Promotionszentrum
Mobilität und Logistik

A Study on Cruising for Parking based on Geodata with Methods from Econometrics and Machine Learning

Siavash Saki

Promotionszentrum Mobilität und Logistik,
Frankfurt University of Applied Sciences,
Frankfurt am Main, Germany

Frankfurt am Main, 2024

Author: Siavash Saki
Frankfurt University of Applied Sciences
Research Lab for Urban Transport
Nibelungenplatz 1
60318 Frankfurt am Main
siavash.saki@engineer.com

Supervisors: Prof. Dr. Tobias Hagen
Professor of Economics and Quantitative Methods
Faculty for Business and Law
Frankfurt University of Applied Sciences

Prof. Dr. Matthias Kowald
Professor of Mobility Management and Mobility Behavior
Department of Architecture and Civil Engineering
RheinMain University of Applied Sciences

Reviewers: Prof. Dr. Anne Lange
Professor of General Business Administration,
in particular Logistics and Air Transport
Faculty for Business and Law
Frankfurt University of Applied Sciences

Prof. Dr.-Ing. Dr.Techn. H.C. Kay W. Axhausen
Director of Institute for Transport Planning and Systems
Department of Civil, Environmental and Geomatic Engineering
ETH Zürich

Acknowledgements

I would like to extend my deepest gratitude to Prof. Dr. Tobias Hagen, whose patience and unwavering belief in me have been pivotal throughout this PhD journey. His guidance has been an invaluable resource, providing me with the direction and support necessary to navigate the complexities of this research. Prof. Dr. Hagen's insightful feedback and high standards have continually pushed me to refine my work and expand my understanding, encouraging me to pursue excellence in both my methodology and analysis. I am immensely grateful for his trust in my abilities and for his commitment to fostering my development as a researcher.

Throughout this process, Prof. Dr. Hagen has been a steadfast source of encouragement, reminding me of the importance of resilience and diligence in the face of challenges. His dedication to my success and his willingness to share his vast knowledge have enriched my experience and significantly shaped the trajectory of my work. I am sincerely thankful for his mentorship, which has not only guided my research but has also profoundly influenced my personal and professional growth.

Siavash Saki
Berlin, October 2024

Abstract

Parking search behavior is a significant component of urban mobility, contributing to congestion, emissions, and inefficiencies within transportation systems. Despite its impact, accurately quantifying parking search behavior remains challenging due to methodological limitations in existing research. Traditional approaches, such as surveys and field experiments, often suffer from biases and assumptions that limit the reliability and generalizability of their findings. The recent advancement of GPS technology offers new possibilities for capturing real-world data on parking search behavior, yet these studies frequently lack a clear framework for defining search starting and ending points, leading to inconsistencies in the reported data.

This dissertation aims to develop an empirical framework for understanding parking search behavior, focusing on capturing accurate data with minimal assumptions. The research is structured around three key objectives: First, to explore the initiation of parking search by identifying factors that influence when and where drivers begin searching for parking. Second, to determine the key factors affecting parking search duration and quantify their effects using data-driven methodologies. Third, to develop a machine learning model capable of classifying GPS trajectory data to identify and analyze parking search behavior.

Through an integrated approach, this research leverages GPS data to collect detailed, ground-truth information across the entire parking search process—from driving to parking and walking to the final destination. This methodology not only addresses existing gaps in the literature but also provides practical insights for urban planners and policymakers. By offering a comprehensive understanding of parking search behavior, the findings of this dissertation contribute to developing targeted strategies for reducing cruising for parking and improving urban mobility. The goal is to advance both theoretical knowledge and practical applications in urban transportation, paving the way for smarter, more sustainable city planning.

Keywords: Parking Search Behavior, Urban Mobility, GPS Data Analysis, Empirical Framework, Transportation Planning

List of Publications

This thesis is organized into several sections, beginning with an introductory review, followed by three scientific articles, and a final conclusion. Below is a list of the three included publications:

I. Saki, S., & Hagen, T. (2024c). What drives drivers to start cruising for parking? modeling the start of the search process. *Transportation Research Part B: Methodological*, *188*, 103058. https://doi.org/10.1016/j.trb.2024.103058

II. Saki, S., & Hagen, T. (2024a). Cruising for parking again: Measuring the ground truth and using survival analysis to reveal the determinants of the duration. *Transportation Research Part A: Policy and Practice*, *183*, 104045. https://doi.org/10.1016/j.tra.2024.104045

III. Saki, S., & Hagen, T. (2024b). Parking search identification in vehicle gps traces. *Journal of Urban Mobility*, *6*, 100083. https://doi.org/10.1016/j.urbmob.2024.100083

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation and problem statement

Urban mobility is a cornerstone of modern city life, essential for supporting economic activity, social interactions, and overall quality of life. As urban populations continue to expand, cities face mounting challenges in managing transportation demands. This increased pressure on transportation infrastructure often results in congestion, pollution, and inefficiencies that compromise environmental sustainability and reduce urban livability (Giffinger, 2021).

To address these challenges, cities have increasingly focused on optimizing their transportation systems to ensure the efficient movement of people and goods. However, as transportation systems evolve, so too do the complexities associated with managing urban traffic. Urban road networks must accommodate a diverse array of users, from personal vehicles and public transit to bicycles and pedestrians, all of whom compete for limited space (Gössling et al., 2016). As a result, traffic congestion is an inevitable outcome, particularly in dense city centers where the demand for road space frequently exceeds the available capacity (Afrin & Yodo, 2020). Managing limited resources like road capacity and parking spaces is critical for ensuring smooth traffic flow and minimizing environmental impacts (Biswas et al., 2017).

Parking management, in particular, is a vital component of urban traffic management (Sándor & Csiszár, 2015). Effective parking strategies alleviate congestion and reduce emissions by ensuring that parking spaces are used efficiently (Litman, 2016). However, in many urban areas, the demand for parking far outstrips available supply, leading to a widespread issue known as "cruising for parking" (Shoup, 2006). As drivers navigate through busy city streets in search of available curbside parking, they inadvertently add to traffic density, fuel consumption, and emissions. Even when parking supply and demand are balanced, the lack of real-time information on available parking spaces forces drivers to cruise, leading to inefficient traffic flows (Arnott & Rowse, 1999). Although this behavior may seem brief compared to the overall trip, it has substantial cumulative impacts on urban environments.

Cruising for parking is particularly prevalent in cities where curbside parking is more affordable or more convenient than off-street options (Shoup, 2006). This preference drives a considerable portion of urban traffic, as drivers circulate through neighborhoods, adding to both local congestion and environmental degradation. Beyond its direct impacts on traffic, the parking search process is often frustrating for drivers, who tend to perceive this portion of their journey as more stressful and time-consuming than other parts of their trip (Weis et al., 2021).

Understanding parking search behavior is essential for devising effective urban traffic management strategies (Verhoef et al., 1995). Historically, research in this area has been hindered by data limitations, relying on surveys or assumptions that lack precision and often fail to capture the real-time dynamics of parking search behavior. With advancements in technology, particularly the increased availability of GPS data, new avenues for understanding parking search patterns have emerged (Montini et al., 2012). GPS data provide an unprecedented level of detail on vehicle movements, allowing for a comprehensive analysis of parking search behavior across diverse urban contexts (Mannini et al., 2017).

The scientific motivation for this research stems from significant gaps in the literature regarding the empirical measurement and modeling of parking search behavior. Existing studies often rely on estimating parking search durations and impacts, potentially leading to biased or incomplete conclusions. Furthermore, few models adequately address the nuanced transition from driving to parking search, or how contextual factors—such as driver demographics, time of day, and urban density—affect parking search patterns.

This research aims to address these gaps by utilizing a novel data collection approach that captures detailed, ground-truth GPS data on parking search behavior. Through a custom-built app, this study records the starting point of each parking search, tracks the entire search process, and collects data on the complete journey, including walking to the final destination ("Start2Park Research Project", 2024). This data enables a more accurate and holistic understanding of the factors influencing parking search behavior, the typical duration of searches, and the decision-making processes involved.

In summary, the motivation for this research is driven by both the need to advance scientific knowledge on parking search behavior and the practical challenges of urban mobility. By developing robust models and frameworks, this research seeks to inform urban planning and traffic management policies, ultimately contributing to a reduction in the negative impacts of cruising for parking on urban environments.

1.2 Theoretical Background and State of Research

1.2.1 What Is Parking Search and Why Does It Occur?

Cruising for parking is the practice of driving around in search of an available curbside parking spot (Shoup, 2005). As can be seen in Figure 1.1, it has three main drivers: 1) the scarcity of readily available parking spaces given the demand for parking at the given price (in the sense of generalized costs) of parking, 2) the characteristics of parking spots not matching drivers' preferences, and 3) the lack of complete information about parking availability. These drivers are influenced by various factors, including supply and demand, human behavior and decision-making, psychological factors, environmental and urban planning factors, technological and

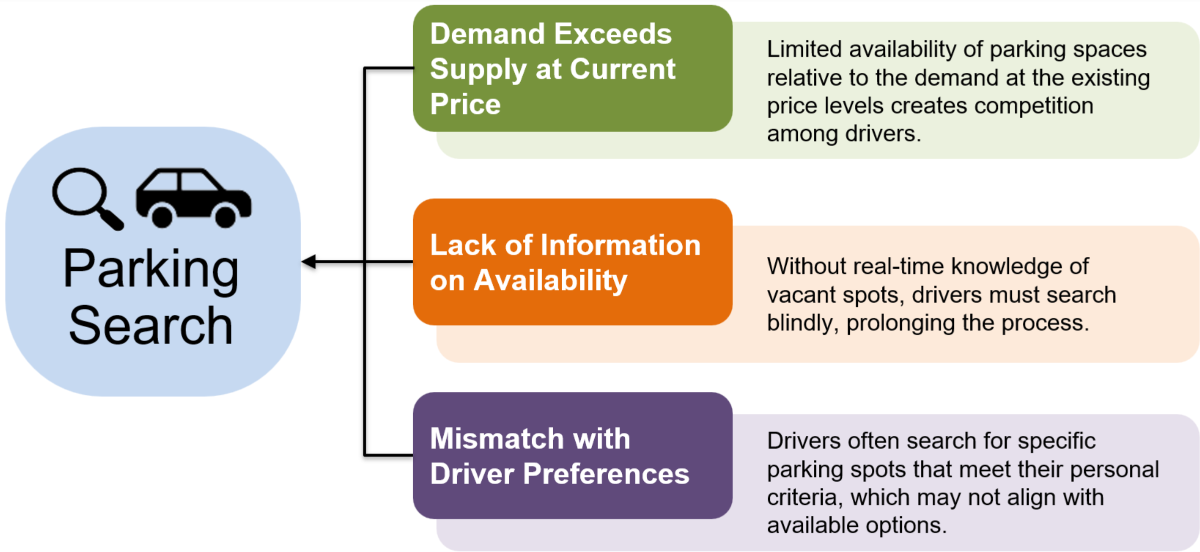


Figure 1.1: Key Drivers of Parking Search

information factors, traffic flow and congestion, economic and social costs, as well as policy and management.

The parking search process is fundamentally governed by the principles of supply and demand (Shoup, 2005). The curbside is a monopoly good managed by local governments, making curbside parking pricing a government decision rather than a market-driven one (Manville & Pinski, 2021). The imbalance between the supply of parking spaces and the demand for them at its given price is the first cause of parking search (Inci et al., 2017). In areas with high demand, such as city centers and popular venues, the number of available parking spaces is often insufficient to meet the number of vehicles looking to park. This scarcity leads to a competition for limited parking spots, resulting in drivers cruising around to find an available space (Fulman & Benenson, 2018). This phenomenon is exacerbated in densely populated areas where the supply of parking cannot keep up with the demand.

In cases with sufficient parking supply, parking searches can still occur due to several factors. Uneven distribution and varying driver preferences lead to local mismatches between demand and supply, prompting drivers to search for available spaces (Brooke et al., 2014; Arnott & Inci, 2006). These preferences are not static; they differ from one driver to another and from one journey to the next, and can even change during a trip as search times extend. This dynamic nature adds complexity to the parking search process, making it more challenging to explore and predict.

Additionally, even when parking supply meets or exceeds demand and the available spots align with drivers' preferences, the lack of real-time information about vacant spots forces drivers to continue searching (Arnott & Rowse, 1999). Without knowing the exact locations of available spaces, drivers must engage in a trial-and-error process to find parking.

1.2.2 The Dimensions of Individual Parking Search Behavior

Human behavior and decision-making also play a crucial role in parking search. Almost in all parking search models, drivers are assumed to be rational actors who seek to maximize their utility. When searching for parking, they weigh the costs, such as time spent searching, (walking) distance from their destination, and parking fees, against the benefits, such as proximity to their destination and the safety of the area (Thompson & Richardson, 1998). This decision-making process can lead to parking searches, especially when drivers prioritize free or low-cost parking options (Shoup, 2005).

Psychological factors such as expectation and frustration may further influence parking search behavior (Ponnambalam & Donmez, 2020). Drivers' expectations of quickly finding a parking spot can lead to frustration when these expectations are not met, sometimes resulting in suboptimal parking choices. The uncertainty of finding a parking spot creates a scenario where drivers must constantly reassess their strategies, adding stress and frustration (Vohs et al., 2018). This behavior often disrupts regular traffic flow, as drivers tend to slow down, stop, or circle while searching, contributing to congestion and impacting overall traffic efficiency.

Environmental and urban planning factors significantly impact parking availability, which means parking supply and demand. Urban design and zoning regulations dictate the number and type of parking spaces available. For instance, in Frankfurt am Main in Germany, the implementation of so-called "bike-friendly streets" has significantly reduced the number of available parking spots, aiming to encourage the use of public transportation, walking, and cycling (Knese et al., 2024). This approach prioritizes sustainable transportation options but also impacts the ease of finding parking for those who choose the car.

1.2.3 Effects of Parking Search

Parking search significantly contributes to traffic flow and congestion (Arnott & Inci, 2006). It can lead to increased vehicle emissions and reduced traffic efficiency. From the individual perspective, the time spent searching for parking translates to lost productivity and increased fuel consumption, which has both economic and environmental implications. Socially, excessive parking search can lead to stress and frustration, exacerbating road rage and aggressive driving behaviors (Melnyk et al., 2019).

In addition to the direct impacts of parking search on traffic congestion and emissions, there is also a broader, macro-level effect on parking demand and even car traffic. Millard-Ball et al. (2020) suggest that drivers often anticipate long parking search durations in high-demand areas. As a result, they may start searching for parking earlier, lower their preferences, and settle for a spot sooner if it meets their basic needs. In some cases, drivers might avoid parking in these congested areas altogether. These behaviors help alleviate parking demand, reducing parking search traffic and creating a form of self-regulation (Schäfer et al., 2024), where parking demand naturally decreases in areas expected to have high parking pressure.

1.2.4 Policies to Affect Parking Search

Policy and management strategies can mitigate the impacts of parking search. As explained in Section 1.2.1, there are three categories of causes for parking search. Policies that directly or indirectly affect parking search can be systemized accordingly.

For instance, parking supply can be managed through strategies like residential parking permits (Van Ommeren et al., 2011), time-limited parking zones (Simićević et al., 2013), and no-parking areas (Shiftan & Burd-Eden, 2001). Parking demand, in turn, can be reduced by increasing on-street parking prices, especially when the cost becomes relatively higher than parking garages, effectively reallocating demand away from on-street spaces. Policies affecting mode choice can also affect parking demand; for example, offering discounted public transit subscriptions instead of free parking could encourage more employees to choose public transportation over driving (Tchervenkov, 2022). Apart from this, broader policies affecting car use—such as taxes, city tolls, and the pricing and availability of alternative transportation modes—can further reduce parking demand and the associated parking search.

As mentioned in Section 1.2.1, even if supply and demand are quantitatively equal, there could be parking search due to missing information. For this reason, cities began implementing parking guidance systems in the 1990s to help drivers locate spaces in parking garages (Axhausen et al., 1993). Recently, advances in real-time data analysis have further improved these systems through the development of Intelligent Parking Systems (IPS) (Caicedo, 2010; Teodorović & Lučić, 2006; Shin & Jun, 2014). IPS also play a critical role by using sensors, cameras, and mobile applications to provide drivers with real-time information on available parking spots. Different studies have employed simulation models to assess the effectiveness of this innovation, demonstrating its potential to streamline urban mobility by adapting pricing in response to real-time conditions (Mei et al., 2020; Rodríguez et al., 2022). From early parking guidance systems (Axhausen et al., 1993; Thompson & Bonsall, 1997) to more recent smartphone applications (Rong et al., 2018; Arora et al., 2019; Dalla Chiara et al., 2022), the use of real-time data through IPS can help to overcome the information problem and can reduce the time drivers spend searching for parking and cut down on emissions (Caicedo, 2010; Vlahogianni et al., 2016; Alam et al., 2018).

There are also policy options addressing the third category of causes for parking search, which involves the mismatch with driver preferences. This requires a tailored approach, as these preferences can vary widely. For example, individuals may prioritize different aspects of parking: some may prefer a shorter walking distance to their destination, such as older adults or those with limited mobility, while others with a high value of time may favor reduced search durations (van Ommeren et al., 2021), often choosing paid parking garages to minimize the time spent looking for a spot (Harmatuck, 2007).

Additionally, certain groups have distinct parking needs. People with disabilities, parents with young children, electric vehicle (EV) owners, employees, and hospital visitors all benefit from designated parking arrangements tailored to their specific requirements. For example, larger spaces located closer to entrances accommodate individuals with disabilities or parents managing strollers, while EV owners require spaces equipped with charging stations. By providing targeted parking solutions for these groups, cities can meet diverse parking demands, ultimately enhancing accessibility and possibly affecting search times (Costa et al., 2014; Marsden, 2006; Bonges & Lusk, 2016; Ross & Buliung, 2019; Shaheen et al., 2010).

1.2.5 A brief overview of approaches to modeling parking search

The study of parking search behavior has matured from basic theoretical approaches to sophisticated models that integrate cutting-edge technology and big data analytics. This progression mirrors the advancements in urban planning and computational methods. Here, the evolution of these models is traced, highlighting key studies and methodologies that have shaped this field of research.

Initial explorations into parking behavior were fundamentally theoretical, laying the groundwork for subsequent empirical and modeling efforts (Tamari, 1982; Van Der Goot, 1982; Thompson & Richardson, 1998). The pioneering work of Polak and Axhausen (1990) initially framed parking search behavior within the broader context of transportation decision-making, suggesting that drivers' choices were significantly influenced by their knowledge of the spatial and temporal availability of parking spaces, as well as associated costs. These early models typically emphasized the role of driver experience and expected utility in determining parking decisions, positing that drivers aim to minimize the expected costs of time and parking fees.

From early on, parking choice models became a focal point in the literature. These models analyze the selection process among different parking options, such as on-street versus off-street parking, influenced by factors like cost, distance, and the expected time to find a spot. Axhausen and Polak (1991) were among the first to employ a choice model to examine how these factors drive parking decisions, incorporating the duration of the search as a significant determinant in their model. Subsequent studies (Hunt & Teply, 1993; Sattayhatewa & Smith Jr, 2003; Bonsall & Palmer, 2004; Ottomanelli et al., 2011; Waraich & Axhausen, 2012; Ibeas et al., 2014; Soto et al., 2018) expanded these models to include dynamic elements such as the real-time availability of parking spaces and varying price conditions, enhancing the predictive power and relevance of parking behavior models in urban planning.

The adoption of transport demand simulation models marked a significant evolution in the field, offering researchers tools to explore and predict parking behavior under a variety of hypothetical scenarios. These models, including agent-based models (ABM) (Benenson et al., 2008; Martens & Benenson, 2008; Dieussaert et al., 2009; Waraich & Axhausen, 2012; Horni et al., 2013; Levy et al., 2013; Fulman & Benenson, 2018), discrete event simulation (DES) (Surpris et al., 2013; Alghwiri et al., 2017; Soto Ferrari et al., 2021) , microsimulations (Rodríguez et al., 2022), and macrosimulations (Leclercq et al., 2017; Gu et al., 2020; Zhao et al., 2021), help analyze and predict drivers' behaviors in searching for parking, thereby guiding policy and infrastructure enhancements. Moreover, hybrid models combining ABM, DES, and microsimulation offer a comprehensive view by integrating various modeling strengths (Gallo et al., 2011; Gu et al., 2021). Recent trends include optimizing simulation models with algorithms for parking allocation and integrating real-time data with machine learning to adapt simulations to dynamic conditions. Parking allocation refers to determining the optimal location, distribution, and pricing of parking spaces within a city. This includes algorithms and strategies to identify the best locations for new parking facilities, optimize existing spaces, and develop [dynamic] pricing to balance supply and demand across urban areas. (Jha et al., 2023).

With the widespread availability of GPS technology, researchers gained access to datasets that drastically transformed the study of parking behavior. GPS data enabled the analysis of actual parking search patterns and durations, leading to more accurate and representative models. Montini et al. (2012) is among the first studies that used floating car data (FCD) to estimate the prolongation of travel time (the so-called cruising "excess" time (Young et al., 1991; Weinberger et al., 2020; Geva et al., 2022; Milia et al., 2023)) caused by parking search. Subsequently, numerous studies have leveraged GPS data to analyze driver trajectories, thereby identifying real-world patterns and elucidating the challenges associated with urban parking searches (van der Waerden et al., 2015; Hampshire et al., 2016; Mannini et al., 2017; Weinberger et al., 2020; Millard-Ball et al., 2020; Dalla Chiara & Goodchild, 2020; Mantouka et al., 2021; Milia et al., 2023). This advancement in data utilization has significantly enhanced the understanding of parking patterns in congested urban environments and is also the basis of this PhD thesis.

The latest advancements in the evolution of parking search behavior models leverage advanced machine learning and predictive analytics. These models utilize extensive datasets—such as GPS, sensor, geospatial, and land use data—combined with sophisticated algorithms to predict parking occupancy and behaviors, and to enhance parking management systems. The deployment of machine learning models has improved the prediction of parking availability (Bock et al., 2017; Yang et al., 2019; Saharan et al., 2020) and search times (Jones et al., 2017; Mantouka et al., 2021; Bisante et al., 2023). This progressive approach is likely to reshape how urban parking is managed and optimized in real time.

1.3 Thematic Context and Research Questions

1.3.1 Overall Research Question

Understanding parking search behavior requires accurate quantification, which in turn relies on robust methodological foundations. Existing approaches to measuring parking search duration range from simple surveys to complex models, each with its limitations. Surveys, for example, could gather data explicitly on parking search durations by targeting a specific time, area or demographic group (Lee et al., 2017; Qin et al., 2020; Assemi et al., 2020). However, they often suffer from limitations like sample size constraints and biases related to the memory of the respondents. Field experiments provide greater accuracy by tracking real-time data, but they, too, are limited by the researchers' assumptions about when and where parking searches start (Belloche, 2015; Alemi et al., 2018; Zhu et al., 2020).

Various studies have employed different modeling approaches to quantify parking search duration. A sound approach must capture not only the duration but also the precise starting and ending points of the search. This task is challenging because the "starting point" of parking search is usually not observed by the researchers. Hence, several assumptions are common: In some studies, the parking search start point has been ambiguously defined, ranging from the driver's initial departure to their arrival at the destination, depending on assumptions used. For example, some researchers (e.g., Thompson and Richardson (1998)) suggest that the search begins with the driver's strategic decision at journey onset, whereas others (e.g., Brooke et al. (2014)) propose that it starts only upon reaching the destination. Additional studies assume a specific distance radius around the destination to approximate the start (Weinberger et al., 2020; Montini et al., 2012), or use other heuristics such as speed changes to identify the initiation of parking search (van der Waerden et al., 2015). Such discrepancies highlight the need for a consistent framework for measuring parking search from a clear, empirical standpoint.

Different studies estimate search duration through various proxy variables and assumptions. Although simulations offer a controlled environment to explore parking dynamics, they often rely on generalized assumptions that may not reflect real-world conditions (Waraich & Axhausen, 2012; Gallo et al., 2011; Horni et al., 2013; Fulman & Benenson, 2018). Similarly, analytical models attempt to estimate search duration based on variables that often require assumptions about driver behavior and environmental context, such as parking occupancy (Inci et al., 2017; van Ommeren et al., 2021).

Recent advancements in GPS technology have paved the way for more precise and representative studies on parking patterns. By leveraging GPS data, researchers can track actual vehicle movements and derive insights into parking search behavior at a larger scale. However, using raw GPS data in this context also presents challenges, including high computational demands, potential privacy concerns, and, most importantly, missing some crucial information about parking search behavior: the start and end of the parking search as well as the final destination of the trip (Montini et al., 2012; van der Waerden et al., 2015; Weinberger et al., 2020; Mantouka et al., 2021; Milia et al., 2023). Nevertheless, GPS data offers a valuable tool for capturing the real-world complexity of parking search behavior (and route choice behavior in general), something that traditional survey and experimental methods often lack.

A notable issue is the variation in mean parking search durations in previous studies. For example, survey-based studies frequently report much longer search durations than those based on GPS data. This discrepancy may arise from methodological biases, such as surveys capturing data primarily during peak times or in congested areas, thus skewing the results. Although all the mentioned methods contribute valuable insights, each has its inherent limitations that can affect the accuracy and generalizability of findings. In addition, models often struggle with defining the critical start and end points of a parking search—a fundamental issue, as inconsistent definitions lead to varying and sometimes incompatible results. Without clearly defining these points, any attempt to measure parking search duration risks producing results that lack reliability and comparability across studies.

Table 1.1 provides a comparative summary of the primary methodologies used in parking search research, highlighting the advantages, limitations, and key studies associated with each approach. This overview complements the discussion by systematically outlining the strengths and weaknesses of surveys, field experiments, GPS data analysis, analytical models, and simulations. The table also emphasizes the common limitation across all methodologies regarding the ambiguous or heuristic definition of search starting points, underscoring the need for a more precise framework to enhance accuracy and consistency in parking search studies.

Research Gap: Many existing studies on parking search rely on survey data (van Ommeren et al., 2012; Lee et al., 2017; Assemi et al., 2020; Qin et al., 2020; Brooke et al., 2018), which is subject to inaccuracies due to subjective biases. Factors such as memory lapses, rounding errors, and response biases can significantly distort survey findings, particularly when it comes to quantifying parking search duration. Field experiments (Alemi et al., 2018; Zhou et al., 2004), though more controlled, require substantial logistical effort and rely on arbitrary assumptions that may not capture the diverse range of real-world behaviors. GPS-based estimations of cruising durations often determine only cruising excess time rather than the actual search route and duration (Montini et al., 2012; Mannini et al., 2017; Weinberger et al., 2020; Dalla Chiara & Goodchild, 2020; Dalla Chiara et al., 2021, 2022). Even attempts to trace the search trajectory using GPS data may be flawed due to arbitrary assumptions, like defining a fixed radius around the parking spot (Bisante et al., 2023; van der Waerden et al., 2015; Milia et al., 2023; Mantouka et al., 2021). Analytical models (Inci et al., 2017; van Ommeren et al., 2021) and simulations (Waraich & Axhausen, 2012; Gallo et al., 2011; Horni et al., 2013; Fulman & Benenson, 2018) suffer from similar limitations, as they often rely on assumed variables rather than ground truth data. Across these approaches, the overall research gap is a lack of a methodology that clearly defines search starting and ending points and measures and collects the search duration and route without heuristic assumptions.

Research Question: How can parking search behaviors be measured and analyzed empirically to accurately capture and explore search durations, search starting points, and search routes?

Research Objective: The overall objective is to develop an empirical approach to recording the start and end of parking searches without any heuristic assumptions (or realistically with minimum possible assumptions). By minimizing reliance on assumptions, the methodology utilizes GPS technology to capture real-world parking search behaviors across entire journeys, from driving to parking and the final walking segment to the destination. By capturing the full

Table 1.1: Comparison of Methodologies in Parking Search Quantification Research

Methodology	**Advantages**	**Limitations**	**Studies**
Surveys	Easy to administer and can target specific times, areas, or demographics	Subject to memory bias, response bias, and limited accuracy due to self-reporting	Assemi et al., 2020, Qin et al., 2020, Cao et al., 2019, Brooke et al., 2018, Cookson and Pishue, 2017, Lee et al., 2017, Belloche, 2015, Holguín-Veras et al., 2016
Field Experiments	High accuracy with real-time, observed data	Logistically complex, often limited in scope and sample size, and assumptions about search start points	Zhu et al., 2020, Alemi et al., 2018
GPS Data Analysis	Provides high precision, real-world context, and large-scale data	High computational demand, privacy concerns, and limited in defining start and end points of search	Milia et al., 2023, Dalla Chiara et al., 2021, Weinberger et al., 2020, Dalla Chiara and Goodchild, 2020, Mannini et al., 2017, Mantouka et al., 2021, Hampshire et al., 2016, van der Waerden et al., 2015
Analytical Models	Can simplify complex phenomena into quantifiable elements	Often rely on assumptions about driver behavior and parking availability that may not reflect real-world conditions	van Ommeren et al., 2021, Fulman and Benenson, 2021, Inci et al., 2017
Simulations	Allows for controlled testing of various scenarios and can model hypothetical conditions, great tool for assessing policies	Results depend heavily on assumptions, may lack real-world validity, and typically require significant computational resources	Fulman et al., 2020, Cao et al., 2019, Fulman and Benenson, 2021, Waraich and Axhausen, 2012, Horni et al., 2013

General Limitation: In all methodologies, the exact starting point of the parking search is often unknown or defined heuristically.

trajectory of parking searches, this method offers a level of precision that previous approaches have lacked. Ultimately, this approach provides a holistic view of parking search processes and enables a more detailed exploration of parking search behavior. In addition, this framework lays the groundwork for deeper inquiries into parking search behaviors, allowing for further research questions and a better understanding of the factors at play. For example, it could provide insights into factors like the determinants of the search duration and the search starting points. This comprehensive data collection also contributes to practical applications in urban planning and policy-making. For instance, cities could use this information to optimize parking allocation, implement dynamic pricing strategies, and reduce the negative impacts of cruising for parking on urban congestion and emissions.

1.3.2 Specific Research Questions

Paper I: Initiation of Parking Search

Research Gap: Previous studies have not sufficiently analyzed the initiation point of the parking search, a critical aspect that could potentially span from the start of the journey to reaching the destination. This initiation point has only been briefly discussed in theoretical terms (Thompson & Richardson, 1998; Horni et al., 2013; Millard-Ball et al., 2020), but lacks data collection and empirical analysis.

Research Question: What factors (e.g., location, time, and distance to the final destination) influence drivers to start the parking search process?

Research Objective: The proposed data collection method makes it possible to investigate the transition from normal driving to parking search empirically for the first time. By utilizing a logistic regression based on a simple microeconomic model of a rational driver minimizing travel duration, the factors that influence the initiation of the parking search can be examined to understand what affects drivers' decision to start searching for a parking spot sooner or later.

Paper II: Determinants of Parking Search Duration

Research Gap: The overall research gap described above results in empirical studies aiming at identifying the determinants of parking search duration being unreliable. Consequently, models based on such data fail to accurately explain the factors influencing parking search. Furthermore, statistical models that better fit the nature of the data, like survival analysis, could potentially offer better insights than simple regression models (Mantouka et al., 2021; Zhu et al., 2020; Fulman et al., 2020).

Research Question: What are the determinants of the parking search duration, and what are their quantitative effects?

Research Objective: Utilize parking search data collected by the proposed data collection approach to model the duration of cruising for parking. The survival model, known from statistical and econometric analyses, should identify factors influencing parking search duration, including time-varying factors changing during the search. This enables, for example, to estimate the search-duration-dependency of the probability of finding a parking spot. The estimated parameters and derived conclusions are vital for policymakers, enabling them to investigate drivers' parking behavior under different circumstances and develop measures to address cruising for parking.

Paper III: Identifying Parking Search in GPS data

Research Gap: Previous models for identifying parking search in GPS traces have primarily relied on heuristic approaches, such as threshold methods (radius around the parking spot) (Montini et al., 2012; Mannini et al., 2017; Weinberger et al., 2020; Dalla Chiara & Goodchild, 2020; Dalla Chiara et al., 2021, 2022; van der Waerden et al., 2015; Milia et al., 2023; Bisante et al., 2023; Mantouka et al., 2021). These models are based on arbitrary assumptions such as a radius of 200, 400 or 800 m.

Research Question: To what extent can machine learning classify GPS points of trajectories in "parking search" and "normal driving"? Is a prediction model suitable for exploring cruising for parking in historical GPS data?

Research Objective: Develop and validate a machine learning model specifically designed to identify parking search in unlabeled GPS data using ground truth data on parking search behaviors. Test the model's performance against existing heuristic rules and external validation data, ensuring reliability and accuracy. Apply the model to a large-scale historical GPS dataset to derive aggregated statistics for a city as a case study, showcasing its capability to provide insightful urban mobility analytics and its adaptability across different urban contexts.

Figure 1.2 illustrates the overall conception of the research and the specific contributions of each paper within the parking search process. The top bar represents the sequential stages of the journey—Normal Driving, Searching, and Walking—emphasizing the research's overarching goal to adequately conceptualize, model, measure, and collect data across these phases. This comprehensive approach forms the foundation of the overall research framework.

The bottom bar maps out the specific contributions of each research paper, showing how each paper integrates into the broader framework while focusing on distinct aspects of parking search behavior. Paper 1 explores the transition from Normal Driving to Searching, identifying the factors that influence when drivers begin to search for parking. Paper 2 delves into the determinants of cruising time within the Searching phase, examining variables that affect the duration of the search. Paper 3 focuses on predicting cruising time, adding a layer of predictive modeling to understand how long drivers might search based on various conditions.

This layered structure not only highlights the relation of the papers but also demonstrates how each study complements the others by addressing different facets of parking search behavior. Together, they should form a cohesive whole that advances the overall research conception,

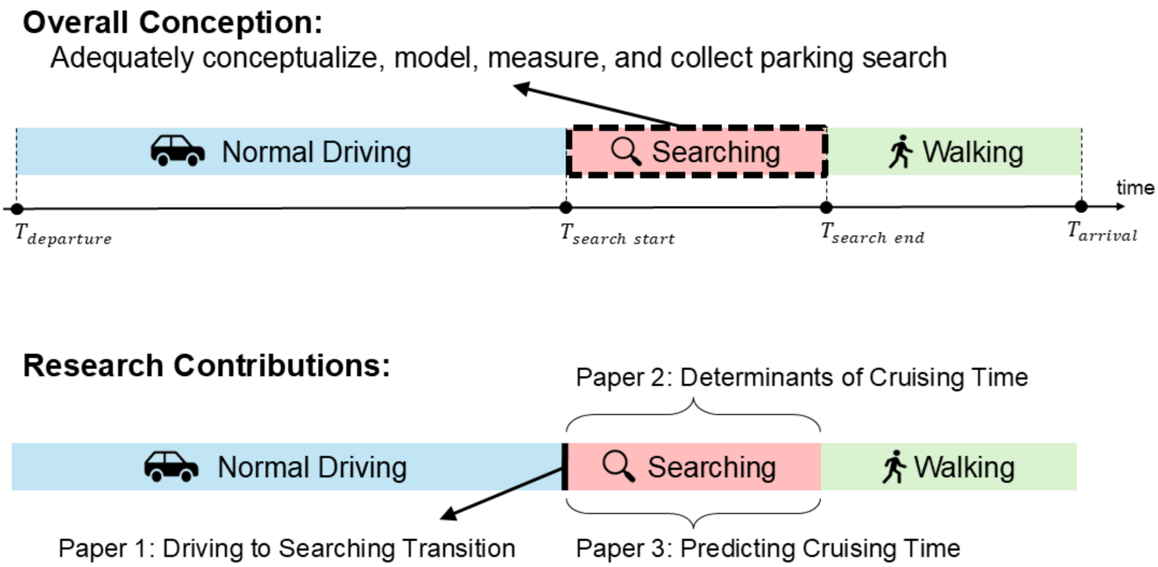


Figure 1.2: Overall Research Conception and Contributions of Research Papers Illustrating the Relationship Between Individual Papers and the Overall Research Framework

enabling a comprehensive understanding of parking search dynamics and facilitating a thorough exploration of this complex process.

The flowchart in Figure 1.3 summarizes the primary objective of this research in another way: to develop a comprehensive empirical framework for understanding parking search behavior. This overarching framework enables a detailed exploration of three key aspects, each represented by one of the research papers.

1.4 Overall Conception and Methodology

1.4.1 Overall Methodology: Data Collection Framework

Development of an app:

To address the overall research question, a new approach is proposed to collect cruising data using a mobile app, called Start2park ("Start2Park Research Project", 2024).

The app operates through volunteer drivers who activate it at the start of their car journey. What sets this app apart is its feature that allows drivers to manually log the initiation of the parking search while driving, mark the end of the parking search upon finding a parking spot, and finally end the recording once they reach their final destination on foot. This method segments the journey into three distinct parts: driving until the search begins, the parking search, and walking to the final destination.

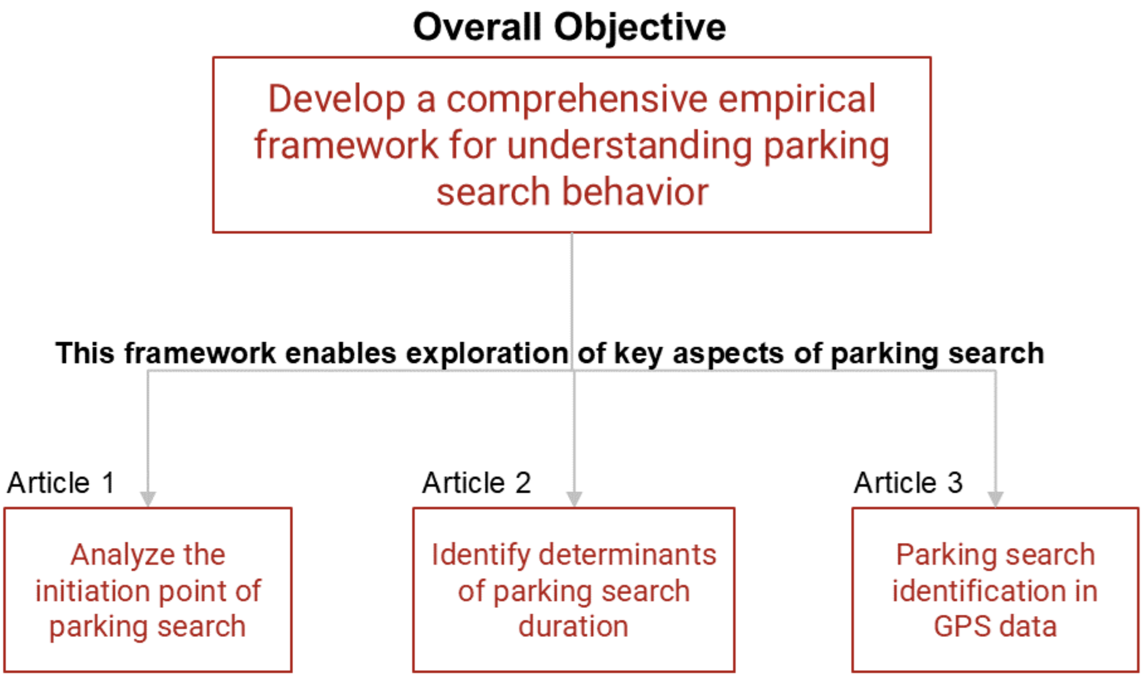


Figure 1.3: Flowchart of Research Objectives for Parking Search Behavior

The Figure 1.4 illustrates the data collection methodology employed through the Start2park app, which is designed to segment a journey into three phases—Normal Driving, Searching, and Walking—allowing for precise tracking of parking search behaviors. On the left, screenshots from the Start2park app showcase the interface that volunteers use to manually log each phase of their journey. Drivers activate the app at the start of their trip and subsequently mark the beginning and end of the parking search, concluding the journey upon reaching their final destination. This results in the journey segmentation timeline, depicted in the top part of the Figure. These phases are marked by time stamps, providing an organized overview of each stage in the journey. To the right, a sample recorded journey visualizes the driver's path through each phase of the process on the map. Each color-coded segment on the map corresponds to a specific phase: blue for Normal Driving, red for Searching, and green for Walking. This allows for deriving key statistics such as parking search duration and parking search length, as shown in the bottom part of the Figure.

The mobile application developed for collecting parking search data aligns closely with the methodologies outlined in the Steinmeyer et al. (2011) (EVE) document, which emphasizes the importance of accurate and empirical data collection methods in traffic studies. The app categorically fits into the "measuring" methodology by recording precise, quantitative data regarding parking search durations. This ensures adherence to EVE's standards of minimizing assumptions and biases in data collection, enhancing the reliability and validity of traffic data pertinent to parking behaviors. By capturing exact times and GPS-coordinated locations from the initiation to the conclusion of parking searches, the app results in objective and replicable data, thereby adhering to the EVE document's emphasis on empirical validity.

Throughout this thesis, the data collected by this app is referred to as "ground truth" data regarding on-street parking searches. In the realm of data science and machine learning, "ground truth" refers to accurate, well-labeled data that serves as a benchmark for validating models and conducting thorough analyses (Jordan & Mitchell, 2015; Stoyanov et al., 2018). This type of data is considered reliable for training algorithms because it is assumed to be the most accurate representation available of what is being measured. However, it's crucial to recognize that the term "ground truth" does not imply that this data captures the absolute and complete reality of parking search behavior. It merely indicates that within the scope of the data collection methodology, the data is as accurate and precise as possible.

This notion of "ground truth" signifies that the data is trustworthy within the limits and context of the technology and methods employed. It's essential for this research because it provides a foundation for the empirical analyses and helps in the development of predictive models. Nonetheless, while this data is invaluable, it's also shaped by the specific interactions and inputs from users—such as pressing a button at certain times—which introduces certain limitations and biases. Therefore, when interpreting the findings and conclusions, it is important to consider these inherent limitations and assumptions. The data is robust within its context, but like all empirical data, it is subject to the constraints and conditions under which it was gathered.

Limitations of the app and its measurements:

Although the app introduces a new method for measuring parking search, the validity of its measurements largely depends on user compliance and behavior. The following discussion explores the challenges associated with the app and the potential biases these challenges may introduce.

- User dependency, delayed interactions, and forgotten interactions: The validity of the data recorded by the app heavily relies on the driver's consistent and timely interaction. Each phase transition—starting the journey, beginning the parking search, marking the parking spot, and signaling the arrival at the final destination—necessitates manual input from the driver. This requirement introduces a variable quality of data collection, heavily dependent on individual user behavior. Real-world scenarios often result in delayed interactions due to distractions, the complexity of driving tasks, or simple forgetfulness, which can lead to inaccurately recorded search durations and, ultimately, biased data. For instance, a driver might begin a trip but forget to activate the parking search function in the app. Upon exiting the parked vehicle and noticing the smartphone app, the driver might rapidly start and then end the parking search, leading to a recorded parking search duration of zero. Identifying and correcting such inaccurate records poses a significant challenge in the data-cleaning process.
- Behavioral variability: Different drivers may have different levels of engagement and reliability when interacting with the app. Some may be meticulous about pressing buttons at the correct times, while others may be sporadic or generally inattentive to these tasks. For example, a distracted driver starts looking for parking but only remembers to press

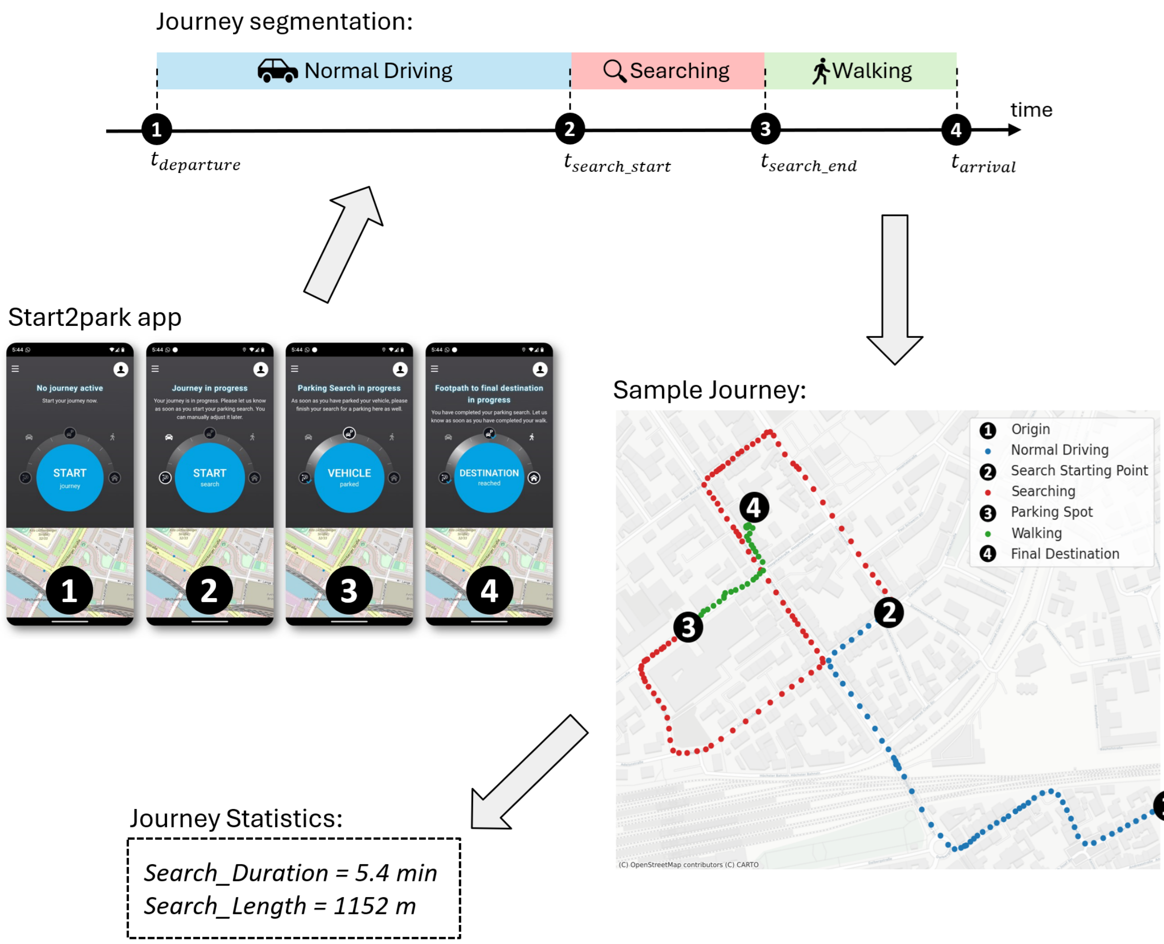


Figure 1.4: Data Collection Framework Using the Start2park App for Recording and Segmenting Entire Journeys and Analyzing Parking Search

the "Start Search" button one minute into their search and thinks this would still be fine. This causes data collected to inaccurately portray a shorter parking search duration. The different behaviors may also be caused by a different understanding of the app. For example, some drivers may end the parking search upon spotting a suitable vacant parking spot, while others may end the search only after parking the vehicle. This variability can introduce a systematic bias where the data quality is correlated with the user's conscientiousness or familiarity with the app rather than solely reflecting parking search behavior.

- Behavioral Modification: Using the app while looking for a parking spot could alter the drivers' behavior in different ways. For example, knowing they need to interact with the app at certain points, drivers might start the parking search earlier than usual to ensure they remember to press the "Start Search" button, or they might choose parking spots that they perceive as simpler options for ending the search phase cleanly (e.g., opting for a parking lot rather than street parking to have a clear "Vehicle Parked" moment). Another factor to consider is the Hawthorne Effect, where individuals modify their behavior because they are aware they are being observed (Adair, 1984). This phenomenon is particularly pertinent in this study as drivers know their actions are monitored and recorded. For example, drivers may avoid illegal parking due to their awareness of being observed. These behavioral changes present a substantial challenge in gathering genuine data in research that involves active participant monitoring. The data collected in such settings may not solely represent natural parking behaviors but also responses to being observed, potentially leading to conclusions and recommendations that do not accurately reflect real-world behaviors.
- Influence on Driving Safety: Efforts have been made to enhance the safety of the app by designing the interface to be as user-friendly as possible. This includes featuring a large, prominent button for ease of use and minimizing the need for interactions while driving. The app requires only a single button press by the driver to initiate the parking search phase, with all other inputs designed to be completed when the vehicle is stationary. Additionally, users are provided with safety instructions, recommending the use of a smartphone holder to secure the device in a position that allows for safer interaction. Despite these precautions, the necessity for any manual interaction with the app during driving presents inherent safety risks. These risks are exacerbated under challenging driving conditions, such as in heavy traffic or adverse weather. The act of interacting with the app—even with a single button press—could potentially divert the driver's attention away from the road. This distraction may lead to unsafe driving behaviors, including removing hands from the steering wheel or eyes from the road, thereby increasing the likelihood of accidents. This, on the other hand, may affect data quality: Drivers in challenging driving conditions may use the app in another way than under "normal" driving conditions. This behavior may systematically bias the results, especially if driving conditions are correlated with parking search durations.

