classified as off-street parking.

2) Determining conversion factor for predicted parking areas

The conversion factor, which is the divisor used to calculate the parking capacity from the area of predicted parking polygons is calculated based on the area of a single slot from the reference dataset. Since the area of a single parking slot varies depending on its orientation to the street (perpendicular, parallel or diagonal), a weighted average of the slot areas is used as the conversion factor. The weight considered here is the share of each parking orientation type in relative to the total number of slots (e.g., how many parallel slots there are relative to the total number of parking slots).

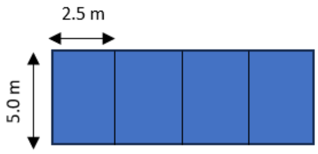
To know the area of a slot for each parking orientation type, the dimensions of the various parking geometries are needed. The dimensions, sourced from the documentation of the reference dataset *(“Kartierung von Straßenparkplätzen innerhalb des S-Bahn-Rings,” 2023)* are as follows:

Perpendicular parking:

Length = 5 m

Width 2.5 m

Area = 12.5 m^2



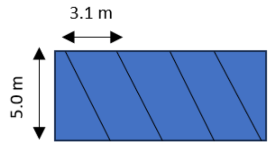
Parallel parking:

Length = 5.2 m

Width = 2 m

Area = 10.4 m^2

Diagonal parking:



Length = 5 m

Width = 3.1 m

Area = 15.5 m^2

Table 3. Share of perpendicular, parallel and diagonal parking in reference dataset

	Perpendicular parking (1)	Parallel parking (2)	Diagonal parking (3)
Capacity (C)	81116	187831	21930
Proportion (P)	0.278	0.645	0.075
Area of individual slots in m2 (a)	12.5	10.4	15.5

Weighted average = (P1*a1) + (P2*a2) + (P3*a3)
= (0.278*12.5) + (0.645*10.4) + (0.075*15.5)
=11.345
~**11.35 m^2**

The conversion factor of 11.35 is now used to calculate the capacity of individual polygons in the parking area predictions. Even though the conversion factor is based on the reference dataset which only consists of on-street parking, the same factor is assumed to off-street parking too as explained in the methodology section.

3) Calculating parking capacity in the predictions using the conversion factor

The conversion factor is applied to the predictions cropped to the boundaries of the reference dataset. After dividing each parking area polygon in the prediction with the calculated conversion factor of 13.5, the following capacity values were distinguished by on street and off-street parking. The capacity values for the reference dataset were already in the exported layer of reference dataset.

Table 4. Comparing parking capacity in the prediction and reference dataset

	Number of on-street parking slots	**Number of off-street parking slots**	**Total number of parking slots**
Prediction (1)	240015	94582	334597
Reference (2)	290120	757	290877
Difference between predicted and reference capacity (1) - (2)	-50105	+ 93825	+43720

These calculations show that the predicted on-street parking is 17.2% less than what is recorded in the reference dataset. However, the predicted off-street parking is significantly higher than the reference dataset, as this dataset focuses on public street parking, making the comparison less relevant for off-street parking. At first glance, the predictions seem to underestimate the total capacity. This underestimation is primarily due to the general limitations of segmentation model which was trained on aerial images where occlusions reduce the surface area of visible parking slots or as the model might be estimating more background. However, it is important to note that conclusions cannot be drawn solely from the capacity figures. Even if the model predicted a higher number of parking locations than the reference dataset, this difference would not be reflected in the comparison if only capacity, derived from predicted surface area is used as the assessment factor.

4) Calculating scaling factor for correcting the shape and size of predictions

Scaling factor is applied to correct the underestimation caused by the limitations of the segmentation model as a possible solution. The calculations of scaling factor are as follows:

Total area of on-street parking in prediction (A1) = 2738027.18 m^2

Total area of on-street parking in reference (A2) = 3311708.37 m^2

Scaling factor, n = $^{A2}/_{A1}$

$$= \frac{3311708.37}{2738027.18}$$

$$= 1.209$$

$$\sim \mathbf{1.21}$$

5) Recalculating parking capacity based on the scaling factor

This scaling factor is then applied to each predicted polygon area to enlarge the size of individual polygons in the prediction to account for the true size of the polygons that was missed. The scaled area is subsequently divided by the conversion factor (the weighted average area of an individual slot) of 11.35 to determine the adjusted parking capacity.

Table 5. Comparing parking capacity in scaled prediction and reference for the reference area

	Number of on-street parking	**Number of off-street parking**	**Total number of parking**
Prediction	240015	94582	334597
Scaled prediction	290849	114529	405378
Reference	290120	757	290877

The number of on-street parking in the scaled prediction and reference is nearly identical as the scaling factor was calibrated using the total area of parking in the reference dataset. Although the calculations of off-street parking are based on certain assumptions such as applying the same conversion and scaling factor as applied for on-street parking due to the lack of reference data for off-street parking, this approach provides a more realistic and accurate estimate of the off-street parking capacity in Berlin. Remarkably, this is the first time off-street parking capacity has been estimated on a statewide scale. The scaled prediction has approximately 21% more parking capacity than the initial unscaled predictions indicating the extent to which parking capacity is underestimated by the segmentation model.

6) Extrapolating parking capacity calculations for entire Berlin

Assuming the same conversion factor and scaling factor estimated for the small region is applicable to the whole state, and by repeating the same steps, the capacity values are scaled for the entire Berlin (the prediction for the entire Berlin is converted into number of parking slots).

Table 6. Comparing prediction and scaled prediction capacity values for the whole Berlin

	Number of on-street parking	**Number of off-street parking**	**Total number of parking**
Prediction	606069	486866	1092935
Scaled prediction	734249	590102	1324351

Just as the scaling factor resulted in a 21% increase in parking capacity than the raw capacity calculations for the smaller reference area, the increase in capacity for the whole Berlin after applying the same scaling factor is also ~ 21%. The traffic area segmentation dataset indicates that there are a total of 1,324,351 (1.32 million) parking slots in Berlin, with **55.42%** being on-street parking and **44.55%** off-street parking. The share of on-street parking and off-street parking is somewhat similar in Berlin and this distribution suggests that modelling approaches that either exclude off-street parking or rely on assumptions are potentially overlooking nearly half of the available parking in Berlin.

7) Excluding the number of HDV parking slots from the estimated capacity

The vehicle detection and classification dataset used in this analysis provides polygons for LDV and HDVs, each assigned a confidence score by the detection algorithm (see figure 34). The confidence scores represent how certain the model is about each detection. These scores range from 0 to 1, where scores closer to 1 indicate higher confidence, while scores closer to 0 indicate lower confidence. As there are vehicles detected with low confidence scores, it is necessary to exclude lower confidence detections. The elimination is done by choosing a threshold value and excluding all the detections with confidence scores below this threshold. To determine the threshold, we need to understand the distribution of the confidence scores for the vehicles.

Maximum confidence score: 0.999863

Minimum confidence score: 0.0500005

Average confidence score: 0.307

A threshold confidence score of 0.4, a value closer to the average was chosen, ensuring that only more reliable data is retained. This threshold balances the need to filter out less accurate detections while considering the average confidence score of the dataset. After applying this filter, approximately 63% of the HDV detections were eliminated.

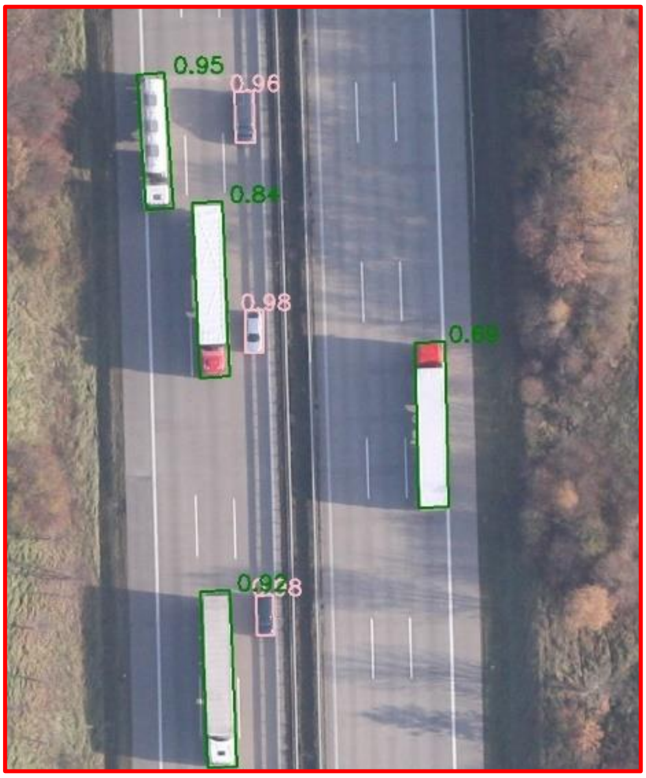


Figure 34. Vehicle detection and classification dataset with confidence scores assigned to each vehicle

The filtered HDV layer was overlaid with the predictions to identify parking slots currently occupied by HDVs. The remaining parking slots in the predictions, after excluding HDVs, represent LDV (primarily passenger cars) parking slots. Since each polygon in the vehicle detection dataset represents one vehicle, the number of polygons in this dataset corresponds to the number of vehicles.

