Assessing the impacts of fleet sizing and matching strategies for Shared Automated Vehicles (SAVs) in mixed mobility systems using an agent-based simulation approach

by

Yiming Wang

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on December 2, 2024 at 10:00 AM.

Student number:	5842387	
Project duration:	April 17, 2024 – December 10, 2024	
Master program:	Transport, Infrastructure & Logistics	
Thesis committee:	Bilge Atasoy,	TU Delft
	Irene Martínez Josemaría,	TU Delft
	Javier Alonso-Mora,	TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Report number: 2024.TIL.9007

Abstract

Due to the challenges of traffic congestion, pollution, and low transportation efficiency in urban areas, Shared Autonomous Vehicles (SAVs) are considered a solution that can integrate into existing transportation frameworks to improve traffic performance and environmental sustainability. Currently, there is a research gap in managing the fleet size of SAVs under different matching strategies. Although the literature on the operational and environmental benefits of SAVs is growing, few studies systematically analyze how different fleet management strategies affect service levels and road congestion. This thesis fills this gap by exploring the impacts of Immediate Decision and Batch Offer strategies on fleet size and their corresponding effects on urban traffic performance.

The thesis focuses on the integration of SAV with mixed mobility systems, using SAVs to entirely replace current Mobility on Demand (MoD) services, employing an agent-based simulation method to study the impacts of fleet size and matching strategies on urban traffic performance. This research employs a case study approach, selecting the Haidian District in Beijing as the research area. It uses an innovative demand generation model based on travel distances as simulation request inputs. The model utilizes travel distance distribution functions based on historical data, road network information, and taxi GPS data to generate traffic demand based on network nodes through traffic zone division, probability density calculation, and travel distribution. The model employs the K-means clustering algorithm to divide traffic areas based on the density of traffic nodes rather than specific road network conditions, a macroscopic, low-resolution method with lower computational complexity suitable for large-scale road networks.

The mode choice model in this study analyzes how passengers make decisions between different modes of travel. The model, using a discrete choice framework, studies how passengers choose between SAVs(Shared Automated Vehicles) and POVs(privately operated vehicles). Employing the Logit model, the study considers several key factors influencing travel choice, including travel time, comfort, reliability, and environmental impact, to provide mode selection in simulations. This part emphasizes the importance of travelers' subjective perceptions.

The application of simulation and the Macroscopic Fundamental Diagram (MFD) are central methodologies of this study. In the simulation analysis part, this research uses the agent-based simulation tool FleetPy, focusing on evaluating two key matching strategies: Immediate Decision Simulation (IDS) and Batch Offer Simulation (BOS). By simulating different fleet sizes and matching strategies, it explores the dispatching and operational performance of Shared Autonomous Vehicles in urban road networks and compares them with traditional taxi systems.

The findings indicate that replacing taxi systems with an SAV system can significantly improve fleet efficiency, achieving the same service level with a smaller fleet size. In both small and large fleet scenarios, the SAV system demonstrates higher operational efficiency. In scenarios with smaller fleet sizes, the Batch Offer (BOS) strategy provides higher service levels. By batch processing and optimizing multiple requests, it can effectively improve the success rate of matching and reduce waiting times, while Immediate Decision (IDS) performs better in terms of service timeliness. Simulation results also show that as the fleet size increases, the advantages of the IDS strategy become apparent but may lead to wastage of vehicle resources.

Contents

1 Introduction

Automated vehicles are progressively advancing toward market adoption (Stoma et al., 2021). These vehicles are characterized by the use of driver assistance technologies that minimize or eliminate the need for human intervention. Vehicle automation is classified into six levels, from stage 0, which represents complete manual control without any assistance, to stage 5, where the vehicle is fully autonomous and capable of self-driving under all conditions without human input (Cole, 2024). On 7 March 2024, Apollo Go announced the launch of a 24/7 autonomous driving travel service in parts of Wuhan, China. The service has now been expanded to cover the entire city, spanning over 3,000 square kilometers (ApolloGo, 2024). Robotaxis are increasingly gaining attention in public discourse, bringing with them a slew of controversies, including their competitive low starting price (4 RMB, approximately 0.5 euros), which is undercutting the traditional taxi and ride-hailing markets.

Over the past few years, the landscape of urban transportation has seen a paradigm shift towards shared mobility, notably car sharing, which has seen a global user increase of 65% and a fleet expansion of 55% between 2012 and 2014 (Shaheen et al., 2016). The concept of combining car sharing with autonomous vehicles was proposed as early as the early 1990s (Parent and de La Fortelle, 2005), and autonomous vehicle technology has played a facilitating role in shared mobility services (Thomas and Deepti, 2018). Shared Automated Vehicles (SAVs) combine autonomous driving technology with shared mobility. It blends features of both traditional public and private transport modes (Haboucha et al., 2017), potentially offering a significant alternative to conventional transportation systems.

Currently, transportation emissions account for approximately 10.4% of China's total carbon emissions, with road transportation specifically contributing over 85% to the national transportation carbon emissions, making it the absolute main source and a key focus for emission reduction (Chi, 2022). Introducing SAVs into urban road networks is a potential systemic mitigation measure. SAVs enable users to "hire" vehicles as needed, potentially leading to a decrease in private vehicle ownership and a subsequent reduction in the total fleet size needed in the transportation system(Martin and Shaheen, 2011).

The motivation of this study is to explore the potential of SAVs in urban development. As cities continue to expand and urbanize, the pressure on traditional transportation systems increases, leading to congestion, pollution, and inefficiencies in mobility. The urban mobility landscape is becoming increasingly complex as SAVs and POVs(privately operated vehicles) share the same road networks. The integration of SAVs into urban mobility systems presents an innovative approach to addressing these challenges by optimizing the fleet size and matching stratiges, while also contributing to environmental sustainability.

1.1 Background

Recently, the Mobility on Demand (MOD) concept has gained prominence for its ability to provide consumers with immediate access to transportation, products, and services. This model enhances convenience and flexibility, allowing users to request and receive various services in real time through the implementation of shared transportation modes, courier services, public transit, and other innovative approaches, the sector has seen rapid growth propelled by advancements in technology (Shaheen and Cohen, 2020).

Transportation Network Companies (TNCs) such as Lyft, Uber and Didi have been at the forefront of revolutionizing travel by offering convenient, on-demand transportation services via a mobile application. Unlike traditional public transport services, which operate on fixed schedules and routes, MOD systems offer flexibility by accommodating user requests for pick-up, drop-off, and timing on an ad hoc basis.

The emergence of autonomous driving technology, combined with innovative ride-purchasing business models, has spurred significant interest in Automated Mobility on Demand (AMOD). AMOD services offer distinct benefits over conventional Mobility on Demand (MOD) systems, such as enhancing road safety by reducing the likelihood of accidents and decreasing operational expenses, particularly through reduced labor costs related to driver salaries.Additionally, the capability of automatic fleet maintenance enhances operational efficiency and increases vehicle utilisation rates (Spieser et al., 2016).

The integration of SAVs into the fabric of Automated Mobility on Demand (AMOD) not only represents a shift toward more efficient transportation models, but also signifies a pivotal transition in societal mobility norms. This evolution is underpinned by technological advances that promise not only to reshape the economics of transportation but also to redefine accessibility and inclusivity in urban mobility. As the potential for SAVs to supplant a significant portion of private vehicle ownership becomes increasingly apparent, it is necessary to examine how these innovations might catalyze a broader transformation in traffic performance. The synergy

between AMOD services and SAVs could herald a new era of transportation, where the emphasis shifts from vehicle ownership to mobility as a service (MaaS) (Wong et al., 2020), fostering a more connected, efficient, and greener urban environment. In this paragraph, three concepts are mentioned: SAVs, AMOD, and MaaS. While these concepts are related and may overlap in some contexts, they have distinct meanings. SAVs specifically refer to autonomous vehicles that are shared among users. AMOD is a broader concept that encompasses various automated mobility services, including SAVs, but also other modes of transportation. MaaS focuses on the integration of different transportation services into a single platform, regardless of whether they are automated or shared.

While research on the effects of SAVs on travel patterns and environmental outcomes is still emerging, early studies indicate that SAVs have the potential to lower transportation expenses and result in only slight increases in overall trip duration, owing to the efficient routing algorithms that enable ride-sharing (Shaheen and Bouzaghrane, 2019). The use of autonomous vehicles can also provide increased mobility for underrepresented groups, such as the elderly and children, enhancing their independence (Harb et al., 2018). According to Fagnant and Kockelman (2014), with a moderate market penetration rate of 3.5% for SAV trips, approximately 12 private vehicles could be replaced by one SAV. More optimistic projections by Zhang et al. (2015) suggested that one SAV could substitute for around 14 private vehicles.

Efficient pooling strategies are essential for optimizing the utilization of SAV fleets. Progress in real-time ride-matching algorithms could greatly contribute to minimizing fleet size, decreasing wait times, and reducing deadheading (i.e., the distance traveled with no passengers). Alonso-Mora et al. (2017) introduced a model for the real-time splitting of high-capacity vehicles (e.g., shared taxis) in Manhattan, finding that a fleet of 3,000 four-passenger SAVs could satisfy 98% of the taxi demand, resulting in a 77% reduction in fleet size. The study highlighted that using higher-capacity vehicles would reduce the number of vehicles required to meet current demand, leading to shorter wait times and less total distance traveled per vehicle. Lokhandwala and Cai (2018) observed that incorporating shared autonomous taxis in New York City (There are both shared self-driving taxis and traditional taxis in the road network) could maintain the same level of service while reducing the required fleet size by 59%. By lowering operational costs for fleet providers and reducing out-of-pocket expenses for passengers, SAVs offer an effective solution for first-mile last-mile (FMLM) connectivity, potentially enhancing public transit ridership (Huang et al., 2022). The emergence of SAVs has notably popularized the "rail plus on-demand shared" mode, especially among younger demographics, offering a novel service form for intercity travelers needing to connect between different modes of transport, such as air or rail (Ren et al., 2023). A reduction in fleet size could lead to lower levels of energy consumption and emissions. It is also hypothesized by researchers that the implementation of SAVs could lead to a 55% reduction in energy consumption and a 90% decline in greenhouse gas (GHG) emissions.