Data Collection Assumptions:

The data collection is based on several explicit and implicit assumptions, which are discussed below.

By construction, the app segments the journey into distinct phases (”normal” driving, searching, walking). Thereby, the initiation of the search is treated as a discrete ”one-way” event. This methodology significantly simplifies a complex and dynamic behavior into a binary condition, where drivers are categorized as either searching for parking or driving ”normally”. Once a search starts, it is by construction of the app not possible to switch back to normal driving. This assumption—that once starting the search, drivers do not return to ”normal” driving—fails to account for drivers cycling between searching and normal driving in response to real-time decisions influenced by traffic, time constraints, or changing priorities during the journey. Such oversimplifications do not fully capture the dynamic aspects, complexities, and nuances of actual parking search behaviors.

In reality, the initiation of a parking search is rarely a clear-cut and binary event. Drivers might start scanning for spots opportunistically as they approach their destination, but not engage in a focused search immediately. For example, a driver might start looking for parking spots casually while they are still mainly focused on getting to their destination. This early phase can’t be completely marked as “searching” or “not searching” and the exact transition is unclear. This behavior resembles a quick check for parking with some willingness to park if they see a good spot. Therefore, this initial phase blurs the lines between general navigation and the active search for parking, challenging the notion of marking a specific “start” point for parking search.

Furthermore, drivers might switch from parking search mode back to normal driving. For instance, this can happen if they realize they have driven too far from their destination, exceeding their preferred walking distance. In response, they would temporarily pause their search, move closer to the destination, and resume the search from a more suitable location. Urban driving conditions can also make this pattern more complex. For example, when drivers find themselves on streets where parking is prohibited or unavailable, they must pause their parking search and focus solely on driving through these areas. Once they return to a location where parking is feasible, they can continue their search. This intermittent searching behavior challenges the concept of a clear distinct parking search from the starting point until the vehicle is parked.

If parking search cannot be captured by a binary variable, a more sophisticated approach would collect and model parking search as a continuous variable, reflecting the dynamic and evolving nature of driver behavior. This approach would recognize the gradations in search intensity, influenced by dynamic factors such as the proximity to the destination, the density of parked cars, or the legality of parking in certain areas.

In Paper III (Saki & Hagen, 2024b), a deep learning model is developed to identify parking searches in vehicle GPS traces. This model generates a probability value for each GPS (observation) point, indicating its likelihood of being a parking search point or a ”normal” driving point. This probabilistic approach provides a more nuanced view of the parking search behavior, allowing for fluctuations in search probability throughout the journey. However, it is important

(a) (b)

Figure 1.5: (a) shows GPS tracking data labeled as "searching for parking" or "not searching" (b) depicts the outcomes of the binary classification model

to recognize that this model is limited by its training on a binary variable collected using the existing methodology.

Figure 1.5a illustrates a journey recorded by the app, showing the actual data labels. This highlights the distinctive data collection methodology. Figure 1.5b demonstrates the results of the binary prediction model proposed in Saki and Hagen (2024b).

Figure 1.6 further explores this by showing the predicted probabilities of being in a state of parking search at various GPS points along a driver's route according to the model in Paper III. Factors such as lower driving speeds and shorter remaining distances to the destination typically increase these probabilities. This figure illustrates the complex interplay between speed, location, and parking search behavior.

In conclusion, parking search was recorded as a binary variable (normal driving versus searching) and also the final result of the prediction model is a binary variable. However, the examination of the predicted (continuous) probabilities of the model provides an impression of the continuous character of parking search decisions, which cannot be deepened within the scope of this thesis.

Data Cleaning:

The data cleaning process is a critical component of this study, serving as the foundation for all subsequent analyses and model-building efforts. Starting with the extraction of raw data from the app's database, meticulous preprocessing steps were undertaken to ensure the data's accuracy and relevance, preparing it for analytical tasks. Below are some of the key steps and strategies employed in the data cleaning process, including the systematic identification and removal of errors.

- Completeness of Journeys: Only journeys that are fully recorded are considered valid.

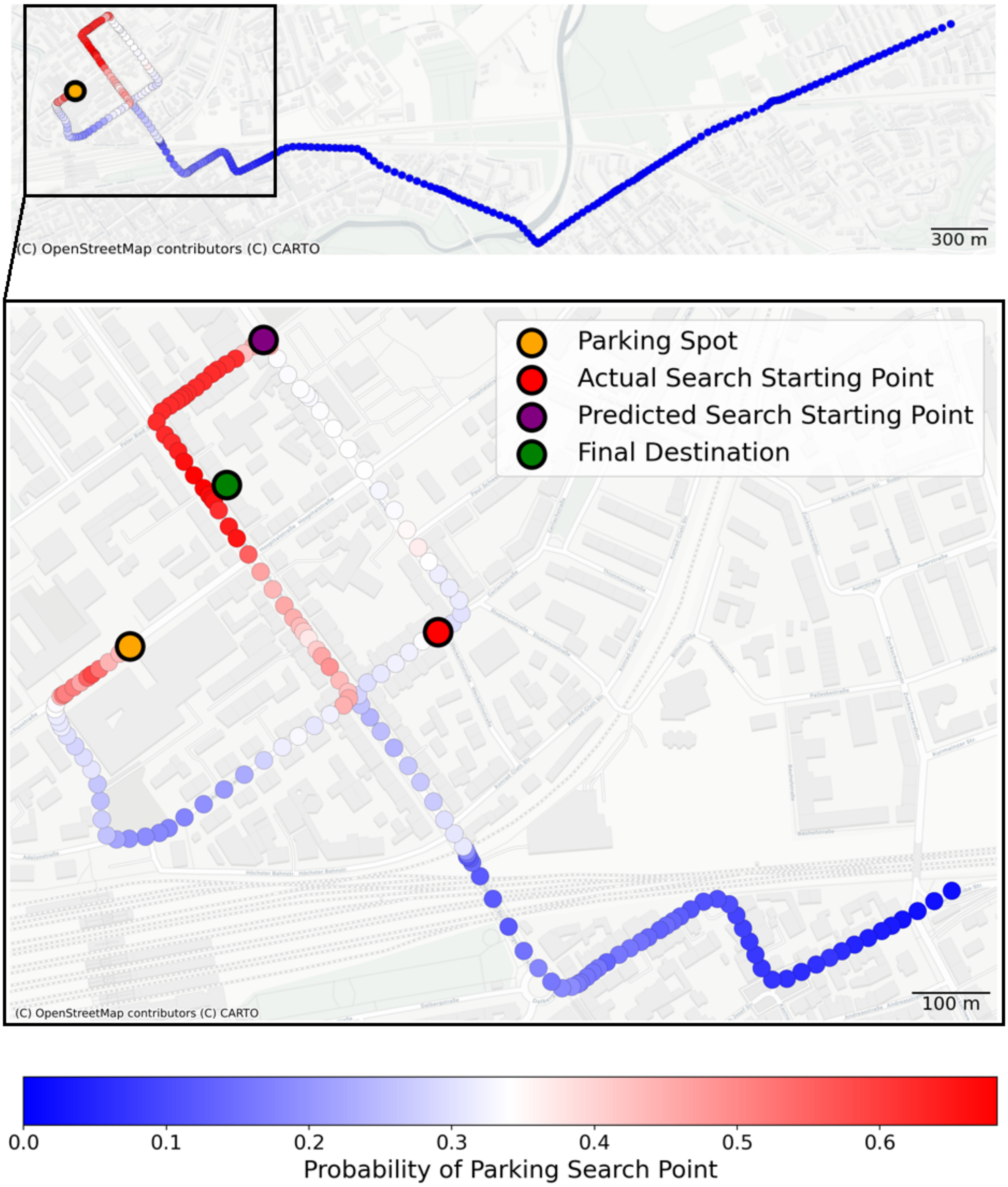


Figure 1.6: Probability of parking search at various points along a route, illustrating how factors like speed and proximity to destination influence parking search behavior.

Any journey that is not complete, perhaps because a user aborted the process midway, is marked incomplete and removed from the dataset. This step ensures that all analyzed data represent full user interactions with the app.

- Removal of Test Data: Some users, often to test the functionality of the app, rapidly click through the record buttons, resulting in unrealistically short journey durations of just a few seconds. These records are easily identified by their short duration and are excluded from the dataset as they do not represent genuine user behavior.
- Exclusion Based on Anomalous Usage: Some users are observed with usage patterns that did not align with the intended purpose of the app. For instance, one user logged over a hundred trips where the durations for parking search and walking were consistently zero, indicating incorrect use. These data points were excluded as they likely represent misuse or misunderstanding of the app's functionality.
- Thresholds Used for Data Cleaning: To ensure the integrity and reliability of the ground truth dataset, specific exclusion criteria are set based on these ratios and additional parameters. Any journeys lasting less than three minutes in journey duration or exceeding thirty minutes in parking search duration were removed, as these were considered either too brief to be meaningful or unrealistically long. In addition, records with a total journey distance less than 500 m are excluded, which indicated an insufficiently substantial journey for analysis, or where the parking-destination clearance distance exceeded 1000 m, which could skew the overall data with unusually long distances.
- Complete Journeys Marked as Searching or Walking: Occasionally, users inadvertently press the record button multiple times at the beginning of their journey before starting to drive. This results in the entire journey being incorrectly labeled as either parking search or walking in the dataset. To address this issue, ratio of time spent searching or walking to the total journey time is analyzed. Specifically, if more than 80 percent of a journey was categorized as searching or walking, this is identified as a likely error due to multiple button presses at the start. Such journeys were considered unreliable for the analysis and were therefore removed from the dataset to maintain the accuracy and quality of the training data.
- Driver forgetfulness: This is a notable challenge in accurately capturing data on parking search behaviors. It's common for users to start recording a journey with the app and then forget its operation during the normal driving phase. They may only remember the app upon reaching their destination, leading to erroneous records. Some users might simply close the app, resulting in an easily detectable incomplete journey. However, others may end the recording by quickly clicking through the app buttons, typically resulting in long driving durations and zero durations for search and walking.

  Similar issues may arise during the parking search and walking phases. For instance, forgetting to log during the parking search results in an extended parking search duration paired with zero walking duration. Alternatively, forgetting during the walking phase might record accurate driving and searching routes, but an incorrect walking path. To identify such instances, average speeds are calculated for each phase by dividing the

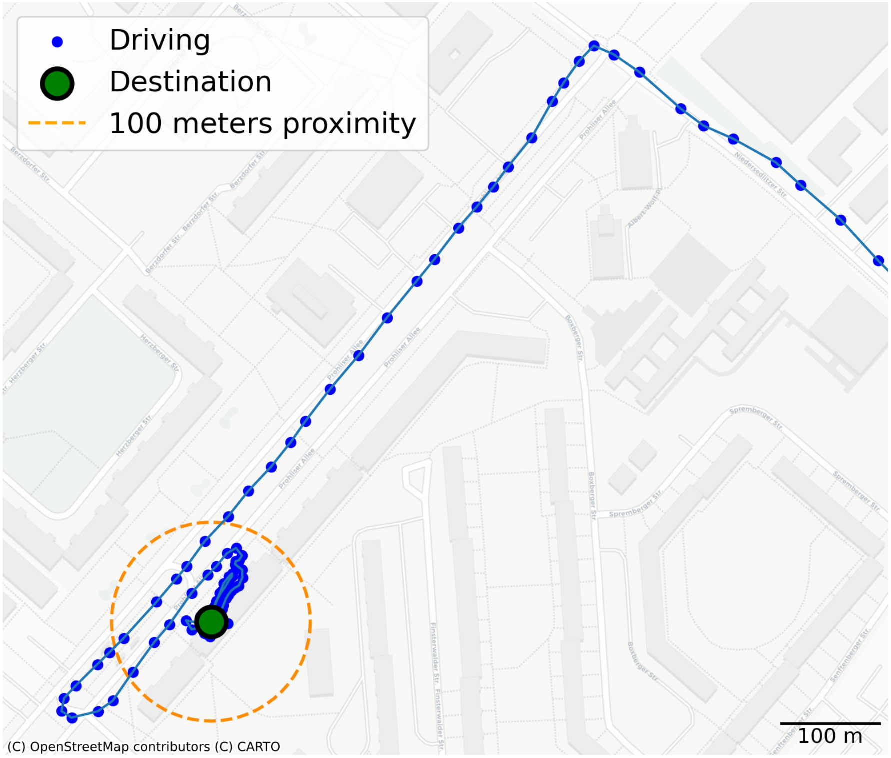


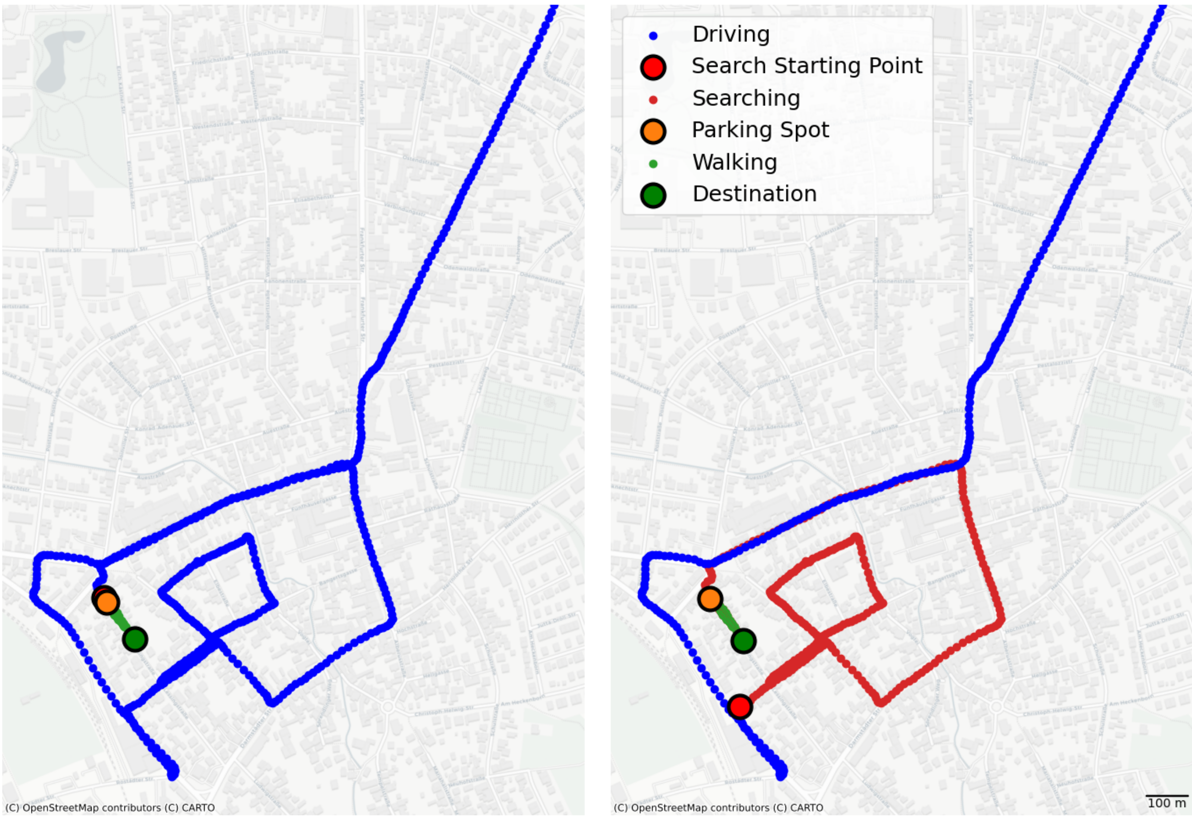
Figure 1.7: A case where a user forgot to log the driving phase, remaining stationary within 100 m of the destination for about 27 minutes.

undertaken and origin-destination (OD) distances by its duration, allowing us to detect anomalies. Additionally, It is assessed how long a user remains within a proximity of the destination, which could indicate forgetfulness. An illustrative case is shown in Figure 1.7, where it's evident that the user forgot to use the app during the driving phase, remaining within 100 m of the destination for approximately 27 minutes.

More complex scenarios include occasions where drivers start a journey but forget to log the beginning of their parking search, only remembering to do so at the moment they park. This results in a reported zero parking search duration, though the driving and walking phases appear normal. Detecting such omissions is challenging, but in these cases, the parking prediction model (Saki & Hagen, 2024b) is used as a tool. By comparing the parking search durations predicted by the model with those actually recorded, significant discrepancies can be identified.

In cases where a notable difference was observed, a manual and visual review of the trips is conducted. Particularly, by examining the driving trajectories towards the end of a trip, especially complex patterns around the destination area, it often became clear that a driver had forgotten to log the search phase. Such instances where the parking search was evidently undertaken but not recorded were removed from the dataset to ensure the integrity of the analysis.

Figure 1.8 provides an illustrative example of this issue. Figure 1.8a displays a trip where no parking search was logged (showing zero duration), whereas Figure 1.8b presents the results from the parking search prediction model, which suggests that a search likely occurred. This contrast helps to highlight trips where logging errors due to forgetfulness might have skewed the original data.



(a) Actual Observations

(b) Predictions

Figure 1.8: (a) A sample journey where no parking search was logged, displaying a zero search duration but a positive walking duration. (b) Results from the parking search prediction model, suggesting that a parking search likely occurred, revealing potential discrepancies due to forgetfulness in logging.

1.4.2 Specific Methodologies

Paper I: Initiation of Parking Search

The methodology employed in this study focuses on understanding the factors influencing the initiation of parking search by utilizing a multinomial logit model. This approach was chosen to capture the decision-making process involved in transitioning from normal driving to actively searching for parking, and in some cases, parking immediately without searching. The primary motivation behind this methodology is the need for an empirical model that can realistically simulate driver decisions based on observable factors such as proximity to the destination and traffic conditions. The multinomial logit model is a widely used tool for analyzing decision-making processes in transportation studies, due to its flexibility in modeling choices among discrete alternatives (Cirillo & Xu, 2011).

This approach aligns with Random Utility Theory (Manski, 1977; Cascetta & Cascetta, 2009), which provides a structured framework for modeling decision-making under uncertainty. It posits that drivers seek to minimize the total costs (which is equivalent to maximizing utility) associated with their journey, with costs being influenced by variables related to driving, searching, and walking. By framing parking search initiation as a utility maximization problem, this study leverages established economic principles, typically applied in microeconomic models of decision-making, to examine how drivers weigh the trade-offs between continuing to drive and initiating a search for parking.

A guiding research hypothesis is that the likelihood of initiating a parking search increases as drivers approach their destination. This hypothesis is grounded in both theoretical and empirical literature, suggesting that factors like traffic congestion, parking availability, and driver characteristics significantly impact the timing of search initiation (Shoup, 2005; van Ommeren et al., 2021; Zakharenko, 2016). Variables such as proximity to the destination, familiarity with the area, and driver demographics are expected to play key roles in this process.

This study contributes to the literature on parking behavior by addressing a gap left by earlier models, which often relied on static assumptions about when a search begins. As highlighted by Millard-Ball et al., 2020, the uncertain nature of the parking search starting point complicates the analysis of cruising behavior. Horni et al., 2013 similarly state that the origin of the parking search "cannot be specified sharply". Different scholars have proposed varied conceptualizations of this starting point. For instance, Thompson and Richardson, 1998 suggest it begins with the driver's decision on a parking strategy, while Brooke et al., 2014 assume it starts upon reaching the destination. Other studies, such as those by Weinberger et al., 2020, Montini et al., 2012, and Kaplan and Bekhor, 2011, have used distance-based thresholds ranging from 100 to 400 m.

Unlike previous models, this approach leverages real-world behavior recorded through the start2park-app described in Section 1.4.1. By using ground truth data on parking search initiation within a multinomial logit framework, this study offers a dynamic and data-driven perspective on parking search behavior. This methodology places the study within an emerging field of research that combines smart data collection, big data, and econometric techniques to enhance the understanding of parking dynamics.

Paper II: Determinants of Parking Search Duration

This study employs survival analysis (Jenkins, 2005; Clark et al., 2003) with a competing-risks model (Austin et al., 2016; Satagopan et al., 2004) to investigate the determinants of parking search duration. This approach was selected to capture the time-dependent nature of parking search behavior and to address multiple possible outcomes—such as finding Free, Paid, or Illegal parking. By enabling a simultaneous examination of these interrelated events, this methodology provides a comprehensive view of how various factors influence not only the duration of the search but also the type of parking ultimately chosen.

The analysis is grounded in utility maximization (Aleskerov et al., 2007), which propose that drivers make parking decisions based on a trade-off between costs (such as time, walking distance, and parking fees) and benefits (such as convenience and perceived safety) (Axhausen & Polak, 1991). The central hypothesis of this study is that the likelihood of a driver selecting a particular parking type, given the duration of the current parking search episode so far, depends on contextual factors, including the time of day, proximity to the destination, and prior parking experiences (Hilvert et al., 2012; Kobus et al., 2013; Hess & Polak, 2004; Ibeas et al., 2014; Qin et al., 2017; Soto et al., 2018). It is expected that as search duration increases, drivers may adjust their preferences, potentially becoming more inclined toward Paid or even Illegal parking options due to factors such as frustration or urgency. The model's time-varying baseline hazard function allows for exploration of this aspect, capturing changes in driver behavior as search duration extends, called "duration dependence of the hazard rate".

This methodology distinguishes itself from previous studies on two fronts. First, it utilizes again high-resolution GPS data to capture real-time parking search behavior, contrasting with research that largely depended on survey data (van Ommeren et al., 2012; Lee et al., 2017; Assemi et al., 2020; Qin et al., 2020; Brooke et al., 2018), field experiments (Alemi et al., 2018; Zhou et al., 2004) or unlabaled GPS data (van der Waerden et al., 2015; Dalla Chiara & Goodchild, 2020; Mantouka et al., 2021).

Second, this study applies advanced modeling techniques that better address the complexities of parking search behavior. Earlier models, such as the one by Dalla Chiara and Goodchild, 2020, calculated cruising excess time from GPS data without a well-defined starting point for parking search and used Ordinary Least Squares (OLS) linear regression to quantify the determinants. While some recent studies have employed survival analysis for parking search duration modeling (Zhu et al., 2020; Fulman et al., 2020; Mantouka et al., 2021), they generally fail to differentiate among parking types or account for the dynamic nature of driver preferences during the search. In contrast, the competing-risks survival analysis with a time-varying baseline hazard model used here is particularly suited to the hypothesis, as it allows for simultaneous analysis of multiple outcomes while accounting for the duration-dependent nature of decision-making. This advanced approach offers a more precise estimation of factors affecting parking search duration and provides insights into the decision-making process that are challenging to capture through simpler models.

Paper III: Predictive Model to Identify Parking Search in GPS Data

The ground truth data collected by Start2park consists of sequential GPS points, each labeled as "normal driving," "parking search," or "walking." To accurately classify each point, it is essential to consider the sequential nature of the data, as well as the attributes of preceding points. Various models are suited to this type of data, including multinomial logit models (Kwak & Clayton-Matthews, 2002), feedforward neural networks (Bebis & Georgiopoulos, 1994), recurrent neural networks (RNNs) (Medsker, Jain, et al., 2001), and long short-term memory (LSTM) networks (Yu et al., 2019). Given that the primary objective is to achieve the highest possible prediction accuracy rather than to quantify the effect of specific attributes, machine learning models emerged as more suitable than statistical / econometric models.

After comparing feedforward neural networks with RNNs and LSTMs, the simpler structure of feedforward neural networks proved advantageous, as they demonstrated similar performance while being easier to configure, train, and deploy in practical applications. Based on these considerations, the final methodology employs a feedforward neural network architecture, which balances predictive power with practical applicability.

The study hypothesizes that variables such as speed and distance to the destination, alongside contextual factors like traffic conditions and journey characteristics, influence the search strategy (Axhausen & Polak, 1991; Hilvert et al., 2012; Kobus et al., 2013; Hess & Polak, 2004; Ibeas et al., 2014; Qin et al., 2017; Soto et al., 2018). These variables should provide sufficient information to build a robust prediction model capable of identifying parking search segments within a GPS trajectory, ideally surpassing the accuracy of existing models.

The final model focuses on speed and distance to the destination based on two main considerations. First, a feature importance analysis revealed that these two variables are the most critical, with speed serving as a proxy for various contextual factors influencing parking search behavior (van der Waerden et al., 2015). Second, by restricting the model to speed and distance, the model's generalizability and transferability are significantly enhanced, as these two variables are commonly found in most GPS datasets and can be easily calculated if not directly available. This ensures that the model can be utilized with different datasets, making it accessible and practical for broader applications. The generalizability and transferability of the model were verified through testing on an independently labeled dataset, collected using a different app and diverse driver profiles.

The increasing availability of GPS data has prompted numerous studies to analyze parking search behavior through various methodologies. Many of these studies focus on estimating cruising excess time without explicitly labeling parts of a journey as "parking search" (Montini et al., 2012; Mannini et al., 2017; Weinberger et al., 2020; Dalla Chiara & Goodchild, 2020; Dalla Chiara et al., 2021, 2022). Among those that attempt to identify parking search segments, heuristic-based methods have been common. These include the fixed radius method (Bisante et al., 2023), where the search is assumed to begin within a certain distance of the destination; the speed threshold method (van der Waerden et al., 2015; Milia et al., 2023), which uses a reduction in speed within a defined radius; and the local minima approach (Mantouka et al., 2021), where a parking search starts when the distance to the destination first increases after decreasing within a set radius.

These heuristic approaches, while useful, are limited by their reliance on fixed assumptions. The proposed ML model was rigorously tested against existing methodologies and demonstrated superior performance in identifying parking search segments within GPS trajectories.

The proposed ML model introduces a framework for detecting parking search behavior in unlabeled historical GPS data, offering value for urban planners, policymakers, and researchers.

2 Article 1: What drives drivers to start cruising for parking? Modeling the start of the search process

Authors	Siavash Saki, Tobias Hagen
Affiliation	Research Lab for Urban Transport, Frankfurt University of Applied Sciences, Frankfurt am Main, Germany
Journal	Transportation Research Part B: Methodological
Volume	188
Article Number	103058
DOI	10.1016/j.trb.2024.103058
Received Date	21 February 2023
Revised Date	15 April 2024
Accepted Date	13 August 2024
Available Online	5 September 2024
License	Open Access, CC BY 4.0

Abstract

This study investigates the starting point of parking search, presenting new findings through empirical and theoretical approaches. It introduces a probabilistic model that describes the transition from normal driving to actively searching for parking, aiming to minimize journey costs. The model is tested using real-world data collected via a smartphone app that tracks the start of parking searches. Results validate the model, showing that drivers are more likely to begin searching for parking earlier when parking spaces are scarce and driving speeds are reduced (e.g., by congestion). Additionally, various factors influence the start of the parking search, including driver age, vehicle class, and familiarity with the destination. Specific conditions such as proximity to amenities, rush hour timing, and destination familiarity prompt earlier search initiation. The study also identifies scenarios where drivers skip the search process and park immediately, influenced by factors like driving home, short parking durations, and destination familiarity.

Keywords: Cruising-for-parking; Parking search behavior; GPS data; Multinomial logit model

2.1 Introduction

Cruising for parking traffic incurs external costs in the form of emissions and congestion (Shoup, 2021). In addition, parking search is perceived as more negative than driving prior to the search start (Weis et al., 2021). The existing empirical literature on parking search predominantly relies on surveys asking people retrospective questions about the start of their search (location and time) as well as the duration and distance of the search (Lee et al., 2017; Assemi et al., 2020; Qin et al., 2020). Some studies use occupancy data representing parking demand to draw a conclusion on the parking situation (Zakharenko, 2016; Inci et al., 2017; van Ommeren et al., 2021). In recent years, there has been a growing body of literature that utilizes Floating Car (GPS) data (Mannini et al., 2017; Dalla Chiara & Goodchild, 2020; Mantouka et al., 2021), which is subject to certain arbitrary assumptions regarding the initiation of the search process. For instance, Weinberger et al. (2020) and Montini et al. (2012) define a radius of 400 and 800 m around the parking spot found and refer to travel within this radius as cruising.

However, to have the complete image of a car journey with the exact parking search route and duration, we require more information. Four crucial aspects are deemed essential for a comprehensive analysis of this phenomenon: Firstly, the origin of the trip. Secondly, the start of the search process, which separates "normal driving" from "cruising". Thirdly, the parking spot found marking the end of the parking search process and the start of walking. Fourthly, the final destination, which is reached after driving, cruising, and walking. While the first information – the origin of the trip – and the third one – the parking spot found – can be derived (more or less accurately) from almost any floating car (GPS) data, the second and the fourth information remain elusive in previous research.

As will be summarized in section 2, there are many studies that concentrate on route choices and parking choices during the search process. They all assume that the search starts somewhere along the way to the destination. Nonetheless, no empirical study has yet investigated what factors influence the driver's choice to start the search at a certain point. The reason is that the exact time and location of this point have been unknown to researchers. Based on a unique dataset containing the ground truth data on parking search, including all four aspects mentioned above, this study aims to model the start of the search process through a statistical model based on a simple theoretical model. The underlying data source of this paper encompasses information on the start of the parking search (time and location) as well as the origin of the trip, the parking spot found, and the final destination of the journey. This information is measured directly by the drivers pressing a button of a smartphone app developed for the purpose of measuring parking search. These four aspects split every journey into three parts: driving until the search begins, parking search, and walking to the final destination. Furthermore, the app collects some personal data of the drivers such as age, gender, vehicle type, and average yearly driven distance, as well as a few journey-related questions about planned parking duration, familiarity with the destination zone, purpose of the journey, and type of the found parking spot.

Our empirical analysis sheds light on the various factors that affect the start of parking search as well as the absence of parking search. Regression results obtained from pooled, fixed effects, and random effects multinomial logit models suggest that driver-related, journey-related, and destination-related variables have a significant impact on the driver's choice of starting the parking search and skipping the search. For example, older drivers, female drivers, and drivers of higher-class vehicles are found to initiate the search process later. Furthermore, familiarity with the destination area, the presence of amenities such as restaurants and bars, and driving during rush hours make drivers start the search sooner. We can generally conclude – in line with our theoretical model – as the difficulty of finding a vacant parking spot increases, drivers tend to initiate the search process at an earlier stage. In addition, the analysis also reveals that, in some instances, the parking search process is skipped entirely since a parking spot can be found immediately. Factors such as driving toward home, having a planned parking duration of less than 30 min, and being familiar with the destination area are found to be significant contributors to this phenomenon.

The paper is structured as follows. In section 2, we provide a brief literature review on cruising for parking, clarifying the search process and investigating behavioral search models. Section 3 expands on the conceptual framework of the paper and elaborates on our theoretical model by building upon existing models. Section 4 contains the empirical methodology, which is used to apply empirical evidence collected using the smartphone app. The data used for the statistical model is presented in Section 5. Section 6 presents the results of the search starting point choice model and starts a discussion to investigate the results. Finally, Section 7 summarizes the key findings of the study and provides a conclusion that highlights the most significant results.

2.2 Literature review

A sound understanding of the parking search process is a prerequisite for modeling the parking search starting point. Previous theoretical works in the field have provided definitions and insights into the nature of the parking search problem. In one of the earliest studies, (Tamari, 1982) defines the walking distance from the parking spot to the destination as a loss. In this case, the driver tends to minimize the expected loss. This perspective was later supported by (Van Der Goot, 1982), confirming that walking time has the greatest influence on drivers' choices with regard to the parking location. Polak and Axhausen (1990) suggest that to fully grasp the parking search process within the context of parking behavior, it is crucial to identify the search strategies employed by drivers. For example, drivers may opt to drive straight to a parking garage that guarantees vacant spots, thereby avoiding the search process completely. (Thompson & Richardson, 1998) provide a definition within a behavioral modeling framework for cases when the drivers do not have a reserved parking spot at the destination. The search is defined as a series of decisions commencing at a specific point, where drivers decide to choose a parking place (a street segment including a group of on-street parking spots or an off-street parking facility) and examine it at each intersection until they find a vacant spot and accept it.

Empirical parking choice models (Hunt & Teply, 1993; Hilvert et al., 2012; Kobus et al., 2013; Ibeas et al., 2014; Soto et al., 2018) focus on the effects of various parking measures, such as alterations in supply and parking fees, on driver behavior, in particular, on parking facility (and sometimes combined with travel mode) choice. One of the main questions is: given some attributes, such as parking cost, egress time, parking duration, parking security, and driving space, will the driver choose a specific parking facility or not? van der Waerden (2012) summarizes the features in which parking choice models differ: the number and type of alternatives (e.g., free on-street, paid on-street, off-street, illegal parking), the number and type of characteristics (e.g., parking fee, egress time, access time, search time, parking duration), the field of application (e.g., shopping trips, leisure trips), the type of data (e.g., revealed preferences and stated preferences), the modeling approach (e.g., multinomial logit, nested logit, mixed logit), and the findings.

Theoretical parking search models (Glazer & Niskanen, 1992; Anderson & De Palma, 2004; Dell'Orco & Teodorović, 2005; Arnott & Inci, 2006; Benenson et al., 2008; Arnott & Rowse, 2009; Krapivsky & Redner, 2019; Fulman & Benenson, 2021) center around search duration and aim to explain parking search behavior based on the time-varying attributes and information gathered during the search effort. A general conclusion based on these theories is that parking search is a continuous decision-making process. The first decision to make is to start the parking search. The driver faces this decision theoretically from the start of the journey until she or he arrives at the street being closest to the final destination. Intuitively, the probability of initiating the search increases as the driver progresses closer to the final destination up to a certain point. Once the search has begun, the driver could at any time point either continue to search, or terminate the search and drive to a parking garage of choice, or park illegally. In some models, the drivers do not have perfect information and they can learn about the system (the situation on site).

In the realm of parking economics, Zakharenko (2016) provides a pivotal exploration of dynamic parking pricing strategies and their impact on driver behaviors related to parking search. His model illustrates that optimal pricing should dynamically adjust according to the real-time flow of arriving parkers and the existing occupancy rates, a concept that departs from traditional static pricing models. This theoretical framework supports the understanding that effective management of parking availability through pricing can significantly influence when drivers decide to start their search for parking. The insights from Zakharenko (2016) study form theoretical foundation for our research, particularly in analyzing how variations in parking space availability and urban congestion (and consequently driving speed) impact the initiation point of the parking search process.

Considering the evolving perspectives on parking search strategies, a compelling framework is provided by the study conducted by van Ommeren et al. (2021), which integrates the dynamic interplay between in-vehicle search and the consequential walking time, elaborating upon the foundational insights initially discussed by Zakharenko (2016). While Zakharenko (2016) elegantly underscores the pivotal role of driving speed and parking space density in determining the cruising for parking, van Ommeren et al. (2021) expands on these ideas by incorporating both linear and circling search strategies into the analysis. This nuanced approach not only refines the theoretical model by acknowledging the often-overlooked component of pedestrian transit from parked vehicles to final destinations but also quantifies the impact through a so-called "walking multiplier" (Ψ), which varies with the ratio of driving to walking speeds. For instance, van Ommeren et al. (2021) adaptation of the model shows that while a linear search strategy might overestimate Ψ due to excessive walking time predicted by naive strategies, a circling strategy, where searching is confined within a single block radius, presents a more realistic and empirically congruent scenario. These theoretical advancements are crucial as they recalibrate our understanding of parking behavior under different urban layouts and demand conditions. Therefore, a more holistic view of parking search strategies, such as those proposed by van Ommeren et al. (2021), can significantly enhance the robustness and applicability of parking economics models initially posited by Zakharenko (2016).

The extensive body of theoretical research on cruising for parking has highlighted the importance of understanding the starting point of the parking search process. However, due to the limitations in the datasets used, the starting point of the search has been left as a blind spot in the research. Millard-Ball et al. (2020) states that the uncertain nature of the parking search starting point makes characterizing cruising complex. Horni et al. (2013) even state that the origination of parking search “cannot be specified sharply”. The origination of the parking search has been variously conceptualized by scholars, with Thompson and Richardson (1998) viewing it as starting from the beginning of the journey when the driver decides on a parking strategy, while Brooke et al. (2014) assume that it begins when the driver reaches the final destination. Meanwhile, other studies assume a certain radius around the parking spot or destination, such as 100 m, 200 m, or 400 m, for starting the search (Weinberger et al., 2020; Montini et al., 2012; Kaplan & Bekhor, 2011). Apart from distance-based methods, other studies proposed alternative methodologies to identify the parking search starting point. van der Waerden et al. (2015) suggest an arbitrary rule based on speed and its decrease rate to identify the search in GPS data. In an attempt to determine this point, Hampshire et al. (2016) record videos of drivers while completing journeys and analyzes their body movements. Finally, Kaplan and Bekhor (2011) address this issue and indicate that pinpointing the start of parking search should be the central concern in future studies. Overall, by building upon the existing parking models, this work fills the gap by proposing a theoretical model for starting point of the parking search and validating it through a unique dataset.

Through evolving transportation infrastructures, new technologies, and new mobility models, the parking search definition could vary. For instance, autonomous vehicles may have entirely different search strategies to minimize the total journey cost (Zhao et al., 2021; Ertekin & Önder Efe, 2021). This work does not extend its analysis to the cruising behavior of autonomous vehicles.

2.3 Theoretical model to explain the start of the search process

We propose a simple sequential model in discrete time, which could be called a “pre-search” model since it is primarily concerned with the decision to start to search depending on various variables such as the remaining distance to the destination and speed.

A rational agent aims to minimize the cost of a journey by finding an optimal search start. To facilitate the analysis, the journey is divided into three distinct phases: ”normal driving”, "searching", and "walking." The normal driving phase represents the journey prior to the initiation of the parking search, the searching phase is from the initiation of the search until parking, and the walking phase lasts from parking to the final destination (Fig. 2.1). The journey's duration is influenced by the speed of the agent, which is assumed to be constant throughout each of these three phases. The speed of normal driving is denoted by v_d, the speed while searching is v_s, and the walking speed is v_w.

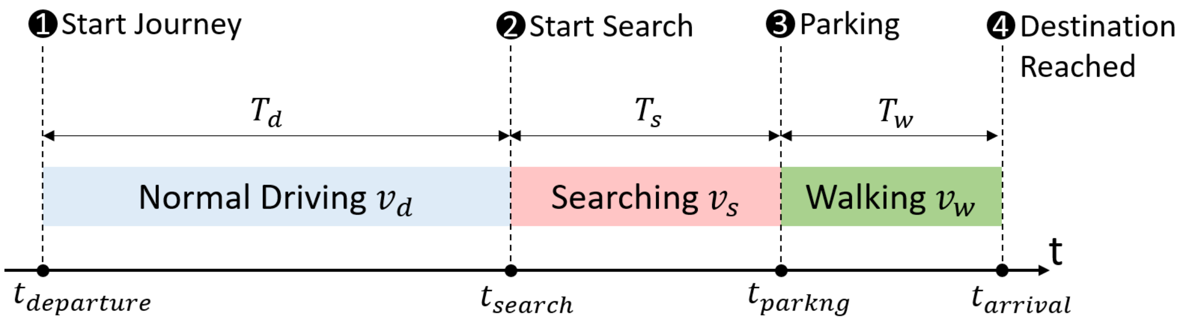


Figure 2.1: Division of the journey into three phases; normal driving, searching, and walking, respectively, at v_d, v_s, and v_w

According to this division, the total cost of the journey is the sum of cost of normal driving, cost of searching and cost of walking,

$$TC = C_d + C_s + C_w$$

TC	Total cost of the journey
C_d	Cost of normal driving
C_s	Cost of searching
C_w	Cost of walking

The costs can be expressed as a function of durations.

$$C_d = \gamma_d T_d$$

$$C_s = \gamma_s T_s$$

$$C_w = \gamma_w T_w$$

γ_d	Weight (cost per unit of time) of normal driving
γ_s	Weight of searching
γ_w	Weight of walking

The total cost of journey can be formulated as:

$$TC = \gamma_d T_d + \gamma_s T_s + \gamma_w T_w$$

Building upon this, further aspects from the model proposed by van Ommeren et al. (2012) could be incorporated. While our model has been primarily concerned with the pre-search phase of the parking process, van Ommeren et al. (2012) focus on the search itself. Nevertheless, some elements fit well into our model since the underpinning goal is similar: to understand the influencing factors in a driver's decision-making process in parking scenarios. Aspects such as parking fees, the number of passengers, the influence of time on travel duration, and parking duration, while typically associated with the search phase, could also be effectively adapted into our model. The following paragraphs outline these extensions:

- **Parking Cost**: Our theoretical framework can be refined further by adding parking cost (PC) to our total cost (TC) equation. PC, which is a function of the parking duration, includes both monetized (e.g., parking fees, overstay fines, and towing fees) and non-monetized cost of parking (e.g., safety, convenience, and weather exposure of the parking spot). The modified equation would henceforth represent TC as the sum of cost of normal driving, searching, walking, and parking cost. Therefore, $TC = \gamma_d T_d + \gamma_s T_s + \gamma_w T_w + PC$.
- **Number of Passengers**: Building on the insight of van Ommeren et al. (2012) that the number of passengers in a vehicle affects the cruising cost, we could extend our model accordingly. As such, the cost functions for each phase of the journey (C_d, C_s, C_w) could be made contingent upon the number of passengers in the vehicle (N). For instance, the cost of searching can be redefined as $C_s = \sum_{n=1}^{N} \gamma_{s,n} T_{s,n}$, thereby capturing the augmented cost associated with multiple passengers undergoing the search phase.
- **Value of time**: van Ommeren et al. (2012) treat the value of time as a variable rather than a constant and assume that it decreases with the progression of travel duration. To align our model with this nuanced perspective, we could render the γ values for each phase depending on the total travel time, excluding cruising.

The rational agent minimizes TC by finding the optimal search starting point. In other words, starting the search means that the driver believes from this moment on, if a vacant parking spot is found, parking the car and walking to the final destination leads to a lower total cost of journey than driving further towards the final destination and starting the search later. Therefore, during the journey (until the start of the search), the agent decides again and again if she or he should start the search now or drive further and start the search later. Given this, the location of the start of the search process indicates, together with the location of the destination, the maximum acceptable walking cost (and distance). We can obtain this cost-minimizing location by comparing the remaining cost of journey in cases which the search starts at current position and the search starts later.

The theoretical analysis of this section provides the basis for the empirical analysis that follows. Since some of the theoretical aspects mentioned cannot be empirically tested with our data, we make the following simplifying assumptions to ensure that our theoretical analysis is helpful for the following empirical analysis. Firstly, the parking cost as defined above remains constant in the destination area. This enables us to focus on driving, searching and walking costs within this context for a certain journey. Secondly, the cost analysis is limited to the driver's perspective, implying that the passenger count is restricted to one ($N = 1$). Thirdly, we assume the value of time to be constant throughout the journey. Finally, we assume that the total cost of the journey is equal to the total duration of the journey, i.e., $\gamma_d = \gamma_s = \gamma_w = 1$, implying that the agent aims to minimize the remaining journey duration.[1] It follows that the remaining journey cost is defined as the driving duration (including searching) plus the walking duration.

$RC = T_d^* + T_s + T_w$

RC	Remaining cost of the journey at a given point along the journey
T_d^*	Remaining duration of normal driving

The remaining duration includes the driving duration from the current position to the location of a parking spot (which may be null if a suitable parking spot is immediately found upon initiation of the search) and the walking duration from the parking spot to the destination. The agent is not fully informed about the availability of parking spots; thus, the estimation of the remaining duration is based on the agent's expectation of how difficult it is to find a parking spot, which in turn is shaped by the known probability of finding a vacant parking spot along the way towards the final destination.

Given a predefined road network, an agent wants to travel from a certain origin point to a certain destination point. The shortest route to the destination is calculated, and it is split into equal segments of length x. Each segment is defined as the route between two points, with the points starting from 0 and incrementing up to infinity to reflect the theoretical possibility of an indefinite search in the absence of a successful outcome. In the example shown in Figure 2.2, the starting position is represented by point 0, while point 8 represents the closest parking spot to the destination.

[1]In some studies, such as Shoup (2021), the weight assigned to walking duration is assumed to be 2 in the overall calculation of cruising cost ($\gamma_w = 2$), as it takes into account the return from the destination to the parking location. We have simplified this assumption by considering only a single walking duration in our computation of journey cost.

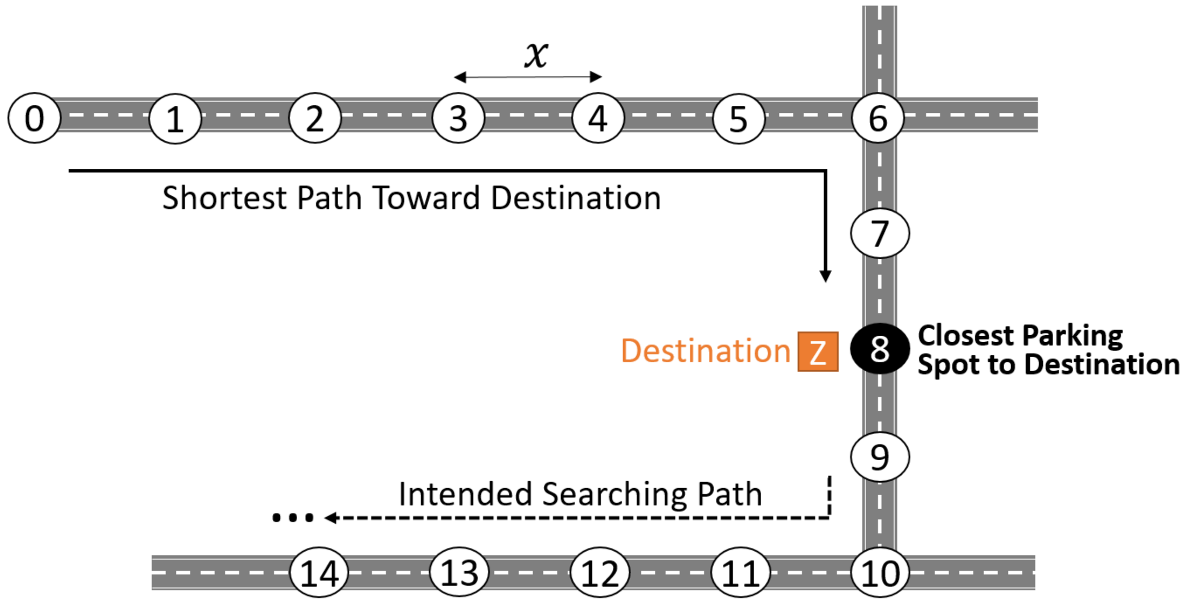


Figure 2.2: Schematic representation of the split route using generated points

At each point, the agent must decide whether to initiate the parking search or to continue driving to the next point before starting the search. From the agent's perspective, the lower the probability of finding a vacant parking space, the sooner the search should start. The agent has then to accept a longer walking path as a trade-off for a long search duration. The agent faces the same decision problem until he or she reaches the final destination at which the search must start (if it has not already) (Millard-Ball et al., 2020).

Let RC_j^q be the remaining cost of the journey when the agent is located at point j and starts the search at point q, where $q \geq j$. In this notation, the subscript shows the position of the agent, while the superscript denotes the search starting point. Let B denote the point at which the parking spot is found and Z represent the destination location. When the agent initiates the parking search at his or her current location j, the remaining normal driving duration T_d^* is zero and the remaining cost RC_j^j is the sum of T_s, which is the cost (duration) of searching from point j to the parking spot B, and T_w, which is the cost (duration) of walking from the parking spot B to the destination Z.

$$RC_j^j = T_s + T_w$$

If the agent finds a parking spot immediately after starting the search, i.e., $B = j$, the searching duration T_s is zero, resulting in the remaining cost being solely the walking duration.