It's important to note that the vehicle dataset only indicates the presence of a vehicle, not whether the vehicle is parked in a designated parking spot. Thus, if there is a dedicated HDV parking, but there was no vehicle detected at that time by the vehicle detection dataset, the HDV parking slot will be misclassified as an LDV parking slot. Additionally, some HDVs temporarily stopped in LDV parking slots may be considered as HDV parking and might be excluded through this filtering approach. To address this issue, only polygons larger than 15.5 m^2 were filtered out to represent HDV parking. The value 15.5 m^2 is chosen because this is the largest size of individual parking area for a car parking (diagonal parking). Furthermore, rather than excluding entire parking polygons in the predictions that may have mixed usage, the predicted polygons were split using QGIS tools to remove the HDV section of the polygon while retaining the remaining area for LDV parking calculations.

After this adjustment, the LDV parking capacity was recalculated using the same conversion factor identified in step 2, accounting for the changes in areas of polygons. The following table shows the distribution of HDV and LDV parking in Berlin. This approach is coarse given the assumptions and limitations, but at least it can exclude the HDV parking spaces which are currently used giving an estimate of the LDV (mostly passenger cars) parking in the prediction.

Table 7. HDV and LDV parking capacity distribution

	Number of on-street parking	**Number of off-street parking**	**Total number of parking**
LDV parking	715404	576949	1292353
HDV parking	2622	10627	13249

It was observed that 31,998 parking slots (2.42%) were removed from the predictions after filtering out HDV slots, indicating that **2.42%** of the total predicted parking slots were occupied by HDVs.

8) Assigning the adjusted capacity to OSM links

The adjusted LDV capacity for the whole Berlin is assigned to the corresponding OSM links in QGIS. Roads classified under the OSM key 'highway', with values ‘primary’, ‘secondary’, ‘tertiary’, ‘residential’, ‘living street’, ‘service’, ‘pedestrian’, and ‘unclassified’ were exported as the ‘road layer,’ and the filtered parking polygons in prediction are assigned to the nearest link.

Figure 35 shows the parking density map attributed to OSM links in zone 92. This type of visualization can be extended to the entire study area, offering a comprehensive view of parking availability across different roads. Such visualizations can also help assess whether parking allocation is equitable and how it varies with land use types, socio-economic characteristics of residents and such other factors.

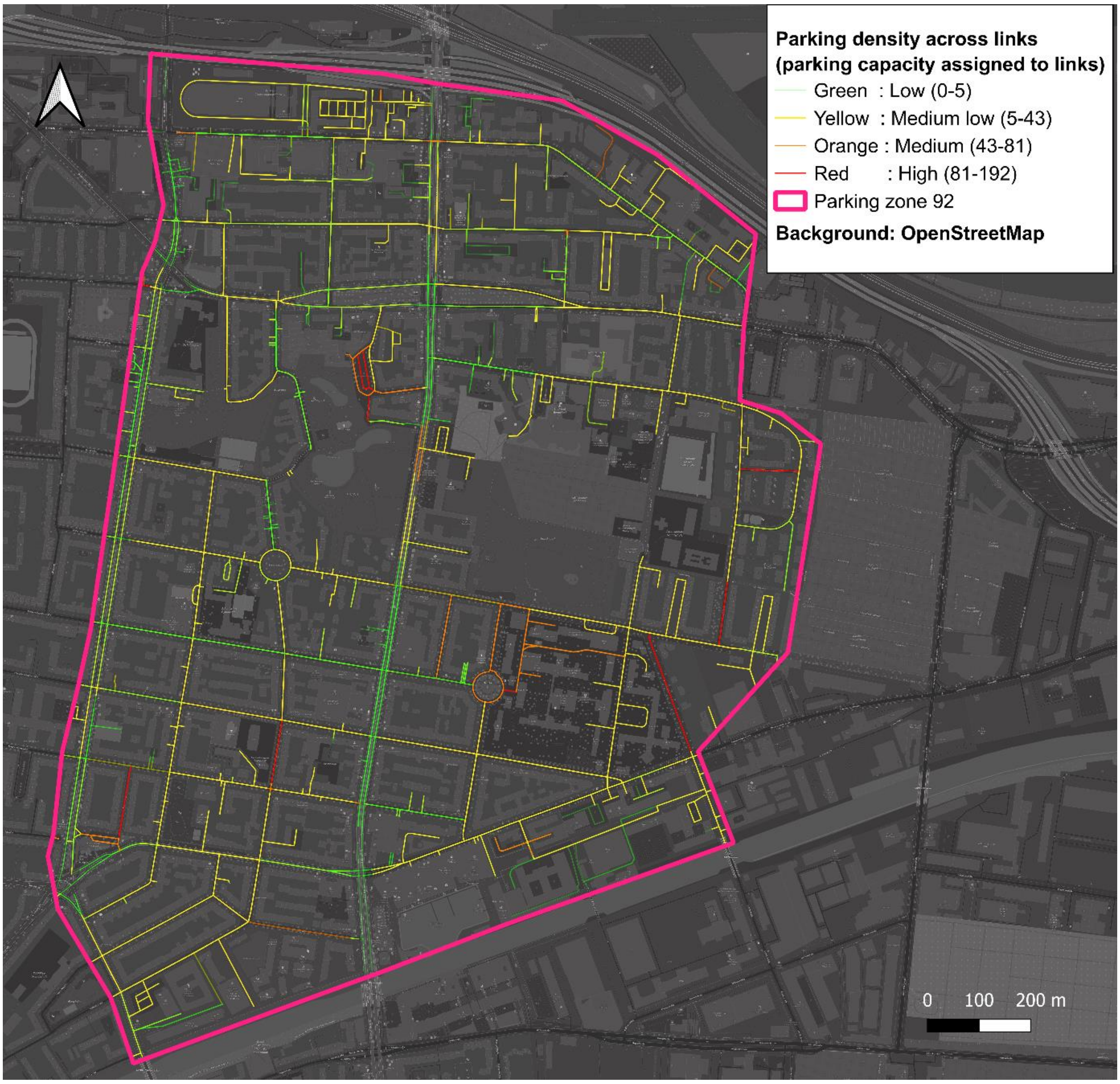


Figure 35. Parking density across OSM links in parking zone 92

7.3. Estimation of illegal on-street parking

By overlaying the prediction cropped to the size of reference dataset, the parking areas that appear only in the predictions can be categorized as illegal parking. Similar to how parking capacity is calculated using conversion factor and scaling factor, the number of illegal parking is also estimated for the reference area. It should be noted that only the information on on-street illegal parking can be obtained in this way as the reference dataset is only for on-street illegal parking. The total number of on-street illegal parking identified for the reference area is 15522 which means that 15522 vehicles were parked illegally at the time of capturing the aerial images in the region of reference dataset.

For clearer visualization and comparisons, the reference area is further cropped to the boundaries of parking zones present in the reference area. The following map (figure 36) shows the distribution of illegal parking across different parking zones in the inner parts of Berlin. The parking zones located outside the S-Bahn ring were excluded, as the reference dataset does not cover those areas. The number of illegal parking in the area cropped to the boundaries of parking zones in Berlin is 9,746. These parking's are aggregated and assigned to the corresponding parking zones.