SAVs could also pose certain challenges. According to Taeihagh and Lim (2019), the data storage and transmission capabilities inherent in autonomous vehicles may raise privacy concerns, and the communication networks could be susceptible to cyberattacks. Additionally, Petit and Shladover (2014) examined safety risks associated with autonomous vehicles and found that the most critical and probable threats include Global Navigation Satellite System (GNSS) spoofing and the injection of false data. Furthermore, due to the introduction of SAVs, the total vehicles distance traveled is expected to increase, primarily because of empty vehicles (Alam and Habib, 2018).

1.2 Research gap

While current literature acknowledges that SAVs can help reduce urban traffic congestion, there remains a gap in understanding how specific fleet management strategies and matching algorithms can be optimized to maximize this potential. Introduced from the following three aspects. Firstly, existing literature primarily focuses on the impact of using SAVs on fleet size, noting that SAVs can maintain the same level of service with fewer vehicles compared to POVs (Fagnant and Kockelman, 2014; Zhang et al., 2015; Alonso-Mora et al., 2017; Lokhandwala and Cai, 2018). Although existing research has explored aspects of congestion reduction, studies comparing the impact of different matching strategies on the efficiency of mixed mobility systems are still limited. Additionally, the adaptability and efficiency of these systems under varying SAV fleet sizes or penetration rates have not been thoroughly examined. Furthermore, no studies have simultaneously evaluated the impact of changes in fleet size from both the perspectives of passenger service level and road congestion level. Therefore, further research is needed to quantify how factors such as fleet size, matching strategy interact to influence traffic performance in mixed mobility systems, and how these systems can create a more efficient urban environment.

To address these gaps, an agent-based simulation framework is utilized to perform scenario analyses, focusing

on metrics such as the pooling ratio and the extra mileage ratio (Jin et al., 2021). These metrics assess the impacts of ride-pooling services and additional miles incurred by factors such as unmatched waiting passengers and detours for ride-sharing. The pooling ratio, which indicates the average number of passengers serviced by a single for-hire vehicle simultaneously, is affected by fare structures, passenger demographics, and matching algorithm efficiency. Extra mileage ratio represents the additional distance traveled by SAVs due to factors like unmatched waiting passengers and detours for ride-sharing. This analytical approach helps delineate the roles of SAVs and POVs within mixed mobility systems and establishes a framework for evaluating the potential effects of various technological advancements and policy interventions on urban traffic dynamics.

1.3 Research scope

This study focuses on exploring the impact of the integration of SAVs and POVs into mixed mobility systems on urban traffic performance. In this study, it is assumed that all car travel demands are met exclusively by SAVs and POVs, without the coexistence of other transportation modes such as traditional taxis.Initial research will be conducted on a single city's network, with network data extracted from the Openstreetmap. The generalizability of these results across different cities and the impact of urban geometry have not been considered. Employing two matching strategies, this research will match passenger demand with SAV fleets or POVs. Both matching algorithms are sourced from the Fleetpy framework and are used directly in this study without optimization. By adjusting the size of the fleet and based on key parameters such as the carpooling ratio and the additional mileage ratio, the study will examine the performance of service levels and the impact on urban road networks under different fleet sizes.

1.4 Research questions

This study seeks to explore the intricate dynamics between fleet sizing, matching strategies, and their collective influence on traffic performance within mixed mobility systems. However, achieving a successful integration of SAVs into existing urban transportation networks necessitates a comprehensive understanding of how varying fleet sizes and matching strategies can influence traffic flow, congestion patterns, and overall system performance within a modeling framework. Against this backdrop, the following research questions (RQ) and sub-questions (SQ) have been formulated to guide this exploration:

RQ

How do fleet sizing and matching strategies for SAVs influence traffic performance in mixed mobility systems?

This primary question aims to investigate the relationship between the strategic deployment of SAV fleets and the operational outcomes in terms of traffic efficiency, congestion mitigation, and mobility enhancements in urban settings that incorporate a mix of traditional and automated mobility services.

SQ

1. How does the modification in the size of the SAVs fleet impact the efficiency of matching strategies and overall traffic dynamics?
2. Can an optimal fleet size value be determined based on the level of service provided by SAVs and the impact on road congestion under different matching strategies?
3. What impact do SAVs have on urban road congestion levels when using different matching strategies?
4. In mixed mobility systems, what kind of preference choices will passengers have between SAVs and other modes of transportation?

This research explores the dynamic integration of SAVs into urban mobility systems, focusing on optimizing traffic performance in mixed mobility settings. Each sub-question addresses a specific aspect of this integration, creating a narrative on the implications and strategies for efficient fleet management and service delivery.

The first sub-question examines how adjustments in fleet size impact the efficiency of matching SAVs with passenger requests and the resulting effects on traffic congestion and flow. The goal is to understand the operational results of the changes in fleet size value and to identify strategies that improve service quality while reducing urban traffic congestion.

The second sub-question investigates how various passenger-vehicle matching strategies influence the optimal fleet size. It looks at how different approaches impact service quality, operational efficiency, and resource use in mixed mobility systems, aiming to identify strategies that maximize the benefits of SAV integration.

The third sub-question considers the impact of different matching strategies on urban road congestion levels when deploying SAVs. This part of the study assesses how different methods of pairing passengers with vehicles affect traffic flow and congestion in urban areas, with the aim of identifying the most effective strategies for reducing congestion while maintaining high service efficiency in SAV deployment.

The fourth question explores the preferences passengers exhibit when choosing between SAVs and other transportation modes within mixed mobility systems, aiming to understand the factors influencing their decisions.

1.5 Report structure

Section 2 reviews existing literature on fleet sizing and matching strategies for Shared Automated Vehicles (SAVs) within mixed mobility systems. It explores the evolution of shared mobility, the integration of autonomous vehicles into these systems, and the theoretical and practical challenges involved in optimizing fleet operations and management. Section 3 describes the research methods used, combining data with simulation analysis to explore the operational performance of SAVs in mixed mobility systems. It details the use of GPS data from taxis to analyze the operational status of existing Mobility on Demand (MoD) services, the demand generation model, the mode choice model, and the simulation model. Section 4 introduces Haidian District in Beijing as the research area and introduced the specific application of the research method. Section 5 introduces the results and analysis of the research. Sections 6 cover the conclusions and discussions of the research.

2 Literature review

The objective of this paper is to evaluate the impacts of fleet sizing and matching strategies for SAVs within mixed mobility systems. During this process, the aim is to address the challenges related to the SAV service fleet sizing and management problem and the modeling of SAVs demand under various matching strategies. Subsequently, we will embark on a literature review focusing on these three critical aspects.

In the current stage of research on SAVs fleet performance, most researchers opt to use simulation techniques for analysis. Simulation models are highly adaptable, making them ideal for studying the performance of SAVs fleets and we do not have empirical data of such systems: they don't really exist yet, although some pilots have been tested in few cities. They can handle a wide range of variables, including different vehicle types, user behaviors, and varying traffic and urban conditions, which is essential for investigating different fleet sizing and matching strategies. This approach offers practical benefits, such as improved understanding of service wait times, vehicle use, and how well different matching strategies work. While optimization-based methods are powerful in identifying theoretical best-case scenarios, they often rely on simplifying assumptions about system behavior and demand patterns. Such assumptions may not hold in the unpredictable and varied environments in which SAV fleets operate. Additionally, optimization models can become computationally infeasible or overly complex when dealing with the full scope of variables and constraints present in mixed mobility systems (Jin et al., 2021).

2.1 Matching strategies

The deployment of extensive carpooling services depends on the use of mathematical models and algorithms capable of efficiently matching numerous passengers with a fleet of shared vehicles in real-time.Mobility on Demand (MoD) is viewed as an integral component of a wider system that addresses demand management and vehicle routing challenges. In these systems, where customers make service requests online and the actual services are provided offline, optimizing fleet operations is essential to maintain the profitability of such services. Matching strategies in pooling scenarios can be classified according to the number of pickup and drop-off locations served by pooling vehicles, as well as the frequency of rider transfers between vehicles. The primary categories include one-to-one, one-to-many, many-to-one, and many-to-many scenarios (Tafreshian et al., 2020). This study focuses on the one-to-many matching scenario, in which a driver is capable of serving multiple passengers, but transfers between vehicles are not permitted.

Fleetpy (Engelhardt et al., 2022) implements two primary matching strategies: Immediate Decisions Simulation (IDS) and Batch Offer Simulation. The fundamental distinction between these strategies lies in how MoD customers, who request trips, interact with the operators. In the IDS approach, the interaction is divided into three stages: first, customers submit service requests to the MoD operators; then, the operators evaluate these requests and decide whether to accept or reject them. Throughout this process, the customer makes a decision and relays it instantly to the operator, who subsequently processes the next incoming request. In the Batch Offer Simulation, the operator only records customer requests without responding to them immediately. Operators do not immediately respond to each inquiry but instead process the open requests in batches within specified time intervals, creating offers for all unanswered requests simultaneously. Here, the user will not immediately agree or disagree with the quote. In the case of multiple operators, they can decide on the waiting time for responses from other operators.

In these two matching strategies, a status update occurs at the beginning of each time step. Fleet control is executed after receiving new requests. In the update phase, vehicles adhere to their existing vehicle plans. In the event-triggered fleet control phase, operators optimize the vehicle plan.

Baldacci et al. (2004) introduced both an exact approach and a heuristic method for addressing the carpooling problem, utilizing two integer programming formulations. The exact approach employs a bounding procedure that incorporates three different lower bounds derived from problem relaxations. Meanwhile, the heuristic approach generates a valid upper bound by converting the solution from a Lagrangean lower bound into a feasible solution. Herbawi and Weber (2011) developed a model for the dynamic one-to-many ride-sharing matching problem, framing it as a pickup and delivery challenge with time constraints. In this approach, drivers' offers are represented using a time-expanded graph, simplifying the multi-objective route planning task into a well-known multi-objective shortest path problem (MSPP). A genetic algorithm is employed to address the ride-matching challenge for each time interval, considering all available requests. Alonso-Mora et al. (2017) introduced a reactive real-time optimization method that efficiently assigns travel requests to vehicles and iteratively refines these assignments over time to approach an optimal solution. The approach begins by constructing feasible trips using a pairwise shareability graph, then allocating these trips to vehicles via an

integer linear programming model, and finally rebalancing any idle vehicles. Simonetto et al. (2019) proposed an algorithm that simplifies the scheduling issue by matching on-demand trips to vehicles through a linear assignment problem. This model restricts new requests from being combined during simultaneous processing, ensuring that one vehicle is only assigned to one new request at a time. Additionally, the concept of joint optimization was introduced, which significantly reduced communication requirements.