Since the costs are defined as durations, we can reformulate the cost function in terms of distance and speed. RC_j^j is the sum of the searching duration from point j to point B and the walking duration from point B to point Z. Therefore, RC_j^j can be rewritten as:

$$RC_j^j = T_s + T_w = \frac{(B-j)x}{v_s} + \frac{D_B}{v_w}$$

with x being the constant segment length, and D_B indicating the remaining walking distance from point B to the final destination.

We assume that the expected probability of finding a vacant parking spot in each point (or its corresponding segment) is constant and equals to p, representing the agent's expectation on how likely it is to find a vacant parking spot in that journey in a segment, given his or her experience and knowledge on the geographical area and the traffic condition. It is also assumed that an unavailable parking spot will be unavailable forever (during the journey). If the agent starts the search at point j, the probability of finding a vacant spot at point $B \geq j$ follows a geometric distribution (Arnott & Williams, 2017):

$$f(j, B, p) = P(\text{starting search@}j\text{, park@}B) = (1-p)^{B-j} \cdot p \quad \text{for } B \geq j$$

RC_j^j is a random variable with the expected value (expected cost of starting the search at point j):

$$\begin{aligned} E(RC_j^j) &= \sum_{B=j}^{\infty} f(j, B, p) \cdot RC_j^j \\ &= \sum_{B=j}^{\infty} (1-p)^{B-j} \cdot p \cdot \left(\frac{(B-j)x}{v_s} + \frac{D_B}{v_w} \right) \end{aligned}$$

RC_j^{j+1} is defined as the remaining cost of the journey when the agent is located at point j and the search starts at point $j+1$ (next point). This cost is represented as the sum of T_d^*, the cost (duration) of driving to the search start point $j+1$ (remaining normal driving duration), T_s, the cost (duration) of searching from point $j+1$ until a parking spot is found at B, and T_d, the cost (duration) of walking from the parking spot B to the final destination Z. Thus, the cost function can be expressed as follows:

$$RC_j^{j+1} = T_d^* + T_s + T_w = \frac{x}{v_d} + \frac{(B-j-1)x}{v_s} + \frac{D_B}{v_w} \quad \text{for } B > j$$

As RC_j^{j+1} is also a random variable, we can estimate the cost of starting the search at the next point (when the vehicle is currently located at point j) by calculating the expected value:

$$\begin{aligned} E(RC_j^{j+1}) &= \sum_{B=j+1}^{\infty} f(j+1, B, p) \cdot RC_j^{j+1} \\ &= \sum_{B=j+1}^{\infty} (1-p)^{B-j-1} \cdot p \cdot \left[\frac{x}{v_d} + \frac{(B-j-1)x}{v_s} + \frac{D_B}{v_w} \right] \end{aligned}$$

Therefore, at point j the following rules must hold:

- if $E(RC_j^j) < E(RC_j^{j+1})$ then the agent starts to search.
- if $E(RC_j^j) > E(RC_j^{j+1})$ then the agent drives to the next point and repeats this decision-making process.

Using this intuition, if we assume typical values for speed, we can show how changes in the probability of finding a vacant parking spot affect the search starting point. Imagine a rational agent is driving from an arbitrary origin point toward the destination. An example of this hypothetical journey is visualized in Figure 2.3a. For better visualization, we visualize only the last section of the journey, which is reasonably long enough to contain the parking search starting point. The solid blue line in this figure represents the shortest path to the destination, along which the parking search must start. The end of this route is the closest parking spot to the destination, which is theoretically the last reasonable opportunity to start the search. Afterward, we split the route into equal lengths using the generated points as shown in Figure 2.3b.[2] These are the points (or steps) where the agent must make a decision on when to initiate the parking search. In the next step, the costs, RC_j^j and RC_j^{j+1}, are calculated to find the optimal search starting point. In this illustration, we assume a driving speed of 20 km/h, a searching speed of 16 km/h and a walking speed of 4 km/h.

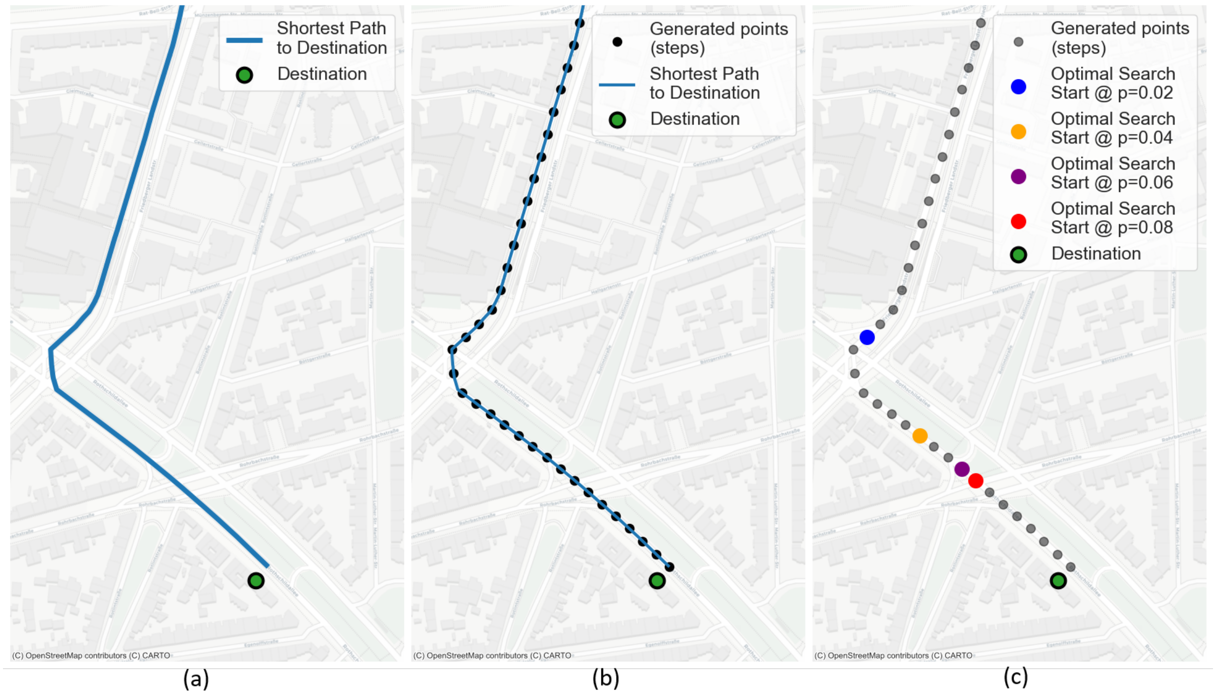


Figure 2.3: (a) agent's route toward the destination (b) route split into equal segments (c) optimal starting search points for different probabilities of finding a parking spot p

The remaining walking distance to the final destination D_B can be calculated at each point. Before reaching the destination, it is proportional to the number of remaining segments. After

[2]This is done as follows: (1.) The shortest path is calculated using Valhalla Open-Source Routing Engine (2.) Line split is done using shapely, a python library for spatial analyses.

reaching the destination, the distance undergoes periodic fluctuations, which reflect the circulation around the block in search of an available parking spot. In addition, as the agent's search for a parking spot continues to be unsuccessful, the radius of the search area expands over time, implying that the agent becomes increasingly willing to endure a longer walking distance (Martens & Benenson, 2008). This is visualized in Figure 2.4.

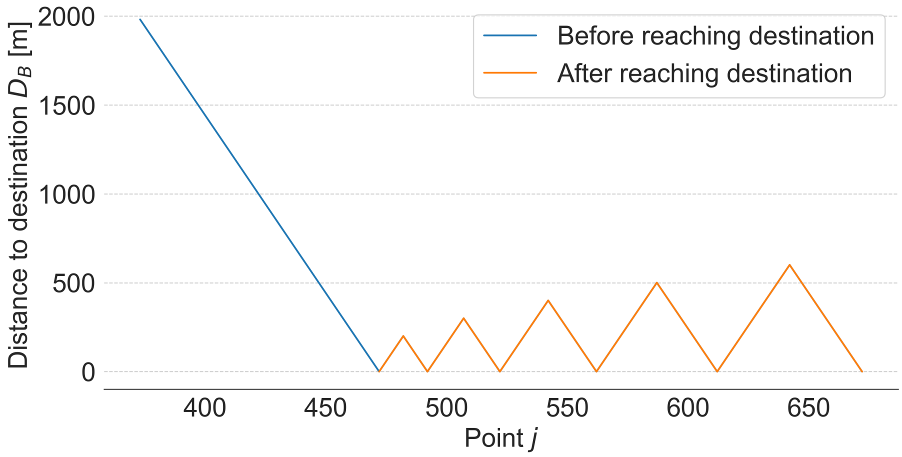


Figure 2.4: The remaining distance to the destination visualized for the selected range of points

We can now calculate $E(RC_j^j)$ and $E(RC_j^{j+1})$ for each generated point along this journey for arbitrary values of p. An example at $p = 0.04$ is visualized in Figure 2.5. Initially, after the start of the journey, $E(RC_j^{j+1})$ is lower than $E(RC_j^j)$ since driving speed is higher than walking speed, by assumption. However, as the agent approaches the destination, the difference between the two costs decreases. At a certain point, $E(RC_j^j) < E(RC_j^{j+1})$, which triggers the start of the search.

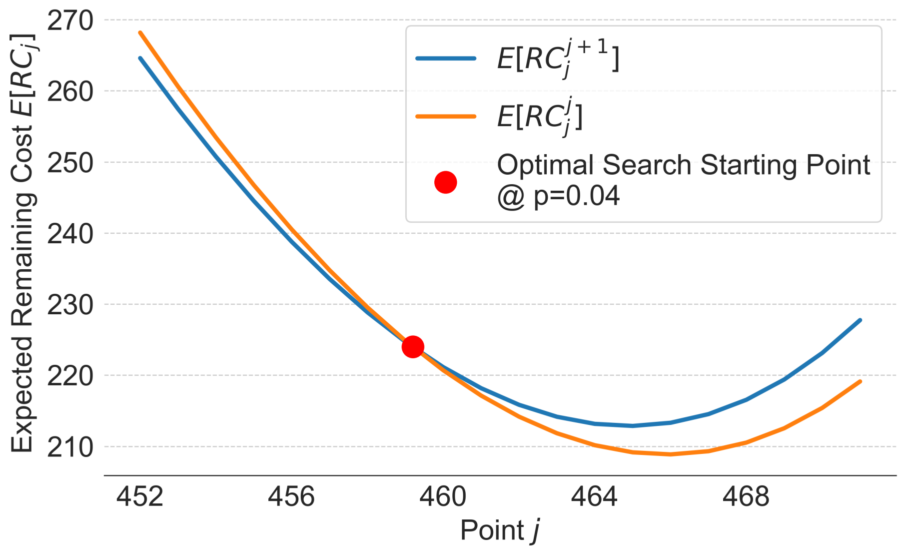


Figure 2.5: Expected costs at generated points for the example journey

For different values of p, we can calculate $E(RC_j^j)$ and $E(RC_j^{j+1})$ and, hence, the optimal search starting point. Figure 2.3c presents these results, demonstrating that an increased probability of finding a parking spot results in an optimal search starting point closer to the final destination.

Based on our simplistic theoretical model, several insightful conclusions can be drawn. The main implications of the model are as follows:

- The model allows for an evaluation of the optimal search starting point in relation to the value of p, which serves as a measure of the parking demand pressure given a certain supply. When the probability of finding a parking spot decreases, the optimal starting search shifts further away from the destination. In other words, when finding a parking spot is more difficult, the driver tends to start the parking search sooner in order to minimize the total journey duration. This supports the hypothesis formed by Millard-Ball et al. (2020) that parking search starts sooner in areas with higher parking occupancy.
- The model enables an analysis of the effects of changes in driving speed, searching speed, and walking speed on the optimal search starting point:
  - A decrease in driving speed v_d or searching speed v_s moves the optimal search starting point further away from the final destination. This means, for example, in situations with low vehicle speed due to high traffic congestion or low speed limits, the driver tends to start the search sooner to minimize the journey duration.
  - An increase in walking speed v_w moves the optimal search starting point further away from the final destination, i.e., implying a sooner starting of the search. Increasing the v_w decreases both RC_j^{j+1} and RC_j^j, with the rate of decline of RC_j^j being higher.

The proposed simplified model minimizes the journey duration. The assumption is that walking and driving have the same cost for a given agent. This assumption is relaxed in the general total cost equation. Personal characteristics of the driver (e.g., age, health status, and preferences), characteristics of the vehicle (e.g., vehicle size), the trip (e.g., purpose), geographical attributes (e.g., type of road, and points of interest), traffic situation, weather and other factors affect the costs. Looking at the general equation to calculate the total cost of the journey, we can draw more conclusions about the different aspects affecting the starting point of the parking search.

For example, for an elderly person, a person with small children or a person with health issues, γ_w may be higher. Additionally, depending on the trip purpose a walking distance may be more or less acceptable. For example, a person who wants to transport groceries may also have $\gamma_w > 1$. The same is true for bad weather conditions or inconvenient footpaths. In these cases, since γ_w increases and the agent seeks to minimize the TC, he or she tries to reduce C_w by reducing T_w. This means moving the starting point of the search toward the destination and also undertaking a long search duration time to find a parking spot close to the destination.

In principle, our theoretical model fits well with the recorded GPS data: The generated points in the theoretical model can be seen as the waypoints along a GPS trajectory. However, in the recorded GPS data, the waypoints recording frequency and segment length are not constant and the speed varies along the way. Section 5 presents the details of how an empirical model can be built based on our GPS dataset.

2.4 Empirical methodology

To test the hypotheses derived in the previous section empirically in a multivariate setting, we specify a (first-order Markov) regression model for the transition from "driving" to "searching". This can be accomplished using binary choice models, such as the logit model. In journey i, at time point t, the agent has the discretion to either initiate the search process or to continue driving, deferring the start of the search to a later point in time. In our theoretical model, the agent aims to minimize the total cost of the trip, which can easily be transformed into a utility maximization problem, with utility of traveling being negative. According to the principles of Random Utility Theory, the utility of alternative k is determined by the following equation:

$U_{ikt} = X_{it}\beta_k + u_{ik} + \epsilon_{ikt}$

U_{ikt}	Utility of alternative k in the journey i at time t
X	Vector of explanatory variables
β_k	he set of coefficients associated with outcome k
u_{ik}	Journey-level error term
ϵ_{ikt}	Observation-level error term

The utility can be considered as a random variable, modeled as a linear combination of an observed and an unobserved term. First, the observed part is represented by the deterministic component $X_{it}\beta_k$, where X_{it} is a row vector denoting the observed variables (measuring or being correlated with γ, T, and p) and β_k is a column vector of parameters that is to be estimated and describes the effect of the observed variables on utility. Second, the unobserved part, or error term, comprises two components. Note that floating car data can be seen as time series data since the various attributes of a journey are collected along the journey. This results in a sequence of outcomes for each journey. Even after controlling for X_{it}, the choices leading to these outcomes for each journey may, therefore, not be independent. Considering this unobserved heterogeneity, an additional error component u_{ik} is introduced at the journey level that accounts for the underlying unobserved and time-constant drivers' preferences and journeys' characteristics. This constitutes the first component of the error term. The second component ϵ_{ikt} is the observation level error term, which is assumed to have a type I (Gumbel) extreme-value distribution.

However, as expected from the theoretical model, there are many journeys with a searching duration (and distance) near zero, implying parking spots are immediately found after starting to search. After speaking with individuals who used the data collection app, it was discovered that this phenomenon occurs predominantly in two circumstances. Firstly, there are enough vacant parking spots at the final destination location. In these situations, drivers drive to the nearest parking spot and park the vehicle there. This corresponds to the minimum possible journey duration. The second case occurs when the driver has not yet started searching, but on the way comes across a free parking spot. This could make the driver park the vehicle immediately and skip the search process. This can be advantageous for a risk-averse driver, as he or she then knows the remaining journey duration (=walking duration) with certainty, whereas starting the search later leads to an unknown total journey duration.

To account for the aforementioned possibility, we specify a multinomial logistic model with three possible outcomes:

- $k = 0$: The driver has neither started the search nor parked the car. This mode is characterized as "normal driving", which indicates a steady state compared to the previous time point.
- $k = 1$: The driver starts the parking search. This indicates the transition from "normal driving" to "searching" and signals the initiation of the parking search.
- $k = 2$: The driver parks the vehicle immediately without searching. This is the case when the search duration is zero and the journey does not include a parking search effort. This outcome indicates the transition from "normal driving" to "parking".

Given this, P_{ict}, the probability of outcome c in journey i at time point t is calculated as:

$$P_{ict} = \Pr(Y_{it} = c | Y_{it-1} = 0, X_{it}, \beta_k, u_{ik}) = \frac{e^{X_{it}\beta_c + u_{ic}}}{\sum_{k \in S} e^{X_{it}\beta_k + u_{ik}}}$$

where

$Y_{it} = k$	Mode of the driver in journey i at time t
$S = \{0, 1, 2\}$	Outcome set

We model unobserved heterogeneity (the error component) by random and fixed effects at the journey level. Note that an error component at the journey level is more flexible than at the driver level since the number of journeys is higher than the number of drivers. The fixed effects and random effects estimators differ only in their assumptions about the error component u_{ik}. The random effect estimator assumes that u_{ik} is uncorrelated with X_{it} and has a normal distribution. In contrast, in case of the fixed effects estimator u_{ik} can be correlated with X_{it} and no distributional assumptions are necessary. However, this advantage comes with some costs. Firstly, it is not possible to identify the effects of time-constant variables. Secondly, the fixed effects estimator cannot be used for predictions of combinations of marginal effects that account for u_{ik} since they are not estimated explicitly.

For these reasons, the fixed effects estimator is just used as a robustness check here. If there are major differences between the results of the fixed effects and the random effects estimators, then this is an indication that the assumptions of the random effect estimator are violated. This comparison between both approaches is the idea of the Hausman test. The H_0 is that u_{ik} are uncorrelated with X_{it} in the model, while H_a is that the u_{ik} are correlated with X_{it} are correlated with the covariates. The fixed effects estimator is consistent under both H_0 and H_a, while the random effects estimator is inconsistent under H_a but efficient under H_0. In case of the random effects estimator, it is possible to test the relevance of the (journey) random effects using an LR-test against the simple pooled multinomial logit (without modeling unobserved heterogeneity at the journey level). In order to facilitate the interpretation of the estimation results, average marginal effects, as well as predicted probabilities depending on specific variables are calculated from the estimated coefficients.

2.5 Data and descriptive evidence

For collecting ground truth data on cruising for parking, we have developed a mobile application that has been available on both the Google Play Store and Apple App Store since August 2021 in Europe. The app allows users to log their car journeys and monitor their cruising duration. After downloading and installing it, users are prompted to provide some personal information. These are gender, birth year, yearly driven distance, and vehicle type. Answering these questions is entirely voluntary, and the users can leave them blank. In addition, at the end of each journey, the users are asked to fill in journey-related questions. These are journey purpose, parking type, planned parking duration, and familiarity with the destination.

The functionality of the app is intentionally easy, with the main screen featuring a large blue button, which must be pressed in four steps to record a journey.

- **1st step - start journey**: As the drivers start a journey, they press this button that triggers the recording process. The app records the GPS coordinates of the undertaken route and the corresponding timestamps and speeds.
- **2nd step - start search**: Once the drivers start the search, i.e., when they start to actively look for a vacant parking spot, they press the button. This records the location and time of the beginning of the parking search effort along the driving route.
- **3rd step - vehicle parked**: Upon finding a vacant parking spot and parking the vehicle, drivers press the button for the third time. This records the location of the parking spot and marks the end of the parking search process. This also signifies the start of the walking path toward the final destination.
- **4th step - destination reached**: After reaching the final destination on foot, drivers press the button for the last time. This marks the location of the destination and the time of reaching it. This also finalizes the journey and stops recording.

These four steps divide the recorded journey into three distinct parts: driving until the parking search begins, parking search process, and walking to the final destination. An illustration of a recorded journey can be seen in Figure 2.6. The app is only used when the driver does not have a pre-reserved parking space at the end of the trip. That means we do not collect any journey with drivers parking in their home garage or when they plan to go to a shopping mall and decide

Figure 2.6: An example journey recorded via the app

beforehand to park in the parking garage of the shopping mall. Nevertheless, the journeys may still end in parking garages or underground parking when a driver terminates the search after a certain amount of unsuccessful searching and proceeds to an off-street parking facility.

The app users are volunteer drivers who have expressed interest in contributing to research on parking search. To promote the app, an advertising campaign was launched utilizing various mediums, including publication in newspapers, television interviews, and website postings by municipalities, as well as presentations at conferences. While the majority of the data collection has taken place in Germany due to a concentrated user acquisition effort, data has also been obtained from other countries, though it has been excluded from this analysis. It is noteworthy that a significant proportion of the recorded journeys terminate in Frankfurt am Main due to a higher concentration of user acquisition efforts in that location.

As of the time of writing this paper, in February 2023, there are over 1,500 individuals who have downloaded and utilized the app for the purpose of collecting ground truth data on cruising for parking. Despite this, only 103 of these users have recorded valid trips in Germany, which serve as the basis for the data analysis.

A trip is regarded as "non-valid" primarily when users interact with the application for reasons other than its intended use, such as testing its functionalities or acquainting themselves with its interface, without implementing it in an actual on-road situation. This often takes place at a stationary location, such as a user's residence, with actions like pressing the button four consecutive times merely to study the application's reaction, rather than to collect genuine driving or parking data. It is not uncommon for users to exhibit this behavior. Such non-valid instances can be detected through GPS data analysis.

On the other hand, valid trips, which are the focal point of our analysis, are ascertained and authenticated using several critical parameters, including the journey duration and the Origin-Destination (OD) distance. The decision of a trip's validity is governed by specific minimum values: a journey must last at least 10 minutes and cover an OD distance of at least 1.5 km.

The users have conducted 1795 valid trips using the app between August 2021 and January 2023 in Germany, which are used for parking search modeling. 91 of 103 users have completed their profile in the app and provided their personal information. In 1,770 out of 1,795 trips, drivers answered the journey-related questions completely. Variables based on user-related and journey-related questions are time-invariant, i.e., they do not vary during the journey. Table 2.1 and Table 2.2 provide a summary of the available data, including the possible answers to the questions and the corresponding descriptive statistics. Note that, in some cases with few available data points, classes are merged as indicated in the tables.

Table 2.1: Driver characteristics – 103 total drivers

Variables	**Values**	**Count (rate)**	
Age	[Integer Between 18 and 85]	42.2	Mean
		20.0	5%-Q
		36.0	Median
		72.9	95%-Q
Gender	Male	58	(56%)
	Female	34	(33%)
	Divers	0	
Average Yearly Driven Distance	Less than 9000 km	30	(29%)
	Between 9000 km and 30000 km, Larger than 30000 km	62	(60%)
Vehicle Type	Mini car, Small car	29	(28%)
	Compact car	35	(34%)
	Family car, Luxury car or SUV	25	(24%)
	Van	3	(24%)
User information missing		11	(11%)

Table 2.2: Journey characteristics – 1795 total journeys

Variables	Values	Total Journeys		Journeys with Parking Search		Journeys with Immediate Parking	
		count	proportion	count	proportion	count	proportion
Familiarity with the destination area	Known	1285	(72%)	1043	(70%)	242	(81%)
	Unknown	485	(27%)	436	(29%)	49	(16%)
Planned parking duration	Less than 30 minutes	544	(30%)	440	(30%)	104	(35%)
	Between 30 minutes and 3 hours	425	(24%)	374	(25%)	51	(17%)
	Longer than 3 hours	121	(7%)	107	(7%)	14	(5%)
	Parking associated with work and home trips	680	(38%)	558	(37%)	122	(40%)
Journey purpose	Shopping	530	(30%)	435	(32%)	95	(32%)
	Entertainment	457	(25%)	404	(27%)	53	(18%)
	Home	428	(24%)	359	(24%)	69	(23%)
	Business Trip, Work	355	(20%)	281	(19%)	74	(24%)
Journey information missing		25	(1%)	16	(1%)	9	(3%)

The primary time-based variables for this study are parking search duration (PSD) and walking duration. The median parking search duration is 58 seconds and the median walking duration is 69 seconds. The arithmetic means are 101 seconds and 142 seconds. Further statistics can be seen in Table 2.3. The discrepancy between the means and medians may indicate right-skewed distributions and/or outlier data points, i.e., instances where users experienced particularly extended search or walking times. Figure 2.7 visually represents the distribution of parking search durations. A clear distinction can be made between cases with immediate parking, depicted by the first bar (green), and the remaining cases by the right-skewed distribution (blue). These numbers show that the captured parking search durations resonate more with the figures presented by Hampshire et al. (2016), van Ommeren et al. (2021), and Alemi et al. (2018), as opposed to the relatively high values reported in Shoup (2006). However, a comprehensive comparison is beyond the scope of this paper.

Table 2.3: Parking search and walking duration statistics in MM:SS

Variable	count	5%	25%	median	75%	95%	mean	std
All Journeys								
Parking Search Duration	1,795	00:00	00:22	00:58	01:56	06:08	01:41	02:22
Walking Duration	1,795	00:00	00:12	01:09	03:01	09:42	02:22	03:27
Journeys Including a Parking Search (PSD>0)								
Parking Search Duration	1,495	00:16	00:39	01:11	02:19	06:41	02:00	02:27
Walking Duration	1,495	00:00	00:26	01:23	03:11	09:46	02:23	03:27

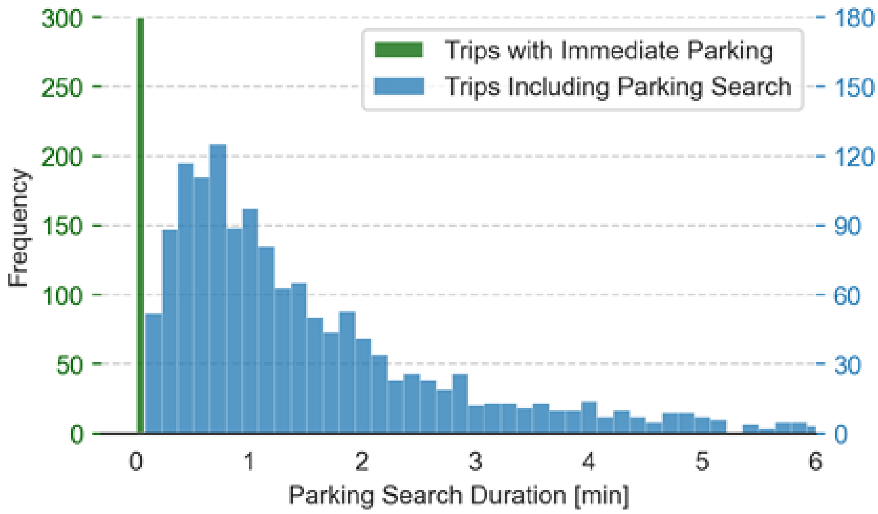


Figure 2.7: Parking Search Durations Distribution

In the data, drivers always start their parking search or park immediately only within 1500 m walking distance of the final destination. Hence, all the points with a remaining walking distance to the final destination larger than 1500 m are dropped. The time interval for sampling GPS points along a journey could vary within and between journeys with a mean of approximately 7 seconds; however, it is about 5 seconds for most of the data (median and mode). The final data used in the multinomial logit model comprises 73,159 GPS points. The count and proportion of points of each outcome can be seen in Table 2.4. The number for $k = 1$ indicates that within the remaining walking distance to the final destination, the average sample probability of switching from driving to searching is approximately 2.0 % per 7 seconds. 1/6 of "normal driving" ends with an immediate transition to parking ($k = 2$).

Table 2.4: Outcomes shares and counts of GPS points in the estimation sample

Outcome	**Points Count**	**Share**
$k = 0$ (driving before starting the search)	71,364	97.55%
$k = 1$ (transition to search)	1,495	2.04%
$k = 2$ (transition to immediate parking)	300	0.41%
Total	73,159	100.00%

Each GPS point in the collected FCD is associated with positional attributes, speed, and timestamp. Speed is recorded using smartphone sensors and can be included in the model directly. For journeys that include a parking search, the mean and median speed values at the point of starting the parking search are found to be 18.5 km/h and 18 km/h, respectively, with a standard deviation of 13 km/h. To control for the effect of time of day, hour dummies are constructed from the timestamp. Night hours are merged from 22:00 to 07:00.

Further data sources are used to enrich the model. Weather information is retrieved from the

German Weather Service (Deutscher Wetterdienst DWD).[3] For each GPS point, we retrieved temperature and wind speed from the nearest weather station and nearest time point. Since the weather information is provided on an hourly basis, the collected weather information remains constant during each journey.

As indicated by our theoretical model, the hypothetical walking distance at each point of the driving phase should have an influence on starting the parking search. These hypothetical walking distances must be distinguished from the actual walking route the driver may have chosen according to the data collection. While the former exist for every single GPS point during the driving phase, the latter exists only once. This variable (hypothetical walking distance) is calculated using the Valhalla[4] open-source routing engine for each of the 73,159 GPS points. The GPS points are first map-matched to the underlying road network from OSM.

The variables "search radius" and "accepted walking distance" are essential factors for the parking behavior. The search radius, representing the air distance from the initial search point (captured when the drivers pressed the search button) to the final destination location for journeys including a parking search has a mean value of 145 m, a median of 114 m, and a standard deviation of 125 m. This represents the usual range of the drivers' search area. The accepted walking distance is the hypothetical route a driver would willingly traverse on foot from the transition point (initiation of the search or immediate parking) to the final destination. The mean and median of this variable amount to 186 m and 146 m, respectively, with a standard deviation of 161 m.

In addition, Valhalla provides information about the street intersections along the calculated shortest path. This could help us to look into the hypothesis that the number of remaining intersections has an influence on the search starting point. Figure 2.8a represents an example journey after filtering out the unnecessary points, visualizing the GPS points of the journey and the destination point, which are used for modeling the search starting point. Figure 2.8b shows the intersections to be encountered along the route to the final destination.

[3] DWD is a public institution under the Federal Ministry of Transport and Digital Infrastructure of Germany.

[4] https://github.com/valhalla/valhalla

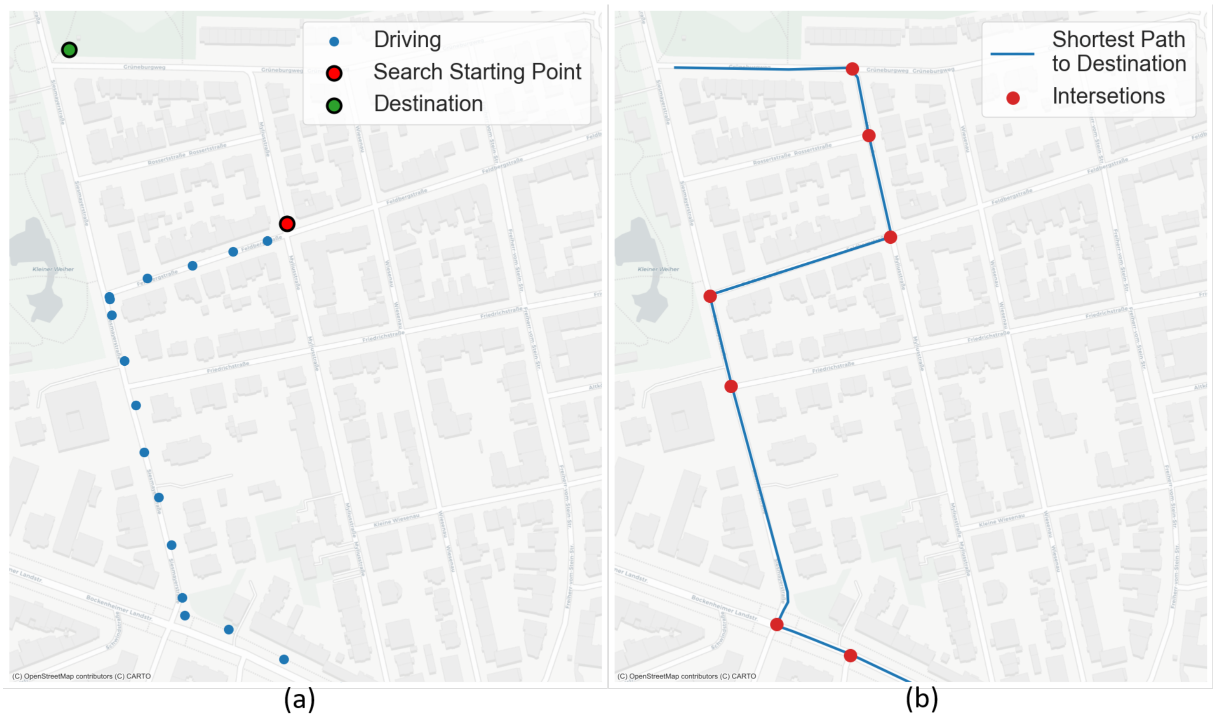


Figure 2.8: GPS points of an example journey used in modeling the search starting point and the intersections along the route to the destination

We also incorporate spatial data from Open Street Map (OSM), which is introduced as counts of Points of Interest (POIs) around each GPS point. To accomplish this, a 200-meter radius was chosen for the calculation of surrounding POIs. This value was selected as it corresponds to the 0.75 quantile of the search radius around the final destination in the unrestricted dataset. The POIs were then grouped into categories based on their primary features in OSM[5]. The POI categories that are used in the model are *amenity*, *building*, *office*, *shop*, *leisure*, and *natural*. The parking occupancy near the journey's final destination impacts the driver's decision to start the search according to our theoretical model as it affects p. For instance, a high occupancy is likely in business districts during working hours or in residential areas overnight. The POI counts serve as proxy variables for different types of neighborhoods that correspond to different parking occupancies.

Finally, please note that the parking type (*free*, *paid*, or *illegal*) is a dynamic variable that can change during the search process, influenced by factors such as available parking options, time constraints, and personal preferences. Due to its endogenous nature, we have not included it as an explanatory variable in the regression models.

[5]https://wiki.openstreetmap.org/wiki/Map_features

2.6 Results and discussion

In order to analyze the influence of the variable speed, it is essential to address an inherent issue in our dataset with the recorded variables at the moment of initiating the parking search. App users first decide to start the search and then press the button on the app to label this point. This action may result in the endogeneity (in the sense of reverse causality) of the recorded variables at the exact time of pressing the button, as the search process may have actually started several seconds prior to the button press. Only the variable speed is affected. To mitigate this issue, we have employed the strategy of utilizing the speed of the preceding GPS point instead of the current speed. This may solve the endogeneity problem but introduces another complexity.

In our dataset, the time interval between recorded GPS points is not constant. Therefore, we need to adjust for the fact that a longer time interval means a higher probability of transition. If the duration to the previously recorded point is longer at a certain point, the driver simply has more time (and opportunity) to start the search or park immediately. The transition probability is proportional to the time interval, and the regression model can capture this using a so-called "offset variable". Any generalized linear model (GLM) can include an offset variable; in our case, since the model contains a log link function, the variable "time span from the actual to the previous GPS point in seconds" is logged. This logarithmic offset is then constrained to 1.

To identify the proposed empirical MNL model, the outcome probability equation must be normalized according to a base category. We select $k = 0$, i.e., steady-state in "normal driving" mode, as the base outcome and restrict the respective parameters to 0, i.e., $\beta_0 = 0$, $u_{i0} = 0$.

We compare three models:

- Model I is the pooled multinomial logistic regression model.
- Model II is the multinomial logistic regression model with journey fixed effects.
- Model III is the multinomial logistic regression model with journey random effects.

Table 2.5 presents a comparison between these three models, accompanied by statistics of goodness-of-fit. It should be noted that the number of estimated parameters in Model II is less than in the other models, as the time-invariant variables are excluded in the fixed effects model. It is important to mention that the number of GPS points is equal across all models. The results indicate that Model II has the lowest AIC and BIC values and the highest Pseudo R^2. A likelihood ratio test comparing Model III and Model I indicates highly statistically significant random effects.

Table 2.5: Model comparison

Statistics	**Model I: pooled mlogit**	**Model II: mlogit fixed effects**	**Model III: mlogit random effects**
Number of Estimated Coefficients	40	9	40
Sample size	73,159	73,159	73,159
Number of groups (=*journeys*)	-	1,795	1,795
Log-Likelihood	-7021.1	-2,506.2	-6,926.7
AIC	14,206.1	5,048.3	14,021.4
BIC	14,960.6	5,213.9	14,794.2
Pseudo R2	0.240	0.614	0.250

Almost all estimated coefficients show the same signs for all three models. Despite this, the result of the Hausman test indicates that there are systematic differences between the coefficient estimates of the fixed effects and the random effects model, favoring the fixed effects model. However, this does not mean that the estimation results of the other two models are without merit. First, the fixed effects model involves fewer assumptions, yet it is not necessarily more robust in the event of a violation of those assumptions (Townsend et al., 2013). Second, variables without within-journey variation (e.g., driver-related and journey-related information) are not estimable with the fixed effects model. Therefore, the results of Models I and III remain valuable and contribute to our interpretation, at least in a qualitative and comparative way.

The estimated coefficients for outcomes $k = 1$ (transition to searching) and $k = 2$ (transition to parking immediately) are in Table 2.6.

Table 2.6: Coefficient estimation results

Variable	Model I: pooled mlogit		Model II: mlogit fixed effects		Model III: mlogit random effects	
	Transition to searching (k = 1)	Transition to parking (k = 2)	Transition to searching (k = 1)	Transition to parking (k = 2)	Transition to searching (k = 1)	Transition to parking (k = 2)
Remaining walking	-0.006***	-0.014***	-0.031***	-0.017***	-0.008***	-0.014***
distance [m]	(0.000)	(0.001)	(0.001)	(0.002)	(0.000)	(0.001)
Remaining intersections	-0.008**	0.007	-0.027***	0.017	-0.011***	0.006
	(0.003)	(0.008)	(0.008)	(0.014)	(0.004)	(0.009)
Speed (t-1) [km/h]	0.027***	-0.232***	-0.218***	-0.549***	-0.026***	-0.277***
	(0.007)	(0.026)	(0.014)	(0.050)	(0.009)	(0.031)
Age	-0.032***	-0.024***	—	—	-0.032***	-0.019**
	(0.004)	(0.008)			(0.006)	(0.009)
Gender (male)	0.201**	-1.164***	—	—	0.354***	-1.204***
	(0.092)	(0.194)			(0.129)	(0.217)
Average yearly driven	-0.132	-0.651***	—	—	0.023	-0.540**
distance less than 9000 km	(0.082)	(0.213)			(0.112)	(0.239)
Vehicle type						
Mini and small: base						
Compact	-0.199**	-0.038	—	—	-0.169	0.051
	(0.088)	(0.200)			(0.120)	(0.225)
Family, luxury, and SUV	-0.224*	0.600**	—	—	-0.144	0.769***
	(0.130)	(0.243)			(0.177)	(0.275)
Van	0.414	-0.390	—	—	1.316***	-0.576
	(0.293)	(1.063)			(0.444)	(1.174)
User information missing	0.453**	-1.945*	—	—	0.671***	-2.171**
	(0.180)	(1.037)			(0.260)	(1.096)
Familiar with the	0.238***	0.808***	—	—	0.247**	0.737***
destination area	(0.074)	(0.199)			(0.105)	(0.224)
Planned parking duration						
Less than 30 min: base						
30 min to 3h	0.070	-0.608***	—	—	0.360***	-0.526**
	(0.080)	(0.201)			(0.116)	(0.230)
Longer than 3h	-0.003	0.120	—	—	0.134	0.175
	(0.129)	(0.325)			(0.181)	(0.367)
Journey purpose						
Entertainment: base						
Home	-0.121	-0.660***	—	—	0.110	-0.639**
	(0.105)	(0.244)			(0.150)	(0.280)
Shopping	-0.194**	-0.110	—	—	-0.083	-0.125
	(0.086)	(0.212)			(0.122)	(0.239)
Work and business	-0.479***	-0.326	—	—	-0.165	-0.239
	(0.104)	(0.241)			(0.148)	(0.282)
Journey information	-0.104	0.503	—	—	-0.108	0.385
missing	(0.289)	(0.456)			(0.393)	(0.524)
Temperature	0.004	0.016*	—	—	0.009*	0.022**
	(0.004)	(0.009)			(0.005)	(0.011)
Wind speed [m/s]	-0.022	-0.022	—	—	-0.045*	-0.032
	(0.019)	(0.038)			(0.027)	(0.043)
POI counts: amenities	0.000	0.003*	0.011***	0.012	0.000	0.002
	(0.001)	(0.002)	(0.004)	(0.010)	(0.001)	(0.002)
POI counts: buildings	-0.091	-0.384**	-0.105	1.018*	-0.073	-0.370**
	(0.064)	(0.155)	(0.198)	(0.525)	(0.080)	(0.181)
POI counts: leisure	0.031**	0.042	-0.060	0.101	0.028	0.047
	(0.014)	(0.039)	(0.049)	(0.085)	(0.017)	(0.042)
POI counts: natural	-0.000	-0.001	-0.013***	0.002	-0.001	0.000
	(0.001)	(0.001)	(0.003)	(0.007)	(0.001)	(0.001)
POI counts: office	-0.017**	0.022	-0.071*	0.055	-0.031***	0.019
	(0.008)	(0.019)	(0.037)	(0.083)	(0.011)	(0.021)
POI counts: shops	-0.000	-0.012***	-0.015***	-0.040***	0.000	-0.012***
	(0.001)	(0.004)	(0.006)	(0.012)	(0.002)	(0.004)
Logarithmic time offset (time span between t-1 and t in sec.)	1 (con.)	1 (con.)	1 (con.)	1 (con.)	1 (con.)	1 (con.)
Const.	-2.923**	-16.763	—	—	-4.572***	-10.963***
	(1.329)	(1,718.969)			(1.466)	(3.345)

Note. Estimated standard error in parentheses, $***p \leq 0.01$; $**p \in (0.01, 0.05]$; $*p \in (0.05, 0.1]$

Only the sign and the statistical significance of the coefficients in all three models can be directly interpreted in multinomial logit models. For a quantitative interpretation, we calculate Average Marginal Effects (AMEs) in units of seconds. Due to the unmanageable computational effort, no standard errors of the AMEs can be calculated. Only AMEs based on statistically significant estimated coefficients are interpreted. The AME table can be found in Table 2.7 in the Appendix. The fixed effects estimator has the drawback of not allowing for constructing predictions that account for journey-level unobserved heterogeneity, so that AMEs cannot be calculated. An AME can be interpreted as follows: it gives the effect of the respective variable on the transition probability per second to parking search or immediate parking in percentage points.

Some of the implications of the estimated coefficients and the AMEs are as follows:

- As anticipated, the remaining walking distance to the destination plays a significant role in determining the timing of starting the parking search, as drivers tend to defer the search as long as possible to minimize journey duration. The quantitative effect is relatively consistent across all three models, with Model III indicating that for every 100 m closer the driver gets to the destination, the probability of starting the search increases by 0.49 percentage points per second. Furthermore, decreasing walking distance also increases the likelihood of parking immediately. Figure 2.9 illustrates this graphically. This finding is consistent with the theoretical model.
- Different levels of risk acceptance of the drivers could affect starting the parking search, and different properties of the destination area (e.g., parking occupancy and prices around the destination) could affect parking immediately without searching. These and other unobserved journey and driver characteristics make Model II better suited for studying the impact of (lagged) speed. In this model, a slower driving speed is associated with a higher probability of starting the search and parking immediately. Assuming that the speed in $t-1$ is a proxy variable for total driving speed (v_d), this result is in line with the prediction of our theoretical model. Figure 2.10 presents the average adjusted predictions in response to lagged speed based on Model III. It not only shows the negative effect of driving speed on the transition to searching, but it also indicates that transition to immediate parking occurs especially if the speed is very low, for example, during traffic jams.
- For a one-year increase in the driver's age, the probability of starting the search decreases by 0.02 percentage points per second, implying that older people tend to start the search later. A possible explanation in line with our theoretical model is that younger people may have a higher walking speed or a generally lower walking cost. The effect of age on transition probability is visualized in Figure 2.11.
- A driver's familiarity with the destination zone is positively associated with an increased probability of starting the search early as well as parking immediately, with an increase of 0.15 and 0.04 percentage points per second, respectively. This implies that drivers who are familiar with the area are more likely to initiate the search sooner or park without searching. Conversely, when drivers are new to the destination area, they may drive further to familiarize themselves with the parking situation and then start their search. In

our theoretical model, the probability of parking availability (p) as well as the remaining distance and the speed of both walking and driving is known by the agent. An extension that could explain these empirical results is the possibility that drivers unfamiliar with an area first try to gain knowledge about p as well as the remaining distances and speed. In contrast, "experienced" drivers already possess this knowledge and therefore tend to initiate their search earlier.

- A planned parking duration of less than 30 minutes yields a decreased search transition probability and an increased immediate parking probability. A possible explanation is that short parking and the activities associated with it (e.g., drop-off or pick-up) allow parking spots to be "imperfect" with respect to the driver's preferences. As a result, drivers may start their parking search later or immediately settle for the first available space they encounter.
- Journey purpose also seems to influence the search and parking transition, albeit the evidence is weak. Journeys toward home have a higher probability of ending without a search, which can be explained by the fact that in residential areas, parking demand may be lower than in areas where entertainment activities (base category) occur and that in residential areas, free-of-charge resident parking is common. Additionally, drivers are observed to initiate their parking search sooner for journeys with entertainment purposes compared to other journey purposes. Possible interpretations are that drivers may have a higher tolerance for walking and a (slightly) lower value of time for this journey purpose (Wardman et al., 2016).
- A likelihood ratio test of the hour dummies indicates that these variables are highly significant (Table 2.8 in the Appendix). In contrast, day-of-the-week dummies are jointly statistically insignificant and are therefore dropped. The probability of starting the search soon is higher during rush hours in the mornings and afternoons compared to night hours. In the context of our theoretical model, the reason could be a slower traffic speed due to congestion and a possible high parking occupancy (low p) during these hours.
- Given a constant walking distance to the destination, a higher count of remaining intersections decreases the probability of initiating the search process. One may infer that drivers exhibit a preference for fewer crossways during the walk. This may arise from the perceived difficulty (waiting times at traffic lights) of passing intersections on foot as opposed to in a vehicle.
- Looking at POI categories, the coefficients in Model II reveal interesting insights. In areas with more amenities, such as restaurants and bars, drivers start the search earlier. Conversely, in areas with many natural objects, such as trees and grasslands, drivers start the search later. The reason may lie in the correlation with parking occupancy (and hence p). Generally, in line with the conclusion of our theoretical model, the empirical results indicate that as finding a vacant parking spot gets more difficult, drivers tend to start the search sooner and vice versa. Additionally, in areas with many buildings (mainly residential buildings in OSM), the probability of parking immediately is higher, possibly due to free parking and higher supply in residential neighborhoods. This may eliminate

the need to search.