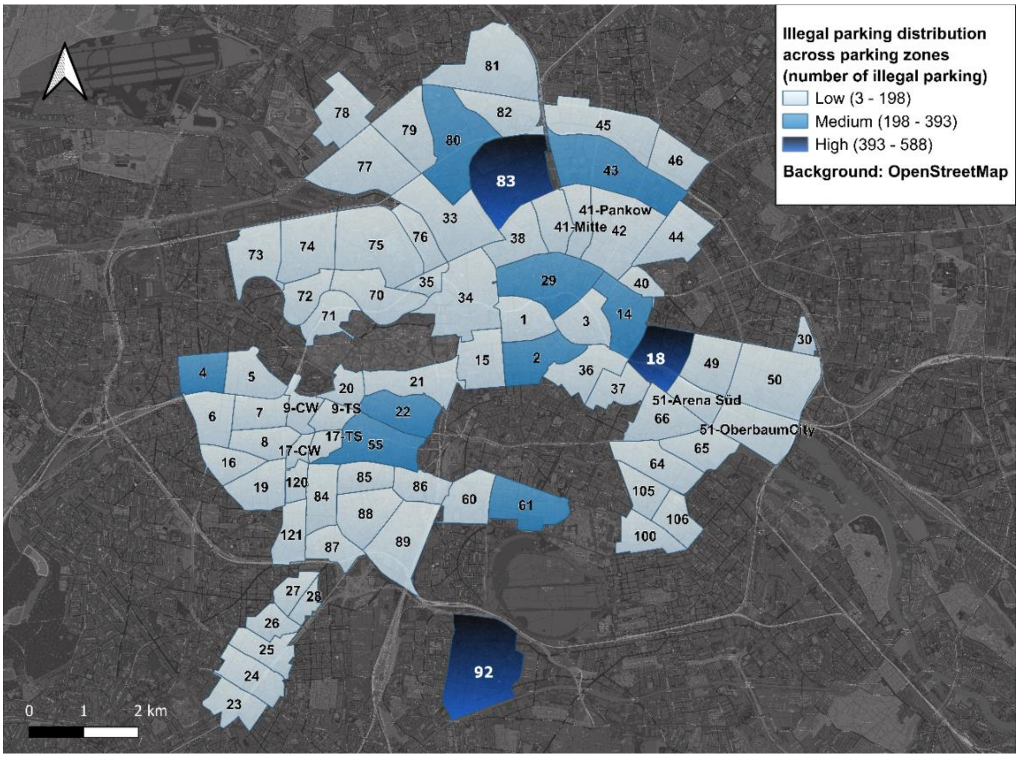


Figure 36. Illegal parking distribution across parking zones in the inner parts of Berlin

As it can be seen, the highest number of illegal parking are concentrated in Zone 92, Zone 83 and Zone 18. These zones are likely experiencing higher illegal parking due to their proximity to busy areas with limited parking availability or strict parking regulations. Additionally, Zone 92 may see higher illegal parking due to its recent establishment as a parking zone, potentially leading to lower compliance rates. To reach more comprehensive conclusions, further information is needed on factors such as the amount of fines and the intensity of enforcement, which influences the likelihood of being caught.

7.4. Summary of results

The traffic area segmentation dataset estimated 1.32 million parking slots in the state of Berlin, with 55.42% on-street and 44.55% off-street parking. After filtering out the currently occupied HDV parking spaces, the total parking capacity was reduced by 2.42%, resulting in 1.29 million parking slots for light-duty vehicles (LDVs). In addition, 15,552 illegal parking spots were identified in the reference area, with the highest concentrations in Zones 83, 92, and 18.

An initial comparison of the raw predictions against the reference dataset revealed that predicted parking underestimates on-street parking capacity by 17.2%. This underestimation can occur due to the limitations of the segmentation model (e.g. occlusion by trees) which makes the predicted polygon sizes smaller than their original size. To address this underestimation, a scaling factor of 1.2 was applied to enlarge the predicted parking polygons, increasing the estimated capacity by 21% (including both on-street and off-street) in the scaled prediction compared to the unscaled prediction.

7.4.1. Parking availability across the districts of Berlin

The map below (figure 37) illustrates the parking capacity distribution across Berlin's districts, with colour shading indicating the total capacity (low, medium, and high) and pie charts showing the share of on-street versus off-street parking. For specific capacity of each district, refer to the Appendix F.

It can be observed from the map that the districts Pankow, Treptow-Köpenick, Marzahn-Hellersdorf and Templehof-Schöneberg has the highest capacity share among other districts. As majority parts of these districts are situated outside the S-Bahn ring, they likely have lower public transport service quality, further increasing dependence on private vehicles and driving up the demand for parking spaces. In contrast, districts such as Friedrichshain-Kreuzberg and parts of Mitte exhibit the lowest parking capacity, which could be attributed to their central locations and higher population densities. These areas tend to benefit from better public transportation networks and the scarcity of space and higher real estate values in the city centre may result in fewer allocated parking slots.

To draw more conclusions about parking capacity distribution, further data is required, including socioeconomic characteristics of residents, land use patterns, vehicle ownership rates, employment trends, and public transportation access across districts. Socioeconomic factors, such as income levels and employment types, likely play a role in parking demand. For instance, higher-income districts may have more off-street parking options, including private garages, whereas lower-income areas might depend more on on-street parking. Furthermore, districts with larger residential or commercial areas may allocate more space for parking, while central business districts or densely developed areas might prioritize other infrastructure, leaving limited room for parking.

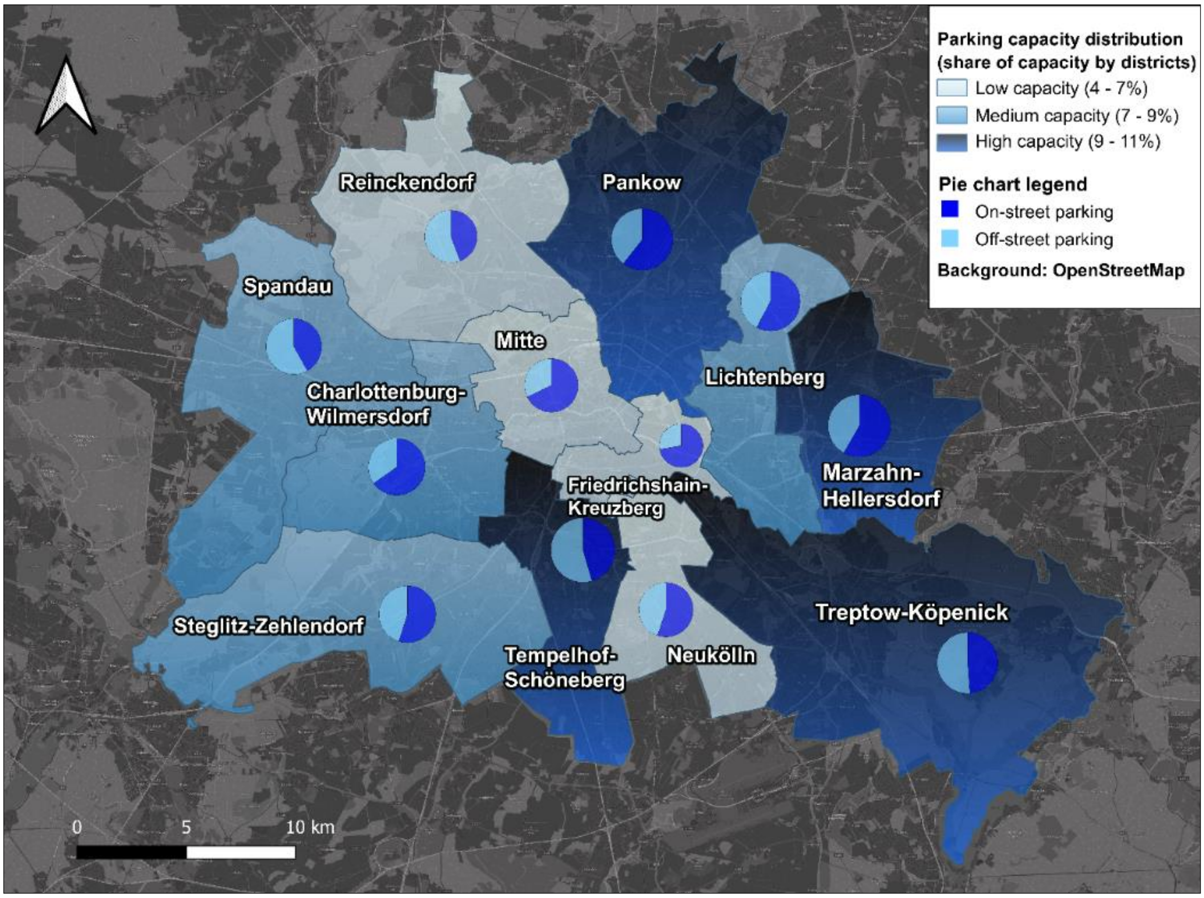


Figure 37. Parking capacity distribution across Berlin distinguished by on-street and off-street parking

7.4.2. Scope and limitations of the approach

The results demonstrate that excluding off-street parking in modelling approaches significantly underestimates the total available parking supply. In almost all districts, the share of on-street versus off-street parking is relatively balanced. This indicates that models which assume infinite or negligible private parking are missing nearly half of the actual parking capacity.

This approach represents one of the first attempts to evaluate off-street parking on such a large scale for an entire city. However, it has some limitations. Notably, it cannot capture multi-story parking garages or underground parking facilities, both of which are common in Berlin. Many parking garages in the city, especially in urban areas, are underground and therefore cannot be estimated by this method. The predicted parking polygons are smaller than their actual size due to occlusions caused by vegetation, shadows, and viewing angles, the limitations inherent in the semantic segmentation model. Despite these limitations, this approach still provides a much more accurate estimate of parking availability than traditional methods that often exclude off-street parking entirely.