2.2 Mode choice modeling

The exploration of passengers' preferences for different modes of transportation has been extensively examined in the existing literature. Discrete choice models, traditionally employed to analyze selection behaviors, describe the process by which individuals make choices among a finite set of alternatives. These models are implemented in various forms, such as: (a) The Multinomial Logit Model (So and Kuhfeld, 1995), which is grounded in Random Utility Theory (RUT), posits that each passenger will choose the transportation mode perceived to provide the greatest utility. This utility is conceptualized as a quantification of the passenger's satisfaction with the mode, determined by factors including travel time, cost, comfort, among others. (b) The Nested Logit Model (Wen and Koppelman, 2001) extends the Multinomial Logit Model by allowing for varying degrees of substitutability among different transportation modes. This model segments the choice set into "nests" or groups of modes that are more closely related, capturing the similarities between modes and allowing for a more nuanced understanding of mode selection behavior. (c) The Mixed Logit Model (Hole, 2013) offers greater flexibility in addressing the diversity of passenger preferences compared to both the Multinomial and Nested Logit Models. It accommodates variability in model parameters with respect to demographic characteristics or other variables, thereby capturing the heterogeneity in passengers' choice behavior more accurately.

Selecting appropriate parameters to represent the utility of different modes in a discrete choice model is a critical consideration in utilizing this model. Fan et al. (2023) integrated the logit model into a trip-based optimization problem, employing the monetary cost for group travelers and the value of travel time for travelers using a specific mode within the group to represent the utility of the mode.

2.3 Simulation framework

Several simulation frameworks are available for modeling vehicle fleets, with a few leading open-source simulators that incorporate MoD solutions being particularly noteworthy.MATSim (W Axhausen et al., 2016) is a framework for simulating transportation systems using an agent-based approach, developed on the Java platform, known for its extensibility. In an iterative simulation process, travel bookings generated from an activity-based pre-day demand model are handled by a mobility fleet during the subsequent transportation simulation. The outcomes are then assessed before starting the next simulation cycle, which helps to improve the model's accuracy over time.Ben-Dor et al. (2019) demonstrated the simulation of service requests and rejections for SAVs within the Tel Aviv Metropolitan Area (TAMA) using MATSim. Hörl et al. (2021) employed the Dynamic Vehicle Routing Problem (DVRP) (Maciejewski et al., 2017) extension of MATSim to simulate pricing, customer behavior, and the systemic impact of a cost-covering automated taxi system in Zurich. They integrated AMoD simulation into experiments with dynamic demand. Yan et al. (2020) also used the Dynamic Vehicle Routing Problem (DVRP) extension of MATSim to conduct a microsimulation of SAVs trips across seven counties in the Minneapolis-Saint Paul region to evaluate how travel density and parking limitations affect mobility and transportation patterns on the performance of shared autonomous fleets.

SimMobility is an agent-based model designed to provide detailed decision-making capabilities for agents, including tasks like scheduling, mode choice, and route planning within a multi-modal network, all based on an activity-based demand model. The framework employs a three-tiered approach to facilitate analyses from long-term to short-term perspectives and includes a sophisticated interface to support interactions among agents. In a study by Nguyen-Phuoc et al. (2023), the SimMobility Long-term (yearly) and Mid-term (daily) simulators were used to simulate three different scenarios: the Baseline Case (BC), Partial Automation (PA), and Full Automation (FA). The research aimed to examine how automated on-demand mobility services could influence public transportation systems . Additionally, Oh et al. (2020) carried out a study in Singapore, utilizing the SimMobility Mid-term simulator to evaluate the impacts of automated on-demand mobility (AMoD). Their study incorporated an activity-based modeling system that used smartphone-based stated preference surveys for data collection. This research was organized into three primary modules: Pre-day for generating travel demand, Within-day for selecting departure times and routes, and Supply for simulating multi-modal network performance.

FleetPy (Engelhardt et al., 2022), designed in Python, is an agent-based simulation framework and it is purpose-

fully designed to intricately model MoD services. It is particularly attuned to capture the nuanced interactions between users and operators, while its versatile framework supports the seamless incorporation of various operators within the broader transportation ecosystem. The modular architecture of FleetPy facilitates the easy transferability of previously developed elements and enables users to choose the level of modeling detail that best suits their needs.

2.4 Research challenges

In current research, carpooling optimization problems are predominantly based on time windows, optimizing scenarios that simultaneously involve multiple user requests. During simulations, user requests often originate from different timestamps. Batch processing these requests can lead to additional waiting times between when requests are generated and when they are processed by the operator. The Immediate Decision strategy avoids this delay but may not achieve global optimization in terms of vehicle resource utilization during the pooling process. From the perspective of the entire network, this could potentially lead to additional waiting times. Balancing these factors remains a major challenge in current research.

3 Research method

This section describes the research methods used, combining empirical data with simulation analysis to explore the operational performance of SAVs in mixed mobility systems. Initially, the study utilizes GPS data from taxis waffic zones. The mode choice model employs a discrete choice analysis framework to explore the trade-offs passengers consider between SAVs and POVs. Subsequently, the simulation model transforms the outputs of demand generation and mode choice into actual traffic flow and vehicle dispatch behaviors, simulating the effects of these strategies on traffic efficiency and congestion by adjusting fleet sizes and passenger matching strategies.

To assess the overall performance of the transportation system at a more macroscopic level, this study uses the Macroscopic Fundamental Diagram (MFD). Through the analysis of the MFD, we can visually observe the variations in traffic flow under different traffic density conditions, particularly noting the density points at which traffic flow peaks and the subsequent sharp decline due to congestion. This macroscopic insight helps evaluate the effectiveness of SAVs and their matching strategies in real urban environments from a systemic perspective.

3.1 Research process

This study adopts a three-part methodology to explore the impacts of fleet sizing and matching strategies of SAVs on urban traffic dynamics. First, the Demand Generation Model uses taxi GPS data and travel distance distribution to create traffic demand through traffic zone division, Inter-zone travel probabilities calculated based on the travel distance distribution function, and traffic demand calculated based on the total travel volume within the research area, providing necessary input for simulations. Next, the Mode Choice Model assesses travelers' preferences among transport modes based on attributes such as travel time, environmental impact, comfort, and reliability, derived from surveys. Finally, the simulation integrates two matching strategies—Immediate Decision and Batch Offer—using a network model containing nodes and data, comparing the MFD generated through simulation with the MFD derived from taxi GPS trajectories to evaluate different fleet management strategies. Figure 1 illustrates the research process employed in this study. In Figure 1, the algorithms for the two matching strategies are integrated within the fleetpy simulation framework. The aggregated results for the four KPIs can be directly obtained from the simulation output. The main tasks of this study include building a demand generation model, a mode choice model, acquiring network data, and plotting the MFD.

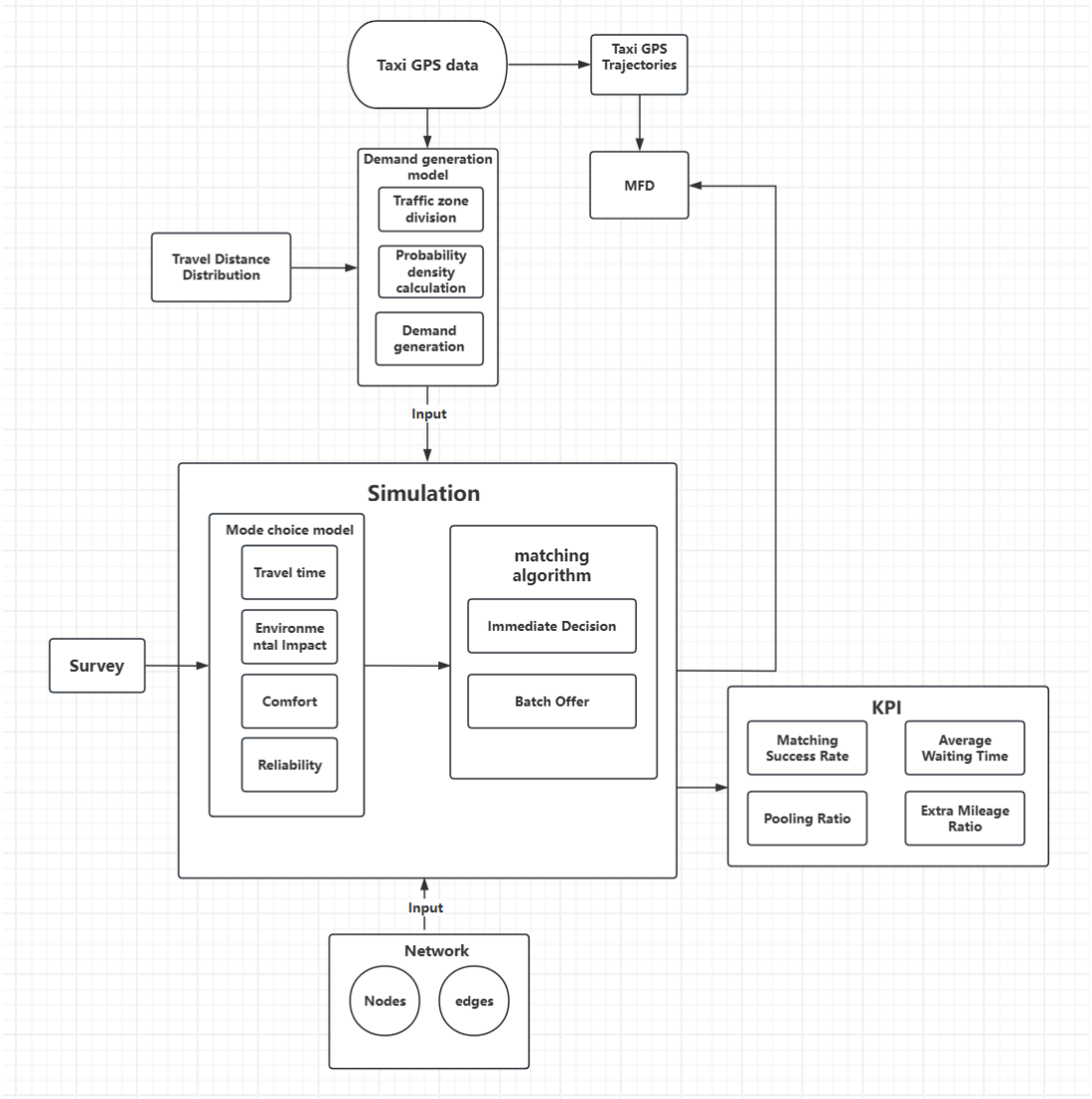

Figure 1: Research Process

3.2 Travel demand generation method based on travel distance distribution function

This section introduces the demand generation model. The travel distribution forecasting presented in this section refers to the process of allocating the initial traffic demand, which constitutes a total travel volume, to node pairs on the network according to certain rules.