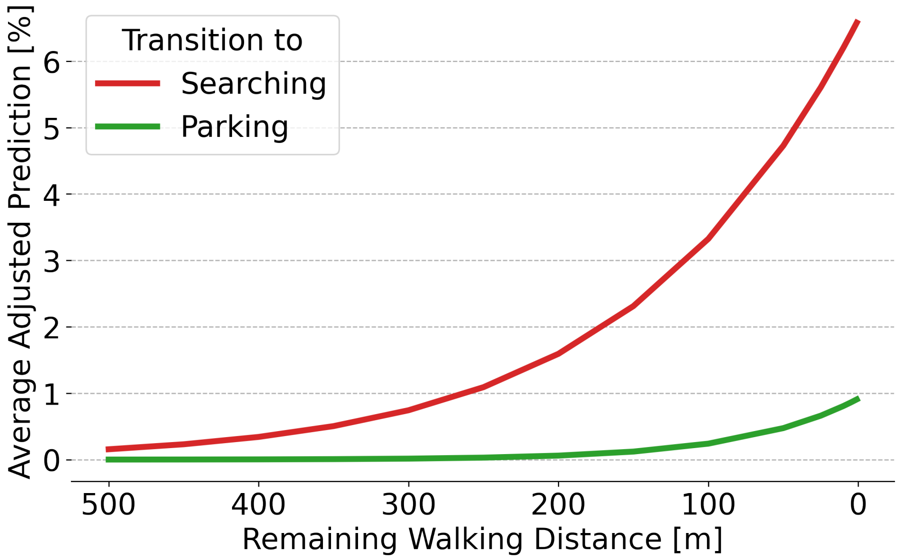


Figure 2.9: Predictive margins of walking distance on the transition probability per second in percent (based on Modell III)

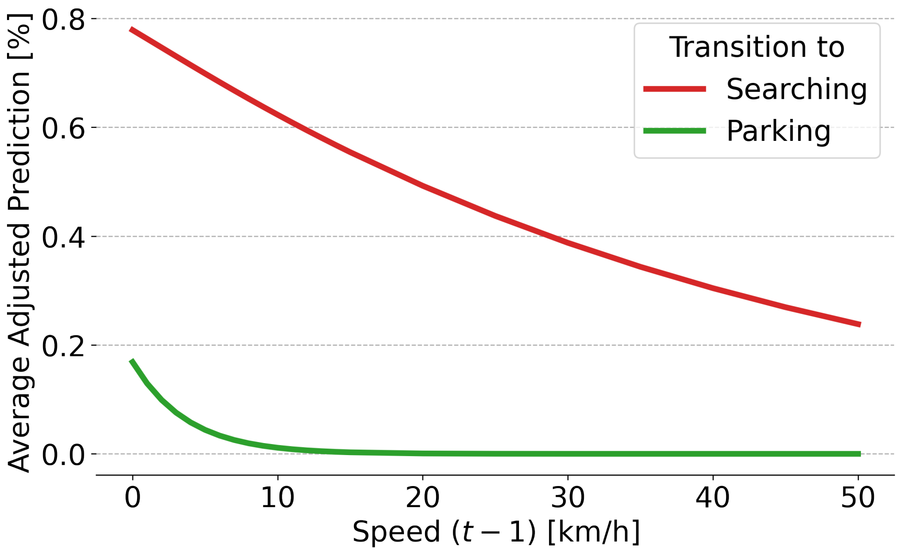


Figure 2.10: Predictive margins of (lagged) speed on the transition probability per second in percent (based on Modell III)

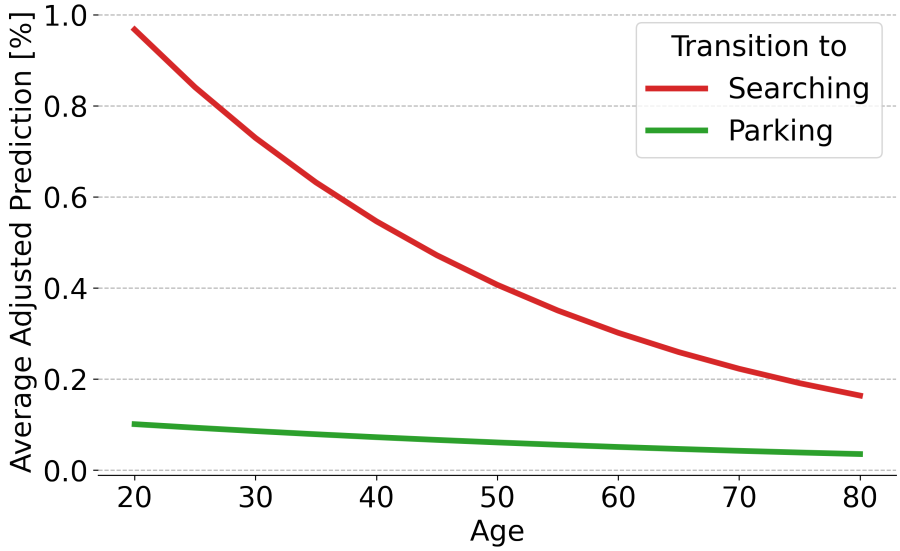


Figure 2.11: Predictive margins of age on the transition probability per second in percent (based on Modell III)

The probability of starting the search at each point, on average, is 0.21 percentage points larger per second for men than for women, indicating that men tend to start the search sooner. On the other hand, men are 0.11 percentage points per second less likely to park their vehicle immediately without searching. We cannot provide a gender-based explanation for the disparity between men and women. This outcome could also stem from a potential selection bias resulting from diverse types and objectives of journeys that are undertaken by men and women even after controlling for journey purpose and other variables as well as random effects in the model.

Identifying distinct impacts of weather conditions on parking search behavior is associated with a higher degree of uncertainty due to the relatively large standard errors. The evidence hints at a potential association between adverse weather circumstances, characterized by higher wind speeds and lower temperatures, and a later parking search start. This outcome is likely attributable to the increased costs of walking relative to driving in these conditions, as vehicles furnish drivers with a protective enclosure.

A positive correlation is observed between frequent drivers with an average yearly driven distance of more than 9000 km and their propensity to immediately park their vehicles without engaging in a search for parking. This could be attributed to the increased familiarity that frequent drivers may have with various driving settings, along with their greater experience in navigating the parking search process.

It is noticeable that a higher-class vehicle may lead to a later start of the parking search. Given that we do not control for income, a possible hypothesis is that drivers of higher-class vehicles may be more financially capable of paying for parking in a garage or lot, which eliminates the need for a time-consuming parking search. Thus, they may initiate their search closer to their final destination, and if they cannot quickly locate an available on-street parking spot, they might abandon the search and proceed directly to an off-street parking facility or an on-street parking spot with parking fees.

2.7 Conclusion

This study represents a unique contribution to the field of cruising for parking, as it focuses specifically on the starting point of the parking search process, which is measured empirically for the very first time. Firstly, a theoretical model has been proposed which seeks to explain the transition of a driver from the state of "normal driving" (not searching) to "searching". This model posits that a rational agent strives to determine the optimal starting point for the search process, with the goal of minimizing the total cost of the journey. According to this probabilistic model, the parking search starts if starting the search results in a lower expected journey duration than driving further away toward the destination.

In support of this theoretical model, we have collected ground truth data on parking search using a smartphone application developed for this purpose. This app collects floating car data of volunteer users. In addition to the GPS trajectory, four key points in each journey are labeled in order to identify the search route and duration. These are the starting point of the journey, the starting point of the parking search, the location of the parking spot, and the final destination. Based on this empirical evidence, hypotheses derived from the theoretical model were tested through a first-order Markov regression via a multinomial logit model with random and fixed effects that explores what factors to what extent contribute to this transition.

The results of this analysis indicate that driver-related, journey-related, and destination-related variables significantly impact the starting point of the parking search. For example, younger drivers, male drivers, and drivers of lower-class vehicles are found to initiate the search process sooner, possibly due to their higher tolerance for a longer walking distance and/or lower values of time. Furthermore, familiarity with the destination area was found to lead to an earlier start of the search process, whereas unfamiliarity with the area makes drivers proceed further toward their destination to learn about the parking situation. In rush hours and areas with many amenity points of interest, such as restaurants and bars, drivers are also observed to start the search sooner.

It is essential, however, to clarify the scope of this study in the context of the specific data collection method used. Our research was predicated on drivers using the app to start data recording when they anticipated the need to search for parking, which primarily occurs in on-street environments. Instances where users had predetermined off-street parking facilities, such as a shopping mall or a reserved spot at the destination, like a home garage, were typically not incorporated in the study. Users did not activate the app under such circumstances. It is noteworthy, though, that there are situations where drivers end up in an off-street parking facility, but it is generally a result of failing to find an on-street spot. The choice of this particular data scope was based on the interest of probing into the initial phase of the parking search process, predominantly in on-street situations.

In conclusion, our findings are generally consistent with the theoretical model, indicating that as the difficulty of finding a vacant parking spot increases or the driving speed decreases, drivers tend to initiate the search process sooner. It was also observed that in certain circumstances, the parking search process could be omitted entirely, with the vehicle being parked immediately. Factors such as driving toward home, having a planned parking duration of less than 30 minutes, and being familiar with the destination area are found to increase the probability of immediate parking without searching.

The unique dataset, containing the ground truth data on cruising for parking, offers many new opportunities to explore various facets of cruising for parking from another perspective. In future works, we plan to provide an econometric model of parking search duration that helps us gain a more accurate estimation of the marginal effect of various determinants of cruising and leads to a more precise assessment of the cruising behavior. Ultimately, this could result in new insights specifically important for policy-makers to improve the parking search situation.

2.8 Appendix

Table 2.7: Average marginal effects on the transitions per second in percentage points.

Variable	**Model I: pooled mlogit**		**Model III: mlogit random effects**	
	Transition to searching (k = 1)	**Transition to parking (k = 2)**	**Transition to searching (k = 1)**	**Transition to parking (k = 2)**
remaining walking distance [m]	-0.00202***	-0.00081***	-0.00491***	-0.00093***
	(0.00009)	(0.00008)	(0.00052)	(0.00013)
remaining intersections	-0.00280**	0.00044	-0.00705***	0.00047
	(0.00113)	(0.00048)	(0.00257)	(0.00060)
speed (t-1) [km/h]	0.00922***	-0.01353***	-0.01514**	-0.01864***
	(0.00239)	(0.00173)	(0.00636)	(0.00312)
age	-0.01041***	-0.00139***	-0.01984***	-0.00121**
	(0.00147)	(0.00048)	(0.00004)	(0.00061)
gender (male)	0.06462**	-0.09297***	0.20758***	-0.11184***
	(0.02726)	(0.02202)	(0.07119)	(0.02882)
average yearly driven distance less than 9000 km	-0.04246	-0.03336***	0.01609	-0.03270**
	(0.02618)	(0.01014)	(0.07050)	(0.01309)
vehicle type				
mini and small: base				
compact	-0.06360**	-0.00197	-0.10316	0.00402
	(0.02727)	(0.01151)	(0.07214)	(0.01555)
family, luxury, and SUV	-0.06926*	0.04148**	-0.08921	0.06358**
	(0.03658)	(0.01996)	(0.10160)	(0.02878)
van	0.16851	-0.01932	1.42669*	-0.03393
	(0.14368)	(0.04217)	(0.75767)	(0.04205)
familiar with the destination area	0.07541***	0.03935***	0.14732**	0.04196***
	(0.02275)	(0.0087)	(0.06262)	(0.01183)
planned parking duration				
less than 30 min: base				
30 min to 3h	0.02432***	-0.03028***	0.24534***	-0.03215***
	(0.0276)	(0.00875)	(0.08979)	(0.01205)
longer than 3h	-0.00126	0.00735**	0.08710	0.01225
	(0.04241)	(0.02097)	(0.12513)	(0.02856)
journey purpose				
entertainment: base				
home	-0.03826	-0.03404***	0.07248	-0.03907**
	(0.03292)	(0.01175)	(0.09904)	(0.01599)
shopping	-0.06214**	-0.00613	-0.05103	-0.00807
	(0.02681)	(0.01198)	(0.07442)	(0.01558)
work and business	-0.13887***	-0.01717	-0.09832	-0.01481
	(0.027)	(0.01227)	(0.08486)	(0.01705)
temperature	0.00128	0.00092*	0.00534*	0.00148*
	(0.0012)	(0.00057)	(0.00322)	(0.00076)
wind speed [m/s]	-0.00734	-0.00125	-0.02828*	-0.00204
	(0.00635)	(0.00223)	(0.01689)	(0.00294)

Note. Estimated standard error in parentheses, ***$p \leq 0.01$; **$p \in (0.01, 0.05]$; *$p \in (0.05, 0.1]$

Table 2.8: Coefficient estimation results – Time of the day

Variable	**Model I: pooled mlogit**		**Model III: mlogit random effects**	
	Transition to searching (k = 1)	**Transition to parking (k = 2)**	**Transition to searching (k = 1)**	**Transition to parking (k = 2)**
Time of the day				
Night hours: base				
07:00 - 08:00	1.153***	0.454	1.217***	0.409
	(0.286)	(0.553)	(0.398)	(0.626)
08:00 - 09:00	0.401**	-0.814	0.513*	-0.909
	(0.204)	(0.555)	(0.292)	(0.615)
09:00 - 10:00	0.394**	-0.833	0.387	-0.707
	(0.183)	(0.566)	(0.261)	(0.609)
10:00 - 11:00	0.124	-0.323	0.164	-0.335
	(0.179)	(0.417)	(0.252)	(0.472)
11:00 - 12:00	0.292*	-0.308	0.401	-0.380
	(0.173)	(0.434)	(0.245)	(0.488)
12:00 - 13:00	0.198	0.248	0.355	0.295
	(0.175)	(0.405)	(0.246)	(0.454)
13:00 - 14:00	0.087	-0.632	0.156	-0.534
	(0.176)	(0.403)	(0.245)	(0.455)
14:00 - 15:00	-0.205	-0.605	-0.004	-0.356
	(0.178)	(0.398)	(0.249)	(0.455)
15:00 - 16:00	0.141	0.161	0.213	0.170
	(0.175)	(0.396)	(0.245)	(0.446)
16:00 - 17:00	0.072	-0.520	0.170	-0.482
	(0.172)	(0.415)	(0.242)	(0.464)
17:00 - 18:00	0.079	-0.295	0.072	-0.317
	(0.170)	(0.388)	(0.238)	(0.434)
18:00 - 19:00	0.246	-0.100	0.245	-0.123
	(0.169)	(0.402)	(0.239)	(0.453)
19:00 - 20:00	0.320*	-0.281	0.396	-0.302
	(0.171)	(0.452)	(0.244)	(0.505)
20:00 - 21:00	0.190	-0.237	0.276	-0.231
	(0.183)	(0.471)	(0.260)	(0.533)
21:00 - 22:00	0.063	0.087	0.061	0.132
	(0.196)	(0.452)	(0.275)	(0.514)

Note. Estimated standard error in parentheses, $^{***}p \leq 0.01$; $^{**}p \in (0.01, 0.05]$; $^{*}p \in (0.05, 0.1]$

3 Article 2: Cruising for parking again: Measuring the ground truth and using survival analysis to reveal the determinants of the duration

Authors	Siavash Saki, Tobias Hagen
Affiliation	Research Lab for Urban Transport, Frankfurt University of Applied Sciences, Frankfurt am Main, Germany
Journal	Transportation Research Part A: Policy and Practice
Volume	183
Article Number	104045
DOI	10.1016/j.tra.2024.104045
Received Date	15 September 2023
Revised Date	24 January 2024
Accepted Date	20 March 2024
Available Online	8 April 2024
License	Open Access, CC BY 4.0

Abstract

To effectively address the challenges posed by parking searches, accurate quantification is essential. This paper introduces a novel methodology to collect unbiased ground truth GPS data on parking search behaviors. Using a mobile application, we tracked the exact starting point, chosen parking spot, and final walking destination of over 3,000 trips in Germany from 2021 to 2023. Our findings indicate a mean parking search duration of 1.5 minutes (in large city centers 1 minute and 53 seconds), a figure notably lower than previous survey-based estimates. This discrepancy suggests potential biases in traditional parking search surveys, possibly stemming from negativity bias. Our research employs a competing-risks survival analysis model to investigate factors affecting parking search duration simultaneously across different categories: Free, Paid, and Illegal parking. This enables the model to examine how the driver's choice between Free, Paid, and Illegal parking interacts with and influences the parking search duration. Furthermore, our duration dependency analysis, facilitated by a time-varying baseline hazard, reveals that prolonged search duration tends to make drivers more flexible toward less optimal parking options. This inclination is particularly evident for Paid parking, indicating that as the search duration extends, drivers increasingly consider Paid parking alternatives. Typically, drivers prioritize Free parking and opt for Paid options only after unsuccessful initial searches. This behavior underscores the notion that the coexistence of Free and Paid parking in urban areas might grow parking search traffic. Additionally, our data shows that approximately 5% of journeys end in Illegal parking. In these journeys, drivers often predetermine their choice of Illegal parking, as evidenced by the significantly shorter average search duration for these

cases and the result that the likelihood of Illegal parking is not affected by the previous search duration.

Keywords: Cruising-for-Parking, Parking Search Behavior, GPS data, Smart Data Collection, Survival Analysis

3.1 Introduction

Parking search is perceived as a common problem in densely populated urban areas and has far-reaching ramifications for transportation planning. The negative effects of parking search can be classified into four broad categories, namely economic, environmental, social, and safety impacts, based on the nature of the issues they precipitate (Brooke et al., 2014).

From an economic perspective, parking search poses a substantial burden on both individuals and the society. For individuals, the cost of searching for a suitable parking spot can be significant, particularly in urban areas where demand for parking is high (Shoup, 2005). According to Cookson and Pishue (2017), the average German driver loses 41 hours annually due to the time spent searching for a parking spot. For the society as a whole, the impacts of parking search are even more significant. The increased congestion and decreased mobility resulting from drivers searching for parking can lead to a loss of productivity and increased transportation costs for businesses (Albalate & Inci, 2018).

From an environmental standpoint, parking search has adverse impacts on air quality, noise pollution, and the natural environment (Millard-Ball et al., 2014). The increased congestion resulting from drivers searching for parking can lead to elevated levels of air pollution, particularly in urban areas where the concentration of vehicles is already high (Caicedo, 2010). This can have a detrimental impact on public health and can contribute to climate change (Alemi et al., 2018). Furthermore, the increase in traffic on roads during parking search can also result in elevated levels of noise pollution (van Ommeren et al., 2012), which can have negative impacts on local communities and wildlife. Attempts to reduce parking search traffic by constructing new parking facilities can lead to the destruction of natural habitats and the loss of green spaces (Chester et al., 2011).

From a social perspective, parking search can have several negative impacts on the quality of life in urban areas. The phenomenon can obstruct individuals' access to essential services and institutions, such as schools and hospitals, hindering mobility throughout the city (Dalla Chiara et al., 2020). Furthermore, the stress and annoyance experienced by drivers during their parking search endeavors can have a negative impact on their overall well-being (Melnyk et al., 2019).

From a safety perspective, parking search can result in distractions for drivers, which is often caused by the need to scan the curbside, assess the availability of parking spots, and read parking restrictions and road markings (Ponnambalam & Donmez, 2020). Driver distractions are well documented as a major contributor to automobile accidents, accounting for a staggering 68%

of crashes according to a study for 6 cities in the US (Dingus et al., 2016). Furthermore, the sudden decreased velocity by initiating the parking search and the sudden braking by observing a vacant spot can increase the risk of accidents (Sisiopiku, 2001).

In order to effectively tackle the challenges caused by cruising for parking, we need a thorough understanding of its quantitative significance. This is also crucial for the design and the evaluation of various policies on the issue. Commonly used approaches to measure search duration include surveys, experiments, analytical techniques, simulations, and the use of GPS data. In the following section, we present a summary of recent studies published after 2015 that report search duration and compare the collected values. Although each of these methods offers unique benefits, and the studies have yielded valuable insights into the cruising phenomenon, to date, no study has collected ground-truth data regarding parking search duration and path.

In this context, we propose and apply a novel approach for smart data collection that captures the complete journey trajectory, encompassing the driving phase up to the search initiation point, the parking search process, and the walking route to the final destination. We developed a mobile application that precisely captures the search initiation point and ultimate destination of a journey. These data, in conjunction with the initial point of the journey and the parking location, typically available in all GPS data, provide a comprehensive framework for analyzing parking search behavior within the context of car travel. This application has been collecting data in Germany since 2021.

This study aims to achieve two primary objectives. Firstly, it endeavors to present a descriptive analysis based on the ground-truth cruising data, which has been collected for the very first time. Secondly, the study seeks to investigate the factors that contribute to prolonged parking search durations.

The structure of this paper is as follows: Section 2 offers a concise review of prior research on cruising for parking, detailing various parking data collection methods and categorizing previous studies accordingly. Section 3 introduces our innovative data collection methodology. Section 4 presents a comprehensive descriptive analysis of the data, including metrics such as average search and walking durations, and outlines the dataset prepared for survival analysis. Section 5 explains the reasoning behind our model choice and delves into the model employed to examine the determinants of parking search. Section 6 showcases the results of our model, discussing its implications. Section 7 wraps up the paper by summarizing the primary insights and emphasizing their broader implications.

3.2 Literature Review

3.2.1 Modelling Parking Search Behavior

For several decades, the study of parking behavior has garnered significant interest among transportation experts and urban planners. In earlier research, various methods were employed to comprehend the underlying factors that influence drivers' decision-making regarding parking type and location (Van Der Goot, 1982).

One of the pioneering studies that underscored the crucial role of parking search behavior in the context of parking behavior is Polak and Axhausen (1990). This study posited that the search for parking is intricately linked to drivers' experience and knowledge, including their understanding of the spatial and temporal availability of parking spaces and associated costs. The researchers identified seven distinct search strategies, with one such approach involving circling around the final destination within a limited radius to locate an on-street parking space, and resorting to an off-street parking spot only as a last resort following a prolonged unsuccessful search.

Axhausen and Polak (1991) stands out as one of the initial studies to employ a choice model in examining the selection process of drivers among on-street (Free or Paid), off-street, and Illegal parking options while incorporating search duration as a significant model determinant. Following this influential work, numerous studies have adopted choice models, mostly employing Stated Preferences (SP) data, to scrutinize parking behavior (Hess & Polak, 2004; Ibeas et al., 2014; Soto et al., 2018). Another influential contribution to the field was the seminal work of Shoup (2005), which shed light on the fact that low-priced or free on-street parking incentivizes drivers to engage in "cruising" behavior, thereby exacerbating traffic congestion. Consequently, this insight has resulted in a heightened integration of the choice between off-street and on-street parking options into choice models.

Thompson and Richardson (1998) presented a choice model founded on the principle of utility maximization for the search process, which integrates a sequence of parking type, location, and route choices. This model postulates that the driver possesses prior knowledge of the area and is familiar with all available parking facilities. At each intersection, the driver decides by selecting the street segment with the highest utility based on various parking characteristics, such as fees, duration, and distance, or decides to terminate the search and proceed to an off-street facility.

Axhausen et al. (1994) and Arnott and Rowse (1999) propose a simple model that utilizes the reciprocal of the expected occupancy rate as an approximation for the search time. Empirical survey data for on-street parking search conducted by Belloche (2015) corroborates these approaches. Arnott and Inci (2006) and Arnott and Rowse (2009) have devised a parking behavior model that incorporates both traffic congestion and parking supply and explores pricing policies. Their findings reveal that curbside parking is considerably underpriced.

Various studies have employed simulation and analytical modeling to explore parking search behavior. Benenson et al. (2008), Horni et al. (2013), and Levy et al. (2013) have developed agent-based models for parking, while Cao and Menendez (2015), Liu and Geroliminis (2016), and Leclercq et al. (2017) have established dynamic macroscopic models. A relatively recent approach is incorporating computer games in simulations. For example, Geva et al. (2022) used PARKGAME serious game to model the cruising experience. Simulation modeling is an efficacious tool for generating distributions for variables of interest, such as parking search duration and walking distance, and examining the effects of different policies. Fulman and Benenson

(2021) have estimated the cumulative probability of search duration for a given destination based on parking demand and supply using an approximation method and validated this approach with results from an agent-based parking simulation. Ogulenko et al. (2022) have proposed a deterministic model that provides an analytical approximation for parking search characteristics, including cruising duration and the number of cars engaged in cruising. In a micro-simulation, Rodríguez et al. (2022) combines parking choice and searching models to assess dynamic pricing policies.

A complex aspect of parking search is the determination of the starting point of the search process, resulting in a proliferation of divergent hypotheses, which have made the understanding of the overall parking search process a challenging endeavor (Millard-Ball et al., 2020). The various assumptions that have been proposed range from the initiation of the car trip by the driver (Thompson & Richardson, 1998) to the point of arrival at the final destination (Jones et al., 2017). In addition, arbitrary rules have been established to identify the search process in floating car data. Notably, threshold approaches (distance-based rules), such as defining an initial search radius of 400 m around the parking spot, have been employed in some studies (Montini et al., 2012; Weinberger et al., 2020), whereas others have employed speed-related rules that incorporate speed and acceleration thresholds (Milia et al., 2023). In simulations, the starting point can be set through the use of a distance-to-destination parameter derived from a predetermined distribution or set as a specific value, such as 250 m, as suggested by Benenson et al. (2008).

van der Waerden et al. (2015) divides the previous studies into studies with a focus on empirical insights and studies with a focus on model development. In the following, we will deal with the former.

3.2.2 Measuring Parking Search Duration in Existing Studies

Many studies have reported a mean parking search duration as an indicator of cruising for parking. We can put the studies that reported a search duration in two categories based on how parking search duration is collected. The studies have used either **direct** data collection or **indirect** estimation. In the first method, direct data collection, the variable of interest is directly measured using tools or techniques that are specifically designed for that purpose. By indirect estimation, the variable of interest is estimated using other variables. This can be done by using analytical models, statistical models, or simulations.

Direct measurement is generally considered more accurate, but it can also be more expensive and time-consuming. Indirect estimation can be more efficient, but it may be less accurate if the underlying assumptions of the model are not met.

Parking search duration is collected directly mainly by:

- **Surveys**: In these studies, the search duration of drivers was obtained through survey reporting. The utilization of surveys to collect data on parking search behavior is probably the most straightforward and efficient method, commonly employed in various studies. Such surveys allow for targeted data collection from specific populations and facilitate policy-makers in prioritizing transportation investments. Furthermore, they offer an effective tool for evaluating policy outcomes through the collection of pre and post-policy implementation data. However, surveys also pose certain limitations, including a limited sample size and high costs, while their relevance may rapidly become obsolete. Additionally, potential confounding factors such as drivers' perception of parking search, recall bias, rounding to minute, or response bias could significantly impact the outcomes of such surveys. Lee et al. (2017), Qin et al. (2020), and Assemi et al. (2020) are several examples of studies that have employed surveys as a data collection method.

- **Field experiments**: In a typical experiment (usually referred to as "park-and-visit"), a driver initiates the search for a parking spot in a designated area, and the duration of the search is recorded once a vacant spot is found. This approach provides a higher degree of accuracy compared to survey-based methods, albeit with limitations of sample size and cost. However, these field experiments are restricted in their ability to account for the diverse search strategies employed by drivers due to individual preferences, as they do not replicate actual journeys. Additionally, these studies may suffer from arbitrary assumptions such as a vague search starting point. Several examples of this type of study include Belloche (2015), Alemi et al. (2018), and Zhu et al. (2020).

Parking search duration is calculated indirectly mainly by:

- **Simulations**: Simulation models are a commonly employed tool in transportation research that may involve the inclusion of real-world elements, such as street segments, parking lots, and on-street parking. These models have proven to be cost-effective and easily replicable, allowing researchers to evaluate different transportation policies across various regions and settings. However, their accuracy may be limited due to their reliance on assumptions that may not accurately reflect real-world conditions. Furthermore, simulations are typically calibrated using input data that may not be entirely accurate, such as an initial search radius. For instance, Waraich and Axhausen (2012) state that previous parking search models exhibit a bias towards overestimating search durations. Noteworthy studies in this realm include Benenson et al. (2008), Gallo et al. (2011), Horni et al. (2013), Arnott and Williams (2017), and Fulman and Benenson (2018).
- **Analytical modeling**: This approach uses mathematical modeling and statistical analyses to estimate parking search duration, providing potentially accurate results based on detailed input data. Such methods are advantageous in that they can be both cost-effective and flexible, allowing for the integration of various factors related to parking search, such as spatial characteristics. However, analytical methods are built on simplified assumptions that may not hold true in the real world. Additionally, their generalizability is often limited, since the data required for these methods, such as occupancy rates, may not be readily available in different geographic regions. Notably, Inci et al. (2017) and van Om-

meren et al. (2021) employ administrative parking data, encompassing parking arrivals and vacancies, to estimate parking search duration. Fulman and Benenson (2021) introduces an approximation model based on parking demand and supply.

- **GPS data**: Trajectories datasets can be vast and potentially representative. They can contain millions of real-world journeys, which can be utilized to extract mobility and parking patterns. This method has garnered significant attention in recent years due to the increasing collection of floating car data across geographical regions. However, working with GPS data can also pose challenges such as high computational power requirements, lack of contextual information, and data privacy concerns. In addition, one notable limitation of GPS data is the lack of information regarding the final destination of the journey and the walking path from the parking spot to the destination, posing considerable difficulties for researchers. As a result, it is not uncommon for researchers to make a false assumption that the found parking spot is the final destination, which can lead to erroneous conclusions (Montini et al., 2012). When utilizing GPS data to estimate search duration, two common approaches are employed. The first approach, namely the threshold approach, involves identifying the search process by determining the starting point of the search using arbitrary rules. Mantouka et al. (2021) posits an indicator for the initiation of the search when the remaining distance to the destination increases for the first time within a 400-meter radius around the parking spot. Alternatively, van der Waerden et al. (2015) suggests that speed could serve as an indicator for the start of the search and specifies a rule whereby the search begins when the average speed falls below 23 km/h with a change rate of less than 5 km/h. In a similar vein, Milia et al. (2023) utilizes speed thresholds from previous studies. In addition to speed, Hampshire et al. (2016) analyzes drivers' body movements in videos recorded during trips to determine the starting point of the search. The second approach involves defining a radius, such as 400 m around the parking spot, and assuming that the parking search takes place within this area. The excess time, defined as the difference between the actual path taken and the shortest path to the parking spot, is then calculated and referred to as cruising time (Montini et al., 2012; Mannini et al., 2017; Weinberger et al., 2020; Dalla Chiara & Goodchild, 2020).

Table 3.1 provides a summary of studies. The upper part shows studies that have employed direct data collection methods to report mean parking search duration (MPSD). The lower part of Table 3.1 summarizes studies that have used indirect estimation methods.

Table 3.1: Studies reporting mean parking search duration (MPSD in MM:SS)

Study	**City**	**Year**	**Method**	**MPSD**	**Further results / notes**
			Direct Data Collection		
Zhu et al. (2020)	Ningbo	2020	Experiment	06:02	Park-and-visit experiments
Assemi et al. (2020)	Brisbane	2019	Survey	0 (25%) ¡05:00 (40%) ¿05:00 (35%)	Focus on the central business district
Qin et al. (2020)	Beijing	2016	Survey	01:00	Focus on the central business district
Cao et al. (2019)	Zurich	2016	Survey	0 (61%) ¡05:00 (27%) ¿05:00 (12%)	Longest MPSD observed at noon being 13 min
Alemi et al. (2018)	San Francisco	2013	Experiment	[00:30, 02:00]	Park-and-visit experiments conducted by bikers
Brooke et al. (2018)	East Midlands, UK	2014	Survey	01:33	
Cookson and Pishue (2017)	Frankfurt	2017	Survey	10:00	Surveys done in multiple cities in the US, UK, and Germany
	Berlin	2017		09:00	
Lee et al. (2017)	Brisbane	2015	Survey	15:00	
Holguín-Veras et al. (2016)	New York	2015	Survey	26:00	Focus on fleet vehicles with a small sample size (n=16).
Belloche (2015)	Lyon	2008	Survey	[00:50, 11:06]	Various neighborhoods resulting in a wide range of MPSD
			Indirect Estimation		
Milia et al. (2023)	Vesterbro	2019	GPS	02:29	Lower MPSD during the COVID-19 pandemic
	Frederiksberg	2019		01:59	
	Islands Brygge	2019		01:49	
Dalla Chiara et al. (2022)	Seattle	2021	GPS	01:30	Focus on fleet vehicles
Mantouka et al. (2021)	Athens	2018-2019	GPS	02:11	Reported median PSD at 57 s is significantly lower than the reported MPSD
Fulman and Benenson (2021)	Bat Yam		Simulation/Analytical	[00:00, 10:00]	Geographic distribution of MPSD for Bat Yam, revealing broad MPSD variations across different areas
Dalla Chiara et al. (2021)		2019-2020	GPS	03:48	Focus on fleet vehicles
van Ommeren et al. (2021)	Melbourne	2014	Analytical	01:00	MPSD calculated at 90-95% occupancy rates
Weinberger et al. (2020)	Ann Arbor		GPS	02:06	
	San Francisco (1)			01:57	
	San Francisco (2)			02:48	
Dalla Chiara and Goodchild (2020)	Seattle	2018	GPS	02:18	Focus on fleet vehicles
Fulman et al. (2020)		2017-2018	Simulation	[00:00, 07:00]	PSD collected by a parking computer game on a simulator
Cao et al. (2019)	Zurich		Simulation	[00:00, 14:00]	MPSD between 19:00 and 09:00 almost zero. Longest MPSD between 12:00 and 16:00 being 6-14 min.
Mannini et al. (2017)	Rome		GPS	[00:00, 06:00]	Divided Rome into five macro-zones, finding higher MPSD in the city center.
Hampshire et al. (2016)	Michigan, US		GPS	00:45	Body movement analysis to identify search initiation in recorded videos
van der Waerden et al. (2015)	Turnhout	2012	GPS	01:18	

A striking observation is the considerable difference in the MPSD reported by studies using surveys versus those employing GPS data. Specifically, the average MPSD reported by surveys is approximately 8.4 minutes, while that reported by GPS data stands at 2.1 minutes. Notably, analytical and simulation models do not typically report a specific MPSD; instead, search duration is calculated for diverse spatial and temporal settings, yielding a range of realistic durations for various hours of the day or regions of the city. This disparity in reported MPSD between survey and GPS studies may be attributed to biases in both methods. One reason could be that surveys in this domain are often conducted during peak times and in areas where finding a parking spot is particularly challenging. This approach tends to highlight scenarios where parking occupancy is near or at capacity, such as in bustling downtown areas or residential neighborhoods during evening hours. As a result, the surveys may disproportionately represent the more extreme cases of parking search durations, like those exceeding 10 minutes.

Survey results may also be affected by response biases. It is well established that people tend to recall highly emotional memories more easily (Tully & Bolshakov, 2010), and when asked to provide an overall assessment of an experience, they may fail to recognize that an extreme event is not necessarily representative. An extreme experience of cruising may evoke intense emotions, further influencing individuals' judgments and perceptions, ultimately leading to an exaggerated report of MPSD instead of a more realistic and representative average. The results of the GPS studies may be biased due the underlying assumptions on the search radius, or the parking search behavior described above.

Finally, it is noteworthy that median search durations were generally found to be lower than mean values due to the presence of excessively long searches as outliers. Unfortunately, not all studies report median durations, so they could not be included in Table 3.1.

Each of the methods previously discussed for estimating parking search duration exhibits sound reasoning and logical coherence. Nonetheless, none of them can offer genuine ground truth data on parking search duration for real-world journeys. In the subsequent section, we elaborate on our approach to tackling this challenge by devising a novel method that enables the collection of exact parking search durations while circumventing the need for any assumptions. In a sense, our method combines GPS data with direct data collection.

3.3 Data Collection

As argued above, there are reasons to believe that the parking search duration has not yet been measured validly for real-world journeys. How can it be measured empirically avoiding assumptions and response biases? A valid empirical approach is to record the exact time of starting the parking search and finding a parking spot in an actual journey without affecting the search behavior.

We developed a mobile application especially for this purpose. This app records the whole journey from the moment that the driver starts the vehicle to the moment of reaching the final destination on foot. The app is designed with a deliberate focus on simplicity and ease of use. Its main page consists of a centrally located blue button, which must be pressed four times to completely record a journey (Figure 3.1). When users intend to initiate a car journey, they must press the "Start Journey" button immediately before starting driving. This action triggers recording, with the app beginning to collect Floating Car Data (FCD) from that point. FCD comprises location data associated with timestamps at short time intervals ranging from two to five seconds, with additional GPS data collected, such as location accuracy, heading, speed, and accuracy.

As soon as drivers initiate the search for an available parking spot, they press the "Start Search" button, which marks the location and time of the parking search starting point. After successfully securing a parking spot, drivers press the "Vehicle Parked" button, which labels the location and time of the end of the search, i.e., the parking spot.

The final step entails the driver disembarking the vehicle and walking to their intended destination. Upon reaching this final destination, drivers press the "Destination Reached" button, which marks the end of the journey data collection process.

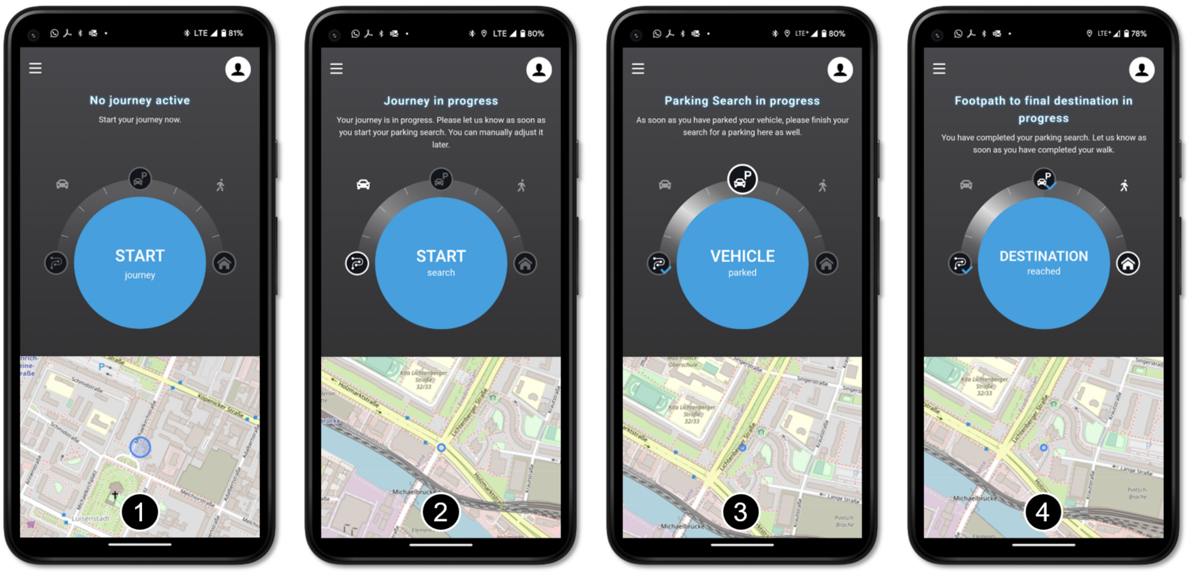


Figure 3.1: Screenshots of the start2park app showing the four steps of recording a journey

The proposed approach for data collection allows to collect valid parking search data of actual journeys. A sample journey collected using this app and its fundamental four points are visualized in Figure 3.2. As illustrated in this example, every journey is divided into three distinct phases: driving until search begins, parking search effort, and walking to the final destination. To facilitate ease of reference, the phase of driving before initiating the parking search is referred to as "normal driving." The recorded sampling rate for most of the data in this study is 1 second, ensuring a detailed and precise capture of each phase of the journey.

This methodology is based on the assumption that the initiation of a parking search is a distinct point in time, not a continuous transition. We acknowledge, however, that this approach may oversimplify the complex, subjective decision-making process inherent in initiating a parking search. Different drivers may perceive the start of a parking search at varying points based on individual criteria and situational awareness. While our method offers consistency and practicality in data collection, it might not fully capture the nuanced transitions that mark the commencement of a parking search for every individual. In future studies, more intricate methods that integrate behavioral and contextual factors could be explored to define and record the onset of parking searches with greater precision, better reflecting the drivers' experiences (Hampshire et al., 2016).

Furthermore, we must consider the potential behavioral changes induced by the monitoring process itself. The awareness of being tracked by a mobile application might influence the drivers' parking search behavior and decision-making patterns. Participants aware of the data collection might alter their natural parking search habits, e.g., avoiding certain types of parking, such as illegal spots. This Hawthorne effect (McCarney et al., 2007), where subjects modify an aspect of their behavior in response to their awareness of being observed, is an inherent limitation in studies involving active participant monitoring.

Compared to alternative approaches, particularly threshold-based methods, it is important to acknowledge the potential similarities in outcomes between these methodologies. Our method offers significant accuracy when the parking search is brief. However, longer parking search durations, which are more critical from a practical standpoint, would likely be identified by both our approach and threshold-based methods. It is worth noting that while our method provides detailed granularity and a high accuracy in the initial search phase, the overall distribution of search times may not drastically differ from threshold-based approaches. This similarity suggests that both methods are valuable in their contexts.

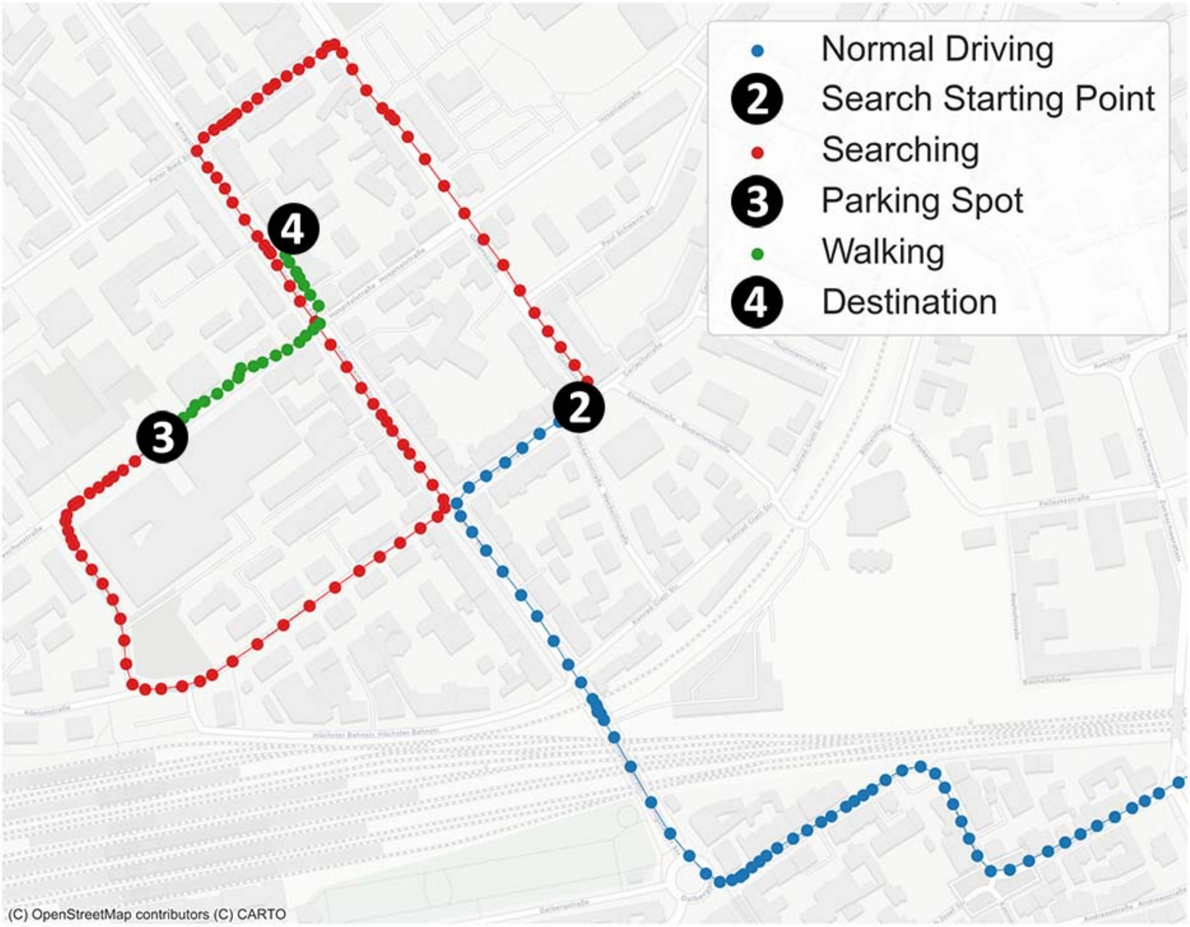


Figure 3.2: Real-world example of an actual journey recorded by start2park app. Driving until search begins (blue), parking search route (red), and walking to the final destination (green)

The application is also designed to gather demographic data, activated upon its initial usage following installation. It requests personal data from the user, including gender, age, annual average driving distance, and vehicle type. These details are subsequently connected with each journey the user records via their device. Providing such information, however, is not obligatory. Users have the right to bypass this step, supply the details at a later time, or simply record journeys without providing this demographic information.

Besides demographic data, the app also collects information about the journey. Upon completion of each trip, the driver is prompted to answer four queries concerning the purpose of the journey, the planned parking duration, his or her familiarity with the destination, and the type of parking spot. The subsequent chapter provides a descriptive analysis of the potential values for these driver- and journey-related variables.

Since August 2021, the app has been accessible on both the Google Play Store and Apple App Store, with interfaces being available in English and German. From the beginning of its launch, the application's advertisement campaign has leveraged a multitude of channels, ranging from municipal websites and academic conferences to social media platforms, television, radio, and YouTube channels. These promotional efforts have significantly supported the app's popularity, evidenced by its download count in both stores, which has reached several thousand. The user base comprises primarily volunteers eager to contribute to research on parking search. In return, the app enables users to log their parking search patterns and durations, providing detailed analysis for them.

It is crucial to note that the app does not provide additional services, such as facilitating the search for parking spots. This deliberate exclusion is designed to prevent any potential bias that could potentially affect the results.

As expected, the app records numerous instances of invalid journeys, resulting in unusable data, necessitating the implementation of filtering processes. For example, a common instance of invalid data is when a user, seeking to acquaint themselves with the app's functionalities, clicks multiple times consecutively to observe the application's response and progression. This action typically yields a rapid succession of data points concentrated in the (almost) same location within a brief period.

Moreover, standard data cleaning procedures are enacted to remove further instances of invalid data. These include, but are not limited to, eliminating incomplete trips, journeys exclusively categorized as 'walking' or 'searching', and those exhibiting significant signal loss. Upon the conclusion of these preliminary data cleaning steps, an additional criterion is applied to the remaining data. Specifically, a journey must encompass an origin-destination distance of a minimum of 1.5 km and last at least 10 minutes to be considered valid.

To enrich our dataset of valid journeys, we used additional data sources. Meteorological data, sourced from the German Weather Service (Deutscher Wetterdienst), were integrated as a single variable using the temperature degrees measured at the nearest weather station in Germany. Moreover, each journey is assigned a specific regional classification corresponding to the destination's location. The regional classification was accomplished using the RegioStaR (RegioStaR Creators, 2021) spatial typology, which divides land into distinct levels of urban and rural areas employing a hierarchical scheme.

For reasons described below, we calculated the hypothetical walking distance from the search starting point to the final destination using an open-source routing engine called Valhalla (Valhalla Development Team, 2023). This variable can be interpreted as the hypothetical walking distance that the driver is willing to go on foot at the end of trip.

3.4 Descriptive Analysis

3.4.1 Behavioral Patterns in Parking and Walking Durations

In our dataset, recorded from August 2021 to August 2023, 162 users conducted a total of 3564 journeys. These journeys represent the valid trips, constituting approximately 52% of all initiated journey recordings. Therefore, the average number of valid journeys per driver, based on this dataset, is 22. Examining the distribution of these journeys among the drivers, we find that the 5th percentile, median, and 95th percentile for the number of journeys per driver are 1, 3, and 87, respectively. This distribution indicates that a large number of users have logged fewer than a handful of journeys. In fact, the majority of the journeys are accounted for by nearly half of the users in the dataset.

The distribution of the Parking Search Duration (PSD) across these journeys is exhibited in Figure 3.3. Upon close inspection, a clear distinction separating journeys with a PSD nearly equal to zero from those with a PSD greater than zero is noticeable. This distinction effectively divides the dataset into two distinct categories. Approximately 18% of the journeys have a PSD$\approx$0, indicating that the drivers found immediately an acceptable parking spot.

Regarding the peak at 0 seconds observed in Figure 3.3, this phenomenon likely reflects a specific user behavior associated with the activation of the start button in our app. This peak represents instances where drivers record the parking search only after they have already parked their vehicle. Our interviews with volunteer users of the app revealed that most instances of journeys logged with a duration of 0 seconds occur when enough free parking spaces are available at the intended destination. In such cases, the driver proceeds directly to the final destination, parks the vehicle, and then records the duration as zero. However, there are also instances where the driver, en route to their destination, encounters a vacant parking spot and wants to take it. In these cases, the observation of the available parking space essentially triggers the initiation of the parking search, but the duration of this search is recorded as zero since the driver secures a parking spot immediately and records the duration thereafter.