8. Conclusions

8.1. Main findings and contributions

Parking is a highly complex and heterogeneous transportation phenomenon, making it challenging to fully understand and manage. Effective parking management not only improves transportation efficiency but also significantly impacts traffic flow and congestion. However, studies have shown that parking has not been adequately integrated into transport models, which affects both the accuracy and practical application of these models. A key finding of this thesis, also supported by literature, is that one of the primary reasons for this gap is the lack of comprehensive parking data. For instance, the absence of data on private parking forces modelers to either ignore it or assume unlimited parking capacity in areas outside cities, potentially impacting the accuracy of the model's results depending on the application. This thesis addresses these gaps by identifying the specific data needs from a modeling perspective and exploring a relatively unexplored data source to help close them.

Interviews with transport modeling experts revealed that models are capable of doing much more than what they are currently doing, provided they have access to better data. Similarly, investigation and interviews with parking data providers revealed that while efforts are being made to improve parking data through new technologies, private parking remain unexplored in most cities. A potential solution, often overlooked, lies in utilizing unexplored data sources like remote sensing, particularly high-resolution aerial imagery. Traffic area segmentation datasets generated from aerial imagery can capture detailed parking data at a high spatial resolution, down to the level of individual parking slots, including lane markings. This study's exploration confirms that aerial image datasets can be used to extract comprehensive parking information, and with further analysis, it is possible to classify parking into on-street, off-street, and even estimate the number of parking slots for large areas, such as the state of Berlin. Furthermore, by combining it with other datasets, it’s also possible to extract information on illegal parking. This represents a significant contribution to the transportation modelling, as it offers a novel approach to address identified gaps in parking data, especially regarding private parking.

In answering the first research question - what are the data requirements necessary for integrating parking into travel demand models or improving current practices? - the interviews highlighted that data is needed across several dimensions: parking location, type of parking (by access), parking capacity, cost, occupancy, search time, and egress distance. Each of these dimensions requires varying levels of granularity depending on the specific application, and current practices often fall short in delivering this.

The second research question- which requirements from the modeling side are inadequately addressed, and to what extent can aerial image datasets meet these needs? - is also addressed through this research. To properly model parking, data is required at high spatial and temporal resolutions, both for input data and validation purposes. Aerial image datasets can directly address spatial data needs and when combined with other data sources, it can also contribute to fulfilling temporal data requirements, to an extent. By combining various data types from different sources, such as parking cost and occupancy fluctuations, which are not provided by aerial imagery alone, a more comprehensive parking dataset can be developed to support the modeling needs identified in this study.

While not all data gaps have been fully addressed, this research successfully identified and extracted the parking data that can be obtained using aerial image datasets, through simple analysis. This study contributes to the research community by presenting an approach that combines features from aerial images with cadastral data and road network to filter out parking spaces with additional attributes. Using the TIAS dataset – a traffic area segmentation dataset and Berlin as a case study, approximately 1.3 million individual parking slots were estimated and classified as either on-street or off-street. The dataset provides polygons for each parking slot, which can be integrated into models requiring point-level parking data. For instance, as Interviewee 3 noted, parking slot dimensions and orienttaions are only used for visualization purposes in the model SUMO, and this dataset provides precise visualization of each parking area. Moreover, the estimated parking capacity was assigned to OSM links to create density maps to identify the allocation of parking spaces on link level. Also, the capacity information was aggregated at the district level to analyse the distribution of parking spaces, including the respective proportions of on street and off-street parking. Additionally, the number of illegal parking was identified by combining with other dataset and aggregated within the parking zones of Berlin.

The use of aerial imagery presents a novel solution to addressing the data gaps identified in this study, with the added benefit of offering standardized data collection methods across different regions. The automated segmentation of aerial imagery into different traffic area classes, combined with the geospatial filtering and analysis discussed in this study, enables the development of statewide parking maps for structurally similar regions. This method ensures consistency in parking data collection and usage for transport modeling, enabling the transfer of modeling approaches across structurally similar regions - something other data sources often lack.

8.2. Limitations and future work

As with any research, this study faced certain limitations, particularly in terms of data collection, accessibility and data validation. While the spider diagram was an effective tool for identifying gaps in parking data, it could not capture all the possible levels of spatial and temporal granularity. Although the levelling scheme was based on literature, additional levels could have been included to reflect finer resolutions. However, this would have added complexity during interviews, as it would be difficult for interviewees to precisely align their responses with a specific level. Some interviewees suggested adding more levels, but implementing these changes would have impacted the results given by other modelers While the spider diagram was reasonably effective for representing demand-side data, it was less suitable for the supply side. The interviews, which focused on the predefined levels, did not gather comprehensive information beyond what was already present. Representing supply-side results solely through the spider diagram may not have been the best approach. Additionally, due to the academic nature of this research, direct access to commercial data providers was not possible. While these providers may have more detailed datasets than investigated in this study, they were unwilling to share specific information regarding the scope, or comprehensiveness of their data when approached.

Although traffic area segmentation is a promising approach for extracting parking data, the segmentation model has inherent limitations. These include occlusions and shadowing caused by buildings and trees, which prevent the accurate classification of parking areas beneath them. As a result, parking area sizes are often underpredicted. Some of these limitations are documented in the literature (Hellekes et al., 2023; Henry et al., 2021), and while the segmentation model shows potential, these challenges must be considered. Furthermore, integrating parking data with OSM posed challenges due to the accuracy and completeness of OSM, which is a volunteered information source. The data extraction methods in this thesis also relied on several assumptions, which, while necessary for exploration, may have influenced the overall quality of results. However, given that the focus of this thesis is on exploration rather than delivering the most refined methodologies, these assumptions were necessary for a broader investigation into new data sources. One significant limitation is the lack of available data to validate the results. It was challenging to assess data quality across the larger region due to the absence of official databases for comparison, particularly for private parking, which made it difficult to verify the accuracy of extracted data across larger regions.

The potential of remote sensing, especially aerial imagery extends far beyond what has been explored in this thesis. For example, the TIAS dataset includes multiple classes and attributes such as ‘shared’, ‘construction’, ‘elevated’, ‘difficult’, and ‘unsure’ as shown in the figure below.

Figure 38. Class 'access way' shared with 'parking area' with associated attributes

These attributes allow to differentiate between whether a parking area is shared with other traffic area classes, under construction, part of a multistorey car park, difficult to classify due to occlusions, or uncertain in terms of its classification. This further classification would be beneficial, such as distinguishing parking lots on top of a building as a part of a multistorey car park or identifying instances where cars are parked on footways or in backyards. This opens up opportunities to utilize the dataset for a wide range of applications in transport modelling. The ongoing work with TIAS holds promise, and as more results are extracted, it could lead to additional insights and applications to show the overall potential of aerial image dataset in transport applications. Furthermore, the scope of this research could be expanded to analyse parking occupancy at the time the aerial images were captured, which couldn't be completed within the timeframe of this thesis. Additionally, the dataset could be enriched by incorporating other data sources, which would provide a more comprehensive analysis of parking behaviour and dynamics.

References

Axhausen, K. W., & Polak, J. W. (1991). Choice of Parking: Stated preference approach.

Bagloee, S. A., Asadi, M., & Richardson, L. (2012). Methodology for Parking Modeling and Pricing in Traffic Impact Studies. Transportation Research Record: Journal of the Transportation Research Board, 2319(1), 1–12. https://doi.org/10.3141/2319-01

Benenson, I., Martens, K., & Birfir, S. (2008). PARKAGENT: An agent-based model of parking in the city. Computers, Environment and Urban Systems, 32(6), 431–439. https://doi.org/10.1016/j.compenvurbsys.2008.09.011

Berlin Geoportal. Berlin Geoportal - FIS Broke.

Bischoff, J., Maciejewski, M., Schlenther, T., & Nagel, K. (2019). Autonomous Vehicles and their Impact on Parking Search. IEEE Intelligent Transportation Systems Magazine, 11(4), 19–27. https://doi.org/10.1109/MITS.2018.2876566

Bischoff, J., & Nagel, K. (2017). Integrating explicit parking search into a transport simulation. Procedia Computer Science, 109, 881–886. https://doi.org/10.1016/j.procs.2017.05.414

Bonsall (1991). The changing face of parking-related data collection and analysis: The role of new technologies.

Bonsall, P., & Palmer, I. (2004). Modelling drivers' car parking behaviour using data from a travel choice simulator. Transportation Research Part C: Emerging Technologies, 12(5), 321–347. https://doi.org/10.1016/j.trc.2004.07.013

Christiansen, P., Engebretsen, Ø., Fearnley, N., & Usterud Hanssen, J. (2017). Parking facili-ties and the built environment: Impacts on travel behaviour. Transportation Research Part a: Policy and Practice, 95, 198–206. https://doi.org/10.1016/j.tra.2016.10.025

Dieussaert, K., Aerts, K., Steenberghen, T., Maerivoetb, S., & Spitaels, K. (2009). SUSTAPARK.