In this study, an innovative approach is employed that utilizes taxi GPS data and its associated travel distance distribution function to generate travel demand forecasts, rather than relying on traditional four step transportation planning models. The approach is derived from a demand generation method based on travel distance proposed by Zhu (2020). Traditional models typically require extensive data on socio-economic factors, demographics, income, and vehicle ownership. In contrast, this method leverages real-world taxi trip data, which reflects actual travel behavior and the dynamic nature of urban traffic flows. By analyzing taxi GPS trajectory data, it becomes possible to directly observe the distribution characteristics of travel demand and the specific patterns of travel distances.

This method utilizes taxi GPS trajectory data to map the total volume of automotive travel within the road network, serving as a basis for estimating the overall travel demand attributed to private cars and Shared Autonomous Vehicles (SAVs) in subsequent simulations. Zhu (2020) conducted a frequency analysis of taxi passenger travel distances across different cities in China and used the least squares method to fit the travel distance distributions to a log-normal distribution. The specific findings are illustrated in Figure 2: different colors in the figure correspond to different cities (with the red circular data points representing Beijing), and the solid lines represent the fitted probability density distribution of travel distances. The legend includes the R-squared values of the fitting errors for different cities, as well as the parameters of the distance distribution function. The horizontal axis is the travel distance, and the vertical axis is the corresponding probability.

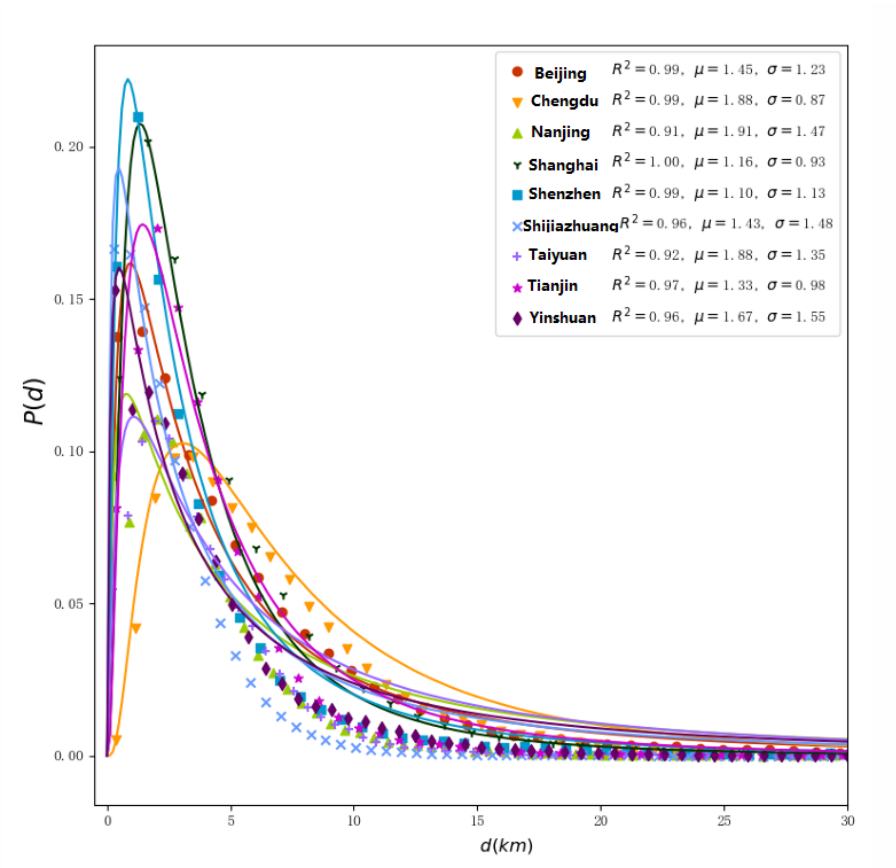

Figure 2: Probability Density Distribution of Travel Distance of Taxi Passengers (Zhu, 2020)

Upon examining the travel distance probability density distribution fitting results for Beijing in Figure 2, the travel distances of taxi passengers closely follow the log-normal distribution shown in Equation 1, with R-squared

values all being 0.99. For Beijing, the parameters are μ=1.45 and σ=1.23 (Zhu, 2020).

$$p(d) = \frac{1}{d\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln d - \mu)^2}{2\sigma^2}\right) \tag{1}$$

Where:

μ represents the average value of the natural logarithm of the variable.

σ represents the variability or spread of the natural logarithm of the variable.

3.2.1 Demand generation model

To generate the travel demand, this study designs a travel demand generation method that aligns with the probability density function of the travel distance distribution. First, the nodes of the road network are divided into several zones, and then, using these zones as basic units, the travel demand is generated based on a travel distance distribution and the total travel volume T. Subsequently, The travel demand between zones (OD) is randomly assigned to pairs of nodes. The model includes two parameters: the distance distribution function $p(d)$ and the total travel volume T. The detailed steps are as follows:

Traffic zone division: This paper employs the K-means clustering algorithm to perform a clustering analysis on the roads networks nodes of Haidian District, Beijing, thus delineating traffic zones from a macroscopic perspective (without considering the actual characteristics of the road).

The selection of the number of clusters is determined qualitatively, as this step functions primarily as a computational simplification. It is feasible to compute distances on the network from one node to all potential nodes using node positions and edge lengths data. Additionally, the probability density distribution mentioned earlier can be used to calculate the likelihood of trips occurring between each pair of zones. This enables the estimation of demand generation from one node to all others based on the overall traffic volume. However, such a process requires substantial computational resources, necessitating an initial clustering of the nodes.

The Elbow Method and the Silhouette Coefficient Method are two traditional approaches for assessing the quality of clustering. The Elbow Method (Cui et al., 2020) is a heuristic technique used to determine the optimal number of clusters, k, in cluster analysis. This method involves calculating the cost associated with each cluster number k — typically the total within-cluster sum of squares errors(SSE) — and then plotting these values as a function of k. The "elbow" point on the curve, where the curve begins to bend significantly, is identified. This elbow point is generally considered to be the point beyond which adding more clusters does not significantly reduce WCSS, indicating that further increases in the number of clusters contribute minimally to improving the model.

Equation 2 shows how to calculate SSE.

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \tag{2}$$

where:

k is the number of clusters. C_i represents the set of data points in cluster i.

x is a data point.

μ_i is the centroid of cluster i.

$\|x - \mu_i\|$ is the Euclidean distance between the data point x and the centroid μ_i.

The Silhouette Score Method offers a refined approach to assess the quality of clustering by integrating measures of cohesion and separation (Shahapure and Nicholas, 2020). Each cluster is evaluated with a silhouette score, which ranges between -1 and 1.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{3}$$

where:

$a(i)$ refers to the average distance between a given sample point and all other points within the same cluster.

$b(i)$ refers to the average distance between a given sample point and all points in the closest cluster that is not the cluster the sample point belongs to.

The distance here refers to the Euclidean distance. The average silhouette score across all samples provides a metric for the overall quality of the clustering, with values closer to 1 suggesting better clustering quality, as it implies that the sample is well-matched to its own cluster and well-separated from other clusters and values closer to -1 signifying inadequate clustering, as it shows that the sample is either significantly distant from its own cluster or too near to points in other clusters, which suggests a lack of distinct separation between groups. The optimal number of clusters, k, is determined by the maximizes the average silhouette score, offering a quantitative measure to assess the rationality of cluster assignments.

Probability density calculation: In the previous step, traffic zones were delineated, and the distance between every two zones can be considered as representative of the distance from any point within one zone to any point within another zone. The method for calculating distances between zones involves matching the centroid of each zone's cluster to the road network, and computing the distance between these centroids using the shortest path algorithm. Subsequently, the distances between centroids are considered as the distances between zones. By calculating based on the probability density function of travel distances, the probabilities of trips occurring between zones can be determined. As shown in the equation 4.

$$R_{pq} = f(D_{pq}) = \frac{1}{1.23 D_{pq} \sqrt{2\pi}} \exp\left(-\frac{(\ln D_{pq} - 1.45)^2}{3.0258}\right) \tag{4}$$

Where:

R_{pq} represents the probability of trip occurring between zone p and zone q.

D_{pq} represents the distance between zone p and zone q.

Next, normalize the probabilities of trips occurring between zones. $R_{pq}^{\text{normalized}} = \frac{R_{pq}}{\sum_{p,q} R_{pq}}$.

Demand generation: In the previous step, the traffic volume ratio between each zone was generated. Next, by multiplying the traffic volume ratio between zones by the total daily car trips, the specific traffic volume between each zone can be obtained.

After completing the generation of traffic demand between traffic zones, the traffic volume T_{pq} is randomly distributed across all node pairs (i, j), where $i \in Z_p$ represents the origin and $j \in Z_q$ represents the destination, denoted as T_{ij}. This generates the initial OD matrix, completing the daily demand generation.

3.3 Mode choice model

Within the theoretical framework of random utility maximization, discrete choice models are instrumental in analyzing the mode choices of various traveler groups. To evaluate the willingness or preference of individuals to utilize particular travel modes, it is essential to calculate the utilities associated with each option. The market penetration rate of Shared SAVs remains uncertain, and representing the trade-offs between SAVs and POVs is crucial in refining the passenger demand model. The focus of this research is to explore network performance under mixed traffic conditions, rather than delving into detailed analyses of individual passenger travel behaviors. This study employs static parameters to determine the utilities of different modes, specifically: Travel Time, Comfort, Reliability, and Environmental Impact. By conducting surveys, respondents are prompted to rate the aforementioned aspects of SAVs and POVs based on their personal perceptions.