Our dataset reveals that the MPSD for all journeys is 1 minute and 29 seconds, whereas the MPSD for journeys with a PSD greater than zero is measured to be 1 minute and 49 seconds. Additionally, the mean walking duration (WD) is recorded as 2 minutes and 31 seconds for all journeys and 2 minutes and 52 seconds for journeys where PSD is greater than zero. Detailed statistical analyses of these time variables can be seen in Table 3.2. Figure 3.3 indicates a right-skewed distribution implying that medians are smaller than the mean values (Table 3.2). Accordingly, an extended PSD is a rather rare event, as only 5% of the journeys with a parking search have a duration of more than 6 minutes and 7 seconds. Furthermore, Figure 3.4 illustrates the distribution of walking durations. It is noted that the average total time from starting the parking search to reaching the final destination is 4 minutes and 9 seconds.

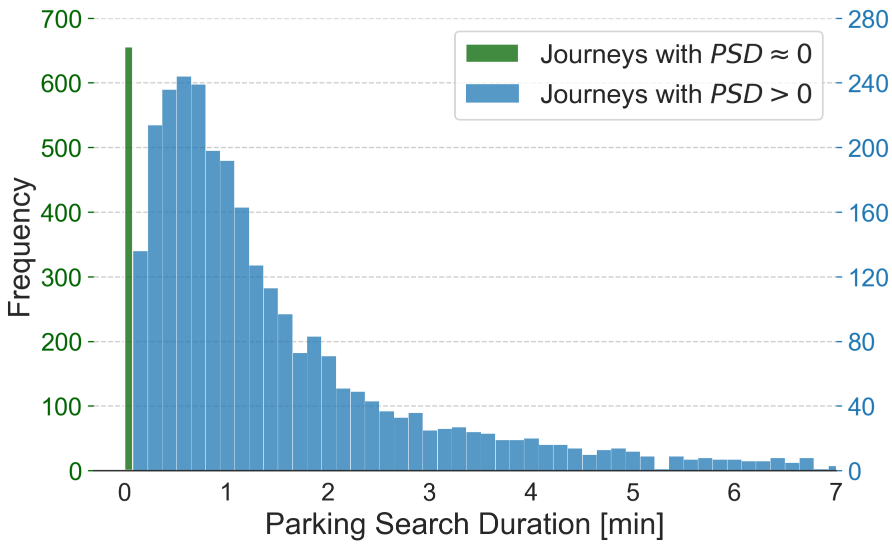


Figure 3.3: Parking Search Duration Distribution

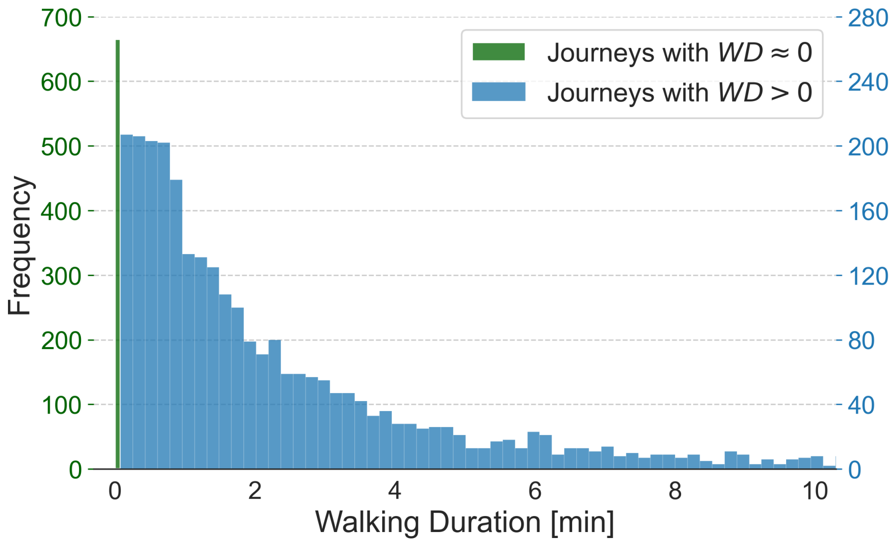


Figure 3.4: Walking Duration Distribution

The mean values in our study are closely aligned with previous research that utilized GPS data to quantify PSD. This alignment may be attributed, in part, to the fact that GPS-based studies typically encompass a more comprehensive scope, capturing journeys across various times of the day and from different areas within a city. This broader approach allows for a more inclusive representation of parking behaviors, as opposed to surveys that might focus on specific, high-demand times and locations. Nonetheless, it is noteworthy that our results still fall on the lower end of the spectrum. This observation confirms the reliability of GPS data as a valuable resource for investigating parking search behavior, reinforcing its applicability and relevance in this study area.

Additional variables to examine are the initial search radius and accepted walking distance. The initial search radius is the straight-line distance between the start of the search and the final destination of the journey. The accepted walking distance represents the hypothetical distance a driver is willing to traverse on foot from the initiation of the search to the final destination. These variables have been investigated exclusively for journeys that include a parking search process (PSD>0) in Table 3.3. In this subset of the dataset, the average initial search radius is 140 m, while the average accepted walking distance is 187 m.

Table 3.2: Parking search, walking, and journey duration statistics in MM:SS

Variable	count	5%	25%	median	75%	95%	mean	std
All Journeys								
Parking Search Duration	3546	00:00	00:18	00:50	01:45	05:27	01:29	02:10
Walking Duration	3546	00:00	00:16	01:07	02:55	10:27	02:31	03:58
Journey Duration	3546	06:23	11:00	16:32	24:24	44:09	19:49	13:51
Journeys with PSD > 0								
Parking Search Duration	2908	00:13	00:35	01:04	02:04	06:07	01:49	02:16
Walking Duration	2908	00:01	00:28	01:19	03:10	10:30	02:52	05:34
Journey Duration	2908	06:29	11:08	16:48	24:27	42:57	19:42	13:01

Table 3.3: Initial Search Radius and Accepted Walking Distance for Journeys with PSD > 0 in Meters

Variable	count	5%	25%	median	75%	95%	mean	std
Initial Search Radius	2908	13	52	106	188	368	140	126
Accepted Walking Distance	2908	13	66	139	245	509	187	193

After examining the general trends in parking and walking durations, as well as initial search radius, we continue by exploring the influence of location types on parking search behavior. This analysis is pivotal in understanding how urbanization impacts the parking search dynamics. To this end, we have utilized the RegioStaR classification, a standard regional typology in Germany, to categorize each journey based on its urban context.

In Figure 3.5, we adopt a non-parametric method, i.e., Kaplan-Meier estimator for visualizing the survival functions of parking search duration across different regional types in Germany, as defined by RegioStaR. The survival function gives the proportion of the sample still searching for a parking after time t. For instance, after two minutes of searching, about 35% of drivers in large city centers and 21% in the suburban areas of large cities are still looking for parking. In contrast, in small cities, only 13% of drivers are in search at the same time. The visualization reveals that PSD tends to increase with the degree of urbanization. Notably, large city centers exhibit the longest MPSD, standing at 1 minute and 53 seconds, a duration significantly longer than that observed in other regions, such as small cities with an MPSD of only 50 seconds.

Building on our understanding of regional influences on parking search duration, we proceed to integrate time of the day into our analysis. This step is crucial as it adds a temporal layer to our

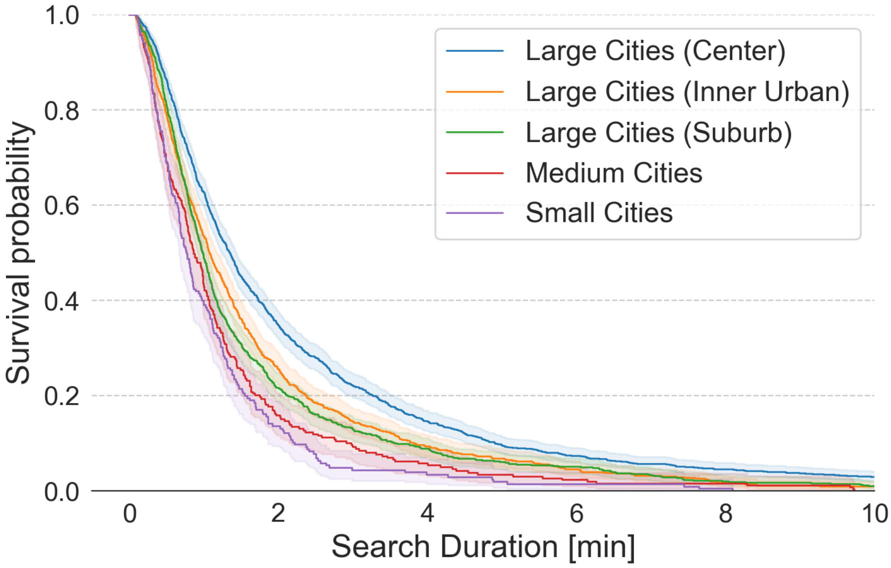


Figure 3.5: Kaplan-Meier estimate of survival function of parking search durations by regional types in Germany and corresponding 95% confidence intervals

spatial observations, offering a comprehensive view of how parking search behavior varies not only across different locations but also at different times. To achieve this, we created Table 3.4, which combines both the regional typology and time categories. Since for some combinations the number of observations can be small, the corresponding figures should be interpreted with caution.

In large city centers, we observe a pronounced trend of longer parking search durations, especially prominent during morning and noon hours. This phenomenon can be attributed to the high volume of traffic and limited parking availability typical of bustling urban centers during working hours. This trend reverses at night, with significantly shorter durations, suggesting a reduction in parking demand as the city quiets down.

Large cities' inner urban and suburban areas display a different pattern with moderately long parking search durations. The suburban areas, in particular, show a peak in parking search durations at night, possibly linked to residents returning home and competing for limited parking spaces in residential neighborhoods. During this time, the parking occupancy is probably near capacity with a low turn-over rate, which makes finding a vacant parking spot challenging.

Contrasting with the urban centers, medium and small cities consistently exhibit shorter parking search durations. This trend highlights the less congested traffic conditions and more readily available parking spaces in these less densely populated areas. This suggests that parking search may not be a relevant problem in those areas.

When it comes to walking durations, for the walk from the parked car to the destination, a clear distinction emerges between urban and less urban settings. Larger urban areas, especially city centers, show longer walking durations. This pattern indicates that, in dense urban spaces, finding a parking spot close to one's final destination can be challenging. Conversely, in medium and smaller areas, the walking durations are notably shorter, pointing to a higher likelihood of finding parking spots closer destinations.

Another interesting aspect is the initial search radius, particularly during morning hours in large and medium cities. The larger search radius observed during these hours suggests that drivers might be more willing to park further from their destination, possibly due to morning work commutes. As the day progresses into evening and night, we notice a decrease in search radius in large city centers and suburban areas, with a substantial increase in the latter at night. This pattern could be indicative of a self-regulatory mechanism in parking search behavior (Millard-Ball et al., 2020). As parking occupancy increases and finding a spot becomes more challenging, drivers may start their search earlier, effectively preventing parking search durations from escalating excessively. This pattern relates to a higher acceptance of the search radius and consequently increases walking duration. This trade-off reflects a strategic adaptation by drivers to the evolving parking conditions throughout the day.

The data reveals an adaptive nature of parking search strategies over different times of the day. Drivers in urban areas, especially during peak hours, appear to navigate a more constrained environment, adjusting their search strategies accordingly. The variability in both search radius and parking duration across different times of day underscores the dynamic nature of urban parking. Longer walking durations in these areas, particularly during peak hours, reflect the compromises drivers make in response to limited parking availability.

Table 3.4: Comparative Analysis of Urban Parking Behavior by Time and Location – Search Radius in Meters, Searching and Walking Durations in MM:SS

Location Type	Variable	Early Morning (04-08)	Morning (08-12)	Noon (12-16)	Afternoon (16-20)	Evening (20-24)	Night (00-04)
Large Cities	Journeys Count	18*	198	336	426	274	42
(Center)	Mean Initial Search Radius	119	162	145	134	115	73
	Mean Parking Search Duration	01:21	02:09	02:11	01:54	01:28	00:50
	Mean Walking Duration	01:56	02:55	03:02	02:52	02:40	02:54
Large Cities	Journeys Count	18*	129	225	269	133	25*
(Inner Urban)	Mean Initial Search Radius	69	134	101	129	101	52
	Mean Parking Search Duration	00:24	01:16	01:30	01:23	01:40	01:11
	Mean Walking Duration	02:01	01:44	02:25	02:59	02:56	02:13
Large Cities	Journeys Count	20*	191	275	286	72	15*
(Suburb)	Mean Initial Search Radius	63	131	129	114	114	178
	Mean Parking Search Duration	00:33	01:34	01:12	01:14	01:39	02:42
	Mean Walking Duration	01:06	02:14	02:20	02:11	02:45	02:24
Medium Cities	Journeys Count	24*	69	111	102	31	2*
	Mean Initial Search Radius	103	131	96	90	81	0
	Mean Parking Search Duration	00:40	01:09	00:57	00:54	01:24	00:01
	Mean Walking Duration	01:38	02:50	01:32	02:18	03:21	00:01
Small Cities	Journeys Count	9*	75	86	86	15	2*
	Mean Initial Search Radius	128	102	96	64	92	29
	Mean Parking Search Duration	00:40	01:06	00:46	00:44	00:48	00:18
	Mean Walking Duration	04:31	01:54	02:04	01:28	01:06	00:20

*Please interpret carefully, as the number of observations is low (< 30).

3.4.2 Parking Dynamics in Frankfurt am Main

Given the geographic focus of our advertising campaign in the city of Frankfurt am Main, Germany, a significant majority of our user base, and therefore, most of the recorded journeys, are located within this city. Frankfurt am Main, a major metropolis situated in the heart of Europe, is renowned as one of the continent's largest transportation hubs. The city's extensive network of roads, its bustling airport – one of the world's busiest – and its status as a key railway junction all contribute to its vibrant and dynamic traffic landscape. In terms of parking, Frankfurt presents a challenging environment, particularly in the city center. The high demand for parking spaces is driven by the influx of commuters, tourists, and local residents, all competing for limited parking availability. This situation is exacerbated in central areas like the banking district and around popular destinations such as the Zeil shopping street, where finding a parking spot can be particularly difficult. Parking search times tend to be longer in these areas, reflecting high parking occupancy rates.

Among the total dataset, 2363 journeys were collected in Frankfurt, resulting in an MPSD of

1 minute and 39 seconds across all journeys and 1 minute and 58 seconds specifically for 1982 journeys with a PSD greater than zero. These findings are substantially lower when compared with the mean value of 10 minutes reported for Frankfurt in Cookson and Pishue (2017) based on a survey.

To further illustrate the city-specific parking search conditions, Figure 3.6 presents a heatmap of the MPSD across different areas within the city of Frankfurt. The city center, shown in this figure, aligns with the RegioStaR regional typology. This visual representation confirms the existence of significant variations in parking search situations throughout the city. More explicitly, the heatmap reveals that in particular districts, such as the city center, the MPSD can rapidly increase, even exceeding 5 minutes. This analysis underscores the profound influence of location on parking search durations and highlights the differential parking search experiences across various regions within the same city.

3.4.3 Investigating Driver Behavior Regarding the Distance to Destination

Another metric employed in this analysis is the linear distance to the destination (DtD), computed over time spent searching for parking. The average trend demonstrated in Figure 3.7 suggest intriguing stages of parking search behavior that deserve detailed exploration. However, before going into detail, it must be stressed that the aggregate data in Figure 3.7 results from two processes, which cannot be distinguished without deeper analysis:

1. **Changing behavior over time / adaption of the search strategy:** Drivers may change their behavior with increasing duration of the parking search. For example, they may increase their initially chosen search radius (and hence the DtD) after not having been successful.

2. **Sample selection effect:** In examining a diverse group of individuals over a specified duration, the composition of the sample is expected to change. This is because those with the highest success probability (so-called "hazard") often leave the sample of "searchers" earliest. As a result, the overall probability of remaining in the sample ("survival") for the entire group changes as time progresses. The sample of trips after 4 or 8 minutes is not identical to the sample of trips at the beginning of the search. It becomes smaller with increasing time (resulting in broader 95% confidence intervals) as more trips have ended after finding a parking spot. This is often called "weeding out" or "sorting effect" in the survival analysis literature (van den Berg 2001). The presence of this phenomenon can produce survival patterns for sub-populations that are entirely different than the whole population (Hess & Persson, 2012). Therefore, an increasing DtD may simply reflect the phenomenon that after some time primarily trips remain in the sample, which have – for unknown reasons – a larger DtD.

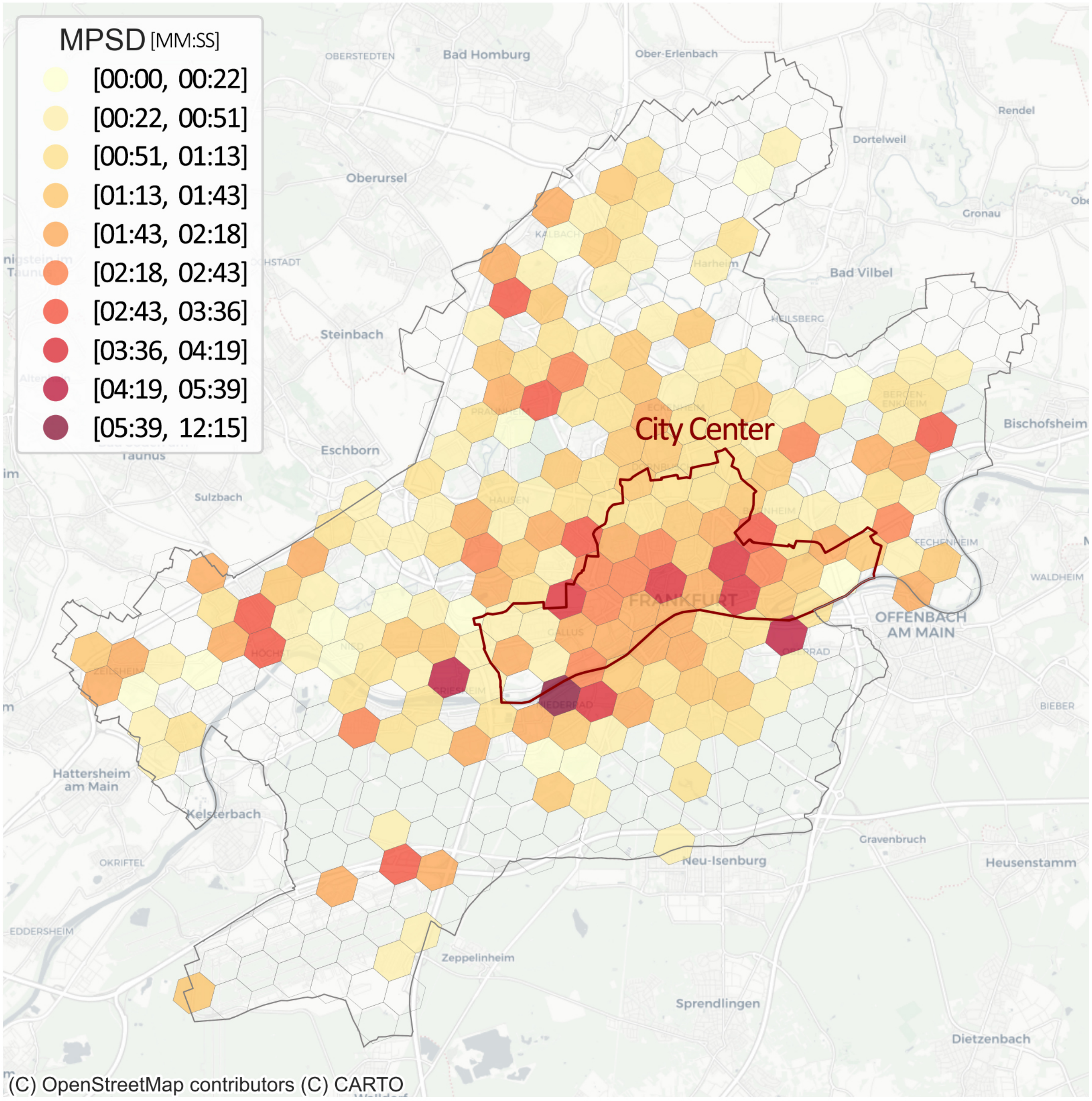


Figure 3.6: Heatmap of Parking Search Duration in Frankfurt am Main - Highlights Zones of High and Low Average Search Durations

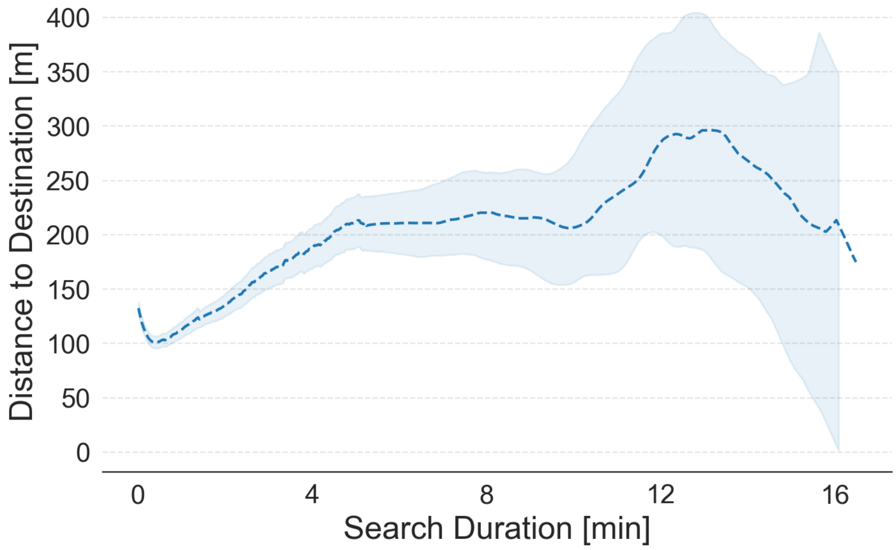


Figure 3.7: Average distance to destination over searching time. The light blue area shows the 95% confidence interval.

This ambiguity in the descriptive analysis with respect to the distinction between the two processes is one reason to use a survival regression analysis in Chapter 3.5.

According to Figure 3.7, initially the average DtD starts at an approximate 140 m, which intuitively aligns with the fact that the majority of drivers begin their parking search prior to reaching their final destination (Millard-Ball et al., 2020). As anticipated, the DtD shrinks, over the initial seconds, to approximately 100 m, indicative of drivers nearing their destination. This trend emerges thereafter, with the DtD progressively increasing to approximately 220 m over a period of 5 minutes of continued search time. To the extent that this is not based on the sample selection effect described above, this subsequent expansion of the DtD could be attributed to the failure of initial parking searches near the destination. It is plausible that drivers, unable to secure a parking spot in the immediate vicinity, are compelled to expand their search radius, thereby leading to an increased DtD.

Following this, the DtD remains relatively steady in the next phase, fluctuating minimally until the 11-minutes mark. It is likely that during this stage, drivers are circling the same general area in hopes of a parking spot becoming available. This period of stability is abruptly terminated by a sharp rise in the DtD to around 300 m, persisting until about 13 minutes into the search. This might represent a change in strategy, possibly influenced by growing urgency or frustration. Drivers might be driving further from their target destination, perhaps in search of less congested areas. However, also due to the reduced sample size at this time, these changes are not statistically significant.

This rapid increase is followed by a rapid decrease back to the initial average DtD of 150 m. The final decrease in DtD towards the end of the search could suggest a return to the vicinity of the destination. They might have learned about the parking situation in the area which results in securing a better parking spot with a shorter walking duration.

3.4.4 Explanatory Variables of the Estimation Sample

We have selected a subset of journeys exhibiting a PSD greater than zero that contains the driver-related and journey-related information provided by the user. A descriptive analysis of this filtered sample is presented in Table 3.5 and Table 3.6. This dataset provides the basis for the analysis in Chapter 3.5. Note that there is a difference between commuting and business-related in journey purpose. Commuting refers to regular travel between one's residence and their primary place of work, like daily drives to an office. Business-related refers to trips made for professional reasons, but not regular work commutes, such as a craftsman traveling to a site for repairs.

In advancing our understanding of parking search duration, survival analysis offers a robust and effective approach. Survival analysis is a statistical methodology often used in medical, social and economic sciences to analyze the time until the occurrence of a specific event, referred to as "failure". In the context of our study, the "failure" event is the completion of the parking search process. Utilizing survival analysis can facilitate a more comprehensive investigation of the factors affecting parking search duration, considering the time-varying factors that affect drivers parking search behavior and success.

Table 3.5: Journey Characteristics Separated by Parking Type – Values of Categorical Variables (Absolute Numbers)

Variable	Free Count	Paid Count	Illegal Count	All Journeys Count
All Journeys	2419	249	146	2814
Gender				
Female	264	76	13	353
Male	2155	173	133	2461
Vehicle Type				
Compact and Small	2087	209	136	2432
Medium and Large	310	39	5	354
Van	22	1	5	28
Journey Purpose				
Home	599	30	18	647
Leisure	692	59	40	791
Shopping	703	89	66	858
Commuting	286	40	1	327
Business-related	139	31	21	191
Area Familiarity				
Known	1759	172	86	2017
Unknown	660	77	60	797
Parking Duration				
Shorter than 30 min	736	74	116	926
Longer than 30 min	1683	175	30	1888
Area Type				
Village or Small City	173	6	6	185
Medium City	210	31	6	247
Large City: Suburban	629	30	22	681
Large City: Inner Urban	554	43	60	657
Large City: Center	853	139	52	1044
Time				
Night Hours: 22-07	195	7	14	216
07-10	212	42	8	262
10-13	421	66	25	512
13-16	526	64	27	617
16-19	626	49	49	724
19-22	439	21	23	483
Day of the Week				
Weekday	1862	214	104	2180
Saturday	289	19	17	325
Sunday or Holiday	268	16	25	309

Table 3.6: Journey Characteristics Separated by Parking Type – Values of Numerical Variables

Variable	Free	Paid	Illegal	All Journeys				
	Mean	Mean	Mean	Mean	Std	5%	Median	95%
Age	30	31	24	30	10	21	25	52
Temperature (K)	285	287	284	285	7	273	286	299

Parking search behaviors vary across different parking categories. Specifically, Paid parking exhibits a longer overall search duration with an MPSD of 2 minutes and 6 seconds compared to

Free parking with an MPSD of 1 minute and 27 seconds, while Illegal parking demonstrates the shortest search duration among the three categories with an average standing at 1 minute and 6 seconds. It is important to recognize that the strategy for the type of parking is not perfectly predetermined before the initiation of the parking search process. As the search unfolds, drivers may modify their initial strategy and begin to seek alternative parking types, especially switch from Free to Paid parking. It appears that Free parking is often the first preference for drivers, with the search for Paid parking options starting only after unsuccessful attempts to locate a Free parking spot. This observation aligns with the findings reported by Shoup (2005). This flexibility in strategy underscores the dynamic nature of the parking search process.

In contrast, the majority of Illegal parking instances are characterized by a comparatively shorter search duration, suggesting that, in many cases, the decision to park illegally is made prior to the parking search. This could indicate that drivers, in these situations, chose to skip the search process and park illegally close to final destinations of the journeys. Such events tend to be associated with shorter parking durations.

3.5 Modeling Search Duration

A natural candidate for modeling the search duration is survival analysis, which deals with modeling the time it takes until an event of interest occurs (the so-called "survival time"). This analytical model has recently gained more attention concerning the parking search duration (Fulman et al., 2020; Zhu et al., 2020; Mantouka et al., 2021). The event here is *"finishing the search process by parking the car"*. This event can be further distinguished into (1) *"Free"*, (2) *"Paid"*, and (3) *"Illegal"* parking spaces. The decisions and hence the search processes are likely to be interrelated. For example, people may decide to choose Paid parking after having not been able to find a Free parking spot. Hence, the survival times until the three events are likely to be correlated, so they have to be modeled simultaneously within a so-called competing risk model in order to avoid biased estimates (Schmid & Berger, 2021; Schuster et al., 2020).

Generally, it is distinguished between continuous time and discrete time survival analysis. Since the time frequency of our data is relatively high – 82% of all GPS points are measured every second – one may argue that the time is measured continuously. However, we have several reasons for modeling in discrete time – further advantages are discussed in Tutz, Schmid, et al. (2016):

Non-constant time intervals: Since the GPS points are not sampled continuously, but with changing frequency, this makes the data inherently discrete. Hence, each GPS data point is effectively a discrete event, not part of a continual stream of data. For the analysis, these time intervals are converted into a standardized unit (seconds). In a continuous model, we need to assume that the time to event follows some continuous distribution, which is a harder assumption to make with non-uniform sampling intervals. **Time Lag between decision making and recording:** The time it takes from when the decision to park until the time the button is pressed on the app introduces a discrepancy. This type of measurement error is more naturally handled in a discrete time framework, as it allows for certain intervals of uncertainty. **Duration dependence:** Using a discrete time approach facilitates the interpretation of so-called duration dependence of the "parking spot finding probability" within a competing risk model. This means, we can analyze the effect of the search time on probability of finding a Free, Paid or Illegal parking spot. This can be particularly important in the context of our research, where the probability of finding a parking spot may change over the search time. **Model flexibility:** Discrete time models are flexible and can handle time-varying covariates such as distance to destination more easily. **Model interpretability:** Discrete time models tend to be more straightforward and easier to interpret than their continuous counterparts. This can be important when communicating the findings to policy-makers or the general public. Especially the hazard rate (see below) is easier to interpret as conditional probability.

Beyond the reasons outlined previously, this distinction between discrete and continuous time should be of little relevance in the scope of our research since the discrete time logistic hazard rate model converges toward the proportional hazard model when the length of the time intervals approaches zero (Kim, 2014).

We start by describing the single event discrete survival analysis (Tutz, Schmid, et al., 2016). We have $t = 1, 2, \ldots, q$ corresponding to time intervals $(0, a_1), [a_1, a_2), \ldots, [a_{q-1}, a_q)$, with a_q indicating the upper limit of the final interval. T denotes discrete event time (here, parking search duration, when the event occurs) $T \in \{1, \ldots, q\}$, implying that if a failure occurs in the interval $[a_{t-1}, a_t)$ then $T = t$.

The hazard function in discrete time represents the conditional probability that the time period T (=search duration) ends at time t, given $T \geq t$ and the vector of explanatory variables $\mathbf{x}$:

$$\lambda(t \mid \mathbf{x}) = P(T = t \mid T \geq t, \mathbf{x}), \quad t = 1, 2, \ldots, q$$

Here $\lambda(t \mid \mathbf{x})$ is the conditional probability of finding a parking spot at time t, given that the driver has not found a parking spot until then and given $\mathbf{x}$. The values of $\mathbf{x}$ may vary over time.

The survival function gives the probability that the event occurs later than at time t, given $\mathbf{x}$:

$$S(t \mid \mathbf{x}) = P(T > t \mid \mathbf{x}) = \prod_{\tau=1}^{t} (1 - \lambda(\tau \mid \mathbf{x}))$$

It is linked to the cumulative density function $F(\cdot)$:

$$S(t \mid \mathbf{x}) = 1 - F(t \mid \mathbf{x}) = 1 - P(T \leq t \mid \mathbf{x}) = P(T > t \mid \mathbf{x})$$

If the hazard rate is assumed to depend through a logit link on the explanatory variables, we get the logistic discrete hazard rate model:

$$\lambda(t \mid \mathbf{x}) = \frac{\exp(\alpha_{0t} + \mathbf{x}'\boldsymbol{\beta})}{1 + \exp(\alpha_{0t} + \mathbf{x}'\boldsymbol{\beta})}$$

In the context of survival analysis with competing risks, the multinomial logit hazard model is an extension of the logistic discrete hazard rate model (Janitza & Tutz, 2015). This model is particularly useful when dealing with multiple types of events and when the values of the covariates can change over time. We denote the hazard function for experiencing the event $j \in \{1, \ldots, J\}$ at time t given a set of covariates $\mathbf{x}$ as $\lambda_j(t \mid \mathbf{x})$:

$$\lambda_j(t \mid \mathbf{x}) = \frac{\exp(\alpha_{0jt} + \mathbf{x}'\boldsymbol{\beta}_j)}{\sum_{k=1}^{J} \exp(\alpha_{0kt} + \mathbf{x}'\boldsymbol{\beta}_k)}$$

We can account for unobserved heterogeneity among drivers and thus capture the influence of unobserved individual-specific characteristics that are not included in the model as covariates by considering random effects. For example, some drivers may be more patient or more risk-averse than others, which could influence their parking search behavior. It also allows for correlations in the hazards of different parking search events of the same individual.

Here, u_{ij} is the random effect for individual i and event j. $\lambda_{ij}(t \mid \mathbf{x}_{ijt}, u_{ij})$ denotes the hazard function for driver i experiencing event j given the covariates $\mathbf{x}_{ijt}$:

$$\lambda_{ij}(t \mid \mathbf{x}_{ijt}, u_{ij}) = \frac{\exp(\alpha_{0jt} + \mathbf{x}'_{ijt}\boldsymbol{\beta}_j + u_{ij})}{\sum_{k=1}^{J} \exp(\alpha_{0kt} + \mathbf{x}'_{ijt}\boldsymbol{\beta}_k + u_{ik})}$$

The driver random effects u_{ij} are assumed to be normally distributed, with an expected value of zero and the possibility that the covariance of the random effects of the J events may be unequal to zero. The latter reflects the possibility that unobserved driver-specific variables affecting the parking search process with regard to Free parking spots may also affect the parking process with regard to Paid and Illegal options, and vice versa.

Within our framework of survival analysis of parking search duration, allowing for a time-varying baseline hazard can be a crucial aspect of the model. This is because the decision to park is not only influenced by the characteristics of the available parking spots (such as the walking distance to the final destination as well as Free versus Paid versus Illegal), but also by the elapsed search time. The assumption of a constant baseline hazard, which implies that the probability of finding a parking spot remains the same regardless of how long the search has been going on conditional on $\mathbf{x}$ and u, is likely to omit key features of parking search behavior. As the search duration increases, drivers may become more and more frustrated or impatient, which means that the perceived cost of continuing the search (in terms of time, effort, and stress) may increase. As a result, the drivers are more likely to settle for a less-than-ideal parking spot that they would have previously deemed unsatisfactory. This could include parking spots that are, for example, smaller, less safe, or more expensive. Consequently, the probability of a parking event (i.e., the driver finds a parking spot and decides to park) could increase over time, which would be captured by an increasing hazard function given the values of $\mathbf{x}$ and u.

An important aspect is how to model the duration dependence of the hazard rate, which is usually done by assuming a certain functional form of the baseline hazard. Incorporating the log of search duration into the baseline hazard allows for a flexible and potentially non-constant baseline hazard (Jenkins, 2005). This assumed functional form seems appropriate according to the mentioned expected behavior. This can capture complex temporal patterns in the data and improve the fit of the model. In this version of the competing risk model with random effects, the hazard function for individual i experiencing event j at time t given a set of covariates $\mathbf{x}_{ijt}$ and a random effect u_{ij} can be written as:

$$\lambda_{ij}(t \mid \mathbf{x}_{ijt}, u_{ij}) = \frac{\exp\left(\alpha_{0j} + \alpha_{1j}\log(t) + \mathbf{x}'_{ijt}\boldsymbol{\beta}_j + u_{ij}\right)}{\sum_{k=1}^{J} \exp\left(\alpha_{0k} + \alpha_{1k}\log(t) + \mathbf{x}'_{ijt}\boldsymbol{\beta}_k + u_{ik}\right)}$$

Here, α_{0j} and α_{1j} are the parameters of the baseline hazard for event j. The term $\alpha_{1j}\log(t)$ allows the baseline hazard to change over time in a natural logarithmic form. The parameters of the model, α_{0j}, α_{1j}, and $\boldsymbol{\beta}_j$ can be estimated from the data using maximum likelihood estimation (Lee et al., 2018). The likelihood function for the logistic discrete hazard rate model is derived from the probability of observing the event times given the covariates, and it can be maximized using standard numerical optimization techniques.

3.6 Results and Discussion

This section presents the findings from the application of the Discrete Hazard Rate Model to our dataset, focusing on the explanation of PSD by a variety of driver-related, journey-related, and spatio-temporal variables. One important variable is missing: the occupation rate of parking spots in the destination area (Axhausen et al., 1994; Millard-Ball et al., 2014). We try to cope with this problem by controlling for area type, time of the day and day of the week, which are highly correlated with the occupation rate (McCahill, 2017; Muleev, 2020).

The results of the model in Table 3.7 provide insights into how these factors influence the likelihood of finding a parking spot, categorized into Free, Paid, and Illegal parking. The sampling rate for logging GPS points in the dataset is one second. The hazard rate can be understood as the probability of finding a particular type of parking spot per second, assuming the driver is still searching.

The random effects model was compared to the pooled model (multinomial logit without random effects) using a likelihood-ratio test with the null hypothesis that the variance and covariance of the random effects are zero (pooled model). This LR-test is known to be conservative, meaning that the null hypothesis is rejected too rarely (StataCorp, 2023). With a p-value of 1.000, the test results indicated that the random effects were not statistically significant. Given this potential bias, we have decided to show the random effects model anyway. [6]

[6] The outcomes of both models are almost identical, with a minor divergence observed in the estimated effects for

In Table 3.8, we present the estimated standard deviations and correlations for the random effects associated with different types of parking—Free, Paid, and Illegal. The standard deviations capture the extent of the unobserved heterogeneity among drivers. For instance, the standard deviation for Free parking is the lowest, suggesting that individual differences are least pronounced in the likelihood of finding Free parking. The covariance between the random effects for each pair of parking types indicates whether the unobserved factors that make a driver more likely to find one type of parking also make them more or less likely to find another type. For example, a negative covariance between Free and Paid parking would suggest that drivers who are more likely to find Free parking are less likely to find Paid parking, and vice versa. However, the correlations are insignificant. Collecting more data could provide a more robust basis for investigating these relationships further and may yield statistically significant results.

The coefficients in the model represent the effect of the corresponding variable on the logarithm of the hazard rate. Therefore, a positive coefficient indicates that an increase in the variable leads to an increase in the hazard rate, which in turn implies a decrease in the expected search duration, and vice versa for a negative coefficient. It is important to note that these coefficients represent effects on a logarithmic scale due to the model's log-linear structure. This means that the effects of the variables are multiplicative with respect to the hazard rates. For instance, a coefficient of 0.2 for a variable in the model implies that, all else being equal, a one-unit increase in that variable multiplies the hazard rate by $e^0.2$, which is approximately 1.22, or a 22% increase.

To further enhance our understanding of the magnitude of the estimated effects, we calculated Average Marginal Effects (AMEs) on the hazard rates. These are additionally embedded in Table 3.7 with the unit "percentage points per second". AMEs provide a measure of the change in the predicted probability of an outcome due to a one-unit change in a variable, holding all other variables constant. This allows us to quantify the effect of each variable on the probability of finding each parking type (Free, Paid, and Illegal). AMEs are particularly valuable in the context of multinomial logit models, where the coefficients can be challenging to interpret directly.

To calculate an AME, we increase the variable by one unit for each observation in the dataset and compute the difference in the predicted probabilities of the outcomes. This difference is then averaged over all observations to obtain the AME for that variable. This process is repeated for all variables in the model. The interpretation of AMEs is straightforward. For example, an AME of 0.5 percentage points for a particular explanatory variable to explain the hazard to Free parking would imply that a one-unit increase in that variable, on average, increases the probability of finding a Free parking spot in the next second by 0.5 percentage points, assuming that all other factors remain constant.

The coefficients and the AMEs for log(t) are positive and significant for Free and Paid parking,

the baseline hazard $log(t)$. Specifically, the random effects model exhibits a marginally stronger positive duration dependence compared to its pooled multinomial logit counterpart. Detailed results of the pooled multinomial logit model can be made available upon request from the authors.

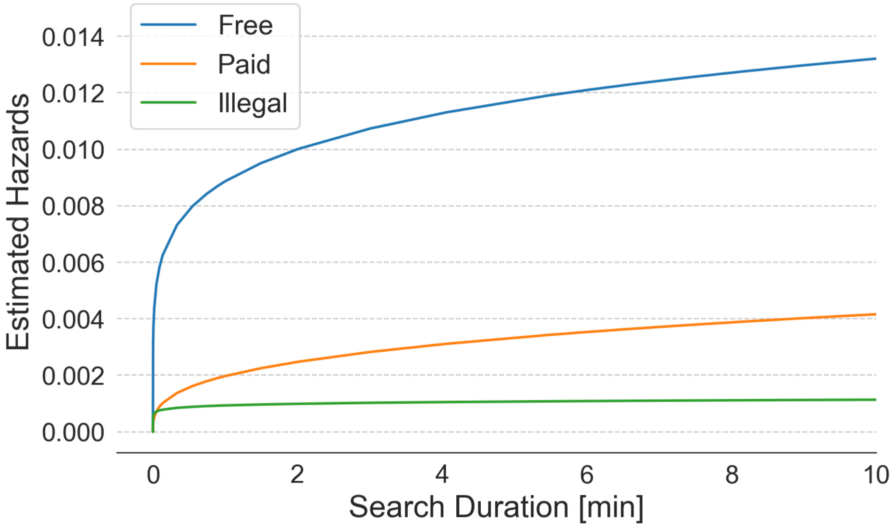


Figure 3.8: The estimated hazard functions for each type of parking (Paid, Free, Illegal) over time

but insignificant for Illegal parking. This means that as the search duration increases, the hazard rate of finding a Free or Paid parking spot also increases after controlling for all other explanatory variables. In other words, the longer a driver spends searching for a parking spot, the higher their instantaneous probability of finding one. This positive duration dependency of the hazard rate can also be seen in Figure 3.8, which displays the estimated hazard functions for each parking type.

The positive duration dependency effect is more pronounced for Paid parking. The higher positive duration dependency effect for Free parking is easier to comprehend when looking at the growth rate of the hazard, shown in Figure 3.9, revealing that with every additional second searching the probability of choosing a Paid parking spot increases faster than finding a Free spot. In summary, the longer the search duration, drivers become more flexible or willing to opt for Paid parking compared to Free parking.

Using the standard assumption in theoretical models that drivers seek to minimize the total cost of journey (van Ommeren et al., 2012), this result can be explained as follows: Drivers tend to look for Free parking initially. However, as search duration extends and the ideal Free parking spots remain elusive, drivers weigh the expected cost of continued searching (e.g., in terms of time, fuel, and frustration) against the cost of Paid parking. At a certain time, continuing the search for a Free parking results in a higher expected total cost of journey for drivers. This is when they switch to choose Paid parking.

Despite the growing flexibility towards Paid parking, drivers remain averse to Illegal parking as the search continues. While there are various strategies for parking illegally (Axhausen & Polak, 1991), one strategy appears to be particularly dominant in our data. Drivers seem to evaluate the potential costs of Illegal parking before they even start searching. Should the perceived cost

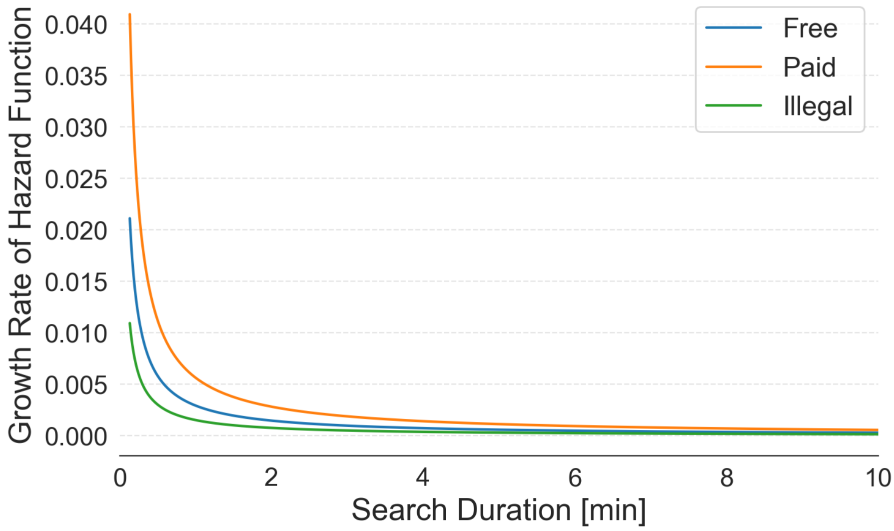


Figure 3.9: Growth rate per second of hazard function across different parking categories

of Illegal parking be lower than searching, they park close the destination and skip the search completely. This behavior is confirmed by the complete dataset as well, which shows that the median search duration recorded for Illegal parking stands only at approximately 39 seconds.

Following, we provide a detailed interpretation of the effect of each variable and discuss their implications for parking search duration.

- **Distance to Destination:**
  - For all types of parking, an increase in the distance to the destination is associated with a decrease in the likelihood of finding an acceptable parking spot. The effect is most pronounced for Illegal parking (AME of -0.16 percentage point), followed by Free parking (-0.57), and least for Paid parking (-0.09).
  - This suggests that the likelihood of finding a parking spot is higher when drivers are closer to their destination. Most likely, this can be explained by a greater willingness to accept closer parking options to avoid a long walking route to the destination.
  - The stronger effect for Illegal parking could indicate that drivers are more likely to resort to Illegal parking when they are closer to their destination, possibly due to convenience or time constraints. These positive aspects of the Illegal parking spot overweight the expected value of the possibility of a penalty for Illegal parking. The weaker effect for Paid parking suggests that as the distance to their destination grows and drivers face difficulties in finding parking, they not only expand their search radius but are also willing to pay for parking.
  - These findings underscore the importance of the drivers' relative location with regard to the final destination during the parking search in their parking choice. They suggest that strategies to guide drivers towards potential parking options closer to their destination, such as real-time parking information systems, could potentially reduce parking search durations (Dalla Chiara et al., 2022).

- **Age:**
  - In the Free and Paid parking category, the coefficients are nearly zero and not statistically significant, suggesting age does not have a strong effect on the likelihood of finding an acceptable parking spot.
  - As for Illegal parking, older drivers are less likely to park illegally with a 0.09 percentage points per second decrease in probability for every 10 years of age. This could be because older drivers are generally more law-abiding, more risk-averse, or less willing to take the risk of getting fined or towed. (In our analysis, "Age in 10 years" represents a transformation of the age variable, where each individual's age is divided by 10, thereby expressing age in terms of decades rather than individual years. This is an approach to avoid estimated effects with too many decimal places.)
- **Vehicle Type:**
  - The data reveals no significant disparities in parking probability among small, medium, and large cars across all parking types. However, a notable exception is observed with vans, which exhibit a marked increase in the likelihood of Illegal parking—specifically, a 0.39 percentage point per second increase compared to the baseline. This heightened tendency for Illegal parking could be attributed to the larger dimensions of vans, which make parking more challenging. Additionally, vans are often used for tasks that involve substantial loading and unloading. This may incentivize drivers to minimize walking distances, leading them to opt for Illegal parking closer to their destinations.
- **Journey Purpose:**
  - The likelihood of opting for Paid parking is highest for business-related trips, showing an increase of 0.32 percentage points per second compared to base level journeys to home. This followed by shopping (0.17 percentage points), leisure activities (0.17), and commuting (0.09). For business-related journeys, this trend may arise as drivers are more open to paying for parking during such trips, as the expense can be considered a business cost, possibly reimbursed by their employer or client. Shopping and leisure activities often occur in high-demand areas where Paid parking is more prevalent, justifying the increased likelihood in these categories. This increased tendency towards Paid parking could be due to shopping areas typically having more available Paid parking options. Generally, the model indicates that drivers looking for parking during any type of journey are more likely to find a Paid parking spot compared to when they are returning home. The lowest Paid parking probability is associated with journeys to home which is reasonable as many people have access to Free parking near their homes.
  - Conversely, the probability of finding Free parking is lower for business-related and shopping trips compared to home journeys, showing a decrease of 0.31 and 0.13 percentage points per second compared to the base category "home". The urgency or time constraints associated with business-related activities may make drivers less willing to spend time searching for Free parking. For shopping trips, the high-demand

nature of commercial areas may limit the availability of Free parking.

  - As for Illegal parking, the data shows that the likelihood is significantly lower for commuters and shopping trips, with a decrease of 0.16 and 0.10 percentage points per second, respectively, compared to home journeys.