Erdmann et al. (2018). A SUMO Based Parking Management Framework for Large Scale smart cities simulations.

European Commission, Directorate-General for Mobility and Transport (2022). EU transport in figures: Statistical pocketbook 2022. Publications Office of the European Union. Retrieved from https://doi.org/10.2832/216553

GeoDPA DLR (Ed.). Poster: TIAS: An aerial traffic infrastructure dataset to study transportation in urban environments. 2024. Retrieved from https://elib.dlr.de/205507/1/GeoDPA_Poster_TIAS.pdf

Gu, Z., Safarighouzhdi, F., Saberi, M., & Rashidi, T. H. (2021). A macro-micro approach to modeling parking. Transportation Research Part B: Methodological, 147, 220–244. https://doi.org/10.1016/j.trb.2021.03.012

Guo, Z. (2013). Does residential parking supply affect household car ownership? The case of New York City. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0966692312002177

Hellekes, J. [Jens], Kehlbacher, A., Díaz, M. L., Merkle, N. [Nina], Henry, C. [Corentin], Kurz, F. [Franz], & Heinrichs, M. (2023). Parking space inventory from above: Detection on aerial images and estimation for unobserved regions. IET Intelligent Transport Systems, 17(5), 1009–1021. https://doi.org/10.1049/itr2.12322

Henry, C., Hellekes, J., Merkle, N., Azimi, S. M., & Kurz, F. (2021). Citywide Estimation of Parking Space Using Aerial Imagery And OSM Data Fusion With Deep Learning And Fine-Grained Annotation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B2-2021, 479–485. https://doi.org/10.5194/isprs-archives-XLIII-B2-2021-479-2021

Horni, A., Nagel, K., & Axhausen, K. W. (2016). The Multi-Agent Transport Simulation MATSim. Ubiquity Press. https://doi.org/10.5334/baw

(2023). Kartierung von Straßenparkplätzen innerhalb des S-Bahn-Rings. Berlin. Retrieved from https://www.berlin.de/sen/uvk/mobilitaet-und-verkehr/verkehrspolitik/forschungs-und-entwicklungsprojekte/laufende-projekte/euvm-1076914.php

Levy, N., Martens, K., & Benenson, I. (2013). Exploring cruising using agent-based and ana-lyt-ical models of parking. Transportmetrica a: Transport Science, 9(9), 773–797. https://doi.org/10.1080/18128602.2012.664575

Lubrich, P. (2023). Analysis of parking traffic in Cologne, Germany, based on an extended Macroscopic transport model and parking API data. Case Studies on Transport Policy, 11, 100940. https://doi.org/10.1016/j.cstp.2022.100940

Mar´ıa L´opez D´ıaz (2020). Comparison of non-fixed and fixed locations for subsistence ac-tiv-ities in destination choice.

Maurer, L., Roca, F., Treffinger, N., Henke, L., Hofmann, V., & Leonhartsberger, O. (2023). Tracking Parking Search and Occupancy in Zurich. Findings. Advance online publication. https://doi.org/10.32866/001c.72793

Nurul Habib, K. M., Morency, C., & Trépanier, M. (2012). Integrating parking behaviour in activity-based travel demand modelling: Investigation of the relationship between parking type choice and activity scheduling process. Transportation Research Part a: Policy and Practice, 46(1), 154–166. https://doi.org/10.1016/j.tra.2011.09.014

PLS Zurich. PLS Zurich - Parkleitsystem. Retrieved from https://www.pls-zh.ch/index.jsp

PTV AG (2020). PTV Visum 2020 Manual.

Rodríguez, A., Cordera, R., Alonso, B., dell'Olio, L., & Benavente, J. (2022). Microsimulation parking choice and search model to assess dynamic pricing scenarios. Transportation Research Part a: Policy and Practice, 156, 253–269. https://doi.org/10.1016/j.tra.2021.12.013

Rybczak, G., Meinhardt, S., Schlenther, T., Rakow, C., Zimeke, D., & Nagel, K. (2024). Agent-based modeling of residential parking zones in Leipzig.

Schiller, c. (2004). Dissertation-Endfassung.doc.

Senatsverwaltung für Mobilität, Verkehr, Klimaschutz und Umwelt (2023). Kartierung von Straßenparkplätzen innerhalb des S-Bahn-Rings. Berlin. Retrieved from https://www.berlin.de/sen/uvk/mobilitaet-und-verkehr/verkehrspolitik/forschungs-und-entwicklungsprojekte/laufende-projekte/euvm-1076914.php

Shoup (1997). The High Cost of Free Parking.

Stadt Zürich, G. Öffentlich zugängliche Strassenparkplätze OGD. Retrieved from https://data.stadt-zuerich.ch/dataset/geo_oeffentlich_zugaengliche_strassenparkplaetze_ogd

SUMO documentation: Microscopic Traffic Simulation Using SUMO (2018).

Tchervenkov, C. (2022). Empirical and simulation studies on parking in Switzerland. ETH Zur-ich. Retrieved from https://doi.org/10.3929/ethz-b-000593365.

Thompson, R. G., & Richardson, A. J. (1998). A parking search model.

Vuurstaek, J., Knapen, L., Kochan, B., Bellemans, T., & Janssens, D. (2018). First steps towards a state-of-the-art parking simulator. Procedia Computer Science, 130, 779–784. https://doi.org/10.1016/j.procs.2018.04.135

Waraich, R. A. (2016). Parking. In A. Horni, K. Nagel, & K. W. Axhausen (Eds.), The Multi-Agent Transport Simulation MATSim (pp. 89–92). Ubiquity Press. https://doi.org/10.5334/baw.13

Waraich, R. A., & Axhausen, K. W. (2012). Agent-Based Parking Choice Model. Transportation Research Record: Journal of the Transportation Research Board, 2319(1), 39–46. https://doi.org/10.3141/2319-05

Waraich, R. A., Dobler, C., & Axhausen, K. W. (2012). Modelling parking search behaviour with an agent-based approach. Advance online publication. https://doi.org/10.3929/ETHZ-B-000051421

Weeraddana, C., Edussooriya, C. U. S., & Abeysooriya, R. P. (Eds.) (2020). Mercon 2020: Moratuwa Engineering Research Conference - 6th International Multidisciplinary Engineering Research Conference: Conference proceedings : 27th, 28th and 29th July 2020, University of Moratuwa, Sri Lanka. Piscataway, NJ: IEEE. Retrieved from https://ieeexplore.ieee.org/servlet/opac?punumber=9179991 https://doi.org/10.1109/MERCon50084.2020

Young, W. (1990). Modelling parking

Young, W., & Thompson, R. (1987). Parksim/1: a computer graphics approach for parking-lot layouts.

[This page is intentionally left blank]

Appendices

Appendix A: Parking data demand from literature review

Model	Dimension	State of practice	Requirements	Source
VISUM	Parking location	Parking data API by INRIX, Building cadastral data (Muncipal GIS), Open data portal	Data for private areas	*Lubrich, Peter (2023): Analysis of parking traffic in Cologne, Germany, based on an extended Macroscopic transport model and parking API data.*
	Type of parking	Parking data API by INRIX, city administration data	-	
	Parking capacity	Parking data API by INRIX, buildings cadastral data	Data for private areas	
	Parking cost	Data from municipality	-	
	Parking occupancy	Parking data API by INRIX	Occupancy fluctuations	
	Parking search time	PSD - OD matrix (master Model) PSR - PSR algorithm	-	
	Egress distance	-	-	
MATSim	Parking location	-	-	-
	Type of parking	Public, private, Reserved and preferred parking	Division of the street parking into several categories could be done as some data on this is available.	*Axhausen, Kay W. (2012): Agent-Based Parking Choice Model.*

Model	Dimension	State of practice	Requirements	Source
	Parking capacity	**Zurich:** Data available for parking garages, limited knowledge for private areas, assumption for unknown areas, counting for small areas. (Supply data from National Surveys and imputation models) **Berlin:** Data for some links, assumption for unknown links	**Zurich:** Parking supply data outside the city and private areas. **Berlin:** Number of parking lots in each link	*Horni, Andreas; Nagel, Kai; Axhausen, Kay W. (2016): The Multi-Agent Transport Simulation MATSim* *Waraich, Rashid A.; Axhausen, Kay W. (2012): Agent-Based Parking Choice Model* *Bischoff, Joschka; Nagel, Kai (2017): Integrating explicit parking search into a transport simulation*
	Parking cost	**Zurich:** No spatial differences in the price, data available about parking permit ownership **Leipzig**: Accurate data from the city administration of Leipzig	**Zurich:** Actual parking prices for garage and on street parking	*Gregor Rybczak (2024): Agent-based modeling of residential parking zones in Leipzig.* *Horni, Andreas; Nagel, Kai; Axhausen, Kay W. (2016): The Multi-Agent Transport Simulation MATSim* *Waraich, Rashid A.; Axhausen, Kay W. (2012): Agent-Based*