3.3.1 Model construction

The model considers influencing factors from four aspects: objective factors (travel time) and subjective factors (comfort, reliability, and environmental impact). Yang et al. (2018) proposed that residents' environmental awareness and subjective attitudes and perceptions toward different transportation modes are critical determinants. For the same mode of travel, individuals may have varying perceptions and preferences regarding safety, comfort, convenience, and other aspects. The study found that these differences in subjective attitudes significantly influence residents' mode choice. At the current stage, the safety of automated driving is a controversial topic, so it is not included in the scope of consideration (Wäschle et al., 2022) (Noy et al., 2018).

Convenience is a relatively abstract subjective perception, while reliability is more concrete and easier for respondents to evaluate. In subjective perception, reliability and convenience share similarities (Meixell and Norbis, 2008), both aiming to make the travel process smoother and more predictable. Reliability can be described as the system's ability to consistently meet user expectations in terms of availability (the vehicle should be readily accessible when needed, with minimal waiting times) and punctuality (it should adhere to expected schedules, ensuring users reach their destinations within a predictable time frame).

3.3.2 Attribute value calculation

In the binary logit model concerning the choice between SAVs and POVs, there are three attributes generated subjectively by travelers: Comfort, Reliability, and Environmental Impact. These attributes cannot be directly represented in subsequent simulations, necessitating a quantitative method to determine their values.

Comfort: Travel comfort is defined as the level of comfort perceived by travelers during their journey, resulting from a combination of factors including comfort, convenience, safety, independence, and physical effort (Ma et al., 2006). Comfort differs in value from travel time and travel cost. Travel comfort is difficult to measure directly as it actually reflects the traveler's perception of the travel state and can be considered as a level of service of the travel service. Zhang et al. (2020) proposes a ranking choice model to analyze the value of comfort. The factors considered include gender, age, mode of travel, and travel time. According to the ordered logistic regression analysis, gender and age were not significant factors; therefore, they were excluded from the model, and the regression was recalculated as shown in equation 5.

$$U_{\text{comfort}} = 4.899 + 2.6X_{0-15\text{min}} + 1.752X_{16-30\text{min}} + 1.287X_{31-45\text{min}} \tag{5}$$

A method for grading comfort levels was proposed(Zhang et al., 2020), defining comfort on a scale Y_n of 1 to 9, where 1 represents the least comfortable and 9 the most comfortable, as shown in equation 6. The comfort levels are categorized based on eight thresholds, as shown in Table 1.

$$Y_n = \begin{cases} 1 & \text{if } U_{\text{comfort}} \leq b_1, \\ 2 & \text{if } b_1 < U_{\text{comfort}} \leq b_2, \\ 3 & \text{if } b_2 < U_{\text{comfort}} \leq b_3, \\ \vdots & \quad \vdots \\ 8 & \text{if } b_7 < U_{\text{comfort}} \leq b_8, \\ 9 & \text{if } U_{\text{comfort}} > b_8 \end{cases} \tag{6}$$

Table 1: Threshold Values for Comfort Levels

Threshold Value
-0.1216
0.8831
1.7274
2.6049
3.3531
4.1355
5.3778
7.0232

Subsequently, the nine comfort levels of the comfort value model will be projected into five levels. This adjustment is made because, in this study, comfort is categorized into five distinct levels for the survey. Apply a K-means clustering algorithm to group the eight thresholds into four clusters. Use the center of each cluster as the new threshold. The projected thresholds are shown in table 2. The reason for choosing k-means clustering over linear interpolation for this dataset is that k-means can effectively divide the data into several clear groups, revealing the natural divisions and patterns within the data. In contrast, linear interpolation is primarily suitable for smoothing and interpolating calculations between known data points, assuming linear changes in the data, and is not suitable for identifying and handling the clustering characteristics of the data.

Table 2: Projected Threshold Values for Comfort Levels

Threshold Value
1.6646
3.4508
5.237
7.0232

Environmental Impact: The environmental impact of vehicle operation is most directly manifested through carbon emissions, which serve as a tangible indicator of the ecological footprint associated with transportation. Measuring the carbon emissions from vehicles provides a quantifiable method to assess their environmental impact. The Vehicle Specific Power (VSP) model is used to estimate the power output of light-duty vehicles and is applied to simulate the connection between vehicle operating conditions and fuel consumption (Cressie, 1990).

$$C_{i,j} = EF_\alpha \cdot F_{i,j} = EF_\alpha \sum_{l=1}^{n} ER_0 \cdot NFCR_l \cdot T_{i,l} \tag{7}$$

The variables are defined as follows:

$C_{i,j}$ represents the trip's carbon emissions by taxi i.

EF_α represents exhaust emission, which is 2.18 kg/L.

ER_0 symbolizes the baseline fuel consumption rate for taxis, established at 0.274 (Zhao, 2009) .

$NFCR_l$represents the normalized fuel consumption rate associated with the average speed range for segment l, as detailed in Table 3.

$T_{i,l}$ represents the time span in seconds that taxi i spends driving within the intermediate speed interval l.

Equation 7 calculates the total carbon emissions for taxi i in segment l. This equation is simplified to determine the hourly carbon emissions $c_{i,j}$ of the taxi over a segment of its journey. The emissions are related to the average speed, as illustrated in equation 8.

$$U_{\text{Environmental Impact}} = c_{i,j} = EF_\alpha * F_{i,j} = EF_\alpha ER_0 * NFCR_l \tag{8}$$

Table 3: NFCR for Various Speed Intervals (Cressie, 1990)

Average Speed (km/h)	**NFCR**
0-2	1.085137
2-4	1.258708
4-6	1.311138
6-8	1.477515
8-10	1.573123
10-12	1.64575
12-14	1.729985
14-16	1.807417
16-18	1.841056
18-20	1.922954
20-22	1.996735
22-24	2.045498
24-26	2.092286
26-28	2.163184
28-30	2.186922
30-32	2.25144
32-34	2.328849
34-36	2.338148
36-38	2.361389
38-40	2.395369
40-42	2.441831
42-44	2.470396
44-46	2.538255
46-48	2.566097
48-50	2.581801
50-52	2.595985
52-54	2.6796
54-56	2.715854
56-58	2.755036
58-60	2.809524
60-62	2.864735
62-64	2.956168
64-66	3.049095
66-68	3.289347
Above 80	3.550955

Linking carbon emissions to environmental impacts presents challenges. By dividing hourly carbon emissions by the emission factor EF_α , the hourly fuel consumption can be determined. There is a significant correlation between this fuel consumption and vehicle model. To obtain data on the fuel consumption of mainstream gasoline vehicles, using survey on the top ten best-selling gasoline vehicles in China in 2023, as shown in Table 4, to determine the range of hourly fuel consumption, which was then categorized into five classes.

Car Model	Average Urban Fuel Consumption (liters/hour)
Nissan Sylphy	1.929
Volkswagen Lavida	2.73
Volkswagen Sagitar	2.07
Toyota Camry	2.64
Volkswagen Passat	4.05
Volkswagen Magotan	2.52
Toyota Corolla	1.746
Audi A6L	3.54
Honda Accord	2.154
Honda Civic	2.211

Table 4: 2023 Top Ten Selling gasoline Car Models and Their Average Urban Fuel Consumption

The vehicle Average Urban Fuel Consumption data in Table 4 was obtained from the China Automotive Energy Consumption Query Platform. By performing K-means clustering analysis with four cluster centers, the Average Urban Fuel Consumption is divided into five levels, which serve as quantitative representations of the environmental impact in the logit model. The classification thresholds obtained are shown in table 5.

Table 5: Threshold Values for Environmental Impact Levels

Threshold Value
2.211
2.73
3.54
4.05

Reliability: For SAVs, reliability is a key indicator of system performance. The reliability of SAVs encompasses not only the technical performance of the vehicles, such as vehicle stability, failure rates, and safety features, but also the stability and efficiency of the entire service system, including response times and the capacity to adapt to traffic conditions. Currently, there is a lack of research on the relationship between reliability, travel time, residents' income, and other attributes during travel decision-making. Reliability cannot be directly quantified. Therefore, reliability is not considered in the construction of the mode choice model in this study. Residents' income levels may influence their needs and expectations regarding the reliability of transportation modes (Ko et al., 2019). Higher-income residents might prefer more expensive but reliable modes of transportation, such as private cars, as they often place a higher value on saving time and travel comfort. In contrast, lower-income residents may rely more on public transportation, despite potentially facing greater uncertainty and inconvenience, because they are willing to accept lower reliability due to cost considerations. Income levels can also affect how sensitive residents are to travel time. Generally, higher-income individuals may opt for faster but more costly transportation options to save time, and vice versa.

3.3.3 Conclusion

The utility function is defined to evaluate the utility of a choice based on several attributes. Each attribute is weighted by a coefficient which represents its influence on the overall utility. The utility function is formulated as follows:

$$U(x) = \beta_0 + \beta_1 x_{\text{travel\_time}} + \beta_2 x_{\text{comfort}} + \beta_3 x_{\text{environmental\_impact}} \tag{9}$$

where:

$\beta_1, \beta_2, \beta_3$ are the parameters of the model,

$x_{\text{travel\_time}}, x_{\text{comfort}}, x_{\text{environmental\_impact}}$ are the attributes of the choice influencing the decision-making process.

β_0 is the error term.