- **Area Familiarity:**
  - Drivers familiar with an area show a significantly increased likelihood of finding a Free parking spot by 0.14 percentage points per second. This can be attributed to the knowledge these drivers have about the area, including understanding the timing and location of available Free parking spots.
  - There is no significant change in the likelihood of Paid and Illegal parking. One possible explanation for this could be that familiarity with an area allows drivers to optimize their search for Free parking, reducing the need to consider Paid or Illegal options.
- **Parking Duration:**
  - For those planning to park their vehicles for more than half an hour, the chance of locating a Free parking space significantly drops by 0.18 percentage points per second, while the likelihood of utilizing a Paid parking space rises. Although it may initially seem unexpected – considering drivers parking for a longer duration could save more if they found a Free spot – this trend might be because short-term parking spots are often Free, while long-term parking typically requires payment. Additional factors contributing to this pattern include the economic benefits of hourly rates in Paid parking for longer durations and the enhanced security features commonly found in Paid facilities.
  - Additionally, when planning for an extended parking period, individuals are less inclined to park illegally by a significant 0.26 percentage points per second since the risk of getting a ticket or being towed increases with parking duration.
- **Area Type:**
  - As urbanization levels rise, the likelihood of finding a Free parking spot declines, while the chances of finding a Paid parking option increase. Often because urban areas are more likely to implement Paid parking systems to manage high demand. Specifically, in small cities, the probability of finding a Free parking spot increases by 0.31 percentage points per second compared to the base category "medium-sized cities". Conversely, in the city centers of large cities, this probability decreases by 0.38 percentage points per second. This trend suggests that drivers are increasingly willing to opt for Paid parking as urban density grows. The scarcity of Free parking in densely populated areas, coupled with the costs of extended search times and fuel consumption, makes Paid parking a more economically rational choice. Additionally, Paid parking facilities are often more conveniently located near popular destinations, minimizing walking duration.
  - Regarding Illegal parking, the data does not reveal a uniform pattern across different types of areas. However, it is notable that the tendency for Illegal parking is higher

in the inner urban sectors and centers of large cities. In these regions, drivers are frequently engaged in time-sensitive activities, making them more inclined to risk Illegal parking for the sake of convenience. The perceived benefits of a prime location in these areas may outweigh the potential penalties, rendering Illegal parking a more attractive option.

- **Temperature:**
  - There appears to be a rise in the probability of locating a Free and Paid parking spots with an increase in temperature. More specifically, the hazard increases by 0.07 percentage point for both Free and Paid parking for every 10 degrees increase in temperature. This may be due to the fact that in pleasant weather, drivers might be more willing to accept a less optimal parking spot, for example, with regard to the walking distance. On the other hand, in adverse weather conditions, drivers might prefer to find a better parking space, such as one closer to their destination, to minimize their walking distance.
  - The effect on Illegal parking is not statistically significant.

Table 3.7: Coefficient and Average Marginal Effect (AME in percentage points) Estimation Results – Discrete Competing Risk Model using Multinomial Logit Hazard with Random Effects and Logarithmic Time-Varying Baseline Hazard

Variable	**Free**		**Paid**		**Illegal**	
	coeff.	**AME**	**coeff.**	**AME**	**coeff.**	**AME**
ln(Search Duration)	0.17754*** (0.01830)	0.15471	0.34080*** (0.06208)	0.07131	0.09245 (0.07693)	0.00793
Distance to Destination						
(in 100 m unit)	-0.65903*** (0.02926)	-0.57418	-0.42651*** (0.07288)	-0.08793	-1.75975*** (0.22350)	-0.15597
Age						
(in 10 years unit)	-0.01647 (0.03186)	-0.01297	-0.07524 (0.12852)	-0.01554	-1.00942*** (0.18610)	-0.08998
Gender						
Male	0.09820 (0.12697)	0.08516	-1.08695*** (0.38134)	-0.33118	0.83436 (0.60837)	0.05627
Vehicle Type						
Ref: Compact and Small						
Medium and Large	0.04999 (0.12538)	0.04465	0.09633 (0.43535)	0.02092	0.35963 (0.67167)	0.03696
Van	-1.09858*** (0.32914)	-0.59901	-1.20639 (1.47950)	-0.15074	1.77373** (0.83489)	0.39329
Journey Purpose						
Ref: Home						
Shopping	-0.12949* (0.07516)	-0.11198	1.10258*** (0.27256)	0.17152	-0.92122** (0.45305)	-0.1034
Leisure	0.03624 (0.06699)	0.03313	1.10921*** (0.27472)	0.17266	-0.48492 (0.42813)	-0.06564
Commuting	-0.01287 (0.08341)	-0.01023	0.75062** (0.30570)	0.09645	-2.39645** (1.04700)	-0.15847
Business-related	-0.30514*** (0.11000)	-0.24468	1.57570*** (0.32756)	0.32021	-0.33391 (0.49442)	-0.04802
Area Familiarity						
Known	0.17127*** (0.05471)	0.14469	-0.00841 (0.17684)	-0.00218	0.26886 (0.19377)	0.02348
Parking Duration						
Ref: Shorter than 30 min						
Longer than 30 min	-0.19990*** (0.05956)	-0.17976	0.20150 (0.18116)	0.04204	-2.49143*** (0.33938)	-0.26137
Area Type						
Village or Small City	0.25714** (0.12294)	0.30929	-0.66928 (0.52208)	0.06485	0.27795 (0.60648)	0.04478
Ref: Medium City						
Large City: Suburban	0.04104 (0.10136)	0.04552	-0.52184 (0.35159)	-0.00412	-0.22995 (0.48709)	0.07535
Large City: Inner Urban	-0.13783 (0.10333)	-0.13739	-0.02074 (0.34689)	-0.08486	0.84737* (0.45900)	-0.01178
Large City: Center	-0.44680*** (0.10115)	-0.38402	0.27227 (0.32226)	-0.10251	0.57695 (0.46395)	0.018
Temperature	0.00826*** (0.00291)	0.00717	0.03371*** (0.01053)	0.00708	-0.00339 (0.01185)	-0.000322
_cons	-6.07541*** (0.84526)		-17.16433*** (3.07735)		-2.15961 (3.51694)	
Mean Hazard Rate per Second (percent)	0.79		0.08		0.05	
Number of Transitions	2,419		249		146	
Number of Drivers	127					
Number of Observations (seconds)	306,214					

- Time of the day and day of the week are embedded in the model as control variables in dummies format to compensate for the missing occupancy rate effect. This includes six time-of-day dummies, one representing night hours between 22:00 and 07:00, and the remaining five covering three-hour intervals throughout the day. However, they are not shown in this table.
- The pooled logit model provides similar results and is available upon request from authors.
- Standard errors in parentheses
- *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 3.8: Standard Deviation and Correlation of Random Effects for Different Parking Types

Random Effect	coeff.	SE	z	p > $\lvert z \rvert$	[95% conf.	interval]
sd (Free Parking)	0.4575883	0.0754974			0.3311581	0.6322872
sd (Paid Parking)	1.563323	0.3468271			1.012063	2.414849
sd (Illegal Parking)	1.193339	0.4405616			0.5787806	2.460446
corr (Free, Paid)	-0.0877948	0.2308639	-0.38	0.704	-0.4960258	0.3522221
corr (Free, Illegal)	0.1169427	0.3877903	0.3	0.763	-0.573762	0.7104409
corr (Paid, Illegal)	0.0418713	0.4764293	0.09	0.93	-0.7131322	0.7519035

3.7 Conclusion

In this paper, we proposed an innovative approach to collect unbiased ground truth GPS data of parking search behavior by recording the exact parking search starting point for the first time besides the parking spot chosen and the final destination after walking. This method is implemented through a mobile application that records the duration and path of the parking search. Over the span of 2021 to 2023, more than 3000 trips were recorded by volunteer drivers in Germany. The data is subsequently employed in a survival analysis model to explore the factors influencing cruising for parking. Our novel methodology employs a competing-risks model to analyze the search durations associated with various parking types: Free, Paid, and Illegal. Another distinctive feature of our methodology is the assumption of a time-varying baseline hazard, allowing for a duration dependency analysis, i.e., the effect of the previous search duration – given all other explanatory variables – on the probability of finding a Free, Paid, or Illegal parking spot can be evaluated.

The application is designed to capture GPS trajectories from the beginning until the end of a journey. Its user interface is deliberately straightforward, featuring a central button on the main screen. Drivers initiate the recording by pressing this button. Upon starting their search for an available parking space, they press the button a second time, marking the exact time and location where the search began. Once parking is found, another press of the button logs the parking spot's time and coordinates. The final button press is made when drivers reach their final destination on foot, marking the walking route from the parking location to the final destination.

Our dataset reveals that the Mean Parking Search Duration (MPSD) across all journeys stands at 1 minute and 29 seconds. Around 18% of these journeys exhibit a Parking Search Duration (PSD) of zero, suggesting immediate parking without any search effort. For journeys with a non-zero PSD, the MPSD is observed to be 1 minute and 49 seconds. The area has a significant effect on MPSD. For instance, in specific regions in the center of the city of Frankfurt, MPSD can rise up to more than 5 minutes. The average walking time is 2 minutes and 40 seconds for all journeys and 2 minutes and 52 seconds for those with a PSD exceeding zero. The data also shows that the average initial search radius (the straight line from the start of the search to the destination) is 140 m. In addition, the average accepted walking distance (from the start of the search to the destination) is roughly 187 m. It is also worth noting that approximately 5% of the recorded journeys end in Illegal parking.

The values presented in our study are significantly lower than those derived from survey-based methods. However, they align closely with findings from prior research that employed GPS data to measure Parking Search Duration (PSD). This observation confirms the reliability of GPS data as a valuable resource for investigating parking search behavior, reinforcing its applicability and relevance in this study area. The discrepancy with survey results might arise from the inherent negativity bias in survey responses. Nonetheless, it is noteworthy that compared to previous studies that used GPS data, our results fall on the lower end of the spectrum. This can be attributed to our comprehensive data collection strategy, which captures all geographic regions, times, and days without limitations. This contrasts with studies that might target specific high-congestion zones or peak times.

To dive deeper into the factors influencing parking search duration, we employed a survival analysis model. Specifically, we used a discrete competing-risks model with a Multinomial Logit Hazard with Random Effects. This model was chosen to study the search durations associated with different parking categories: Free, Paid, and Illegal. One of its strengths is its ability to account for the interrelation between decision-making and the search process. For instance, a driver might initially seek a particular parking type but shift preferences during the search. Additionally, including a time-varying logarithmic baseline hazard enhances the model's flexibility, adjusting the duration-dependent nature of the hazard.

Our findings reveal a positive duration dependency for Free and Paid parking. This trend is intuitive: as drivers extend their search without success, they tend to relax their criteria, becoming more willing to less ideal parking options, whether due to safety concerns, size limitations, or cost. This increasing hazard rate is particularly pronounced for Paid parking, suggesting that as search durations extend, drivers become more flexible toward Paid parking and are willing to pay the parking cost to minimize the total journey cost.

Demographic and situational factors also play a role. Elderly individuals are less inclined to shorten their search times by using Illegal parking. Conversely, for short anticipated parking durations (under 30 minutes), drivers appear more willing to skip the search by parking illegally. Similarly, van drivers are more likely to park illegally. In addition, parking probabilities varies significantly across journey purposes. Business-related trips predominantly lead to Paid parking choices, succeeded by shopping, leisure, and commuting, with journeys to home showing the least inclination towards Paid parking. Expectedly, the search duration increases in larger cities, descending through medium to smaller cities. With rising degree of urbanization, drivers increasingly favor Paid parking and face reduced chances of securing Free spots. Finally, familiarity with a destination area also proves advantageous, typically leading to quicker Free parking spot identification.

Our results suggest that in line with previous research (Shoup, 2005), the coexistence of Free and Paid parking spaces in cities may increase parking search as drivers tend to search for a

Free parking spot first. This may increase the total search duration.

One notable limitation of our study is the representativeness of our sample. While the dataset offers valuable insights into parking search behavior, it is primarily drawn from a specific group of app users and may not fully capture the diverse parking behaviors of the broader population. Additionally, the awareness that they were being tracked might have altered the behavior of participants, a common issue in studies involving human observation (Hawthorne effect), potentially leading to variations in parking patterns.

Another limitation relates to the accuracy and precision in measuring the starting point of the parking search. Due to various factors, such as forgetfulness or delayed interaction with the app by users, it is challenging to pinpoint the exact moment when the parking search begins. These factors contribute to potential measurement errors, which are inherent in studies reliant on user-operated devices and data collection methods.

Despite these limitations, the unique dataset, containing ground truth data on parking search durations, presents various intriguing possibilities. For example, in future works, we aim to develop a prediction model capable of identifying parking search patterns within historical GPS data. Such a model could be integrated with extensive historical Floating Car Data (FCD) datasets, facilitating deeper insights into parking search behaviors across various regions or journey clusters.

4 Article 3: Parking search identification in vehicle GPS traces

Authors	Siavash Saki, Tobias Hagen
Affiliation	Research Lab for Urban Transport, Frankfurt University of Applied Sciences, Frankfurt am Main, Germany
Journal	Transportation Research Part A: Policy and Practice
Volume	183
Article Number	104045
DOI	10.1016/j.tra.2024.104045
Received Date	15 September 2023
Revised Date	24 January 2024
Accepted Date	20 March 2024
Available Online	8 April 2024
License	Open Access, CC BY 4.0

Abstract

The challenge of "cruising for parking" in urban areas has long been a subject of study, but existing research often relies on biased surveys or arbitrary assumptions in the absence of ground truth data. This paper addresses these gaps by introducing the first-ever collection of ground truth data on parking search durations gathered through a self-developed app. The dataset encompasses more than 3,500 journeys collected in Germany, with approximately two-thirds of them ending in Frankfurt am Main. Utilizing this unique dataset, we developed a deep learning neural network model that accurately identifies parking search routes in GPS data and predicts search duration. Our model outperforms existing parking search identification models proposed in previous studies. The model's efficacy is further evaluated on an independent park-and-visit dataset and then applied to a large-scale dataset from Frankfurt/Germany. This generates the first reliable statistics on parking search durations and reveals key insights about parking search patterns in this city. Notably, the predicted mean parking search duration from this extensive dataset, comprising over 860,000 journeys, is approximately 1.5 minutes. This work not only advances the field by providing a new data collection methodology and a superior predictive model but also offers a reusable framework that can be applied to other cities and datasets for broader urban mobility insights.

Keywords: Parking Search, Prediction Model, Neural Network, GPS data, Traffic Management

4.1 Introduction

The rapid urbanization of cities has led to various challenges in transportation and mobility, one of which is the increasing difficulty in finding parking spaces. As cities grow denser and vehicle ownership rates rise, the competition for parking has intensified, leading to what is commonly referred to as "cruising for parking". This phenomenon not only contributes to traffic congestion (Shoup, 2005) but also increases energy consumption and emissions, thereby exacerbating environmental concerns (Brooke et al., 2014).

To analyze parking search behavior, researchers have used various approaches and tools. Several studies have relied on surveys (Assemi et al., 2020; Cao et al., 2019; Brooke et al., 2018; Belloche, 2015) to explore parking search duration and behavior. However, these survey-based approaches are generally not accurate, not representative, and may even be systematically biased (Alemi et al., 2018). Alternatively, other studies have employed analytical models to estimate parking search duration (van Ommeren et al., 2021; Fulman & Benenson, 2021) and have performed simulations (Waraich & Axhausen, 2012; Benenson et al., 2008; Horni et al., 2013; Fulman et al., 2020) to mainly investigate the effects of different parking policies.

More recently, researchers have turned to GPS data as a potentially more robust source for analyzing mobility patterns, including parking search behavior (Mannini et al., 2017; Dalla Chiara et al., 2020; Mantouka et al., 2021; Milia et al., 2023; Bisante et al., 2023). In this framework, a well-designed parking search prediction model can offer numerous benefits, such as identifying parking search routes in comprehensive historical GPS data. This, in turn, allows the generation of statistics about parking search, which may - depending on the historical GPS data source used - even be representative for an underlying population. Moreover, this enables the researcher to explore the determinants of cruising and to analyze parking search in various settings.

Despite the availability of GPS data, existing models for identifying search routes in GPS trajectories are far from being perfect. These models often rely on arbitrary assumptions, which are largely due to the lack of access to ground truth data on parking search, making these models less reliable. This absence of accurate ground truth data hinders the development of effective predictive models and policies aimed at mitigating the parking search problem. Furthermore, the discrepancies in existing aggregated numbers from various studies make it difficult to trust the current state of knowledge in this area.

To address these challenges, we developed a specialized app for collecting ground truth data, marking the first time such data has been systematically gathered. Since the data collection was based on volunteer test drivers, this ground truth data is anything but representative. For this reason, based on this training data, we first develop a predictive model labeling parking search in GPS data of cars. Second, we apply this model to large-scale historical GPS data to overcome the problem of missing representativeness and to generate robust statistics and insights, particularly focusing on the city of Frankfurt/Germany.

The major contributions of this paper are as follows:

- Developing, training, and validating a deep learning neural network model for predicting parking search based on parking search ground truth data. This model is open-source and is published on GitHub [github.com/ReLUT/parking-search-prediction].
- Comparing the model's performance with applied heuristic rules in the literature.
- Applying the model to a large-scale historical GPS data set to generate robust aggregated statistics for the city of Frankfurt.
- Demonstration of the model's reusability and applicability to other data sources.

Large GPS datasets, which include thousands of journeys, offer a rich array of data that can reflect the varied and complex nature of parking search behavior across various cityscapes, times, and driver demographics. Thus, we aim to ensure that the insights obtained are not confined to particular use cases or demographics but are instead indicative of the wider population and numerous urban settings. Understanding parking search behavior is crucial to unraveling the concealed dynamics that drivers maneuver through in their pursuit of a parking spot (Dalla Chiara et al., 2021). Examining this behavior through our model allows us to:

- Identify Patterns: Recognize recurring patterns or trends in parking search behaviors, such as peak search times, preferred search areas, and common search routes (Liu & Geroliminis, 2016).
- Understand Challenges: Gain insights into the challenges drivers face during parking searches, such as prolonged search times or indirect search routes (Dalla Chiara et al., 2022).
- Explore Variabilities: Investigate how parking search behaviors vary across different drivers, times, and urban environments, providing a holistic view of the diverse challenges and strategies employed by drivers (Polak & Axhausen, 1990).

The insights obtained from our model, when applied to a large-scale dataset, can serve as a robust foundation upon which impactful policies and urban planning strategies can be formulated. Understanding parking search behavior better allows us to:

- Optimize Urban Infrastructure: Identify areas where parking infrastructure may be lacking or underutilized, enabling targeted enhancements to parking facilities and urban infrastructure (Cao et al., 2019).
- Inform Policy Decisions: Develop policies that align with the actual needs and challenges faced by drivers during parking searches, such as dynamic pricing, reserved parking zones, or incentivized parking (Shoup, 2021).
- Enhance Mobility Solutions: Innovate and implement mobility solutions that can alleviate parking search challenges, such as smart parking systems, integrated mobility platforms, or alternative transport solutions (Fahim et al., 2021).
- Reduce Congestion and Emissions: By optimizing parking search strategies and infrastructure, reduce the time spent by drivers in searching for parking, thereby mitigating traffic congestion and minimizing vehicular emissions (Millard-Ball et al., 2020).

The rest of the paper is organized as follows: Section 2 provides a literature review, providing a comprehensive overview of existing methodologies to estimate parking search duration using GPS data, highlighting the strengths and limitations of each. Section 3 presents our training data, elaborating on the data collection app and its associated descriptive analysis. In Section 4, we articulate the design of our model and the features it incorporates. Section 5 starts with an assessment of the model's performance, followed by a detailed explanation of the dynamic park-and-visit dataset used for validation. This section finishes with a comparison of existing (heuristic) methods to identify parking search in GPS data and our advanced machine learning approach. Section 6 presents the extensive INRIX dataset, the application of our model to this dataset, and the key insights derived. Finally, Section 7 concludes the paper, summarizing the main findings and pointing towards potential directions for future research.

4.2 Literature Review

The utilization of GPS data in studying parking search behavior has been the subject of numerous studies in recent years. The growing interest in this area is largely attributed to the need to address the challenges posed by urban parking search, particularly in densely populated cities as well as to the increasing availability of GPS data. Several researchers have explored the application of GPS data for identifying and analyzing search routes in the context of urban parking and developed different approaches to estimate cruising time.

A common approach in existing studies involves employing heuristic methods and making arbitrary assumptions to identify parking search routes from GPS data. One straightforward approach is to designate a radius around the destination of the journey, enclosing the parking search area. As the destination of the journey is usually not observable in GPS car data the parking spot finally found is used as the center of the circular search area. Previous studies indicate that drivers typically search for parking within a range of 200m to 800m around their destination (Martens et al., 2010; Leclercq et al., 2017; Khaliq et al., 2018; Weinberger et al., 2020). Bisante et al. (2023), for instance, assume that parking search commences the moment drivers enter a 200m radius around the destination, distinguishing the remainder of the route as normal driving.

Another simplistic approach is to use a speed threshold to find the onset of parking search. van der Waerden et al. (2015) defines parking search based on vehicle speed, positing that a vehicle is engaged in parking search when its speed falls below a certain threshold. In this method, the start of the search is marked when the mean of speed over five GPS points is below 23 km/h and the standard deviation is below 5km/h. In a similar approach, Milia et al. (2023) apply speed thresholds from earlier studies to detect cruising. Taking a novel approach, Hampshire et al. (2016) enhances this technique by combining GPS speed data with video analysis of drivers to pinpoint when parking search begins. In a different heuristic approach, Mantouka et al. (2021) assumes that the start of parking search occurs at the first local minimum of the distance to the parking spot within a 400m radius around it. This method is based on the assumption that as a driver gets closer to a parking spot, the distance should continuously decrease, and any initial increase in this distance indicates the commencement of the parking search.

In the exploration of methodologies for identifying parking search in GPS data, it is crucial to consider a variety of approaches due to the multifaceted nature of the problem. While most methods aim to pinpoint the exact moment parking search begins within a GPS trajectory, a further commonly adopted method computes cruising time, defined as the excess travel time due to parking search. This involves first determining a radius (e.g., 400m) around the parking spot that encloses the search. Subsequently, the shortest path from the entry point within this radius to the parking spot is calculated. The difference between the actual driven path and the shortest path is defined as cruising (Montini et al., 2012; Mannini et al., 2017; Weinberger et al., 2020; Dalla Chiara et al., 2020, 2022).

Instead of identifying the starting point of the parking search, this method estimates the impact of parking search on overall driving duration. This paper takes an inclusive approach by considering this method as well. Although the results of this approach are not directly comparable to the results of other approaches, the inclusion of this method generates a more comprehensive summary of parking search identification methods. Moreover, it enriches our understanding of cruising regarding the estimated excess travel times in a holistic framework containing the exact search durations.

Finally, a more reliable and accurate approach is using a machine learning model to predict the parking search phase within a GPS trajectory. Bisante et al. (2023) and Jones et al. (2017) have employed similar methods to identify cruising for real-time applications, using smartphone sensor data to train their models. However, neither of these studies has access to ground truth data to train their models. Consequently, they applied heuristic methods to identify parking search and used that as training data for their machine learning models. This, in turn, results in unreliable and invalidated models that act arbitrarily in parking search prediction. A summary of these approaches can be seen in Table 4.1.

Table 4.1: Approaches to Identify Parking Search in GPS Data

Approach	Studies	Category	Pros	Cons
Approach 1: Naive 200m Radius	Bisante et al. (2023)	Heuristic	• Simple and easy to implement. • Needs only GPS locations data	• May not accurately represent parking search behavior in all scenarios. • Does not consider variations in driver behavior or different urban layouts.
Approach 2: Speed Threshold	van der Waerden et al. (2015) Milia et al. (2023)	Heuristic	• Utilizes available GPS data. • Relatively straightforward to implement.	• May misidentify other slow-driving scenarios (e.g., traffic, stoplights) as parking search. • Does not consider variations in driving speed in different areas.
Approach 3: First Local Minima	Mantouka et al. (2021)	Heuristic	• Attempts to identify a behavioral change (searching for parking) based on distance. • May work well in certain urban layouts.	• May be sensitive to GPS noise or inaccuracies. • Might misidentify other behaviors (e.g., missing a turn) as parking search.
Approach 4: Actual-Shortest Path	Montini et al. (2012) Mannini et al. (2017) Weinberger et al. (2020) Dalla Chiara et al. (2020) Dalla Chiara et al. (2021) Dalla Chiara et al. (2022)	Model-Based /Heuristic	• Provides a quantitative measure of excess search time. • Can be applied uniformly across various scenarios.	• Assumes all excess time is spent searching for parking, which may not always be true. • Does not account for different reasons for taking a longer route (e.g., traffic, road closures).
Approach 5: Machine Learning Model	Bisante et al. (2023) Jones et al. (2017) Current Study	Data-Driven	• Can learn complex patterns and behaviors from data. • Potentially more accurate if trained and validated on robust, representative data.	• Requires sufficient, high-quality, labeled training data. • Model performance depends on feature selection, model type, and hyperparameter tuning.

The main limitation in all existing studies is the absence of ground truth data on parking search behavior, which significantly hinders the development and validation of methods to detect parking search in existing (unlabeled) GPS car data. Consequently, this leads to arbitrary assumptions, such as defining a specific radius around a parking spot or considering a vehicle to be searching for parking based on its speed, which cannot be validated against actual parking search data. Using these simplistic assumptions may not accurately reflect the complexity and variability of real-world parking search behavior.

4.3 Data

4.3.1 Data Collection App

The necessity for valid empirical measurement of parking search duration (PSD) in real-world journeys promoted the development of a specialized mobile application designed to accurately record the exact times of initiating parking search. The app, characterized by its user-friendly design, records the entire journey, from vehicle start to reaching the final destination on foot, through a four-step process involving the pressing of a single button at crucial journey phases:

1. Starting point of the journey, 2. Starting point of the parking search, 3. Parking spot, and 4. Final Destination.

This innovative data collection method divides every journey into three distinct phases: ”normal driving,” “parking search”, and “walking to the final destination”, providing a first-of-its-kind opportunity to collect valid parking search data from actual journeys. Additionally, the app gathers optional demographic and journey-related data from users, offering insights into various driver and journey variables without mandating user participation in data provision. Two example recorded journey, collected via this app, are illustrated in Figure 4.1.

The app was designed with a large, centrally located button on the main screen to facilitate ease of use and minimize distraction while driving. Drivers were instructed to optimally position their smartphones using a phone mount, ideally situated near the radio or digital media stereo device of the car. This setup ensures that recording the parking search starting point requires just a single tap on the screen, a user interaction comparable to common actions in navigation apps. This design consideration was crucial to ensure that the data collection process did not compromise the safety of the drivers or other road users. Functionality of the app is explored more in depth in Saki and Hagen (2024a).

Building upon the foundation of utilizing mobile applications for GPS data collection, several researchers have previously used similar approaches. Jones et al. (2017) developed ParkUs 2.0, an app that collects GPS trajectories and relevant smartphone sensor data. While it does record the start and end points of journeys, it assumes that the parking search begins as soon as the driver reaches their destination, which may not always mirror actual driver behavior. Additionally, ParkUs 2.0 offers parking information to assist drivers in finding parking spots. While this is useful for the user, it introduces a potential bias into the recorded parking search instances by possibly altering natural search patterns.

Similarly, Bisante et al. (2023) introduced an app that shares more similarities with our approach, recording users’ journeys from the start of driving to the point of finding a parking spot, but it does not record the walking path to the final destination. This app also allows users to manually input their parking search starting point by pressing a button. However, this feature was rarely used, as seen in only 19 records. Therefore, the researchers defined the ground truth as the moment drivers entered a 200m radius of the parking spot, a decision that may not accurately represent varied parking search behaviors.

In contrast, our application carefully records the entire journey without influencing the driver’s natural parking search behavior by providing parking information. It captures the details of the journey from the moment driving begins, through the parking search phase, and concludes with the walking path to the final destination. This approach offers a more complete and unbiased data collection method, ensuring the reliability and authenticity of the parking search data gathered. This distinction is vital in providing a more accurate and comprehensive dataset for subsequent analyses, which will be explored in the following chapters.

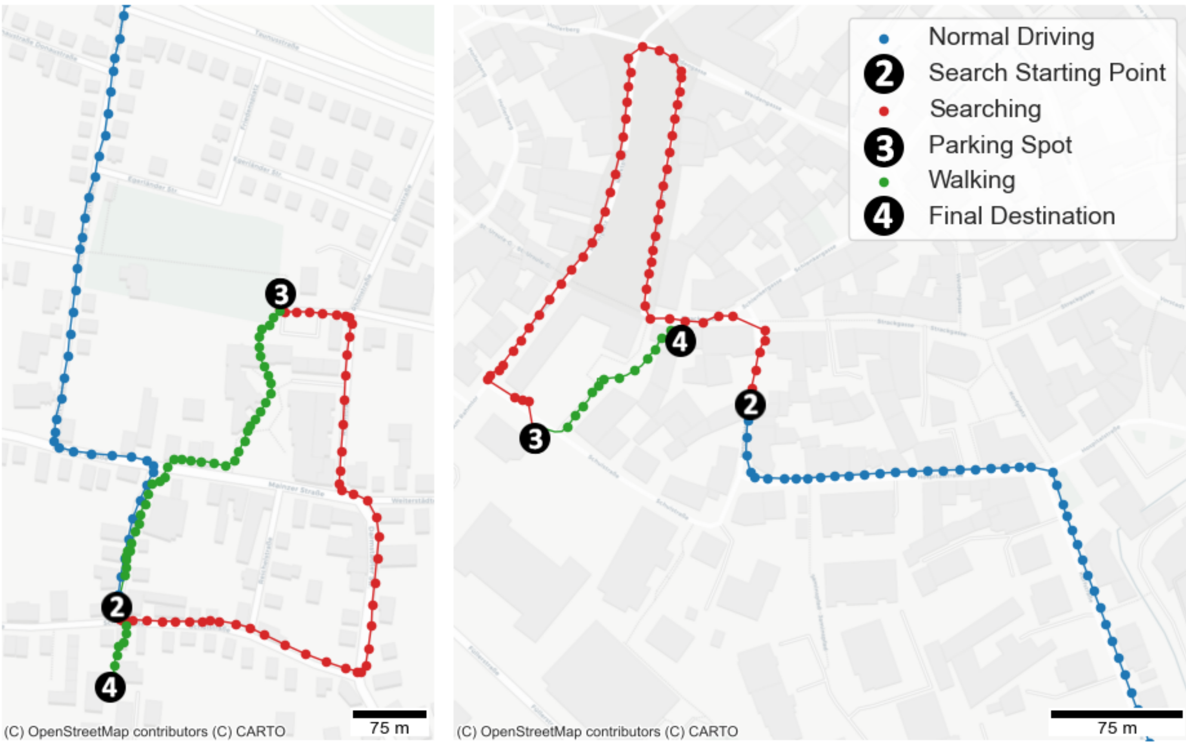


Figure 4.1: Real-world example of two actual journeys recorded via our app - Driving until search begins (blue), parking search route (red), and walking to the final destination (green)

4.3.2 Data Bias and Behavioral Influences in Data Collection

Addressing the challenge of biased and erroneous data is a critical aspect of our study, given the potential for inaccuracies coming from user errors, forgetfulness, or systematic issues within the app's usage. Recognizing the diversity of "junk" data, from fleeting, erroneous trips lasting only a few seconds to systematic errors like double-clicking the parking search initiation button, we undertook a rigorous data cleaning process. These types of errors, often resulting from initial user interactions aimed at understanding the app's functionality or demonstrating it to others, required deep examination and filtering to ensure the integrity of our dataset.

To mitigate these issues, we established strict criteria for considering a journey as valid data. Specifically, a journey needed to have a minimum duration of six minutes and cover an origin-destination distance of at least 1.5 km. Additionally, we imposed limits on the ratio of parking search to journey duration, as well as walking to journey duration, to identify and eliminate implausible records. This approach enabled us to remove a substantial portion of inaccurate data. Journeys that hovered around these thresholds underwent further optical investigation to ascertain their validity.

Another facet of biased data is related to user behavior, particularly forgetfulness in activating or deactivating the parking search mode within the app. For example, drivers may forget to hit the button to indicate parking search and do so when they have parked the car. To address this, we compared the actual parking search durations and estimated durations derived from our machine learning model (and other heuristic methods) for all journeys. Where discrepancies between the actual and estimated values were substantial, we conducted detailed reviews of the journeys. This allowed us to identify and exclude cases where patterns of parking search were evident, but the user likely forgot to interact with the app correctly. However, we acknowledge that for journeys with shorter search durations, accurately identifying such errors becomes challenging, leaving room for some residual bias.

Despite these challenges, our comprehensive data cleaning and validation efforts have been pivotal in refining the dataset. By applying both statistical descriptions and optical investigations, we think that we have curated a reliable set of journeys for our analysis.

Furthermore, we also need to acknowledge an intrinsic source of bias, beyond our direct control or revised measures. This bias comes from the potential alterations in driver behavior as a consequence of their awareness that their actions are being monitored by our app. Such awareness could inadvertently lead to modifications in their natural parking search strategies— for instance, drivers might consciously avoid parking illegally or engage in behavior they perceive as more socially acceptable or in alignment with the study's perceived objectives. This phenomenon, often referred to as the Hawthorne effect, suggests that individuals may change their behavior simply because they know they are being observed (McCarney et al., 2007). This effect represents a fundamental challenge in research that relies on active participant monitoring, introducing a layer of complexity to interpreting our findings. While our study design and data cleaning efforts aim to mitigate biases wherever possible, the potential for such behavioral modifications underscores the need for cautious interpretation of the data and findings, acknowledging this as an inherent limitation of our research methodology.

4.3.3 Descriptive Analysis of the Ground Truth (Training) Data

Available in English and German on both Google Play Store and Apple App Store since August 2021 until September 2023, the app has collected a substantial dataset comprising over 7,000 initiated journeys. After a rigorous process of cleaning the data and filtering out erroneous and biased records, a total of 3550 journeys conducted by 162 drivers in Germany, have been identified as valid trips, which equates to an average of approximately 22 valid journeys per driver. Delving deeper into how these journeys spread across our cohort of drivers, the engagement levels painted a varied picture. Specifically, the distribution's 5th percentile, median, and 95th percentile for journeys per driver stand at 1, 3, and 88, respectively, indicating a heavy concentration of activity among a relatively small group of participants. Furthermore, a significant portion of these journeys, amounting to 2,344 or about 66%, ended within Frankfurt, underscoring the city's prominent role in our study's geographical focus.

Notably, about 18% of journeys exhibit a PSD ≈ 0, indicating immediate parking spot acquisition by drivers. The Mean Parking Search Duration (MPSD) across all journeys is 1 minute and 25

Table 4.2: Descriptive Statistics of Parking Search and Journey Variables (Duration in MM:SS Format)

	Variable	count	mean	std	5%	25%	median	75%	95%
All Journeys	Parking Search Duration	3550	01:25	01:59	00:00	00:18	00:50	01:44	04:59
	Walking Duration	3550	02:40	05:20	00:00	00:18	01:08	02:56	10:26
	Journey Duration	3550	19:22	13:22	06:20	10:51	16:17	23:56	42:20
Journeys with PSD>0	Parking Search Duration	2924	01:44	02:04	00:13	00:35	01:04	02:01	05:32
	Walking Duration	2924	02:52	05:34	00:01	00:28	01:19	03:09	10:28
	Journey Duration	2924	19:27	12:43	06:29	11:04	16:35	24:05	42:11
	Initial Search Radius (m)	2924	137	123	13	51	104	184	360
	Parking Offset Radius (m)	2924	128	118	4	46	97	178	351

seconds, while for journeys with a PSD greater than zero, it stands at 1 minute and 44 seconds. The mean walking durations are 2 minutes and 40 seconds (all journeys) and 2 minutes and 52 seconds (journeys with PSD>0). Observing that only 5% of parking search journeys exceed 4 minutes and 59 seconds indicates that prolonged PSDs are relatively infrequent. Furthermore, the Initial Search Radius is explored for journeys involving a parking search (PSD>0), indicating an average value of 137 m.

The Initial Search Radius refers to the distance between the point where a driver commences their search for parking and their eventual final destination, which is the location they aim to reach on foot after exiting their parked vehicle. This metric is pivotal in understanding a driver's perceived or acceptable walking distance from where they decide to park to where they intend to go. The radius essentially captures the anticipatory mindset of drivers, reflecting their willingness or unwillingness to walk long distances after parking. By examining this radius, we can infer how close drivers desire to park relative to their end destination, offering insights into urban planning needs, potential areas for transportation improvements, and the influence of factors such as weather, physical capabilities, or purpose of the trips on parking decisions.

While the Initial Search Radius provides an indication of a driver's initially intended walking distance and can be calculated in our training dataset, certain limitations in other datasets necessitate the calculation of another derivative metric termed here as the "Parking Offset Radius". A visualization of the Initial Search Radius (ISR) and Parking Offset Radius (POR) can be seen in Figure 4.2. The POR represents the distance between the starting point of a driver's parking search and the location of the parking spot they eventually select. Due to constraints in many GPS datasets, the exact pedestrian route from the parking spot to the final destination might not be available, implying that the exact location of the final destination is unknown. Thus, the POR does not provide a direct measure of the driver's initial walking tolerance, but it may serve as a proxy. Analyzing POR gives an impression of the search scope a driver is willing to undertake before settling on a parking spot. While not as precise as the desired initial walking distance indicated by ISR, this metric sheds light on driver behavior, preferences, and the dynamics of parking search patterns in various urban settings. Table 4.2 delves into the statistical nuances of these variables.

Exploring the aspect of speed during the journeys, especially during parking searches, brings

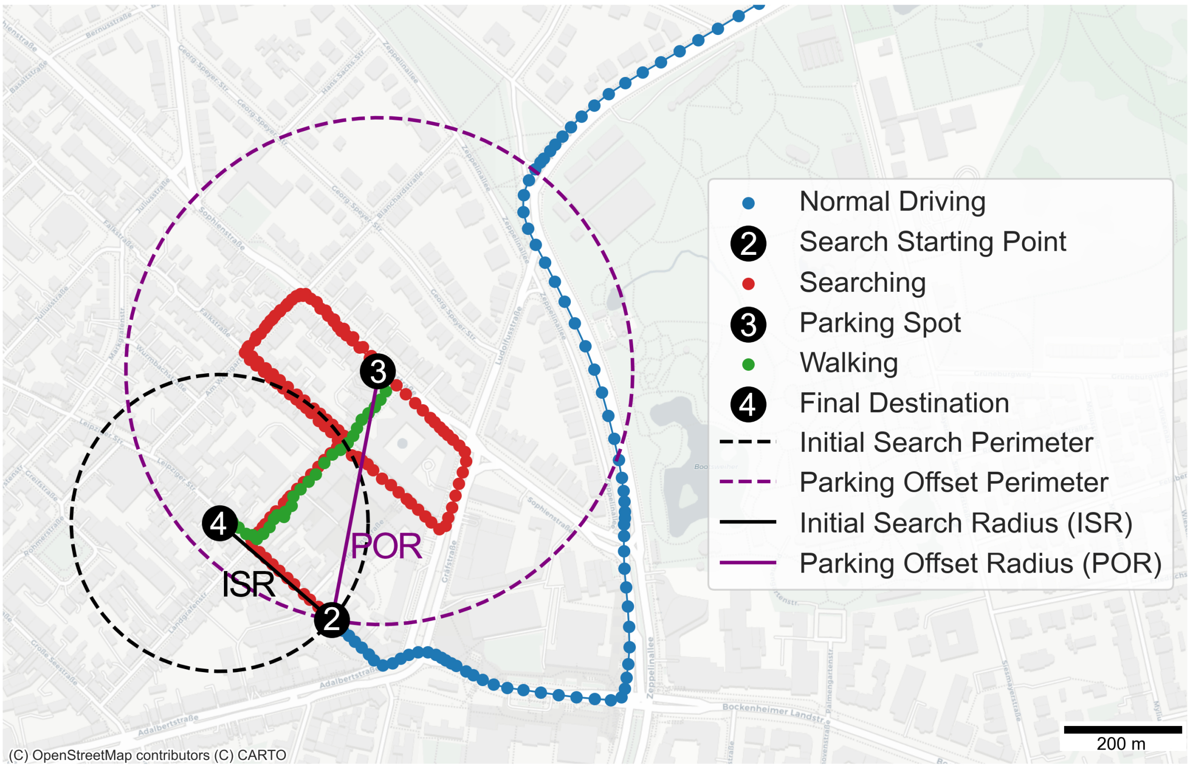


Figure 4.2: Schematic Representation of Parking Search Parameters, illustrating the Initial Search Radius (ISR) as the distance from the search starting point to the final destination and the Parking Offset Radius (POR) as the distance from the search starting point to the chosen parking spot

Table 4.3: Descriptive Statistics of Speed of GPS points During Normal Driving and Parking Search (Speed in km/h)

Speed	count	mean	std	5%	25%	median	75%	95%
All Driving Points	2,293,850	43	29	5	22	37	57	100
Normal Driving (within 1 km of Destination)	492,355	30	16	4	19	29	39	57
Parking Search (within 1 km of Destination)	123,482	16	10	0	8	16	23	36

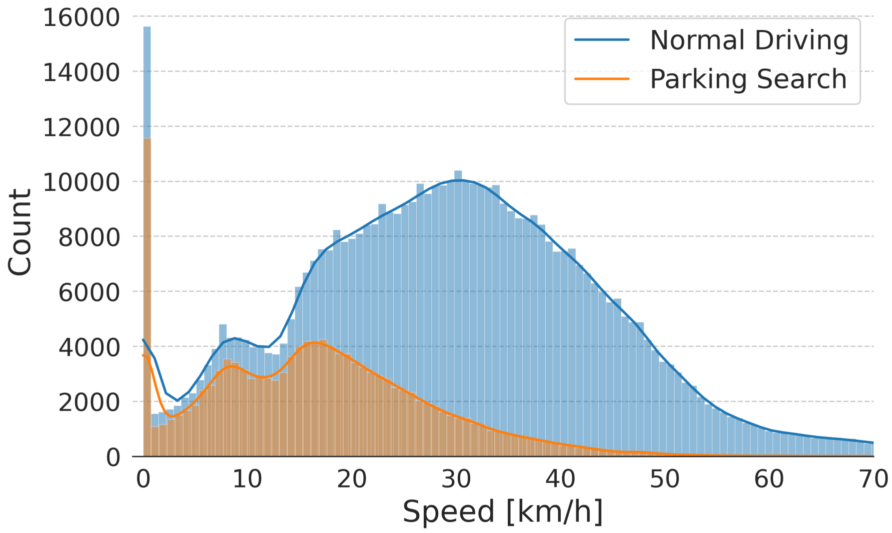


Figure 4.3: Histogram of Speed of GPS Points within 1 km radius of Final Destination, Separated between Normal Driving and Parking Search (Curves showing the estimated KDEs)

forward some insightful data. Across the 3,550 journeys, we analyzed around 2,293,850 GPS points, mostly recorded at a 1-second sampling rate. The speed at each point, captured directly by smartphone sensors (instead of being calculated by timestamp and location), offers a detailed look into how driving behavior shifts during different parts of a journey. Notably, the average speed of GPS points during the search for parking drops from 30 km/h to about 18 km/h in the final kilometer (Table 4.3). This slowdown is a clear indicator of drivers cautiously moving through areas, likely keeping an eye out for available parking spots. At the same time, the ratio of the standard deviation to the mean (coefficient of variation) increases (from 0.53 to 0.63), indicating a less even driving behavior.

Additionally, Figure 4.3 showcases the distribution of speed, distinctly categorized for points marked as "searching" and "normal driving". This detailed analysis of speed during different parts of a journey not only confirms that drivers slow down significantly while searching for parking but also highlights the importance of speed as a key variable in identifying and predicting parking search behavior in GPS data.

To further enrich our understanding of the dataset, additional details regarding journey and driver-related variables, and temporal information have been compiled and analyzed. Table 4.4 offers a snapshot of our dataset, shedding light on the diversity of journeys, driver demographics, and the contextual timing of events. Despite the fact that the contextual data has not been used in the training process, the inclusion of this information helps to better understand the data basis.

The drivers journey purposes —encompassing everything from work commutes to leisure outings— alongside a broad demographic spectrum, paints a vivid picture of the dataset's complexity. Temporal insights, revealing when these journeys typically occur, provide a snapshot of the dataset's dynamics over different times of day and days of the week. More detailed and nuanced analyses about the parking search behavior can be found in Saki and Hagen (2024a).

Obviously, the data is far from being representative for the population of car trips. However, this is exactly one of the reasons for developing the prediction model and for applying it to more comprehensive datasets.

Table 4.4: Descriptive statistics of journey-related and driver-related variables, and temporal information.

Journey-Related		**Driver-Related**		**Temporal Information**	
Variable	**Count**	**Variable**	**Count**	**Variable**	**Count**
Parking Type		Age		Time of the Day	
Free	3043	18-29	1643	Early Morning (4-8)	1173
Paid	280	30-39	1403	Morning (8-12)	1029
Illegal	184	40-49	175	Noon (12-16)	653
Planned Parking Duration		50-59	147	Afternoon (16-20)	522
Less than 30 min	1179	60+	127	Evening (20-0)	88
Between 30 min and 3 h	878	Gender		Night (0-4)	85
More than 3 h	197	Male	2983	Day of the Week	
Area Familiarity		Female	499	Weekday	2731
Known	2615	Diverse	13	Saturday	432
Unknown	892	Average Yearly Driven Distance		Sunday and Holiday	387
Trip Purpose		Less than 9000 km	1153		
Shopping	1082	Between 9000 and 30000 km	2278		
Leisure	932	More than 30000 km	64		
Home	807	Vehicle Type			
Work	446	Mini Car	1515		
Work	446	Compact Car	1472		
		Middle Class	345		
		Family Class	134		
		Van	29		
Missing	43	Missing	55		
Total	**3550**				

4.4 Model Architecture and Input Features

The development of a predictive model, especially in the context of parking search using GPS data, requires a method ensuring both precision and usability across different datasets. GPS car data typically differ with respect to the device with which they are recorded (smartphones, navigation devices, embedded devices) and the sampling rate of the data points. The fundamental goal of this research is to develop a model applicable to large GPS datasets coming from different sources, thereby overcoming the lack of missing representativeness of the ground truth data and generating robust aggregated statistics, even in the absence of detailed information present in our dataset.

Given that most historical GPS datasets primarily contain only driving trajectories from the starting point to the parking location, without the inclusion of walking routes or additional contextual information related to the driver or journey, our model had to be developed with a simplified attribute set. This approach was chosen to ensure that the model remains generalizable

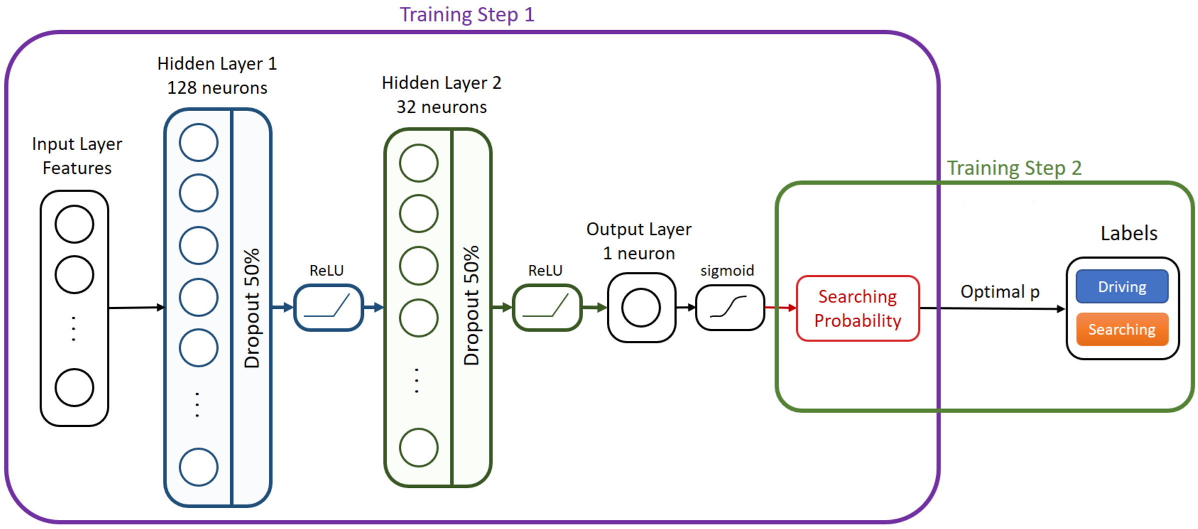


Figure 4.4: Model Architecture – This model identifies parking search route in GPS data

and applicable to a wide range of datasets, thereby enhancing its utility and relevance in practical applications.