Model	Dimension	State of practice	Requirements	Source
				Parking Choice Model.
	Parking occupancy	**Zurich:** Parking Garage: parking guidance system (update frequency: 1 minute)	**Zurich:** Parking occupancy log at the metered on-street parking, accurate occupancy data for nighttime.	*Waraich, Rashid A.; Axhausen, Kay W. (2012): Agent-Based Parking Choice Model.*
	Parking search time	Simulation (output), surveys (input)	-	*Horni, Andreas; Nagel, Kai; Axhausen, Kay W. (2016): The Multi-Agent Transport Simulation MATSim* *Waraich, Rashid A.; Axhausen, Kay W. (2012): Agent-Based Parking Choice Model.*
	Egress distance (input)	**Zurich:** Egress distance - fixed value: 350m	**Zurich:** Actual egress (parking related walking) distance and time instead of fixed values (through surveys)	*Waraich, Rashid A.; Axhausen, Kay W. (2012): Agent-Based Parking Choice Model.*
SUMO	Parking location	OSM	-	*SUMO Documentation,* *Erdmann et al, A SUMO Based Parking Management Framework for Large Scale smart*
	Type of parking	Off-street vs on-street	-	
	Parking capacity	OSM, Monaco parking website	-	
	Parking cost	-	-	

Model	Dimension	State of practice	Requirements	Source
	Parking occupancy	Crowd sensing	-	*cities simulations*
	Parking search time	-	-	
	Egress distance	-	-	
AIMSUN	Parking location	-	-	*Rodríguez, A., Cordera, R., Alonso, B., dell'Olio, L., & Benavente, J. (2022). Microsimulation parking choice and search model to assess dynamic pricing scenarios.*
	Type of parking	Regulated paid on-street parking, private parking and paid off-street private parking.	-	
	Parking capacity	SP survey	-	
	Parking cost	Muncipal parking regulations, public knowledge, SP survey	-	
	Parking occupancy	Muncipal parking regulations, Open data of Santander website, sensor data, SP survey	-	
	Parking search time	SP survey	-	
	Egress distance	-	-	
PARKAGENT (Israel)	Parking location	Field Survey, GIS database	-	*Benenson, Itzhak; Martens, Karel; Birfir, Slava (2008): PARKAGENT: An agent-based model of parking in the city.*
	Type of parking	GIS database, field surveys	-	
	Parking capacity	GIS Database, field surveys for private parking places (estimations)	Data for private off-street parking places	

Model	Dimension	State of practice	Requirements	Source
	Parking cost	GIS database	-	*Waraich, R.A., C. Dobler and K.W. Axhausen (2012): Modelling Parking Search Behaviour with an Agent-Based Approach*
	Parking occupancy	Field surveys	-	
	Parking search time	10 mins as model input (if search time exceeds 10 min, drivers directly go to the paid parking)	-	
	Egress distance	350 m as model input (average computed by comparing car number plate and ICBS data)	-	
SUSTAPARK (Belgium)	Parking location	-	-	*SUSTAPARK: An Agent-based Model for Simulating Parking Search* *Waraich, R.A., C. Dobler and K.W. Axhausen (2012): Modelling Parking Search Behaviour with an Agent-Based Approach*
	Type of parking	Parking lots, private garages, on-street parking (GIS service of the city – G@lileo)	-	
	Parking capacity	On street parking – data available Private parking- estimation	Number of private parking slots	
	Parking cost	-	-	
	Parking occupancy	-	-	
	Parking search time	-	-	
	Egress distance	-	-	

Appendix B: Interview questionnaire - modelers

The interviews were conducted online using Zoom as the platform. Interview questionnaire was send prior via email to familiarize the interviewees with the questions. At the start of the questionnaire, an introduction and background were provided, including information on all dimensions and levels. Once the interviewees were familiar with the research background, they were asked to select a level for each dimension based on their modeling application. A PowerPoint presentation was used to present the background information, and a Miro board, an online whiteboard tool, was used to rank the levels within the spider diagram. Each interviewee was asked the same set of questions. An example questionnaire for the model MATSim is presented below.

General questions

1. Currently, how is parking modeled in MATSim - which all aspects are included and ready-to use in the parking module of MATSim? In the future, how do you want to include parking in MATSim?

2. For what specific applications in transport modeling are parking modelled in MATSim? Based on your application, how does the level of detail in data matter?

3. Based on the applications, under which classification(s) of models - strategic, tactical, and operational models does MATSim belong to?

4. What are the general limitations for integrating parking in MATSim?

5. In the related research section of your paper *(*Agent-based modeling of residential parking zones in Leipzig*)*, you include details about only three explicit parking models, PARKAGENT, SUSTAPARK and SimPark. Even though there are several other explicit parking models, why did you choose to consider only these three models?

Identifying the levels of each dimension

6. Do you agree with the dimensions identified here? If not, do you want to include or exclude one or more dimensions?

7. Based on literature review, four levels for the level of detail in data were distinguished for the dimension **'parking location'**. If you require this data, do you agree with the leveling scheme, or do you want to include or exclude one or more levels than what's defined here?

   Level 1: City level

   Level 2: Zonal level

   Level 3: Link level

   Level 4: Point level

8. Based on your **application**, do you need **'parking location'** in your model, and if yes which level (level 1- 4) describes your current practice? What limits you to not go further beyond this level in the current scenario? If you desire to go further and your model is capable, till what level do you need this data?

9. For the level of detail in data, five levels were distinguished for the dimension **'type of parking'**. If you require this data, do you agree with the leveling scheme, or do you want to include or exclude one or more levels than what is defined here?

   Type of parking is classified here into:

   1) On-street Parking

   2) Off-street Parking

      a. Private garage

      b. Private lots

      c. Public garage

      d. Public lots

   Level 1: Distinction between parking and no parking

   Level 2: Distinction between on street and off-street parking

   Level 3: Distinction between off-street public and off-street private parking

Level 4: Distinction between individual parking facilities

Level 5: Distinction of individual slots based on parking regulations

10. Based on your **application**, do you need **'type of parking'** in your model, and if yes which level (level 1- 5) describes your current practice? What limits you to not go further beyond this level in the current scenario? If you desire to go further and your model is capable, till what level do you need this data?

11. Four levels for the level of detail in data were distinguished for the dimension **'parking capacity'**. If you require this data, do you agree with the leveling scheme, or do you want to include or exclude one or more levels than what is defined here?

    Level 1: General estimate /range

    Level 2: Exact numbers

    Level 3: Previous + parking slot dimensions and orientation

    Level 4: Previous + number of illegal parking

12. Based on your **application**, do you need **'parking capacity'** in your model, and if yes which level (level 1- 4) describes your current practice? What limits you to not go further beyond this level in the current scenario? If you desire to go further and your model is capable, till what level do you need this data?

13. Five levels for the level of detail in data were distinguished for the dimension **'parking cost'**. If you require this data, do you agree with the leveling scheme, or do you want to include or exclude one or more levels than what is defined here?

    Level 1: General estimate/range

    Level 2: Previous + average subscription costs

    Level 3: Time specific detailed cost of individual parking facilities

    Level 4: Previous + detailed subscription costs

    Level 5: Previous + parking fines

14. Based on your **application**, do you need **‘parking cost’** in your model, and if yes which level (level 1-5) describes your current practice? What limits you to not go further beyond this level in the current scenario? If you desire to go further and your model is capable, till what level do you need this data?

15. Seven levels for the level of detail in data were distinguished for the dimension **‘parking occupancy’**. If you require this data, do you agree with the leveling scheme, or do you want to include or exclude one or more levels than what is defined here?

    Level 1: Aggregated average occupancy for TAZ’s /parking zones

    Level 2: Average occupancy of individual facilities over a month/week

    Level 3: Average occupancy of individual facilities over a day

    Level 4: Average occupancy of individual facilities over peak and off-peak hours

    Level 5: Average occupancy of individual facilities over each hour

    Level 6: Real-time occupancy of individual facilities

    Level 7: Real time occupancy of individual parking slots

16. Based on your **application**, do you need **‘parking occupancy’** in your model, and if yes which level (level 1- 7) describes your current practice? What limits you to not go further beyond this level in the current scenario? If you desire to go further and your model is capable, till what level do you need this data?

17. Seven levels for the level of detail in data were distinguished for the dimension **‘parking search time’**. If you require this data, do you agree with the leveling scheme, or do you want to include or exclude one or more levels than what is defined here?