The unit of $x_{\text{travel\_time}}$ is minutes. x_{comfort} and $x_{\text{environmental\_impact}}$ are scores (from 1 to 5) for attributes based on the subjective perceptions of the respondents.

x_{comfort} is determined by calculating the value of U_{comfort} to confirm its corresponding threshold range, thereby establishing its value.

$$x_{\text{comfort}} = \begin{cases} 1 & \text{if } U_{\text{comfort}} \leq 1.6646, \\ 2 & \text{if } 1.6646 < U_{\text{comfort}} \leq 3.4508, \\ 3 & \text{if } 3.4508 < U_{\text{comfort}} \leq 5.237, \\ 4 & \text{if } 5.237 < U_{\text{comfort}} \leq 7.0232, \\ 5 & \text{if } U_{\text{comfort}} > 7.0232 \end{cases} \tag{10}$$

The value of x_{comfort} is also determined based on $U_{\text{Environmental Impact}}$ and its threshold.

$$x_{\text{Environmental Impact}} = \begin{cases} 1 & \text{if } U_{\text{Environmental Impact}} \leq 2.211, \\ 2 & \text{if } 2.211 < U_{\text{Environmental Impact}} \leq 2.73, \\ 3 & \text{if } 2.73 < U_{\text{Environmental Impact}} \leq 3.54, \\ 4 & \text{if } 3.54 < U_{\text{Environmental Impact}} \leq 4.05, \\ 5 & \text{if } U_{\text{Environmental Impact}} > 4.05 \end{cases} \tag{11}$$

The parameters β_i are estimated based on stated choices preferences, and they quantify the relative importance of each attribute in determining the utility. The higher the coefficient, the more significant the attribute is in influencing the decision.

In this study, the parameters β_i are estimated based on the results of stated preference surveys. These surveys gather data on respondents' daily travel times, ratings for different attributes, and preferences between SAVs and POVs to understand the basic travel behaviors and subjective perceptions of attributes among respondents. The values of the parameters are estimated using the maximum likelihood estimation method.

The logit model is used to calculate the probability that a given choice will be selected, based on the utility derived from that choice. The logistic function is employed to transform the utility into a probability, ensuring that it lies between 0 and 1. The probability function is defined as:

$$P(x) = \frac{1}{1 + e^{-U(x)}} \tag{12}$$

3.4 Macroscopic Fundamental Diagram

When building the MFD, the relationship between vehicle accumulation (average network density) and trip production (average network flow) characterizes the performance and level of congestion in the urban traffic system (Geroliminis and Daganzo, 2008). The MFD simplifies the complexity of urban traffic by representing it as a macroscopic reservoir model, it not only demonstrates the maximization of traffic flow at certain density levels but also illustrates how flow sharply declines beyond a specific density threshold due to congestion.

Equation 13 and 14 (Geroliminis and Daganzo, 2008) defines the weighted average of flow and density.

$$q^w = \Sigma_i q_i l_i / \Sigma_i l_i \tag{13}$$

$$k^w = \Sigma_i k_i l_i / \Sigma_i l_i \tag{14}$$

Where:

q^w represents the weighted traffic flow

k^w represents the weighted traffic density.

l represents the length of the road

i indexes the road segments within the network

To derive a functional form of the MFD that effectively models traffic flow and density relationships within an urban road network, it is essential to incorporate empirical data analysis and theoretical traffic dynamics formulations. Ambühl et al. (2020) suggested a mathematical representation for the MFD that possesses inherent physical meaning. By employing a smooth approximation of the minimum operator to model the MFD, as illustrated in equation 15, this approach utilizes a smoothing parameter λ. This parameter quantifies the degree of alignment between the approximation and the true minimum function, with values closer to zero indicating a more precise match. smaller values of λ result in an approximation that is more closely aligned with the true minimum. The λ-trapezoidal MFD function incorporates significant traffic parameters such as the free-flow speed u_f, backward wave speed w, jam density κ, and intersection capacity Q, all reflecting tangible aspects of traffic flow. These parameters all have physical meaning. By using Equation 13 and 14, plot the MFD scatter diagram. Fit the scatter data according to the function of Equation 15. When λ is 0, a trapezoidal diagram can be obtained. By estimating the λ value, the functional form of the MFD can be derived. The parameters can be either actual measured values or estimated values.

$$q(k) = \min\left(u_f k; Q; (\kappa - k)w\right) = \lim_{\lambda \to 0} -\lambda \ln\left(\exp\left(-\frac{u_f k}{\lambda}\right) + \exp\left(-\frac{Q}{\lambda}\right) + \exp\left(-\frac{(\kappa - k)w}{\lambda}\right)\right) \tag{15}$$

In Equation 15, for actual road network data, some parameters can be calculated. The intersection capacity Q can take into account the proportion of different types of roads and obtain the average intersection capacity for the area based on the capacity of intersections on different types of roads. Regarding the congestion density κ, it is assumed that when the network is fully congested, all vehicles in the network are stationary, the distance between adjacent vehicles is zero, and only passenger cars and buses exist in the road network. Under these conditions, κ depends solely on the length of the vehicles. The remaining parameters, free-flow speed u_f, backward wave speed w, and smoothing parameter λ need to be estimated based on curve fitting.

The equation for jam density κ is:

$$\kappa = \frac{1}{p_p \cdot L_p + p_b \cdot L_b} \tag{16}$$

Where:

L_p is length of the passenger car.

L_b is length of the bus.

p_p is proportion of passenger car.

p_b is proportion of bus.

MFD calibration is performed by calculating the Root Mean Square Error (RMSE)(Chai et al., 2014). RMSE measures the square root of the average of the squared differences between predicted values and actual observations. It provides a metric for assessing the size of prediction errors; the smaller the error, the higher the accuracy of the model.

The RMSE formula is as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{17}$$

where:

- n is the number of observations.
- y_i is the actual observation at the i^{th} index.
- $\hat{y}_i$ is the prediction at the i^{th} index.

3.5 Simulation model

Fleetpy is an open-source, agent-based simulation framework designed with a modular architecture that allows for the easy transfer of existing components and the customization of modeling detail to suit various simulation needs(Engelhardt et al., 2022). It is possible to control the matching strategy by rewriting some specific modules. Using microsimulation to capture the driving dynamics of all vehicles in a network for analyzing congestion levels and service quality requires substantial computational costs and runtime.

Figure 3 illustrates the main simulation flow of Fleetpy. Once the essential modules and the data for the network, infrastructure, vehicles, and demand are set up, the Fleetpy framework runs through the simulation step-by-step over time. The framework is organized into various components, such as the Network Module, Traveler Module, Vehicle Module, and Fleet Control Modules. The Fleet Control Modules are crucial for implementing and overseeing the matching strategy and are built around three primary objects:

- Vehicle objects store details about their present conditions, including their location, the count of passengers, and a list of the passengers currently onboard.
- Plan request objects collect data about customers who have made trip requests to the operator. This data encompasses the request time, the origin and destination locations, the size of the group, and any time constraints related to pick-up and drop-off windows, whether communicated by the customer or assumed.
- Vehicle plan objects define a sequential (theoretical) list of tasks for vehicles, serving as potential solutions to the underlying vehicle routing problem. Each task is associated with a specific location and includes timing details, such as the duration or the earliest possible start time, along with the description of the task itself.

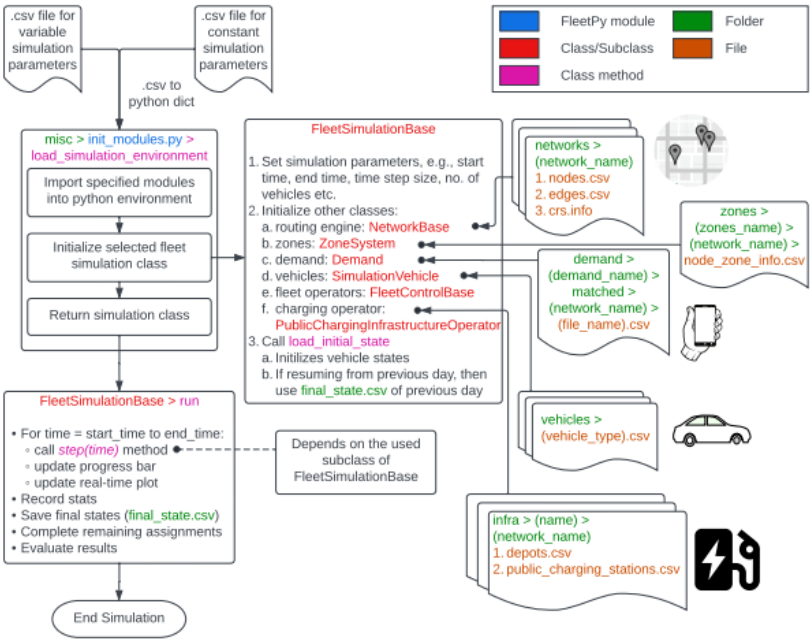

Figure 3: Main simulation flow of Fleetpy (Engelhardt et al., 2022)

3.6 Simulation scenarios

This study employs two different vehicle-passenger matching strategies for simulation, namely Batch Offer and Immediate Decisions. These strategies differ in their approach to handling requests and allocating resources.

3.6.1 Immediate Decision

The Immediate Decision approach handles transportation requests as they are received. When a passenger's request is registered in the system, it is immediately processed to generate an offer and assign a vehicle. This method prioritizes immediacy, addressing requests as they occur. The system evaluates the current vehicle fleet by examining factors such as vehicle location, status, and scheduled itineraries to determine the optimal match for each request (Use the nearest available vehicle). Upon determining a suitable vehicle, the system immediately generates and sends a ride offer to the passenger. Passengers are expected to quickly respond to these offers. For those who accept, the designated vehicle is dispatched to begin the trip.