The model employs a feed-forward neural network, a type of architecture that has demonstrated proficiency in handling complex patterns within data, especially when the relationships between input features are non-linear and multifaceted (Ojha et al., 2017). The architecture incorporates two hidden layers, with 128 and 32 neurons respectively. This specific architecture was chosen based on preliminary testing and adheres to standards in the field, ensuring a balance between model complexity and the ability to generalize to unseen data (Karlaftis & Vlahogianni, 2011).

- **First Hidden Layer (128 Neurons)**: This layer is tasked with capturing the initial patterns and relationships within the input features, providing a foundational basis for subsequent layers to build upon.
- **Second Hidden Layer (32 Neurons)**: This layer further refines the patterns identified by the preceding layer, focusing on more nuanced relationships that are crucial to accurately predicting parking search behavior.
- **Output Layer (Sigmoid Activation)**: The output layer employs a sigmoid activation function, providing a probability score indicating whether a point is a "searching" point or a "normal driving" point.
- **Optimizing Step**: This step labels the point as "searching" or "normal driving" based on the searching probability and the optimal cut-off (optimal p) value associated with the journey's sampling rate.

The complexity of the model is justified by the necessity to accurately recognize the subtle and often complex patterns inherent in parking search behavior, which simpler models might overlook. The architecture of the model is illustrated in Figure 4.4.

For training the model, we test a wide range of variables:

- **Speed**: Speed, obtained from smartphone sensors and commonly featured in Floating Car Data (FCD), is a crucial variable, directly influenced by driver behavior and has been utilized in previous heuristic models to define searching. It is anticipated that variations in speed, particularly reductions, may be indicative of a driver initiating their parking search, navigating through potential parking spots, or maneuvering into a parking space (see Figure 4.3).
- **Acceleration**: Acceleration data has been demonstrated to be crucial for travel mode detection (Feng & Timmermans, 2013). Viewing parking search as a subcategory within the driving mode, this data can significantly enhance the model's accuracy. While the app does not directly collect accelerometer data, incorporating the sampling rate (the time interval between consecutive points) alongside speed allows us to implicitly integrate acceleration information into the model.
- **Time and Location Variables:** Time and location variables are investigated due to their potential correlation with parking occupancy rates (McCahill, 2017). Different times of the day or days of the week may exhibit varying levels of parking occupancy, which in turn, could influence a driver's decision to start their parking search earlier or later as well as search success and duration. Furthermore, specific locations may be associated with distinct parking search patterns due to factors such as proximity to popular destinations, availability of parking infrastructure, or general traffic conditions.
- **Distance to Parking Spot:** The distance to the destination is a crucial variable, as it signifies the driver's acceptable walking distance. Most drivers aim to minimize their total journey duration, and this often involves optimizing the walking distance. However, given that historical GPS data typically doesn't include the exact final destination, we - in line with previous research - opted to use the distance to the parking spot as a proxy. (See the related discussion on "Initial Search Radius" versus "Parking Offset Radius" in Section 4.3.3).
- **Sampling Rate:** Our dataset, which is recorded at a high frequency, predominantly features a 1s sampling rate. However, this level of granularity is not a common characteristic across all datasets. To ensure that the model remains applicable to datasets with varied sampling rates, we used it as an input variable in the model. Additionally, we created data subsets with average sampling rates of 5s, 10s, and 15s. The model was then retrained on these subsets, thereby enhancing its robustness and ensuring its applicability to datasets with different sampling rates.
- **Lagged Variables:** For time-varying variables such as speed and sampling rate, the values of five previous points are incorporated into the input vector reflecting the possibility that the start of parking search is not only triggered by contemporaneous variables but also by lagged variables.
- **Additional Variables:** While additional variables such as weather data are incorporated into the model, their influence on parking search behavior are evaluated during the model training process.

The next section will evaluate the relevance of these variables / input features. Through the design of this model, we have strived to create a tool that is not only accurate in its predictions but also versatile in its applications. By ensuring that it can be applied to various datasets with different characteristics, we enhance its utility and potential impact in the field of parking search analysis, providing a valuable resource for future research and practical applications in urban mobility and planning.

4.5 Model Performance

4.5.1 Training and Evaluation

The training process is divided into two stages. In the first stage, the weights of the neural network are calculated using a binary cross-entropy loss function. The model employs a sigmoid function in its output layer, which yields a probability score for each GPS point, indicating the likelihood that the driver is engaged in parking search behavior at that particular point.

Specifically, the model outputs a probability value between 0 and 1 for each GPS point, where a higher value signifies a higher probability that the point is part of a parking search route. In practical terms, these probability scores act as a subtle tool to decode the start and duration of parking search within a journey. By identifying points with high probabilities, the model outlines the transition from normal driving to parking search. The f1-score, a harmonic mean of precision and recall, was selected as the performance metric during the initial stage of training. The choice of f1-score is grounded in its ability to balance both precision and recall, providing a comprehensive view of the model's performance across both metrics. In our framework, precision is the accuracy of positive predictions, which is crucial to ensure that the identified parking search points are indeed accurate and not false positives. In contrast, recall is the ability of the model to correctly identify all actual instances of parking search points, which is vital to ensure that all relevant data points are captured.

Since both false positives and false negatives have significant implications in the context of parking search behavior analysis, the f1-score becomes a relevant metric, ensuring that the model does not disproportionately prioritize either precision or recall, thus maintaining a balanced predictive ability.

The second stage of the training process involves determining the optimal probability cut-off through grid search, with the objective of minimizing the Mean Absolute Error (MAE) in PSD. Once the model identifies the transition from normal driving to searching, it categorizes the remaining route as the parking search. This is based on the assumption that after there has been a "switch" from "Normal Driving" to "Searching", a switch back is no longer possible. MAE represents the average absolute differences between the observed actual outcomes and the predictions made by the model. In the grid search process, MAE is selected as the primary metric due to its robustness against outliers, which is essential given the right-skewed distribution of PSD that includes extreme values. MAE's straightforward interpretability is advantageous, as it directly reflects the average prediction error in PSD, providing clear practical significance. Moreover, MAE aligns well with the core objective of precisely identifying the complete search route by measuring the direct deviation from the predicted PSD.

The dataset was partitioned into training and test sets, with the test set comprising approximately 10% of the total data. This partitioning ensures that the model is exposed to a diverse range of data during training while also reserving an unbiased subset of data for validation purposes.

To further enhance the robustness of the validation process, a k-fold cross-validation technique was employed with k being set to 10. The choice of k=10 is based on its widespread adoption in similar studies, striking a balance between computational efficiency and obtaining a reliable estimate of the model's performance (Kohavi et al., 1995). In this approach, the dataset is partitioned into 10 subsets; the model is trained 10 times, each time using 9 subsets for training and the remaining subset for testing. This approach not only maximizes the utility of available data but also provides a more reliable and generalized performance estimate by mitigating the potential biases or variances that might arise from a single random data split (Fushiki, 2011).

The model's performance was eventually evaluated by assessing the MAE in PSD and comparing the actual and predicted MPSDs on an aggregate level. The MAE provides a straightforward, interpretable measure of the average error in the predicted PSD, while the comparison of MPSD offers insights into the model's ability to accurately predict average parking search durations across different journeys.

In refining our predictive model, we initially conducted a feature importance analysis using the Leave-One-Out (LOO) method (Lei et al., 2018) to understand the impact of each variable within the model. The LOO method is an approach where, one by one, each variable is excluded from the model, and the performance is assessed without it. This technique helps in identifying the contribution of individual variables to the model's accuracy.

Figure 4.5 illustrates the f1 scores for models trained with one variable excluded at a time, with the excluded variable displayed on the x-axis. The y-axis shows the f1 scores, reflecting the model's performance without the respective variables. The variables are ordered by their impact on model performance, in descending order of the f1 score. As anticipated, "Distance to Parking Spot" and "Speed" emerged as the most influential variables. The other variables, while seemingly similar in their impact, exhibit varied levels of importance, with "Weather" being the least influential.

Given our objective to develop a generalized model with wide applicability to diverse datasets, strategic decisions regarding variable inclusion were necessary. "Heading" and "Weather" data are often not recorded in large datasets, and while there are methodologies to approximate these when missing, the variance from internal readings can pose challenges in cross-dataset application. "Area Type" was also considered for exclusion to prevent potential bias, as it could inadvertently introduce a predisposition towards Frankfurt, where the majority of our data was sourced.

Subsequent models were trained using "Speed" and "Distance to Parking Spot" as base variables, with the inclusion of various combinations of the remaining variables. Through this iterative process, we found that a combination of "Speed", "Distance to Parking Spot", and "Sampling Rate of the GPS Points" yielded the most accurate model, surpassing even those models that included time variables such as "hour of the day" and "day of the week".

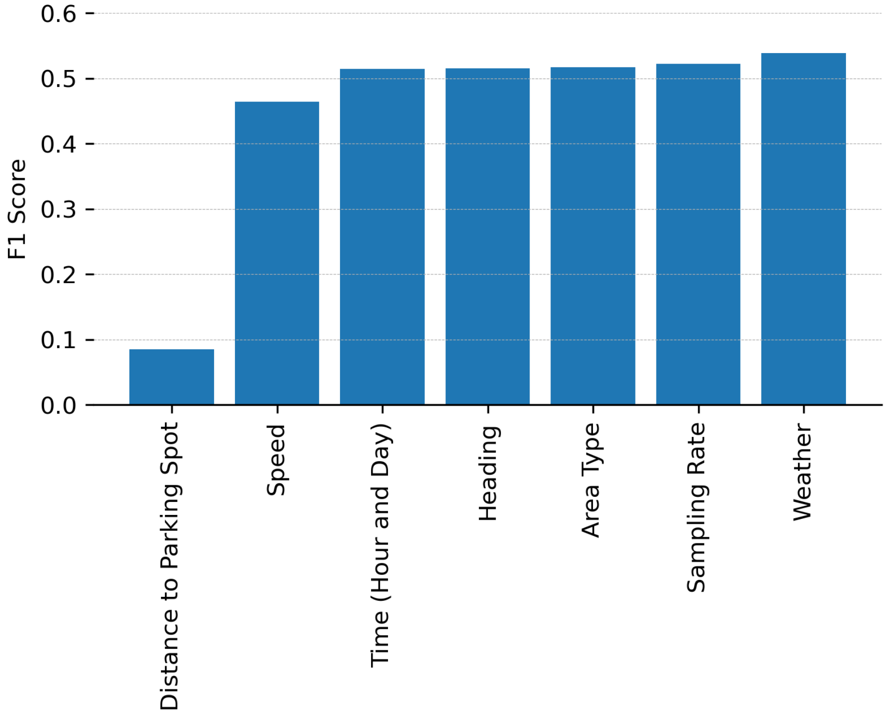


Figure 4.5: Feature Importance Analysis Using Leave-One-Out Method. Each bar represents the f1 score of the predictive model when a specific variable, as labeled on the x-axis, is excluded.

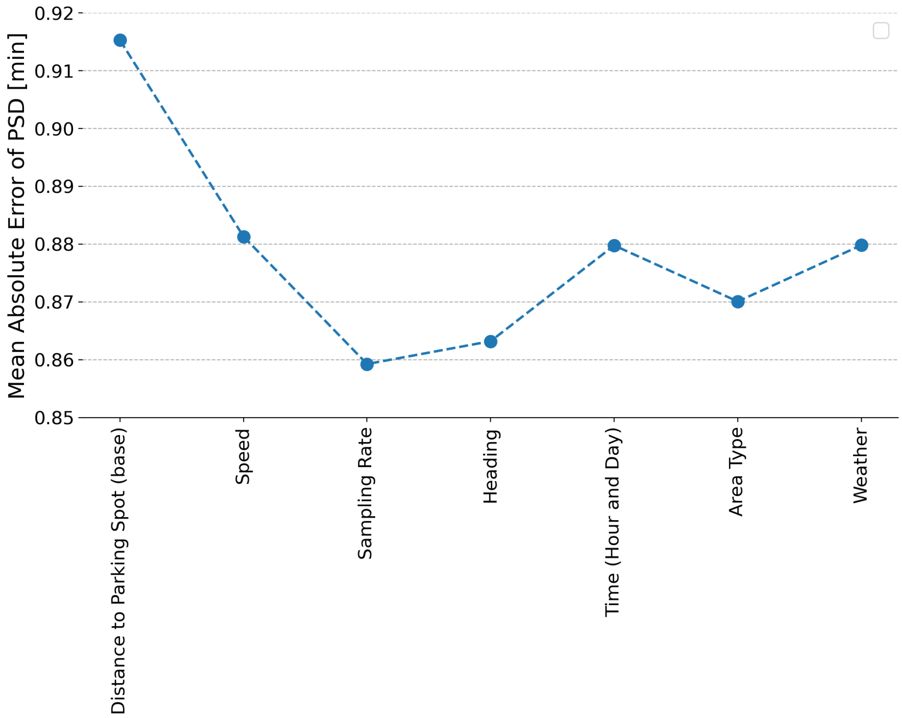


Figure 4.6: Evolution of Model Performance with Added Variables

This observation is visually represented in Figure 4.6, where the model's performance is plotted against the order of variable addition, as represented on the X-axis. It is clear from the figure that the addition of subsequent variables in addition to speed, sampling rate, and distance did not contribute to any significant performance gains.

A possible explanation for this result is that these three variables - distance to parking spot, speed and sampling rate - (as well as their lagged values and also in their combination), may serve as proxy variables for a range of other potentially relevant variables. For example, as mentioned, speed and lagged speed, in combination with the sampling rate, contain information on acceleration. Speed and acceleration of a vehicle may already include information on the traffic situation (hour and day), the area type, and the weather.

Given this observation, the input features for the final model include only the distance to parking spot, the vehicle speed, and the sampling rate of the GPS points. This not only streamlined the model but also improved its simplicity and generalizability. By focusing on a concise set of impactful features, we ensure that our model remains adaptable and applicable across a wide range of datasets and scenarios. These choices result in a set of 13 distinct input features for each data point, ensuring the model remains robust yet efficient. The details of these input features are outlined as follows:

- **Current Distance to Parking Spot:** For a point at timestamp "t", the model considers the vehicle's distance to parking spot. This variable is crucial as it helps predict the likelihood of a driver beginning a parking search, with closer proximity potentially indicating imminent parking behavior.

- **Six Speed Measurements (one contemporaneous and five lagged values):** The speed at the current timestamp "t" and the speeds at the five preceding timestamps ($t-1$, $t-2$, $t-3$, $t-4$, $t-5$) are essential inputs. These speeds reveal dynamic changes in driving behavior, which are key indicators of parking search. For instance, slowing down near potential parking spaces or maintaining a consistent low speed can suggest that the driver is searching for parking.
- **Six Sampling Rate Measurements (one contemporaneous and five lagged values):** The sampling rate at which the speed and position data are captured is integral to the model. Different sampling rates may result in different patterns for speed variation before starting the parking search. This variable is considered at timestamp "t" and for the previous five timestamps ($t-1$, $t-2$, $t-3$, $t-4$, $t-5$). Additionally, by integrating the sampling rate alongside with speed, the model indirectly captures changes in acceleration, which can be beneficial in driving mode recognition (Janidarmian et al., 2017).

This carefully selected set of input features allows the model to effectively analyze and predict parking behavior by leveraging critical aspects of driving data. By focusing on only tree main variables, the model maintains high predictive accuracy and efficiency in identifying parking search patterns while being generalized to other data sources. To make the model usable for others, we have published the model publicly on GitHub [github.com/ReLUT/parking-search-prediction].

The evaluation results, divided by different sampling rates, are presented in Table 4.5. The model shows superior performance across various sampling rates, with a negligible deviation between actual and predicted MPSD and an MAE of less than one minute. This indicates the model's ability to predict individual PSD with an average error margin of under one minute, confirming its reliability and accuracy in predicting parking search.

A more detailed look at the results in Table 4.5 reveals subtle variations in the MPSD predictions across different sampling rates. For instance, while the 1s and 5s sampling rates yield identical MPSD predictions, a slight deviation is observed as the sampling rate increases to 10 s and 15 s. This could be attributed to the reduced granularity of data at higher sampling rates, which might impact the model's ability to capture the complicated details of parking search behavior. The variations in predictions at different sampling rates underscore the importance of selecting an appropriate sampling rate, ensuring that the model is fed with data that is sufficiently detailed to capture the nuances of parking search behavior while being computationally feasible. In general, we can conclude that as the sampling frequency decreases the prediction accuracy decreases.

The model demonstrates significant performance at an aggregated level, closely approximating the actual MPSD in its predictions. To further assess the model's efficacy, particularly at an individual level, we conducted an additional analysis (for the 1s sampling rate dataset) comparing it against a baseline model that consistently predicts the overall MPSD of 01:26 as the parking search duration for every instance. The baseline model, applying a simplistic approach of using the average MPSD as the prediction for all data points (Brownlee, 2016), resulted in a MAE of approximately 80 seconds. In contrast, our neural network model achieved an MAE of 52 seconds. This comparison indicates a significant reduction in prediction error by 28 seconds per instance, underscoring the enhanced accuracy of the neural network model over the baseline approach.

This improvement is particularly noteworthy because it demonstrates the neural network's ability to effectively learn and adapt to the complexities embedded within the GPS data. Such an ability is crucial for applications that require precise predictions of parking search times, such as in the development of smart urban mobility solutions where even small gains in prediction accuracy can lead to better traffic management and enhanced urban planning. The comparison also highlights the model's robustness in handling individual variability in parking behavior, which is often masked by aggregated data.

Given that approximately two-thirds of our dataset originates from Frankfurt, there is a potential concern that our model could exhibit a bias toward this geographic area. However, it is important to note that we have not incorporated any explicit spatial characteristics specific to Frankfurt into the model. Instead, the influence of this city in our model could potentially arise indirectly through other variables, such as driving speed, which might reflect urban driving conditions peculiar to Frankfurt.

To assess and control for possible geographical bias, we undertook an evaluation strategy. Initially, we trained our model on the entire dataset, encompassing all recorded journeys. Subsequent predictions of parking search duration were then analyzed to compare the model's performance for trips within Frankfurt against those outside it. This comparison involved calculating the MAE for both sets of trips.

The results of this analysis were enlightening. The MAE for trips within Frankfurt was found to be approximately 52 seconds, while the MAE for trips conducted outside of Frankfurt was notably lower, at about 40 seconds. This disparity could suggest that the model is slightly more conservative in its error margin for Frankfurt, possibly due to the denser and more complex urban environment. However, the relatively close performance metrics indicate that the model maintains a commendable level of accuracy and reliability when applied to other regions.

This finding is critical as it confirms that the model's predictive capabilities are not overly tailored to the Frankfurt environment. The similar levels of precision achieved both within and outside Frankfurt underscore the model's applicability and effectiveness for other cities in Germany. Such robustness enhances the model's utility as a tool for urban mobility analysis and planning across diverse urban settings, not just confined to the context in which it was primarily developed.

Table 4.5: Model Validation (Duration in MM:SS Format)

Data Sampling Rate	MAE	MPSD (Actual)	MPSD (Predicted)
1s	00:52	01:26	01:27
5s	00:52	01:26	01:27
10s	00:54	01:26	01:26
15s	00:56	01:26	01:23

4.5.2 External Test Dataset

The ultimate goal of our predictive model is to ensure its applicability and reliability across various data sources, thereby enhancing its utility in diverse research and practical contexts. To assess the model's performance and evaluate its predictive capabilities on other data sources, we conducted additional field experiments, namely extended park-and-visit experiments, to gather additional labeled parking search data through a different smartphone application and involving distinct drivers.

Park-and-Visit experiments are foundational tools used to measure how long it takes for drivers to find parking in a controlled setting (Polak & Axhausen, 1992). Typically, these experiments involve selecting a group of volunteer drivers who are asked to find parking in a specified area while their actions are monitored. The core objective is to understand parking search behavior under different conditions—such as during peak traffic hours, in various types of parking lots (like on-street versus off-street), or under different pricing strategies. Researchers use these experiments to gather data on the duration of parking searches, the routes taken by drivers, and the effect of various interventions on parking efficiency.

Traditionally, Park-and-Visit experiments control several variables to standardize the conditions under which data is collected. This often includes setting a predefined starting point for the search, which might be a certain distance from the desired destination or directly at the destination itself. Researchers might also control for the time of the day, the type of parking being searched for, and other factors to ensure that the data can be reliably compared across all participants. This method has been cited in various studies (Zhu et al., 2020; Alemi et al., 2018) and is valued for its ability to provide clear, quantifiable insights into parking behavior under controlled experimental conditions.

While traditional methods offer valuable insights, they also come with significant limitations. The major drawback is that by controlling many aspects of the parking search, these experiments may not accurately represent the true diversity of strategies and decisions made by drivers in real-life scenarios (Belloche, 2015). For example, in real-world conditions, drivers may choose different starting points based on their prior knowledge of the area, anticipated parking difficulties, or even personal preferences for walking distances. Such dynamic decision-making is constrained in a traditional fixed-start experiment.

To bridge the gap between experimental control and real-world variability, we introduced the "Dynamic Park-and-Visit Experiment". This approach allows drivers to choose their starting points for parking searches. In this modified methodology, participants are given real-world tasks, such as driving from a shopping center to an office building, and are free to start their search for parking at any point along their journey, reflecting a typical decision-making process that drivers undergo daily. This approach is particularly valuable in urban studies, where understanding the decision-making process on starting point of the parking search is of significant importance.

The dynamic park-and-visit experiment was conducted between November 2020 and June 2021, resulting in the collection of 161 distinct journeys. These experiments were carried out mostly in cities of Frankfurt and Rostock in Germany, with 96 and 63 journeys each, respectively.

The data collection was facilitated by five drivers, who were either students or colleagues of the authors. The drivers were assigned specific origin and destination points, covering a wide array of locations including shopping centers, office buildings, residential areas, and places for leisure activities. During the experiment, drivers were given the freedom to start their parking search from any point along their journey from the origin to the destination, based on their personal preferences and familiarity with the area. They were also free to choose any available parking, be it on-street or off-street. The precise moment when the driver began the parking search was recorded by the driver. Additionally, the entire route, including the search for parking, was tracked using the GPS functionality of the OsmAnd app[7] on the driver's smartphone.

Apart from tracking the journey, drivers provided further information, including whether the drivers were familiar with the destination area, their reasons for choosing a particular parking spot, and other contextual details. Information such as the type of parking chosen and the prevailing weather conditions were also recorded, adding layers of contextual data to the GPS traces.

Within our dataset, 57 trips, representing approximately 35% of the total, recorded a parking search duration of 0 seconds. The parking search durations in our dataset varied significantly, with an average search time of 1 minute and 20 seconds. The distribution of these times revealed that the median parking search duration was notably shorter, at just 30 seconds, indicating that half of the parking searches were completed quite quickly. The 75th percentile was observed at 1 minute and 40 seconds, suggesting that most parking searches were relatively brief. However, the maximum recorded parking search duration extended up to 22 minutes, highlighting occasional challenges faced by drivers in finding suitable parking spaces.

More information on the parking events are summarized in Table 4.6, which details the types of parking locations used, the familiarity of drivers with their destinations, the legality and cost of parking spots, and the distribution of parking events over different times of the day.

[7] https://osmand.net/

Table 4.6: Summary of Categorical Data from Dynamic Park-and-Visit Dataset

Category	Count	Percentage
Parking Location Type		
On-street	118	73.3%
Off-street	43	26.7%
Familiarity with Destination		
Familiar	21	13.0%
Not Familiar	140	87.0%
Parking Type		
Free	146	90.7%
Paid	12	7.5%
Illegal	3	1.9%
Time of Day		
Morning [06-10]	9	5.6%
Noon [10-14]	53	32.9%
Afternoon [14-18]	30	18.6%
Evening [18-22]	68	42.2%
Night [22-06]	1	0.6%
Total Journeys	161	

This external test data serves as a resource for evaluating the model's predictive accuracy. We made this dataset publicly available and published on GitHub [github.com/ReLUT/parking-search-prediction] to be used for prediction purposes and further investigations by other researchers.

Upon applying our final model to this dynamic park-and-visit dataset, we observed promising results. The model, trained comprehensively using all data from the start2park app, effectively predicted parking search durations for all journeys in the dataset. We calculated the MAE at 39 seconds. Furthermore, the deviation between the actual MPSD and the predicted MPSD was remarkably low, at only 17 seconds. These error metrics are consistent with those achieved during the training and testing phases with the start2park dataset, thereby confirming the model's adaptability and predictive accuracy when applied to datasets from varied sources. Figure 4.7 provides a visual representation of an example journey along with its predicted labels.

The minor difference between predicted and actual values, as indicated by the MAE and MPSD, validates the model's ability to accurately predict PSDs, even when applied to data collected under varied conditions and contexts. This validation not only confirms the model's robustness but also enhances its utility and reliability in further research and practical applications.

4.5.3 Comparison of the ML Model to Previous Heuristic Methods

Section 2 summarized various methods designed to identify parking search within GPS data, each with its own foundational assumptions. A thorough comparative analysis is crucial to uncover

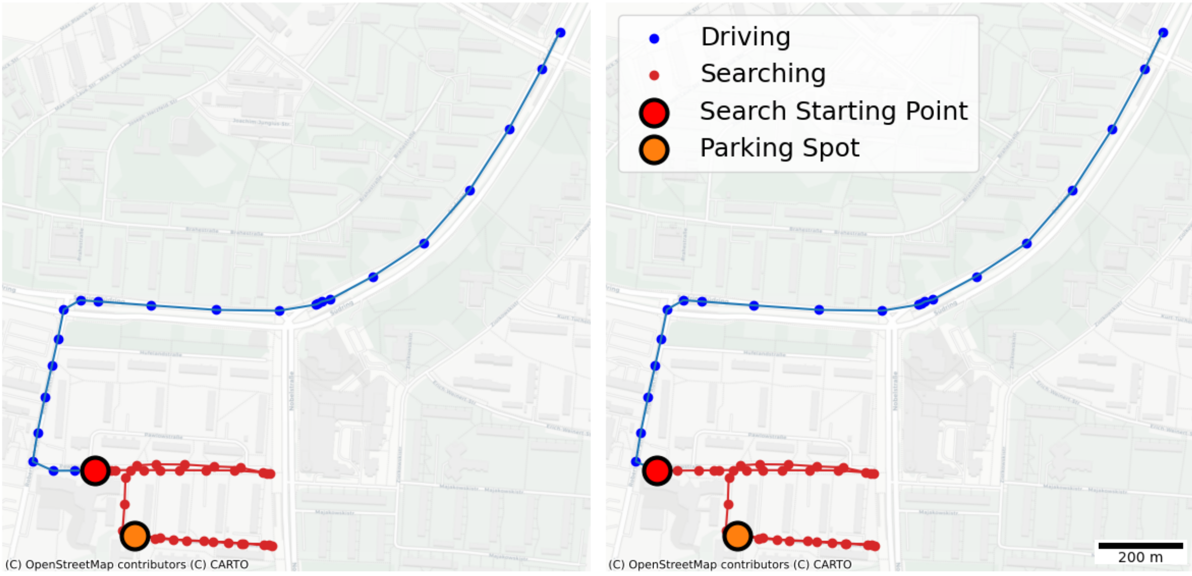


Figure 4.7: An example recorded journey using dynamic park-and-visit approach – Actual labels (left), predicted labels (right). The actual PSD in this example is 03:33 with an absolute error of 12s in prediction.

the effectiveness, precision, and applicability of these models. Furthermore, it sheds light on the advantages, limitations, and best-use scenarios of each. We conduct a detailed comparison of the parking search identification models explained in section 2. The goal is to highlight the relative performance of each method in accurately identifying parking search behavior within GPS data. The models will be evaluated using several key metrics, including:

- **Mean Absolute Error (MAE)**: Measures the average absolute error between predicted and actual parking search durations.
- **Mean Parking Search Duration**: Assesses the average duration of parking searches predicted by the model.
- **Median Parking Search Duration**: Evaluates the median duration of parking searches, providing insights into the central tendency of the predicted durations.
- **Mean Parking Offset Radius**: Analyzes the average distance between the parking spot and where the model predicts parking search begins.

The analysis will involve applying each method to the same dataset, ensuring a uniform basis for comparison. The predictions from each model will be evaluated against our Ground Truth data, providing a solid framework for assessing their accuracy and reliability. The metrics mentioned above will serve as the foundation upon which the performance of each method will be assessed, offering a comprehensive view of their capabilities. In each approach, the parking search is identified as follows:

- **Approach 1 - Naive 200 m Radius:** Assumes parking search begins when the driver reaches a 200 m distance to the parking spot.
- **Approach 2 - Speed Threshold:** Identifies parking search when average vehicle speed

across 5 consecutive points drops below 23 km/h, given the standard deviation remains under 5 km/h.

- **Approach 3 - First Local Minima:** Determines the onset of parking search at the initial local minimum in the distance to the parking spot profile within a 400m vicinity of the parking spot. It assumes that during regular driving, the distance to the parking spot should typically decrease. The first deviation from this trend, i.e., the first time the distance to the parking spot increases, within a specified radius, indicates the commencement of the parking search.
- **Approach 4 - Actual-Shortest Path:** Identifies parking search excess duration by comparing the actual path and the shortest path, noting how much they diverge. The shortest path is estimated using Google Maps Direction API considering the predicted traffic condition during the time and day of the journey.
- **Approach 5 - Machine Learning Model:** Uses the neural network model to predict parking search route based on various GPS point attributes. The model architecture is explained in section 4.5.1.

Table 4.7: Comparative Analysis of Performance in Identifying Parking Search in GPS data of Different Methods
(Duration in MM:SS Format)

	MAE	Mean PSD	Median PSD	Mean Parking Offset Radius (m)
Ground Truth	-	01:25	00:50	120
Approach 1: Naive 200 m Radius	01:06	02:21	01:40	-
Approach 2: Speed Threshold	01:22	02:38	01:56	225
Approach 3: First Local Minima	01:10	01:39	00:47	68
Approach 4: Actual-Shortest Path	-	02:05	01:46	-
Approach 5: Machine Learning Model	00:48	01:26	01:08	87

The results of the comparative analysis are presented in Table 4.7. A clear standout from the results is the Machine Learning (ML) model, or Approach 5, which demonstrates best performance by achieving an MAE of just 48 seconds — notably the lowest among all methods explored. This implies that the ML model has the smallest average error margin, thus providing predictions that are closest to the actual values observed in the ground truth. Moreover, it records a Mean PSD of 1 minute and 26 seconds, which is impressively close to the ground truth's Mean PSD, showcasing its ability to predict parking search durations with a high degree of accuracy and consistency at an aggregate level.

Approaches 1 (Naive 200m Radius) and 2 (Speed Threshold) exhibit higher MAEs and Mean PSDs, hinting at a potential tendency to overestimate parking search durations. This might come from their relatively simple assumptions and a possible inability to adapt to the varied parking search behaviors exhibited by different drivers. Additionally, while Approach 1 does not offer insights into the Parking Offset Radius, Approach 2 seems to overestimate it.

Approach 3 (First Local Minima) yields interesting results. Despite its higher MAE and an overestimated Mean PSD, its Median PSD aligns quite closely with the ground truth. This might suggest that Approach 3 can provide a decent estimate for the typical PSD within a dataset, even though the Parking Offset Radius calculated using this method seems to be significantly underestimated.

Approach 4, which measures the extra time spent driving due to parking search, is not directly comparable to the ground truth data. However, if we assume that any difference between actual and shortest paths should be calculated only after the parking search has begun —since any route taken before that is the driver's initial preference regardless of parking search route choice— the excess duration calculated by Approach 4 should be equal to or less than the actual parking search duration. When we look at the data as a whole, we notice that the average duration estimated by Approach 4 tends to be higher than both the Mean and Median PSD of the ground truth, suggesting it might overestimate the excess travel time due to parking search.

While the ML model shines as the top-performing approach in accurately identifying parking search within GPS data, it is vital to recognize the intricate and data-dependent nature of this method. The effectiveness of the ML model is fundamentally linked to the quality and volume of the ground truth data available for training, which can be notably challenging to obtain in both high quality and substantial quantity. Moreover, Approach 3, despite its limitations, has demonstrated commendable results among the heuristic approaches, securing its spot as the second-best approach. Surprisingly, the next position is occupied by Approach 1, which has outperformed Approach 2 across all compared metrics. This suggests that Approach 1 might serve as a viable first step for an initial exploration of data.

In order to evaluate the performance of the different approaches more comprehensive, Figure 4.8, compares the distribution of predicted search durations by each method against the ground truth data. Our analysis reveals that while our model shows an increment in prediction errors for shorter durations, its overall accuracy across all durations—short, medium, and long—is superior.

Hence, detecting very short parking searches remains a challenge also for the ML model. This aspect is crucial, as short search durations are often less predictable and can vary more dramatically than longer searches, making them inherently difficult to model with high accuracy. The ML model's strength lies in its ability to handle longer search durations effectively, where it matches or exceeds the performance of heuristic methods. As parking search durations increase, the performance gap between our ML model and other heuristic approaches narrows (Figure 4.8). This indicates that both heuristic and our proposed ML model are likely to perform well in scenarios known for extended search times, such as during peak hours or in densely congested areas, with ML model still being superior.

However, the broader application of our ML approach becomes particularly advantageous when considering its use across various urban environments and times. Unlike heuristic methods that may require adjustments or recalibrations based on specific urban contexts or time variations, our ML model maintains consistent performance due to its training across a diverse set of data inputs.

In practical terms, this means that our model offers a robust tool for urban planners and traffic management systems to understand and predict parking behaviors in a comprehensive manner. By excelling in aggregate-level performance, the model provides reliable insights that can inform the development of more effective parking management strategies and urban planning solutions.

4.6 Large-Scale Application

4.6.1 INRIX Dataset Description

As mentioned, one major drawback of our ground truth data is its lack of representativeness. A possible approach to dealing with this issue is to apply the ML model to a large-scale dataset, which may even be representative.

The extensive dataset of GPS trajectories used in this study is provided by the private company INRIX. The initial provided dataset contained approximately 18 million journeys with billions of GPS points either originating, ending, or passing through a bounding box around Frankfurt am Main in 2019. The dataset incorporated a variety of vehicle types, both from individual consumers and fleet operations. The fleet segment comprised mainly of delivery trucks, taxis, ride-sharing vehicles, and essential service vehicles such as police cars, ambulances, and fire trucks. The dataset incorporated diversity not only in terms of vehicle function but also in size, classifying vehicles as light, medium, or heavy based on their weight.

For the purpose of this research, we streamlined the dataset to focus on:

- Consumer Trips: To understand patterns relevant to most daily commuters and city dwellers.
- Light-Weight Vehicles: Offering insights specifically about the majority of vehicles on the road.
- Trips ending in Frankfurt: Ensuring that only trips are taken into account which end as parking within Frankfurt.

After filtering, the dataset encompassed 868,561 journeys, a vast treasure of data that promises to shed light on previously undiscovered aspects of parking search behavior in Frankfurt. The sampling rate ranges from a median of 6 seconds with a 5% percentile of 4 seconds and a 99% percentile of 20 seconds. Journey distance shows a broad range with a median of 15.2km. The shortest 5% of journeys are barely over a kilometer, while the longest 1% stretch over 310 km. On average, a journey lasts 35 minutes. However, 5% of journeys are as short as 3 minutes, while

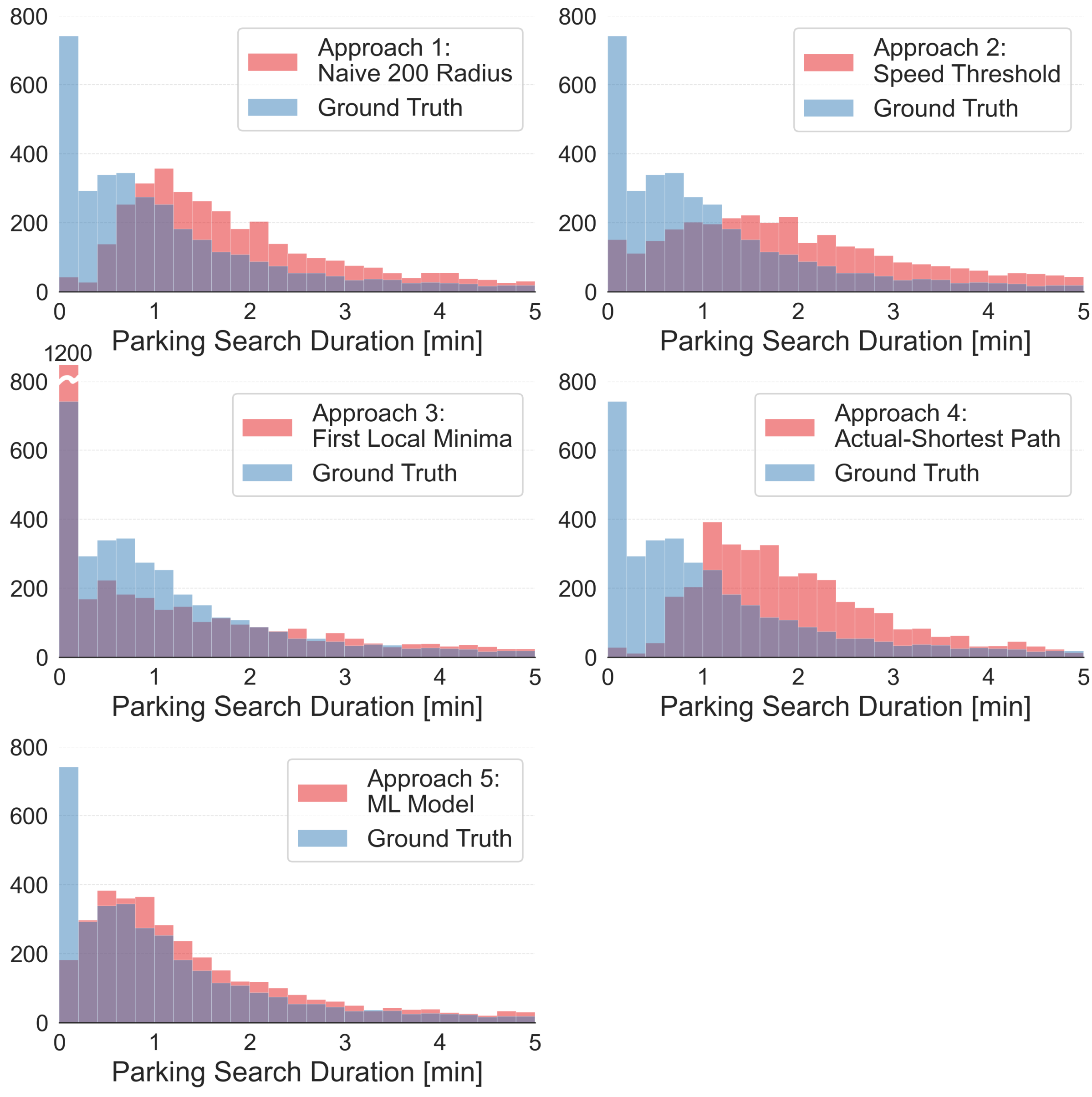


Figure 4.8: Comparative Distribution of Predicted Parking Search Durations for Various Heuristic Approaches and the Machine Learning Model Against the Ground Truth Data. (Purple color indicates overlapping.)

the top 1% can last up to 195 minutes. A mean value of average speed of 49 km/h is noted. The 5% percentile is only 12 km/h, whereas the fastest 1% average speed is approximately 121 km/h. Further details of the dataset can be seen in Table 4.8.

Table 4.8: Descriptive Statistics of the Studied Dataset for Parking Search Behavior Analysis

	count	mean	std	5%	25%	50%	75%	95%	99%
Sampling Rate (sec)	868,561	9	5	4	4	6	14	16	20
Journey Distance (km)	868,561	40	63	1	5	15	41	183	310
Journey Duration (min)	868,561	35	39	3	11	22	42	119	195
Journey Average Speed (km/h)	868,561	49	29	12	24	42	70	104	121

The dataset paints a vivid picture of Frankfurt's hourly and weekly traffic patterns. Early weekdays, generally witness a higher volume of journeys ending in Frankfurt in the mornings, peaking around 8-9 AM. However, as the week progresses towards the weekend, a shift is observed with higher volumes during the later parts of the day. The least number of journeys ending in Frankfurt occur between midnight and 5 AM, with a pick after that, highlighting the city's morning rush. A steady volume of journeys is observed in the afternoon, with a slight peak around 5-7 PM, indicating the evening rush.

The depth and granularity of this dataset offer a remarkable scope to explore parking search behavior in Frankfurt. To the best of our knowledge, no prior research has leveraged such an extensive dataset for this specific purpose.

4.6.2 Insights and Statistics

The parking search model was applied to our dataset, encompassing 868,561 journeys. Within this dataset, a significant 33% (or 285,594 journeys) reported a PSD of zero. This indicates that for nearly one-third of the journeys, drivers found parking immediately upon reaching their destination.

Considering the entire set, the mean PSD is 1 minute and 30 seconds with a median PSD of 15 seconds. Focusing on the 582,967 journeys that did face a non-zero PSD, the mean increases to 2 minutes and 15 seconds with a median of 42 seconds. The data distribution, as depicted in Figure 4.9, is predominantly right-skewed, indicating that while the majority of drivers either found parking immediately or within a short span of time, the top searchers spent a considerable time finding a vacant parking spot. Highlighting the experience of the 95th percentile, these drivers searched for a significant 8 minutes and 20 seconds. A holistic view of these statistics is presented in Table 4.8.

Parking Offset Radius captures the distance between the starting point of the parking search and the eventual parking spot. This variable is a proxy for the acceptable waking distance by the driver (see the discussion in Section 4.3.3). For journeys with a non-zero PSD, the mean

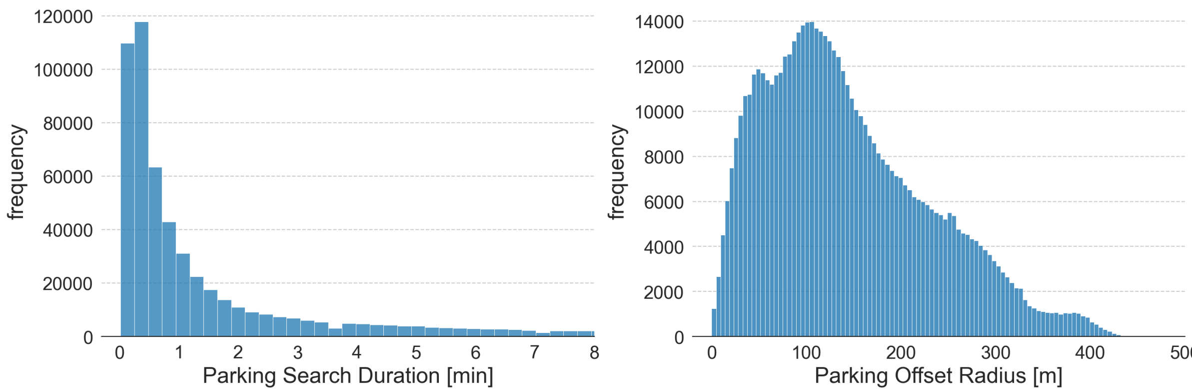


Figure 4.9: Histogram of Predicted Parking Search Duration (left), Histogram of Predicted Parking Offset Radius (right)

Parking Offset Radius stands at 143 m, but this distance stretches to a significant 309 m for the 95th percentile. Peaks observed in Figure 4.9 at the distances 60 and 110 m might indicate varying driver strategies. Table 4.9 reveals that a substantial 50% of drivers embark on starting parking search within a practical range of 76 to 199 m from their final parking spot, which seems to be a plausible range of walking distance for many people.

Furthermore, an examination of the GPS speed data surrounding the parking areas can offer insights into driving behaviors. When we study the average speed of GPS points within a 1 km radius of the parking spot, a marked decrease from 31 km/h to 17 km/h is evident, as illustrated in Table 4.10. This slowing down indicates the transition from regular driving to a more cautious parking search mode and is almost equal to the values in the training ground truth data (Table 4.3).

Using spatial analysis, we can gain new insights into Frankfurt's urban mobility trends. Figure 4.10 presents a heatmap showing the MPSD for street segments based on parking spots found in them in the city center. This visual representation begins with mapping street segments from the Open Street Map, followed by associating nearby parking spots with these segments. The MPSD is then calculated as an average of the parking search durations for trips ending at each street segment. This representation, enhanced and smoothed using the spatial lag technique (Rey et al., 2023), offers a detailed view of the parking landscape. By building a weighted average of the values of all the neighboring street segments, the spatial lag ensures that each street's MPSD value reflects not just its isolated scenario but also the influence of its surroundings. This approach recognizes that drivers usually search for parking over a larger area, not just a single street segment.

For instance, the "Innenstadt" (city center) and "Altstadt" (old town) areas in Frankfurt show notably high MPSD values despite the presence of multiple parking facilities. These regions are central shopping, dining, and other leisure activity zones in Frankfurt, which can explain the heightened parking demand. Encouraging better use of existing parking facilities here might ease the on-street parking situation.

Table 4.9: Descriptive Statistics of Predicted Parking Search Duration and Parking Offset Radius (Duration in MM:SS format and Distance in Meters)

	count	mean	std	5%	25%	median	75%	95%
Parking Search Duration (All Trips)	868,561	1:30	3:16	0:00	0:00	0:15	1:55	8:20
Parking Search Duration (Trips with PSD>0)	582,967	2:15	3:46	0:00	0:20	0:42	3:30	10:30
Parking Offset Radius (Trips with PSD>0)	582,967	143	87	63	80	127	190	309

Table 4.10: Descriptive Statistics of Speed of GPS points within 1km of the Parking Spot during Normal Driving and Parking Search (Speed in km/h)

	count	mean	std	5%	25%	median	75%	95%
All Points	32,156,859	27	21	2	11	24	39	63
Normal Driving	23,324,499	31	21	3	15	29	43	69
Parking Search	8,832,360	17	14	1	5	14	26	43

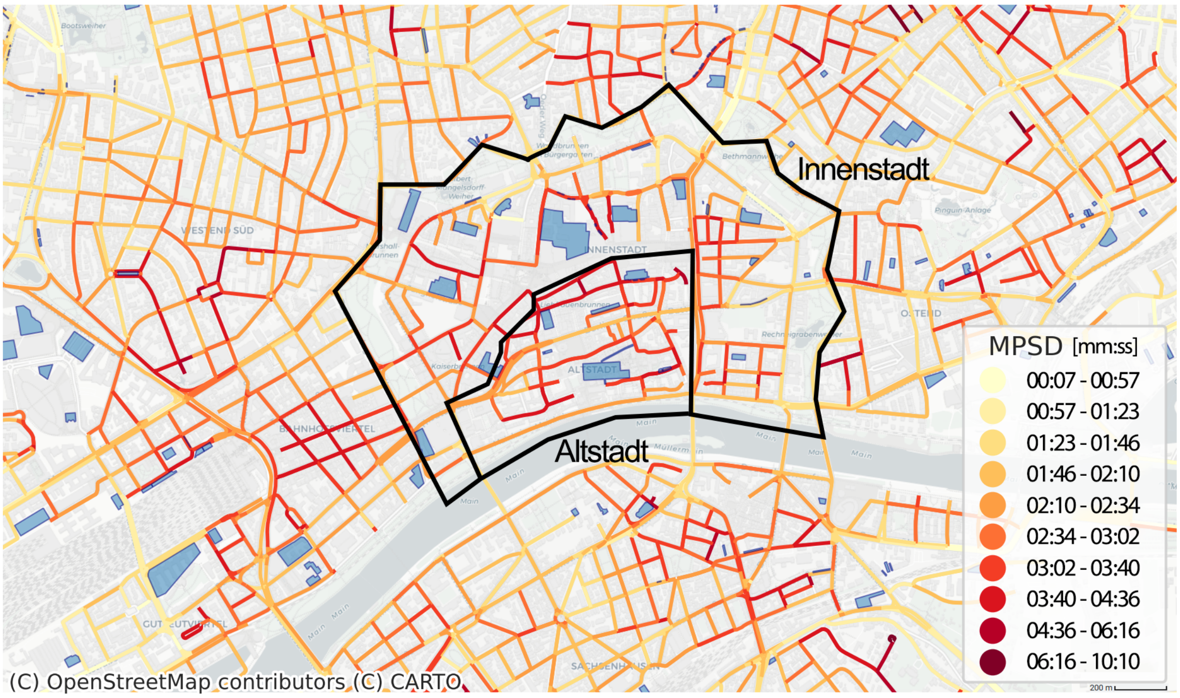


Figure 4.10: Spatial Heatmap of Mean Parking Search Duration Showing the Frankfurt City Center, Parking Garages are Shown in Blue

4.7 Conclusion

In this study, we conducted an in-depth analysis of parking search prediction, a topic with profound implications for urban planning, traffic management, and transportation science. We proposed a novel machine learning model to identify parking search patterns in GPS data, marking the first time such a model has been trained and validated using two sources of ground truth data. Subsequently, our model's performance was benchmarked against existing models from the literature, demonstrating superior performance. Finally, it is applied to a comprehensive dataset, producing aggregated statistics and insights into the parking search patterns of Frankfurt.