    Level 1: Aggregated average search time for TAZ’s /parking zones

    Level 2: Average search time based on type of parking (on-street vs off-street)

    Level 3: Average search time over a week/month

    Level 4: Average search time over a day

Level 5: Average search time over peak and off-peak hours

Level 6: Average search time over each hour

Level 7: Total search time of individual agents including the search time within the parking facility

18. Based on your **application**, do you need **‘parking search time’** in your model, and if yes which level (level 1-7) describes your current practice? What limits you to not go further beyond this level in the current scenario? If you desire to go further and your model is capable, till what level do you need this data?

19. Five levels for the level of detail in data were distinguished for the dimension **‘egress distance’**. If you require this data, do you agree with the leveling scheme, or do you want to include or exclude one or more levels than what is defined here?

Level 1: Aggregated average egress distance for TAZ’s /parking zones

Level 2: Average egress distance based on type of parking (on-street vs off-street)

Level 3: Average egress distance of individual parking facilities

Level 4: Total egress distance for each agent

Level 5: Total egress distance for each agent including the egress distance within a parking facility

20. Based on your **application**, do you need **‘egress distance’** in your model, and if yes which level (level 1- 5) describes your current practice? What limits you to not go further beyond this level in the current scenario? If you desire to go further and your model is capable, till what level do you need this data?

21. After conducting the interviews with several modelers, I will be interviewing parking data providers in major German speaking cities. In that case, do you want to ask any specific questions to the candidates concerning parking data supply?

[Conclusion and presenting summary of result]

Appendix C: Interview questionnaire - supply data providers

The interviews were conducted online using Zoom as the platform. Interview questionnaire was sent prior via email to familiarize the interviewees with the questions. At the start of the questionnaire, an introduction and background were provided, including information on all dimensions and levels. Once the interviewees were familiar with the research background, they were asked to select a level for each dimension based on their modeling application. A PowerPoint presentation was used to present the background information, and a Miro board, an online whiteboard tool, was used to rank the levels within the spider diagram. Each interviewee was asked the same set of questions. An example questionnaire for the city Berlin is presented below.

General questions

1. Currently, which all among the **seven parking dimensions** (parking location, type of parking, parking capacity, parking cost, parking occupancy, parking search time and egress distance) are included and ready-to use in the parking dataset of Berlin?

2. What are the limitations for integrating parking data in your dataset?

Identifying levels of each dimension

3. Based on literature review, four levels for the level of detail in data were distinguished for the dimension **'parking location'**? Which level (levels 1- 4) describes the current state of this data in Berlin? How complete and reliable is 'parking location' data in your dataset (completeness of the data)? If the current state is not the highest level (level 4), what limits you to not collect data further beyond this level?

   Level 1: City level

   Level 2: Zonal level

   Level 3: Link level

   Level 4: Point level

4. For the level of detail in data, five levels were distinguished for the dimension **'type of parking'**. Which level (level 1-5) describes the current state of this data in Berlin? How

complete and reliable is ‘type of parking’ data in your dataset (completeness of the data)? If the current state is not the highest level (level 5), what limits you to not collect data further beyond this level?

Type of parking is classified here into:

1) On-street Parking

2) Off-street Parking

   a. Private garage

   b. Private lots

   c. Public garage

   d. Public lots

Level 1: Distinction between parking and no parking

Level 2: Distinction between on street and off-street parking

Level 3: Distinction between off-street public and off-street private parking

Level 4: Distinction between individual parking facilities

Level 5: Distinction of individual slots based on parking regulations

5. Four levels for the level of detail in data were distinguished for the dimension **‘parking capacity’**. Which level (level 1-4) describes the current state of this data in Berlin? How complete and reliable is ‘parking capacity’ data in your dataset (completeness of the data)? If the current state is not the highest level (level 4), what limits you to not collect data further beyond this level?

   Level 1: General estimate /range

   Level 2: Exact numbers

   Level 3: Previous + parking slot dimensions and orientation

Level 4: Previous + number of illegal parking

6. Five levels for the level of detail in data were distinguished for the dimension **'parking cost'**. Which level (level 1-5) describes the current state of this data in Berlin? How complete and reliable is 'parking cost' data in your dataset (completeness of the data)? If the current state is not the highest level (level 5), what limits you to not collect data further beyond this level?

   Level 1: General estimate/range

   Level 2: Previous + average subscription costs

   Level 3: Time specific detailed cost of individual parking facilities

   Level 4: Previous + detailed subscription costs

   Level 5: Previous + parking fines

7. Seven levels for the level of detail in data were distinguished for the dimension **'parking occupancy'**. Which level (level 1-7) describes the current state of this data in Berlin? How complete and reliable is 'parking occupancy' data in your dataset (completeness of the data)? If the current state is not the highest level (level 7), what limits you to not collect data further beyond this level?

   Level 1: Aggregated average occupancy for TAZ's /parking zones

   Level 2: Average occupancy of individual facilities over a month/week

   Level 3: Average occupancy of individual facilities over a day

   Level 4: Average occupancy of individual facilities over peak and off-peak hours

   Level 5: Average occupancy of individual facilities over each hour

   Level 6: Real-time occupancy of individual facilities

   Level 7: Real time occupancy of individual parking slots

8. Seven levels for the level of detail in data were distinguished for the dimension **'parking search time'**. Which level (level 1-7) describes the current state of this data in Berlin? How

complete and reliable is ‘parking search time’ data in your dataset (completeness of the data)? If the current state is not the highest level (level 7), what limits you to not collect data further beyond this level?

Level 1: Aggregated average search time for TAZ’s /parking zones

Level 2: Average search time based on type of parking (on-street vs off-street)

Level 3: Average search time over a week/month

Level 4: Average search time over a day

Level 5: Average search time over peak and off-peak hours

Level 6: Average search time over each hour

Level 7: Total search time of individual agents including the search time within the parking facility

9. Five levels for the level of detail in data were distinguished for the dimension **‘egress distance’** (egress distance is the distance from the parking location to the activity/end destination). Which level (level 1-5) describes the current state of this data in Berlin? How complete and reliable is ‘egress distance’ data in your dataset (completeness of the data)? If the current state is not the highest level (level 5), what limits you to not collect data further beyond this level?

   Level 1: Aggregated average egress distance for TAZ’s /parking zones

   Level 2: Average egress distance based on type of parking (on-street vs off-street)

   Level 3: Average egress distance of individual parking facilities

   Level 4: Total egress distance for each agent

   Level 5: Total egress distance for each agent including the egress distance within a parking facility

10. In the near future, are you planning to improve the parking data in your dataset in anyway and if yes, how?

Appendix D: Interviewee candidates

The following table gives information on the interviewee candidates who accepted the interview invitation.

Interviewee No.	**Representation**
1	Model - VISUM
2	Model - MATSim
3	Model - SUMO
4	Model - TAPAS
5	Model - PARKAGENT
6	City - Berlin (GeoPortal Berlin)
7	City - Munich (GeoPortal München)
8	City - Hamburg (Landesbetrieb Verkehr (LBV), Hamburg)
9	City - Zurich (Stadt Zürich)

Appendix E: State of practice: parking supply data

City	Dimension	Available data	Source
Berlin	Parking location	On street - Coordinates[1], Address[1,3,4] Garage - coordinates[1], Address[1,2,3]	*1. Geoportal Berlin* *2. APCOA* *3. Parkopedia* *4. OSM*
	Type of parking	Public parking – Underground/surface[1,2,3], number of disabled parking spaces[1,2,3], parking for women[1,2], parking for family[2], charging facilities for EV[1,3] On-street parking – disabled parking[1], P+R facilities[1], car sharing[1], charging facilities for EV[1,] parking stopping bans and reasons[1], opening hours[1,3], parking zones[1], parking guidance systems[1], parking time restriction[1,3] Parking Garage: opening hours[1,2,3], parking zones[1]	
	Parking capacity	Public parking: parking capacity (for some garages)[1,2,3,4] On-street parking: parking capacity[4], parking alignment[1,4], slot length[4], slot area[4]	
	Parking cost	On-street- detailed data[1,3] Public parking – cost data only for parking under parking zones[1], detailed data[2,3]	
	Parking occupancy	-	

City	Dimension	Available data	Source
	Parking search time	-	
	Egress distance/ time	Egress distance can be calculated from map[1], egress time[3]	
Munich	Parking location	Street address, coordinates [1,2,3,4]	*1. Geo Portal Munich* *2. APCOA* *3. Parkopedia* *4. OSM*
	Type of parking	On-street parking[1,3,4] - Resident parking[1,] mixed parking[1], car sharing[1], no parking areas[1], restricted parking[1], sidewalk parking[1], disabled parking[1,4], short-term parking[1], parking for Women[1,4], parent parking[4], P+R facilities[1,4], bus parking[1,] E – parking [1,3,4] , Opening hours of parking areas[1,3,4], presence of parking discs[1,4], time restrictions for parking areas [1,3,4] Parking garage [2,3,4] - disabled parking[2,3], parking for women[2], opening hours [2, 3] Private (incomplete)[3,4]: Pre-Booking, opening hours [3]	
	Parking capacity	On-street: parking capacity [1,3,4] Garages: parking capacity [2,3] Parking orientation for certain parking facilities[4]	
	Parking cost	Hourly parking cost available for parking garages [2,3], cost data for on-street parking [3] cost for private parking [3] (OSM only says parking fees: yes/no for on street parking)	