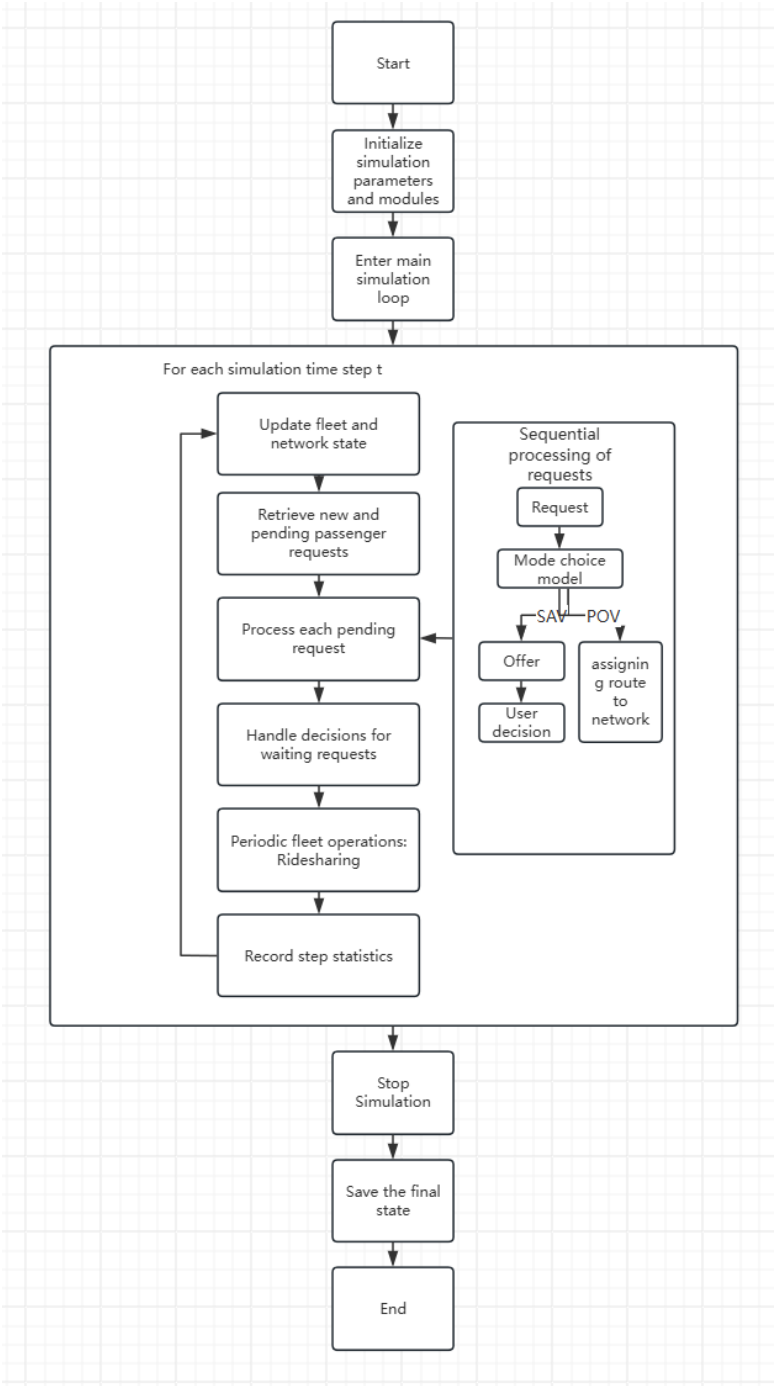

Figure 4: Immediate Decision Simulation Process

Figure 4 is the flowchart of the Immediate Decision Simulation. During the initialization phase, the system establishes the simulation environment and initializes all relevant modules and parameters, including fleet status, network conditions, and the user request handling mechanism. Subsequently, at each time step, the system first updates the fleet and network status, ensuring that vehicle positions and road network conditions reflect the latest information. For processing passenger requests, the system collects newly generated requests as well as those that were unassigned in the previous time step, and presents multiple service options for each request. For each request, determine the mode of transportation (SAV or POV) using a mode choice model. If determined to be POV, directly load the trip into the network. Each SAV request sequentially receives one service proposals from operators, and users make acceptance or rejection decisions based on the maximum waiting time. Simultaneously, the system performs periodic fleet operations, such as pooling optimization(Insertion Heuristic). At each time step, relevant statistical data are recorded for subsequent analysis. When the simulation concludes,

the system saves the final state of the fleet and operators, evaluates performance throughout the simulation.

Table 6: Input Parameters for Immediate Decision Simulation

Input_Parameter_Name	**Parameter_Value**
sim_env	ImmediateDecisionsSimulation
network_type	NetworkBasicWithStore
start_time	0
end_time	86400
time_step	60
nr_mod_operators	1
op_module	PoolingInsertionHeuristicOnly
user_max_decision_time	0
op_min_wait_time	0
op_max_wait_time	300

Table 6 outlines the input parameters required for the Immediate Decision Simulation. This includes the start time (start_time), end time (end_time), and time step (time_step), which together establish the fundamental temporal framework for the simulation. Additionally, the table specifies the number of operators (nr_mod_operators) involved in the simulation and the maximum decision time for uses (user_max_decision_time), which are essential for managing the interactive process of the simulation. In the simulation, assume that users make a decision immediately upon receiving a service proposal, setting user_max_decision_time to 0. Furthermore, to consider user experience, minimum and maximum wait times for operators (op_min_wait_time and op_max_wait_time) are also set to evaluate the efficiency of service responses. Additionally, Table 6 lists the key modules required for the Immediate Decision Simulation. The simulation environment is set as ImmediateDecisionsSimulation, and the process of its algorithm has already been introduced. The introductions for the other two modules uesd in Immediate Decision (Pooling Insertion Heuristic Only and Network Basic With Store) are as follows:

Pooling Insertion Heuristic Only: The PoolingInsertionHeuristicOnly algorithm handles real-time passenger travel requests by optimizing passenger pairing and routing through heuristic-based insertions into existing vehicle routes. Upon initiation, the algorithm processes each received request, calculating the origins and destinations, as well as the temporal and spatial compatibility with other requests. It employs several heuristic methods to filter available vehicles, including assessing the alignment between the vehicle's current route direction and the origin-destination vector of the passenger request, as well as the current workload of the vehicle.

Then outlines the heuristics used for vehicle selection before the insertion of new ride requests in ride-pooling systems. This algorithm begins by defining the mathematical model used to formalize the selection process.

Let $\mathbf{p}_o$ and $\mathbf{p}_d$ represent the coordinates of the request's origin and destination, respectively. The direction vector of the request, $\mathbf{v}_{req}$, is computed as follows:

$$\mathbf{v}_{req} = \mathbf{p}_d - \mathbf{p}_o \tag{18}$$

The unit vector of the request's direction is given by:

$$\hat{\mathbf{v}}_{req} = \frac{\mathbf{v}_{req}}{\|\mathbf{v}_{req}\|} \tag{19}$$

For each vehicle v in the initial set of available vehicles V, the vehicle's direction vector $\mathbf{v}_{veh}$ is calculated similarly, and the dot product with $\hat{\mathbf{v}}_{req}$ determines the alignment of the vehicle's current route with the request's direction:

$$\hat{\mathbf{v}}_{veh} = \frac{\mathbf{v}_{veh}}{\|\mathbf{v}_{veh}\|}, \quad \text{dot} = \hat{\mathbf{v}}_{req} \cdot \hat{\mathbf{v}}_{veh} \tag{20}$$

Vehicles are selected based on the highest values of this dot product, with a predetermined number N_{dir} forming the subset S_{dir}. Additionally, the vehicles with the least number of current tasks, represented by $n_{tasks}(v)$, are selected, with the top N_{lwl} forming the subset S_{lwl}.

The final set of selected vehicles, $S_{selected}$, is formed by the union of these two subsets:

$$S_{selected} = S_{dir} \cup S_{lwl} \tag{21}$$

Once a set of candidate vehicles is determined, the algorithm attempts to insert new requests into their current operational plans. Vehicles are selected based on their current position and orientation relative to the requested origin and destination. For each vehicle, the algorithm generates a set of potential new operational plans.

After identifying the a solution, the algorithm presents passengers with an offer for transportation service, which they can accept or reject. If no feasible vehicle is available or the insertion would result in inefficient routing, the request is denied. Current vehicle plans will not be altered. This means existing vehicle routes and assignments remain unchanged and are not affected by any unsuccessful new insertion attempts. This process is executed in real time, ensuring rapid response to passenger demands and efficient vehicle utilization.

Network Basic With Store: This routing class conducts all routing computations utilizing Dijkstra's algorithm implemented in Python. It stores previously computed travel information in a dictionary, allowing it to quickly retrieve and return these values when they are queried again.

3.6.2 Batch Offer

Batch offer simulation is a matching strategy used to optimize the allocation of services in a ride-pooling or fleet management system. In this type of simulation, passenger requests are collected over a period of time, and rather than processing each request immediately, they are grouped together and handled in batches. The operators periodically run a batch optimization process, which analyzes all pending requests and assigns vehicles based on factors such as travel time, vehicle availability, and route efficiency. The batch processing cycle is 60 seconds.