To facilitate this, we collected 3550 GPS trajectories using a specially designed smartphone app. This app captured the exact time and location of the journey's origin, the onset of the parking search, the parking spot, and the final destination, all based on the driver's input. Using this ground truth data, we trained a deep learning neural network model that considered speed, sampling rate, and distance to the parking spot to pinpoint the parking search route within a trajectory. Through a 10-fold cross-validation, our model achieved an MAE for parking search duration (PSD) of under one minute, with a negligible discrepancy between actual and predicted mean parking search durations (MPSDs) in test sets.

Our model is further evaluated using a dynamic park-and-visit dataset, comprising 161 journeys collected by different drivers and a different application. The predictions mirrored our earlier results, both at the aggregated level for MPSD and at the individual level for MAE of PSD.

We explored distinct approaches to identifying parking search in GPS data. The Naive 200m Radius approach made a straightforward assumption that the parking search commences within a 200m radius of the eventual parking spot. The Speed Threshold method identified the commencement of the parking search based on a specific drop in vehicle speed. The First Local Minima method likely employed a strategy of pinpointing local minimum values of distance to the parking spot indicative of parking searches. The last and arguably the most sophisticated approach was grounded in our machine learning model, aiming to harness the power of data and algorithms to predict PSDs.

Our comparative analysis clearly showcased the strengths and limitations of each model. Among them, our proposed machine learning model outperformed the other methodologies, underscoring the potential of machine learning in addressing complex transportation challenges. However, it is essential to recognize that while machine learning offers promising results, its effectiveness is intertwined with the quality and quantity of the data it is trained on.

Applying our model to a large GPS dataset, which included over 860,000 trajectories ending in Frankfurt in 2019, yielded more insights. Approximately 33% of these journeys had a zero PSD, while the remaining had an MPSD of 2 minutes and 15 seconds. The overall MPSD was 1 minute and 30 seconds, with a median PSD of 15 seconds. Furthermore, spatial and temporal analyses identified areas with high MPSD, as visualized in heatmaps. Such areas could benefit from urban interventions like augmented parking infrastructure or improved parking and traffic management. The insights gleaned from these representations can also guide the establishment of new parking zones, potential pricing adjustments, and even the redirection of traffic flow.

Our parking search prediction model has broader applications. It can be integrated into navigation apps to detect the onset of parking search in real-time, offering drivers immediate updates on parking conditions and alternate routes. In addition, most navigation apps overlook the effect of parking search when estimating journey durations for private vehicles. By accounting for parking search, users can make more informed comparisons between transportation modes. This might lead to a decline in private car usage in favor of alternatives like cycling or public transport, thereby reducing parking search, overall traffic, and emissions.

In alignment with our commitment to open-source and open data, we have made the parking search prediction model publicly available. It is published as open-source on our GitHub account [github.com/ReLUT/parking-search-prediction]. Furthermore, the dynamic park-and-visit dataset is also shared as open data in the aforementioned repository.

Despite our study's advancements in predicting PSDs, it has limitations. Urban landscapes are constantly evolving, with a steady stream of data, emerging technologies, and new infrastructure. Consequently, prediction models like ours may quickly become obsolete and require retraining with fresh ground truth data, a process that can be effortful. Moreover, data bias arising from user forgetfulness and the potential behavioral changes due to the monitoring effect (also known as the Hawthorne effect) present additional limitations. These factors can skew the data, impacting the model's performance. In future research, we aim to integrate our parking search prediction model into navigation apps or in-car systems, offering drivers real-time estimates and transport mode suggestions. Evaluating the impact of such a model within parking information systems will shed light on its potential to benefit individuals and reduce traffic.

5 Conclusion

5.1 Summary

This dissertation aims to develop a comprehensive empirical framework for understanding parking search behavior, addressing the main research question: *How can parking search behaviors be measured and analyzed empirically to accurately capture and explore search durations, search starting points, and search routes?* By leveraging data collection methods and advanced analytical models, the research overcomes limitations in existing studies that rely on heuristic assumptions or incomplete data. The work is structured around three interrelated studies, each contributing to the overarching theme by focusing on specific aspects of parking search behavior.

Paper I: Initiating the Parking Search

- **Contribution:** Provides a method to empirically measure and analyze the starting point of parking searches, directly contributing to the overarching goal of accurately capturing the initiation phase without relying on assumptions.
- **Results and Key Findings:** The results underscore the significance of proximity to the destination in determining search initiation. Drivers tend to delay the start of their search to minimize walking distance. The lagged speed was also influential; slower speeds increase the likelihood of both searching and parking immediately. Age, area familiarity, vehicle type, and journey purpose further modulate these probabilities, reflecting variations in driver behavior and preferences. Importantly, the study highlights that men are more likely to start their search sooner but less likely to park immediately compared to women, suggesting possible gender-based differences in parking preferences.
- **Conclusions:** This paper establishes a foundational understanding of the factors that influence the initiation of parking search. The unique dataset and model enable an unprecedented empirical analysis of this aspect, showing that driver characteristics, journey purpose, and situational variables like traffic and location significantly affect search initiation timing.

Paper 2: Determinants of Parking Search Duration

- **Contribution:** Offers a detailed empirical analysis of parking search durations and their determinants, enhancing the overall framework by providing insights into the temporal dynamics of parking search behavior.
- **Results and Key Findings:** The study finds a clear positive duration dependency in the search for both Free and Paid parking, indicating that as drivers search longer, they tend to adjust their willingness to accept Paid parking. In contrast, drivers are unlikely

to switch to Illegal parking unless they had already decided against a lengthy search at the outset. Demographic factors, journey purpose, and area familiarity also affect parking choices, with drivers on business trips, for example, more inclined toward Paid parking due to potential reimbursement. Interestingly, high temperatures correlate with higher acceptance of Free parking, likely due to increased willingness to walk.

- **Conclusions:** The analysis shows that drivers' flexibility in their parking choices grows with longer search durations. This reflects a trade-off where drivers gradually shift from seeking Free to Paid parking, depending on situational and demographic factors. The findings suggest that cities could potentially alleviate parking searches by adjusting Paid parking availability in areas with high demand for Free parking.

Paper 3: Predicting Parking Search in GPS Data

- **Contribution:** Enables the application of the empirical framework to extensive historical GPS datasets, allowing for city-wide analysis of parking search behaviors and overcoming limitations of small sample sizes.
- **Results and Key Findings:** The model demonstrates superior performance over existing methods, such as fixed-radius and speed-threshold approaches. By applying this model to a large-scale dataset from INRIX, covering over 860,000 journeys in Frankfurt, the study reveals that approximately one-third of journeys involve immediate parking. For the remaining journeys, the average parking search duration aligns closely with real-world experiences, highlighting differences in search behavior by time of day and location.
- **Conclusions:** This study offers a scalable method for parking search detection that is adaptable to various urban settings and datasets. The model's accuracy and reliability position it as a valuable tool for urban planners and policymakers, who can leverage the insights to optimize parking infrastructure and mitigate urban congestion.

Each paper contributes uniquely to the overarching theme of parking search behavior and its implications for urban mobility. The first two papers delve into decision-making frameworks, with the first focusing on initiation and the second on duration, both central to understanding parking search behavior. The third paper complements these findings by providing a practical tool for large-scale application. The combined results from these studies provide a comprehensive view of parking search behavior, bridging theoretical insights and practical implications. The dissertation's findings suggest that parking search behavior is not only a matter of driver choice but is also shaped by urban context and available technology. This research underscores the potential for data-driven approaches to optimize parking management, which could ultimately reduce the negative impacts of cruising for parking on city congestion and improve urban mobility.

5.2 Implications for the Scientific Context

This dissertation contributes to the scientific discussion on parking search behavior by advancing methodologies and providing empirical evidence in transportation research.

- **Advancing Data Collection Methodologies:** By developing data collection methods utilizing GPS technology and a custom-designed smartphone application, the research captures precise, real-time data on driver behavior throughout entire journeys. This approach overcomes the limitations of previous studies that relied on surveys or assumed start and end points, enhancing data accuracy and completeness. It enables researchers to study parking search dynamics with greater precision and reliability. Furthermore, the introduction of dynamic park-and-visit field experiments in Paper III could establish a new standard for conducting field studies by minimizing assumptions and actively incorporating the driver's decision on the parking search starting point. This approach not only enhances the accuracy of data but also provides a more realistic representation of real-world parking behaviors.
- **Contributions to the Scientific Discussion on Parking Search Behavior:** The findings deepen the understanding of parking search behavior by examining both the initiation and duration of the search process. Through empirical analysis, this research reveals how various factors, such as proximity to the destination, driver characteristics, and contextual elements, shape parking search decisions. This not only adds to the existing body of knowledge but also highlights the complexities of parking search behavior that traditional methods have struggled to capture.
- **Advancing Machine Learning and Econometrics Applications:** The dissertation represents an advancement in applying machine learning and econometrics within transportation research. For example, by developing a deep learning neural network model to identify parking search behavior in GPS trajectory data, it demonstrates how artificial intelligence can address complex problems in urban mobility. Survival analysis with competing risks has also rarely been applied in transportation research.

5.3 Policy Implications

While the primary contributions of this research are scientific—advancing methodologies and providing new insights into parking search behavior—current findings also have the potential to inform policy design and decision-making. However, it is important to note that new policies could not be derived directly from these results. Instead, these empirical insights and analytical tools seem to confirm and support existing policy frameworks already discussed in the literature. Several ways in which the findings are aligned with and reinforce existing policies are outlined below. They can be categorized according to the reasons for parking search, which were presented in Section 1.2.1.

- **Increasing the Price of Parking to Reduce Parking Demand**: Basic economic principles suggest that raising the cost of parking can reduce demand. Previous research (Pierce et al., 2015) demonstrates that dynamic pricing is an effective approach for this. Findings from Paper III reveal significant variations in parking search patterns across different city subareas and times of day, supporting the implementation of dynamic pricing based on current demand in specific locations. Additionally, as noted in the literature, the coexistence of both free and paid parking options can inadvertently increase parking search traffic (Shoup, 2005). This notion is further supported by results from Paper II, which indicate that drivers often choose paid parking only after failing to find a free spot. Eliminating this coexistence—by raising the price from zero to a positive amount in cases of excess demand—may help to reduce overall parking search traffic.

- **Real-Time Intelligent Parking Systems to Address Lack of Information on Availability**: As discussed in Section 1.2.1, parking searches may still occur even when demand equals supply due to a lack of information on parking availability. The literature already suggests that providing drivers with real-time information on available parking can reduce search times and help alleviate congestion (Caicedo, 2010; Teodorović & Lučić, 2006; Shin & Jun, 2014). Findings from Paper II support this, showing that drivers familiar with an area are generally more successful in finding free parking and begin their search earlier. By offering real-time parking information, especially to drivers unfamiliar with high-demand areas, cities can reduce inefficiencies in parking searches. This underscores the importance of policies promoting the development of real-time parking information systems, an intervention strongly supported in urban planning research (Dalla Chiara et al., 2022).
- **Addressing Mismatches Between Parking Supply and Driver Preferences**: Certain groups of people, such as those with disabilities, often rely on parking spaces that offer short walking distances or allow brief stops for loading and unloading. Special regulations already exist for these groups, including designated parking for people with disabilities. In the future, it is conceivable that these dedicated parking spaces could be made "smarter," dynamically adjusting availability based on real-time demand to serve specific groups, the general public, or restrict access altogether when appropriate (Fikri & Hwang, 2019).

Beyond directly addressing the causes of parking search, this thesis may also have indirect policy implications by improving transport models for policy simulations. The findings provide key measurements of parking search behaviors, such as initiation points, search durations, and search radius. Additionally, the proposed prediction model offers a framework to identify these parameters in other unlabeled GPS datasets. These insights can be integrated as parameters in traffic simulation models used by urban planners and policymakers. For instance, incorporating the empirically derived initial search radius into traffic simulations can enhance the realism and predictive accuracy of models assessing the impact of parking policies. Improved simulations can better predict outcomes such as traffic flow changes, parking occupancy rates, and potential congestion, enabling policymakers to assess the effects of proposed interventions before implementation.

5.4 Outlook

As urbanization continues to accelerate, the challenges associated with parking search behavior are likely to become more pronounced. With cities growing denser and car ownership remaining high, the need for efficient parking management systems will intensify. The research presented in this dissertation opens several avenues for future exploration and offers a foundation for addressing parking-related issues in the context of evolving urban mobility trends.

One key direction for future research lies in the integration of real-time data into parking management systems. While this dissertation has focused on empirically understanding parking search behavior through the collection and analysis of ground truth GPS data, the next step is to utilize this knowledge in the development of dynamic parking guidance systems. By integrating real-time parking availability data with predictive models of search behavior, cities could offer drivers more accurate and timely information, potentially reducing the time spent cruising for parking and alleviating congestion. This would not only improve the driver experience but also contribute to environmental sustainability by reducing emissions caused by unnecessary driving.

Additionally, the findings from this dissertation could inform the development of smart cities initiatives, where interconnected urban systems work together to optimize traffic flow, parking, and overall mobility. As autonomous vehicles become more prevalent, parking search behavior may shift dramatically, necessitating new models that account for vehicles capable of independently searching for parking or dropping off passengers before parking themselves. The integration of machine learning and artificial intelligence will be crucial in predicting and managing these shifts in behavior, and the models developed in this research provide a starting point for such applications.

Moreover, the rise of shared mobility and micromobility solutions presents both challenges and opportunities for parking management. With more people opting for ride-sharing, car-sharing, and alternative modes of transportation like scooters or bikes, the demand for traditional parking spaces may change. However, these shifts will require careful monitoring, as new forms of parking demand may emerge, particularly for vehicles involved in logistics and delivery services, which are increasingly occupying curbside spaces. Future research should explore how these changes in mobility patterns interact with parking infrastructure and urban design, ensuring that cities remain adaptable to the evolving needs of their residents.

From a policy perspective, the insights gained from this research can inform the development of more equitable and efficient urban parking strategies. Dynamic pricing, congestion charges, and reallocation of parking spaces can be fine-tuned based on empirical data, leading to better outcomes for drivers, urban planners and society. Additionally, as cities strive to reduce car dependency and promote sustainable transportation, there will be opportunities to explore how parking policies can support these goals, perhaps by encouraging the use of public transit or active modes of transportation through strategic parking reforms.

In conclusion, the outlook for parking search research and its practical applications is promising. The findings presented in this dissertation hopefully contribute to a growing body of knowledge that can help shape the future of urban mobility. As cities continue to evolve, so too must the models and strategies used to manage parking and traffic flow. By building on the empirical insights gained here, future researchers, policymakers, and urban planners may be able to develop more responsive, efficient, and sustainable solutions to the complex challenges of parking in modern cities.

References

Adair, J. G. (1984). The hawthorne effect: A reconsideration of the methodological artifact. *Journal of applied psychology*, *69*(2), 334.

Afrin, T., & Yodo, N. (2020). A survey of road traffic congestion measures towards a sustainable and resilient transportation system. *Sustainability*, *12*(11), 4660.

Alam, M., Moroni, D., Pieri, G., Tampucci, M., Gomes, M., Fonseca, J., Ferreira, J., & Leone, G. R. (2018). Real-time smart parking systems integration in distributed its for smart cities. *Journal of Advanced transportation*, *2018*(1), 1485652.

Albalate, D., & Inci, E. (2018). Introduction to the special issue on the economics of parking. *Journal of Transport Economics and Policy*, *52*(3), 181–183.

Alemi, F., Rodier, C., & Drake, C. (2018). Cruising and on-street parking pricing: A difference-in-difference analysis of measured parking search time and distance in san francisco. *Transportation Research Part A: Policy and Practice*, *111*, 187–198.

Aleskerov, F., Bouyssou, D., & Monjardet, B. (2007). *Utility maximization, choice and preference* (Vol. 16). Springer Science & Business Media.

Alghwiri, A., Wang, S., & Coleman, J. D. (2017). University parking system analysis through discrete event simulation. *IIE Annual Conference. Proceedings*, 2039–2044.

Anderson, S. P., & De Palma, A. (2004). The economics of pricing parking. *Journal of urban economics*, *55*(1), 1–20.

Arnott, R., & Inci, E. (2006). An integrated model of downtown parking and traffic congestion. *Journal of Urban Economics*, *60*(3), 418–442. https://doi.org/10.1016/j.jue.2006.04.004

Arnott, R., & Rowse, J. (2009). Downtown parking in auto city. *Regional Science and Urban Economics*, *39*(1), 1–14. https://doi.org/10.1016/j.regsciurbeco.2008.08.001

Arnott, R., & Rowse, J. (1999). Modeling parking. *Journal of Urban Economics*, *45*(1), 97–124. https://doi.org/10.1006/juec.1998.2084

Arnott, R., & Williams, P. (2017). Cruising for parking around a circle. *Transportation research part B: methodological*, *104*, 357–375.

Arora, N., Cook, J., Kumar, R., Kuznetsov, I., Li, Y., Liang, H.-J., Miller, A., Tomkins, A., Tsogsuren, I., & Wang, Y. (2019). Hard to park? estimating parking difficulty at scale. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2296–2304.

Assemi, B., Baker, D., & Paz, A. (2020). Searching for on-street parking: An empirical investigation of the factors influencing cruising time. *Transport Policy*, *97*, 186–196.

Austin, P. C., Lee, D. S., & Fine, J. P. (2016). Introduction to the analysis of survival data in the presence of competing risks. *Circulation*, *133*(6), 601–609.

Axhausen, K. W., Polak, J. W., & Boltze, M. (1993). Effectiveness of parking guidance and information systems: Recent evidence from nottingham and frankfurt am main. *63rd Annual Meeting: Compendium of Technical Papers*, *37*.

Axhausen, K. W., & Polak, J. W. (1991). Choice of parking: Stated preference approach. *Transportation*, *18*, 59–81.

Axhausen, K. W., Polak, J. W., Boltze, M., & Puzicha, J. (1994). Effectiveness of the parking guidance information system in frankfurt am main. *Traffic engineering and control*. https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/24862/5/l48.pdf

Bebis, G., & Georgiopoulos, M. (1994). Feed-forward neural networks. *Ieee Potentials*, *13*(4), 27–31.

Belloche, S. (2015). On-street parking search time modelling and validation with survey-based data. *Transportation Research Procedia*, *6*, 313–324. https://doi.org/10.1016/j.trpro.2015.03.024

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

Bisante, A., Panizzi, E., & Zeppieri, S. (2023). Cruising-for-parking detection on the smartphone based on implicit interaction and machine learning. *Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 93–102.

Biswas, S., Chandra, S., & Ghosh, I. (2017). Effects of on-street parking in urban context: A critical review. *Transportation in developing economies*, *3*, 1–14.

Bock, F., Di Martino, S., & Sester, M. (2017). Data-driven approaches for smart parking. *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2017, Skopje, Macedonia, September 18–22, 2017, Proceedings, Part III 10*, 358–362.

Bonges, H. A., & Lusk, A. C. (2016). Addressing electric vehicle (ev) sales and range anxiety through parking layout, policy and regulation. *Transportation Research Part A: Policy and Practice*, *83*, 63–73. https://doi.org/https://doi.org/10.1016/j.tra.2015.09.011

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.

Brooke, S., Ison, S., & Quddus, M. (2018). Analysing parking search ('cruising') time using generalised multilevel structural equation modelling. *Journal of Transport Economics and Policy*, *52*(3), 202–220.

Brooke, S., Ison, S., & Quddus, M. (2014). On-street parking search: Review and future research direction. *Transportation Research Record*, *2469*(1), 65–75.

Brownlee, J. (2016). *Machine learning algorithms from scratch with python*. Machine Learning Mastery.

Caicedo, F. (2010). Real-time parking information management to reduce search time, vehicle displacement and emissions. *Transportation Research Part D: Transport and Environment*, *15*(4), 228–234.

Cao, J., & Menendez, M. (2015). System dynamics of urban traffic based on its parking-related-states. *Transportation Research Part B: Methodological*, *81*, 718–736.

Cao, J., Menendez, M., & Waraich, R. (2019). Impacts of the urban parking system on cruising traffic and policy development: The case of zurich downtown area, switzerland. *Transportation*, *46*, 883–908.

Cascetta, E., & Cascetta, E. (2009). Random utility theory. *Transportation systems analysis: models and applications*, 89–167.

Chester, M., Horvath, A., & Madanat, S. (2011). Parking infrastructure and the environment. *Access Magazine*, *1*(39), 28–33.

Cirillo, C., & Xu, R. (2011). Dynamic discrete choice models for transportation. *Transport Reviews*, *31*(4), 473–494.

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part i: Basic concepts and first analyses. *British journal of cancer*, *89*(2), 232–238.

Cookson, G., & Pishue, B. (2017). Die folgen der parkplatzproblematik in den vereinigten staaten, großbritannien und deutschland.

Costa, Á., Rocha, C., & Melo, S. (2014). Parking management policies and the effectiveness of public policy solutions [Transportation: Can we do more with less resources? – 16th Meeting of the Euro Working Group on Transportation – Porto 2013]. *Procedia - Social and Behavioral Sciences*, *111*, 965–973. https://doi.org/https://doi.org/10.1016/j.sbspro.2014.01.131

Dalla Chiara, G., Cheah, L., Azevedo, C. L., & Ben-Akiva, M. E. (2020). A policy-sensitive model of parking choice for commercial vehicles in urban areas. *Transportation Science*, *54*(3), 606–630. https://doi.org/10.1287/trsc.2019.0970

Dalla Chiara, G., & Goodchild, A. (2020). Do commercial vehicles cruise for parking? empirical evidence from seattle. *Transport Policy*, *97*, 26–36.

Dalla Chiara, G., Krutein, K. F., Ranjbari, A., & Goodchild, A. (2022). Providing curb availability information to delivery drivers reduces cruising for parking. *Scientific reports*, *12*(1), 19355.

Dalla Chiara, G., Krutein, K. F., Ranjbari, A., & Goodchild, A. (2021). Understanding urban commercial vehicle driver behaviors and decision making. *Transportation research record*, *2675*(9), 608–619.

Dell'Orco, M., & Teodorović, D. (2005). Multi agent systems approach to parking facilities management. In H.-J. Zimmermann, N. O. Attoh-Okine, & B. M. Ayyub (Eds.), *Applied research in uncertainty modeling and analysis* (pp. 321–339, Vol. 20). Springer US. https://doi.org/10.1007/0-387-23550-7{\textunderscore}15

Dieussaert, K., Aerts, K., Steenberghen, T., Maerivoet, S., & Spitaels, K. (2009). Sustapark: An agent-based model for simulating parking search. *AGILE International Conference on Geographic Information Science, Hannover*, 1–11.

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, *113*(10), 2636–2641.

Ertekin, M., & Önder Efe, M. (2021). Autonomous parking with continuous reinforcement learning. *The International Conference on Artificial Intelligence and Applied Mathematics in Engineering*, 25–37.

Fahim, A., Hasan, M., & Chowdhury, M. A. (2021). Smart parking systems: Comprehensive review based on various aspects. *Heliyon*, *7*(5), e07050. https://doi.org/10.1016/j.heliyon.2021.e07050

Feng, T., & Timmermans, H. J. (2013). Transportation mode recognition using gps and accelerometer data. *Transportation Research Part C: Emerging Technologies*, *37*, 118–130.

Fikri, R. M., & Hwang, M. (2019). Smart parking area management system for the disabled using iot and mobile application. *2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*, 172–176.

Fulman, N., & Benenson, I. (2018). Agent-based modeling for transportation planning: A method for estimating parking search time based on demand and supply. *ATT@ IJCAI*, 31–39.

Fulman, N., & Benenson, I. (2021). Approximation method for estimating search times for on-street parking. *Transportation science*, *55*(5), 1046–1069.

Fulman, N., Benenson, I., & Ben-Elia, E. (2020). Modeling parking search behavior in the city center: A game-based approach. *Transportation Research Part C: Emerging Technologies*, *120*, 102800.

Fushiki, T. (2011). Estimation of prediction error by using k-fold cross-validation. *Statistics and Computing*, *21*, 137–146.

Gallo, M., D'Acierno, L., & Montella, B. (2011). A multilayer model to simulate cruising for parking in urban areas. *Transport Policy*, *18*(5), 735–744. https://doi.org/https://doi.org/10.1016/j.tranpol.2011.01.009

Geva, S., Fulman, N., & Ben-Elia, E. (2022). Getting the prices right: Drivers' cruising choices in a serious parking game. *Transportation Research Part A: Policy and Practice*, *165*, 54–75.

Giffinger, R. (2021). Smart city: The importance of innovation and planning. In M. Helfert, C. Klein, B. Donnellan, & O. Gusikhin (Eds.), *Smart cities, green technologies and intelligent transport systems* (pp. 28–39, Vol. 1217). Springer International Publishing. https://doi.org/10.1007/978-3-030-68028-2{\textunderscore}2

Glazer, A., & Niskanen, E. (1992). Parking fees and congestion. *Regional Science and Urban Economics*, *22*(1), 123–132.

Gössling, S., Schröder, M., Späth, P., & Freytag, T. (2016). Urban space distribution and sustainable transport. *Transport Reviews*, *36*(5), 659–679.

Gu, Z., Najmi, A., Saberi, M., Liu, W., & Rashidi, T. H. (2020). Macroscopic parking dynamics modeling and optimal real-time pricing considering cruising-for-parking. *Transportation Research Part C: Emerging Technologies*, *118*, 102714.

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.

Hampshire, R. C., Jordon, D., Akinbola, O., Richardson, K., Weinberger, R., Millard-Ball, A., & Karlin-Resnik, J. (2016). Analysis of parking search behavior with video from naturalistic driving. *Transportation Research Record: Journal of the Transportation Research Board*, *2543*(1), 152–158. https://doi.org/10.3141/2543-18

Harmatuck, D. J. (2007). Revealed parking choices and the value of time. *Transportation Research Record*, *2010*(1), 26–34.

Hess, S., & Polak, J. W. (2004). An analysis of parking behaviour using discrete choice models calibrated on sp datasets.

Hess, W., & Persson, M. (2012). The duration of trade revisited: Continuous-time versus discrete-time hazards. *Empirical Economics*, *43*, 1083–1107.

Hilvert, O., Toledo, T., & Bekhor, S. (2012). Framework and model for parking decisions. *Transportation research record*, *2319*(1), 30–38.

Holguín-Veras, J., et al. (2016). Impacts of freight parking policies in urban areas: The case of new york city. *City University of New York*.

Horni, A., Montini, L., Waraich, R. A., & Axhausen, K. W. (2013). An agent-based cellular automaton cruising-for-parking simulation. *Transportation Letters*, *5*(4), 167–175.

Hunt, J. D., & Teply, S. (1993). A nested logit model of parking location choice. *Transportation Research Part B: Methodological*, *27*(4), 253–265.

Ibeas, A., Dell'Olio, L., Bordagaray, M., & Ortúzar, J. d. D. (2014). Modelling parking choices considering user heterogeneity. *Transportation Research Part A: Policy and Practice*, *70*, 41–49.

Inci, E., van Ommeren, J., & Kobus, M. (2017). The external cruising costs of parking. *Journal of Economic Geography*, *17*(6), 1301–1323. https://doi.org/10.1093/jeg/lbx004

Janidarmian, M., Roshan Fekr, A., Radecka, K., & Zilic, Z. (2017). A comprehensive analysis on wearable acceleration sensors in human activity recognition. *Sensors*, *17*(3), 529.

Janitza, S., & Tutz, G. (2015). Prediction models for time discrete competing risks. *University of Munich*.

Jenkins, S. P. (2005). Survival analysis. *Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK*, *42*, 54–56.

Jha, M. K., Schonfeld, P., & McCullough, F. (2023). A machine learning and simulation-based dynamic parking choice model for airports. *Journal of Air Transport Management*, *111*, 102425. https://doi.org/https://doi.org/10.1016/j.jairtraman.2023.102425

Jones, M., Khan, A., Kulkarni, P., Carnelli, P., & Sooriyabandara, M. (2017). Parkus 2.0: Automated cruise detection for parking availability inference. *Proceedings of the 14th EAI international conference on mobile and ubiquitous systems: computing, networking and services*, 242–251.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260.

Kaplan, S., & Bekhor, S. (Eds.). (2011). *Exploring en-route parking type and parking-search route choice: Decision making framework and survey design*.

Karlaftis, M. G., & Vlahogianni, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, *19*(3), 387–399.

Khaliq, A., van der Waerden, P., & Janssens, D. (2018). Modeling car drivers' on-street parking decisions using the integrated hierarchical information integration approach. *Transportation Research Record: Journal of the Transportation Research Board*, *2672*(49), 23–33. https://doi.org/10.1177/0361198118776524

Kim, S. (2014). *The comparison of discrete and continuous survival analysis* [Doctoral dissertation, Virginia Polytechnic Institute and State University].

Knese, D., Künbet, S., Busse, J., Dehler, K., & Wong, C. (2024). Abschlussbericht oeder weg: Wissenschaftliche begleitung der" fahrradfreundlichen nebenstraßen" in frankfurt am main.

Kobus, M. B., Gutiérrez-i-Puigarnau, E., Rietveld, P., & Van Ommeren, J. N. (2013). The on-street parking premium and car drivers' choice between street and garage parking. *Regional Science and Urban Economics*, *43*(2), 395–403.

Kohavi, R., et al. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*, *14*, 1137–1145.

Krapivsky, P. L., & Redner, S. (2019). Simple parking strategies. *Journal of Statistical Mechanics: Theory and Experiment*, *2019*(9), 093404. https://doi.org/10.1088/1742-5468/ab3a2a

Kwak, C., & Clayton-Matthews, A. (2002). Multinomial logistic regression. *Nursing research*, *51*(6), 404–410.

Leclercq, L., Sénécat, A., & Mariotte, G. (2017). Dynamic macroscopic simulation of on-street parking search: A trip-based approach. *Transportation Research Part B: Methodological*, *101*, 268–282.

Lee, J., Agdas, D., & Baker, D. (2017). Cruising for parking: New empirical evidence and influential factors on cruising time. *Journal of Transport and Land Use*, *10*(1), 931–943.

Lee, M., Feuer, E. J., & Fine, J. P. (2018). On the analysis of discrete time competing risks data. *Biometrics*, *74*(4), 1468–1481.

Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R. J., & Wasserman, L. (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, *113*(523), 1094–1111.

Levy, N., Martens, K., & Benenson, I. (2013). Exploring cruising using agent-based and analytical models of parking. *Transportmetrica A: Transport Science*, *9*(9), 773–797.

Litman, T. (2016). *Parking management: Strategies, evaluation and planning.* Victoria Transport Policy Institute Victoria, BC, Canada.

Liu, W., & Geroliminis, N. (2016). Modeling the morning commute for urban networks with cruising-for-parking: An mfd approach. *Transportation Research Part B: Methodological*, *93*, 470–494.

Mannini, L., Cipriani, E., Crisalli, U., Gemma, A., & Vaccaro, G. (2017). On-street parking search time estimation using fcd data. *Transportation Research Procedia*, *27*, 929–936. https://doi.org/10.1016/j.trpro.2017.12.149

Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, *8*(3), 229.

Mantouka, E. G., Fafoutellis, P., & Vlahogianni, E. I. (2021). Deep survival analysis of searching for on-street parking in urban areas. *Transportation Research Part C: Emerging Technologies*, *128*, 103173. https://doi.org/10.1016/j.trc.2021.103173

Manville, M., & Pinski, M. (2021). The causes and consequences of curb parking management. *Transportation Research Part A: Policy and Practice*, *152*, 295–307.

Marsden, G. (2006). The evidence base for parking policies—a review [Parking]. *Transport Policy*, *13*(6), 447–457. https://doi.org/https://doi.org/10.1016/j.tranpol.2006.05.009

Martens, K., & Benenson, I. (2008). Evaluating urban parking policies with agent-based model of driver parking behavior. *Transportation Research Record: Journal of the Transportation Research Board*, *2046*(1), 37–44. https://doi.org/10.3141/2046-05

Martens, K., Benenson, I., & Levy, N. (2010). The dilemma of on-street parking policy: Exploring cruising for parking using an agent-based model. In B. Jiang & X. Yao (Eds.), *Geospatial analysis and modelling of urban structure and dynamics* (pp. 121–138, Vol. 99). Springer Netherlands. https://doi.org/10.1007/978-90-481-8572-6{\textunderscore}7

McCahill, C. (2017). Factors affecting residential parking occupancy in madison, wisconsin. *Transportation Research Record*, *2651*(1), 71–79.

McCarney, R., Warner, J., Iliffe, S., Van Haselen, R., Griffin, M., & Fisher, P. (2007). The hawthorne effect: A randomised, controlled trial. *BMC medical research methodology*, *7*, 1–8.

Medsker, L. R., Jain, L., et al. (2001). Recurrent neural networks. *Design and Applications*, *5*(64-67), 2.

Mei, Z., Feng, C., Kong, L., Zhang, L., & Chen, J. (2020). Assessment of different parking pricing strategies: A simulation-based analysis. *Sustainability*, *12*(5), 2056.

Melnyk, P., Djahel, S., & Nait-Abdesselam, F. (2019). Towards a smart parking management system for smart cities. *2019 IEEE international smart cities conference (ISC2)*, 542–546.

Milia, K., Hedengran, M. D., Jansson, T., Rodrigues, F., & Azevedo, C. L. (2023). Forecasting parking search times using big data. *Proceedings of the 12th International Scientific Conference on Mobility and Transport: Mobility Innovations for Growing Megacities*, 175–197.

Millard-Ball, A., Hampshire, R. C., & Weinberger, R. (2020). Parking behaviour: The curious lack of cruising for parking in san francisco. *Land Use Policy*, *91*, 103918. https://doi.org/10.1016/j.landusepol.2019.03.031

Millard-Ball, A., Weinberger, R. R., & Hampshire, R. C. (2014). Is the curb 80% full or 20% empty? assessing the impacts of san francisco's parking pricing experiment. *Transportation Research Part A: Policy and Practice*, *63*, 76–92.

Montini, L., Horni, A., Rieser-Schüssler, N., & Axhausen, K. W. (2012). Searching for parking in gps data. https://doi.org/10.3929/ETHZ-A-007319185

Muleev, Y. (2020). Exploring associations between parking occupancy rate at residential estates and spatial characteristics. the case of ekaterinburg. *The Case of Ekaterinburg (February 28, 2020). Higher School of Economics Research Paper No. WP BRP*, *9*.

Ogulenko, A., Benenson, I., & Fulman, N. (2022). The nature of the on-street parking search. *Transportation Research Part B: Methodological*, *166*, 48–68.

Ojha, V. K., Abraham, A., & Snášel, V. (2017). Metaheuristic design of feedforward neural networks: A review of two decades of research. *Engineering Applications of Artificial Intelligence*, *60*, 97–116.

Ottomanelli, M., Dell'Orco, M., & Sassanelli, D. (2011). Modelling parking choice behaviour using possibility theory. *Transportation Planning and Technology*, *34*(7), 647–667.

Pierce, G., Willson, H., & Shoup, D. (2015). Optimizing the use of public garages: Pricing parking by demand. *Transport Policy*, *44*, 89–95.

Polak, J., & Axhausen, K. (1992). Possibilities and requirements for including the effects of parking in a strategic transport model for london. *A Report to Departmant of Transport*.

Polak, J., & Axhausen, K. W. (1990). Parking search behaviour: A review of current research and future prospects. *TSU Ref*, *540*.

Ponnambalam, C. T., & Donmez, B. (2020). Searching for street parking: Effects on driver vehicle control, workload, physiology, and glances. *Frontiers in psychology*, *11*, 574262.

Qin, H., Gao, J., Zhang, G., Chen, Y., & Wu, S. (2017). Nested logit model formation to analyze airport parking behavior based on stated preference survey studies. *Journal of Air Transport Management*, *58*, 164–175.

Qin, H., Yang, X., Wu, Y.-J., Guan, H., Wang, P., & Shahinpoor, N. (2020). Analysis of parking cruising behaviour and parking location choice. *Transportation Planning and Technology*, *43*(7), 717–734.

RegioStaR Creators. (2021). Regiostar: Regional statistical spatial typology for mobility and transport research. *BMDV*.

Rey, S., Arribas-Bel, D., & Wolf, L. J. (2023). *Geographic data science with python*. CRC Press.

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.

Rong, Y., Xu, Z., Yan, R., & Ma, X. (2018). Du-parking: Spatio-temporal big data tells you realtime parking availability. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 646–654.

Ross, T., & Buliung, R. (2019). Access work: Experiences of parking at school for families living with childhood disability. *Transportation Research Part A: Policy and Practice*, *130*, 289–299. https://doi.org/https://doi.org/10.1016/j.tra.2019.08.016

Saharan, S., Kumar, N., & Bawa, S. (2020). An efficient smart parking pricing system for smart city environment: A machine-learning based approach. *Future Generation Computer Systems*, *106*, 622–640.

Saki, S., & Hagen, T. (2024a). Cruising for parking again: Measuring the ground truth and using survival analysis to reveal the determinants of the duration. *Transportation Research Part A: Policy and Practice*, *183*, 104045.

Saki, S., & Hagen, T. (2024b). Parking search identification in vehicle gps traces. *Journal of Urban Mobility*, *6*, 100083.

Saki, S., & Hagen, T. (2024c). What drives drivers to start cruising for parking? modeling the start of the search process. *Transportation Research Part B: Methodological*, *188*, 103058.

Sándor, Z. P., & Csiszár, C. (2015). Role of integrated parking information system in traffic management. *Periodica Polytechnica Civil Engineering*, *59*(3), 327–336.

Satagopan, J., Ben-Porat, L., Berwick, M., Robson, M., Kutler, D., & Auerbach, A. (2004). A note on competing risks in survival data analysis. *British journal of cancer*, *91*(7), 1229–1235.

Sattayhatewa, P., & Smith Jr, R. L. (2003). Development of parking choice models for special events. *Transportation research record*, *1858*(1), 31–38.

Schäfer, P. K., Hagen, T., & Saki, S. (2024). Beeinflussung des parksuchverkehrs durch parkraummanagement. *MobilityAgenda*, *4*(4), 42.

Schmid, M., & Berger, M. (2021). Competing risks analysis for discrete time-to-event data. *Wiley Interdisciplinary Reviews: Computational Statistics*, *13*(5), e1529.

Schuster, N. A., Hoogendijk, E. O., Kok, A. A., Twisk, J. W., & Heymans, M. W. (2020). Ignoring competing events in the analysis of survival data may lead to biased results: A nonmathematical illustration of competing risk analysis. *Journal of clinical epidemiology*, *122*, 42–48.

Shaheen, S. A., Cohen, A. P., & Martin, E. (2010). Carsharing parking policy: Review of north american practices and san francisco, california, bay area case study. *Transportation Research Record*, *2187*(1), 146–156. https://doi.org/10.3141/2187-19

Shiftan, Y., & Burd-Eden, R. (2001). Modeling response to parking policy. *Transportation Research Record*, *1765*(1), 27–34.

Shin, J.-H., & Jun, H.-B. (2014). A study on smart parking guidance algorithm. *Transportation Research Part C: Emerging Technologies*, *44*, 299–317.

Shoup, D. (2005). *High cost of free parking* (1st). Routledge. https://doi.org/10.4324/9781351179539

Shoup, D. (2021). Pricing curb parking. *Transportation Research Part A: Policy and Practice*, *154*, 399–412. https://doi.org/10.1016/j.tra.2021.04.012

Shoup, D. C. (2006). Cruising for parking. *Transport policy*, *13*(6), 479–486.

Simićević, J., Vukanović, S., & Milosavljević, N. (2013). The effect of parking charges and time limit to car usage and parking behaviour. *Transport Policy*, *30*, 125–131.

Sisiopiku, V. P. (2001). On-street parking on state roads. *ITE Annual Meeting Compendium, Washington DC, USA*.

Soto, J. J., Márquez, L., & Macea, L. F. (2018). Accounting for attitudes on parking choice: An integrated choice and latent variable approach. *Transportation Research Part A: Policy and Practice*, *111*, 65–77.

Soto Ferrari, M., Romero Rodriguez, D., Escorcia Caballero, J. P., Daza Escorcia, J. M., & Chams-Anturi, O. (2021). A discrete-event simulation approach to assess the benefits of parking technologies in hospitals. *International Journal of Operations and Quantitative Management*, *27*(2), 141–152.

Start2park research project [Accessed: 2024-06-15]. (2024). https://www.frankfurt-university.de/de/hochschule/fachbereich-1/ffin/fachgruppen-des-ffin/relut/forschungsprojekte/start2park/

StataCorp. (2023). *Stata 18 longitudinal data / panel data manual*. Stata Press.

Steinmeyer, I., Bäumer, M., Fahnberg, C., Hahn, W., Kagerbauer, M., Kathmann, T., Krause, J., Leerkamp, B., Ließke, F., Mayer-Kreitz, M., Moik, P., Pohl, J., Preising, W., Sommer, C., & von Zadel, E. (2011). *Empfehlungen für verkehrserhebungen* (2011th ed.) [Arbeitsausschuss: Erhebung und Prognose des Verkehrs]. Forschungsgesellschaft für Straßen- und Verkehrswesen, Arbeitsgruppe Verkehrsplanung. Germany.

Stoyanov, D., Taylor, Z., Kia, S. M., Oguz, I., Reyes, M., Martel, A., Maier-Hein, L., Marquand, A. F., Duchesnay, E., Löfstedt, T., et al. (2018). *Understanding and interpreting machine learning in medical image computing applications*. Springer.

Surpris, G., Liu, D., & Vincenzi, D. (2013). Evaluating the effect of smart parking technology on campus parking system efficiency using discrete event simulation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 1948–1952.

Tamari, M. (1982). An optimal parking problem. *Journal of Applied Probability*, *19*(4), 803–814.

Tchervenkov, C. (2022). *Empirical and simulation studies on parking in switzerland* [PhD thesis]. ETH Zurich [Doctor of Sciences ETH Zurich (Dr. sc. ETH Zurich)]. https://doi.org/10.3929/ethz-a

Teodorović, D., & Lučić, P. (2006). Intelligent parking systems. *European Journal of Operational Research*, *175*(3), 1666–1681.

Thompson, R. G., & Bonsall, P. (1997). Drivers' response to parking guidance and information systems. *Transport Reviews*, *17*(2), 89–104.

Thompson, R. G., & Richardson, A. J. (1998). A parking search model. *Transportation Research Part A: Policy and Practice*, *32*(3), 159–170. https://doi.org/10.1016/S0965-8564(97)00005-0

Townsend, Z., Buckley, J., Harada, M., & Scott, M. A. (2013). The choice between fixed and random effects. *The SAGE handbook of multilevel modeling*, 73–88.

Tully, K., & Bolshakov, V. Y. (2010). Emotional enhancement of memory: How norepinephrine enables synaptic plasticity. *Molecular brain*, *3*, 1–9.

Tutz, G., Schmid, M., et al. (2016). *Modeling discrete time-to-event data*. Springer.

Valhalla Development Team. (2023). Valhalla [Available online, accessed in 2023].

Van Der Goot, D. (1982). A model to describe the choice of parking places. *Transportation Research Part A: General*, *16*(2), 109–115.

van der Waerden, P., Timmermans, H., & van Hove, L. (2015). Gps data and car drivers' parking search behavior in the city of turnhout, belgium. In I. Ivan, I. Benenson, B. Jiang, J. Horák, J. Haworth, & T. Inspektor (Eds.), *Geoinformatics for intelligent transportation* (pp. 247–256). Springer International Publishing. https://doi.org/10.1007/978-3-319-11463-7{\textunderscore}18

van der Waerden, P. (2012). Pamela, a parking analysis model for predicting effects in local areas. https://doi.org/10.6100/IR739991 van Ommeren, J., McIvor, M., Mulalic, I., & Inci, E. (2021). A novel methodology to estimate cruising for parking and related external costs. *Transportation Research Part B: Methodological*, *145*, 247–269. https://doi.org/10.1016/j.trb.2020.12.005

Van Ommeren, J., Wentink, D., & Dekkers, J. (2011). The real price of parking policy. *Journal of Urban Economics*, *70*(1), 25–31.

van Ommeren, J. N., Wentink, D., & Rietveld, P. (2012). Empirical evidence on cruising for parking: Transportation research part a: Policy and practice, 46(1), 123-130. *Transportation Research Part A: Policy and Practice*, *46*(1), 123–130. https://doi.org/10.1016/J.TRA.2011.09.011

Verhoef, E., Nijkamp, P., & Rietveld, P. (1995). The economics of regulatory parking policies: The (im) possibilities of parking policies in traffic regulation. *Transportation research part A: policy and practice*, *29*(2), 141–156.

Vlahogianni, E. I., Kepaptsoglou, K., Tsetsos, V., & Karlaftis, M. G. (2016). A real-time parking prediction system for smart cities. *Journal of Intelligent Transportation Systems*, *20*(2), 192–204.

Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2018). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. In *Self-regulation and self-control* (pp. 45–77). Routledge.

Waraich, R. A., & Axhausen, K. W. (2012). Agent-based parking choice model. *Transportation Research Record*, *2319*(1), 39–46.

Wardman, M., Chintakayala, V. P. K., & de Jong, G. (2016). Values of travel time in europe: Review and meta-analysis. *Transportation Research Part A: Policy and Practice*, *94*, 93–111.

Weinberger, R. R., Millard-Ball, A., & Hampshire, R. C. (2020). Parking search caused congestion: Where's all the fuss? *Transportation Research Part C: Emerging Technologies*, *120*. https://doi.org/10.1016/j.trc.2020.102781.

Weis, C., Kowald, M., Danalet, A., Schmid, B., Vrtic, M., Axhausen, K. W., & Mathys, N. (2021). Surveying and analysing mode and route choices in switzerland 2010-2015. *Travel Behaviour and Society*, *22*, 10–21. https://doi.org/10.1016/j.tbs.2020.08.001

Yang, S., Ma, W., Pi, X., & Qian, S. (2019). A deep learning approach to real-time parking occupancy prediction in transportation networks incorporating multiple spatio-temporal data sources. *Transportation Research Part C: Emerging Technologies*, *107*, 248–265.

Young, W., Thompson, R. G., & Taylor, M. A. (1991). A review of urban car parking models. *Transport reviews*, *11*(1), 63–84.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: Lstm cells and network architectures. *Neural computation*, *31*(7), 1235–1270.

Zakharenko, R. (2016). The time dimension of parking economics. *Transportation Research Part B: Methodological*, *91*, 211–228.

Zhao, C., Liao, F., Li, X., & Du, Y. (2021). Macroscopic modeling and dynamic control of on-street cruising-for-parking of autonomous vehicles in a multi-region urban road network. *Transportation Research Part C: Emerging Technologies*, *128*, 103176.

Zhou, D.-m., Xu, J.-c., Radke, J., & Mu, L. (2004). A spatial cluster method supported by gis for urban-suburban-rural classification. *Chinese Geographical Science*, *14*(4), 337–342. https://doi.org/10.1007/s11769-004-0038-7

Zhu, Y., Ye, X., Chen, J., Yan, X., & Wang, T. (2020). Impact of cruising for parking on travel time of traffic flow. *Sustainability*, *12*(8), 3079.