City	Dimension	Available data	Source
	Parking occupancy	-	
	Parking search time	-	
	Egress distance/ time	Egress distance can be calculated from the map (on street) [1] or in minutes (on street, garages) [3]	
Hamburg	Parking location	Coordinates[1,4,] address - on street[1], garage[1,2,3,4]	*1. Geoportal Hamburg* *2. APCOA* *3. Parkopedia* *4. OSM*
	Type of parking	Parking garage – Underground/surface[1,2,4], number of disabled parking spaces[1,2,3,4], parking spaces based on type of vehicles[1], parking for women[1,2], parking for family[2], charging facilities for EV[3], parking management types[1], opening hours[1,2,3] On-street parking: based on vehicle type[1], disabled parking[1,3,4], P+R facilities[1], bus parking[1], presence of parking ticket machines[1], car sharing[1], charging facilities for EV[3], parking management type (e.g. parking disc) [1], parking stopping bans[1], construction sites[1], opening hours[1,3] Private[4] (incomplete)	
	Parking capacity	Parking garage: parking capacity (for some garages)[1,2,3,4], parking area[4]	

City	Dimension	Available data	Source
		On street parking: capacity[3,4], orientation to the street[1,4], length[4], parking angle[4], parking area[4] Private (incomplete): parking area[4]	
	Parking cost	Parking cost for certain parking garages[1,2,3], on -street[3]	
	Parking occupancy	Parking garage- Number of free parking spaces (for some garages) [1]	
	Parking search time	-	
	Egress distance/ time	Garage, on-street: egress time (in minutes)[3]	
Cologne	Parking location	Public parking: Address[3,4,5] On street: Address[4,5]	*1. Open data cologne (csv-link)* *2. GeoData cologne(map-link)* *3. APCOA* *4. Parkopedia* *5. OSM*
	Type of parking	On street parking[1,5], Reserved resident parking[1], opening hours[3,4,5], time restrictions[4,5] Public parking: Surface/underground parking[3,4,5], disabled parking (numbers) [2,3,4,5] parking for Women[3], EV charging stations[4,5], P+R[5], opening hours[4,5], time restrictions[4,5] Private parking[5] (incomplete)	

City	Dimension	Available data	Source
	Parking capacity	On street parking: parking capacity,[5], orientation of certain parking facilities[5] Public parking: parking capacity,[5], orientation of certain parking facilities[5]	
	Parking cost	Public parking: detailed cost[3,4], (fee: yes/ no by OSM)[5] On street parking: cost data[4,] (fee: yes/no by OSM)[5]	
	Parking occupancy	-	
	Parking search time	-	
	Egress distance/ time	Egress time in minutes (on street, garages) [4]	
Zurich	Parking location	On street parking: coordinates[2], address[3,4] Parking garage: address[1,3,4], coordinates[2]	*1. Parking Guidance system - parkleit system, stadt Zurich* *2. OSM (incomplete information)* *3. Geoportal Zurich* *4. Parkopedia*
	Type of parking	On street parking[2,3]: disabled parking[2], P+R[2,4], EV charging station[3], opening hours[2,3,4], time restrictions[2,3,4], Residents card[3], parking meters[3], parking dics[3] Parking garage[1,2,3] –disabled people[1,4], women[1], parking with EV charging station[1,4], time restrictions [1,3,4], opening hours[4]	

City	Dimension	Available data	Source
		Private parking[2] (incomplete)	
	Parking capacity	On street parking: capacity[2,4] Parking garage: capacity[1,4]	
	Parking cost	On street parking: detailed cost data[4], fees (yes/no) [2] Parking garages: timely tariffs[1,4]	
	Parking occu-pancy	Parking garage- number of free parking spaces[1]	
	Parking search time	-	
	Egress distance/ time	Egress time in minutes (on street, gar-ages) [4]	

Appendix F: Parking distribution across Berlin districts from predictions

District	On-street parking	Off-street parking	Total number of parking	Share of each district (%)
Charlottenburg-Wilmersdorf	69810	37204	107014	8.27
Reinickendorf	40248	50581	90829	7.023
Spandau	42839	59479	102318	7.911
Steglitz-Zehlendorf	58918	48492	107410	8.305
Pankow	75486	49169	124655	9.638
Friedrichshain-Kreuzberg	45318	17383	62701	4.848
Tempelhof-Schöneberg	62206	74500	136706	10.57
Mitte	64567	30474	95041	7.349
Lichtenberg	67834	49753	117587	9.092
Treptow-Köpenick	59945	62866	122811	9.496
Neukölln	53871	43930	97801	7.562
Marzahn-Hellersdorf	74906	53525	128431	9.93

Appendix G: Validation metrics for TIAS model predictions

There are typically 3 metrics used in semantic segmentation to measure the quality of the predictions versus the ground truth:

The recall (also known as sensitivity in other domains): It measures how complete the predictions are, by calculating the amount of predictions made where expected, over the total amount of ground truth, i.e.: true positives / (true positives + false negatives)

The precision (also known as specificity): It measures how much the predictions do NOT overflow passed the ground truth, by calculating the amount of predictions made where expected, over the total amount of predictions, i.e.: true positives / (true positives + false positives)

The Intersection Over Union (IoU): it measures a mix between the recall and precision which serves as a good decision metrics to judge which model is the most balanced between both criteria, by calculating the intersection of the predictions and the ground truth over the union of both, i.e.: true positives / (true positives + false negatives + false positives)

The performance metrics of the model predictions in the TIAS validation set received are as follows:

Roads

IoU: 77.84%

Precision: 88.17%

Recall: 86.93%

Parking area

IoU: 48.12%

Precision: 58.37%

Recall: 73.25%

Access ways

IoU: 41.36%

Precision: 70.85%

Recall: 49.85%

These values can be better understood with specifying some success and failure cases. The success cases include:

- The roads are well predicted with good coverage matching their outline in the aerial images, with regular shapes and good connectivity throughout.
- The access ways are also well predicted around large freight or parking areas (i.e. without any clearly marked lane) and around clearly marked parking areas.
- The parking areas are well predicted with regular shapes in the case of parking marked by paint or delimited by pavement stones.
- Roadside parking areas are also well identified by the model.

Some of the failure cases include:

- False positives for roads on keep-out areas as there were only few samples to train from.
- Lack of continuity in roads and access ways in areas without clear markings or asphalt change.
- Access ways in residential driveways are ignored by the model.
- For large, unmarked parking areas, the model predicted parking areas only around vehicles.

Appendix H: Table of dimensions and levels

Please note that, level 0 for all the dimensions means ‘no data’.

Parking location Level 1: City level Level 2: Zonal level Level 3: Link level Level 4: Point level	**Type of parking** Level 1: Distinction between parking and no-parking Level 2: Distinction between on-street and off-street parking Level 3: Distinction between off-street public and off-street private parking Level 4: Distinction between individual parking facilities Level 5: Distinction between individual parking slots based on parking regulations
Parking capacity Level 1: General estimate /range Level 2: Exact number of parking slots Level 3: Previous + parking slot dimensions and orientation Level 4: Previous + number of illegal parking	**Parking cost** Level 1: General estimate/range Level 2: Previous + average subscription costs Level 3: Time-specific detailed cost of individual parking facilities Level 4: Previous + detailed subscription costs Level 5: Previous + parking fines
Parking occupancy Level 1: Aggregated average occupancy based on zones Level 2: Average occupancy of individual facilities over a month/week Level 3: Average occupancy of individual facilities over a day Level 4: Average occupancy of individual facilities over peak and off-peak hours	**Parking search time** Level 1: Aggregated average search time based on zones Level 2: Average search time based on type of parking (on-street vs off-street) Level 3: Average search time over a week/month Level 4: Average search time over a day

<table>
<tr><td>Level 5: Average occupancy of individual facilities over each hour

Level 6: Real-time occupancy of individual facilities

Level 7: Real time occupancy of individual slots</td><td>Level 5: Average search time over peak and off-peak hours

Level 6: Average search time over each hour

Level 7: Total search time of individual agent/person including the search time within the parking facility</td></tr>
<tr><td colspan="2">Egress distance

Level 1: Aggregated average egress distance based on zones

Level 2: Average egress distance based on type of parking

Level 3: Average egress distance of individual parking facilities

Level 4: Total egress distance for each agent/person

Level 5: Total egress distance for each agent/person including the egress distance within a parking facility</td></tr>
</table>