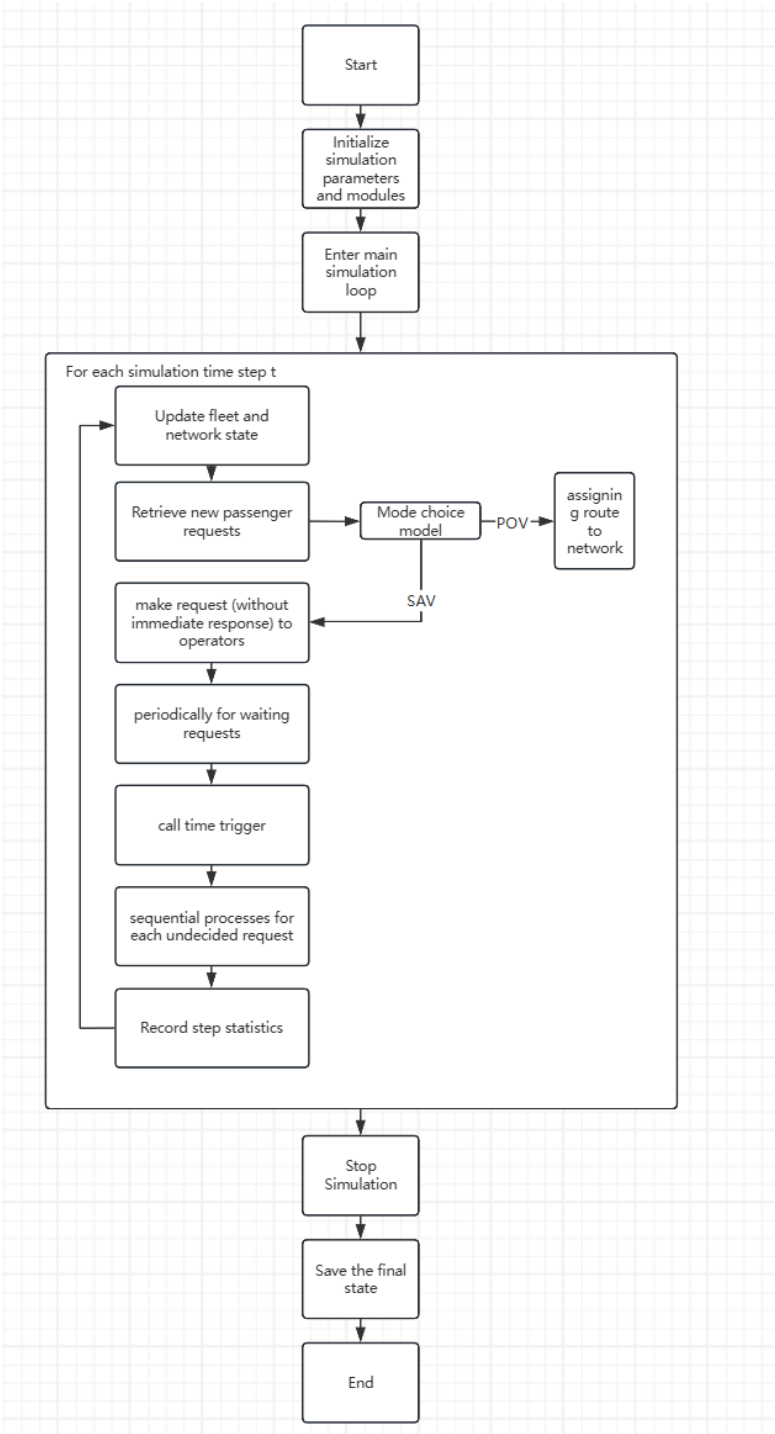

Figure 5: Batch Offer Simulation Process

Figure 5 illustrates the primary process of the Batch off simulation. In Batch off, during the initialization phase, the system configures the simulation environment and initializes all modules and parameters as Immediate Decisions Simulation. However, the main distinction arises in how passenger requests and operator responses are handled. Upon receiving a request, the system performs a mode choice. If the selected mode is POV, the trip is directly loaded into the network. If SAV is selected, the process continues. Instead of receiving service proposals immediately, passengers submit requests to operators but must wait for the operator's batch optimization process. This batch process is triggered periodically, during which operators perform optimizations and generate service offers for all pending requests at once. Users then make decisions based on the offers received after this batch processing. Following the batch optimization, the system continues to monitor accepted but unexecuted requests, checking for any modifications or cancellations by users. Like the Immediate Decisions

approach, fleet operations such as vehicle repositioning and charging management occur periodically. Data are recorded at each step for analysis, and when the simulation concludes, the final states of the fleet and operators are saved, and the system evaluates overall performance.

Table 7: Input Parameters for Batch Offer Simulation

Input_Parameter_Name	**Parameter_Value**
sim_env	BatchOfferSimulation
network_type	NetworkBasicWithStore
start_time	0
end_time	86400
time_step	60
nr_mod_operators	1
op_module	RidePoolingBatchAssignmentFleetcontrol
user_max_decision_time	0
op_min_wait_time	0
op_max_wait_time	300
op_rp_batch_optimizer	AlonsoMora

Table 1 lists the Input Parameters for Batch Offer Simulation, which differs from the Immediate Decision Simulation primarily in terms of the simulation environment (sim_env) and the operator module (op_module). Figure 5 illustrates the process flow of the Batch Offer Simulation. Subsequently, a introduction to the Ride-Pooling Batch Assignment Fleet control system is provided.

Ride-Pooling Batch Assignment Fleet control: It is a fleet control system designed for ride-pooling services, primarily focused on the batch optimization of passenger and vehicle assignments. This system periodically optimizes the matching of passengers and vehicles based on a set time interval, employing batch assignment algorithms to process passenger requests. It allows for a retry in assignment by adjusting the maximum waiting time for passengers if the initial attempt fails, thus enhancing the success rate of assignments.

Upon receiving a passenger request, the system first attempts to allocate a vehicle. If the initial attempt is unsuccessful, the system may consider extending the passenger's maximum waiting time in the next optimization step to facilitate reassignment. Additionally, the system permits adjustments to the specific pickup times based on predetermined pickup time windows, enhancing scheduling flexibility and passenger satisfaction.

When passengers confirm their bookings, the system updates the pickup time constraints based on previously generated proposals, ensuring that the timing aligns with operational needs. If a passenger cancels a request, the system accordingly updates the vehicle plan and releases the associated resources.

In the Ride-Pooling Batch Assignment Fleet control module, it is necessary to specify the algorithm used to assign and re-optimize vehicle scheduling for ride-pooling scenarios. Fleetpy provides two algorithms: Batch Insertion Heuristic Assignment and AlonsoMora Assignment. The Batch Insertion Heuristic Assignment utilizes a simple insertion heuristic method to batch assign requests that were not previously allocated. AlonsoMora Assignment algorithm is a dynamic and high-capacity ride-pooling service algorithm that optimizes and manages the matching between vehicles and passengers through dynamic trip-vehicle assignment. The algorithm is a variant of the publicationAlonso-Mora et al. (2017). In this simulation, the AlonsoMora Assignment is chosen for optimization. The AlonsoMora Assignment performs global optimization by considering all combinations of vehicles and passengers to find the optimal overall solution. In contrast, the simple insertion heuristic focuses only on local optima. Moreover, the AlonsoMora Assignment demonstrates relatively higher efficiency in handling large-scale systems, specifically in scenarios with a large number of vehicles and requests.

In the simulation, the process of the AlonsoMora Assignment is as follows: Before optimization, the system prepares by loading essential parameters such as vehicle initial states, environmental changes, and demand forecasts, while configuring parallel processors to enhance efficiency. It synchronizes data like vehicle locations and passenger requests, ensuring accuracy through data preprocessing. The system employs a series of simple heuristics—distance-based, insertion cost-based, and a mixed approach—to select candidate vehicles, thereby simplifying the computational effort. A decision tree is constructed to detail all potential vehicle-passenger combinations and optimize using linear programming with Gurobi solver. The results directly inform operational adjustments, refining routes and allocations, with a focus on local optimizations like boarding points. System

updates and performance evaluations are based on these results, using operational data and user feedback to continuously improve the algorithm's effectiveness.

3.7 KPIs

This study uses four KPIs and MFD to evaluate the service level and road congestion level of SAVs: Matching Success Rate, Average Waiting Time, Pooling Ratio (denoted as π), and Extra Mileage Ratio (denoted as ϵ) (Jin et al., 2021). Matching Success Rate, Average Waiting Time are directly provided in the output of Fleetpy simulation and they are aggregated for the whole simulation.

Pooling Ratio (π) measures the efficiency of an SAV serving multiple passengers simultaneously. In the simulation output, each vehicle's occupancy for every individual trip is provided. To extract trips with occupancy greater than or equal to one, average the occupancy numbers and subtract one. This process will yield the pooling ratio for the simulated scenario. The formula for its calculation is:

$$\pi = \text{Average number of passengers served by an SAV simultaneously} - 1$$

This indicator represents the increase in the number of passengers that an SAV serves in a given period compared to serving a single passenger. If a vehicle serves more passengers, it indicates higher efficiency in shared rides, leading to better resource utilization, lower operational costs.

Extra Mileage Ratio (ϵ) evaluates the additional distance traveled by the SAV to complete a passenger's trip. The simulation output provides the length of all trips. By classifying each trip based on whether it is occupied, they can be divided into passenger-carrying trips and empty trips. This allows for the calculation of the Extra Mileage Ratio. The calculation is as follows:

$$\epsilon = \frac{\text{Total distance traveled by SAV} - \text{Distance with passengers}}{\text{Distance with passengers}}$$

This formula expresses the ratio of the additional mileage traveled by an SAV to complete order i to the actual passenger-carrying distance. Distance with passengers can be considered as the travel distance generated by passengers choosing POVs as their mode of travel. This relates to the extra distance compared to the distance that would have been traveled if the users took their POVs. A lower extra mileage ratio indicates that less idle distance is added during ride-pooling services, thereby improving vehicle operational efficiency.

4 Case study

This chapter introduces Haidian District in Beijing as the research area, which is distinguished by its significant demographic and transportation network characteristics. Haidian is not only one of the most densely populated districts in Beijing but also a critical center for scientific research and education in China, hosting numerous top-tier higher education and research institutions. Moreover, the district's transportation network is notably complex, designed to accommodate the demands of high urbanization and dense population distribution.

Haidian District is located in the northwest of Beijing's urban area, and is one of the six major urban districts of the city, as well as being among the most developed. The district spans approximately 426 square kilometers, stretching about 30 kilometers from north to south, and reaching up to 29 kilometers at its widest point from east to west, comprising approximately 2.53% of Beijing's total area. The topography of Haidian District gradually elevates from east to west. The eastern part is relatively flat and highly urbanized, being close to the center of Beijing. This area boasts well-developed infrastructure and urban services. The eastern sector features densely packed residential and commercial facilities, alongside major roads and transportation hubs such as the North Third and North Fourth Ring Roads. In contrast, the western part of Haidian District exhibits more varied terrain, preserving extensive natural landscapes and ecological resources, with numerous hills and large green spaces. The terrain and natural scenery of the west provide the area with abundant natural resources and recreational spaces, including major tourist and leisure destinations like Xiangshan Park and the Beijing Botanical Garden.

4.1 Population characteristics

Haidian District has a permanent population of 3,133,469, which has decreased by 4.5% since 2010. Of this population, 35.7% are people from other provinces or cities who have moved to Beijing, showing an 11% decrease in this demographic. The district is divided into family and collective households, with the former accounting for approximately 2,602,895 people. Among the 29 towns and streets in the district, five have a population exceeding 150,000, namely Xueyuan Road Street, Xibeiwang Town, Beitaipingzhuang Street, Sijiqing Town, and Xisanqi Street; 12 areas have populations between 100,000 and 150,000, and another twelve have populations under 100,000. The administrative structure of Haidian District in Beijing is shown in the Figure 6. The gender distribution is nearly balanced with males making up 49.9% and females 50.1%. In terms of age structure, 11.8% of the population is aged 0-14, 69.7% are between 15 and 59 years old, and 18.5% are 60 years and older. Compared to 2010, the proportion of the population aged 0-14 has increased, the 15-59 age group has decreased, and the 60+ age group has grown, reflecting an aging population trend to some extent.

Figure 6: The administrative division of Haidian District in Beijing (Yuan et al., 2015)

4.2 Road network

The road network in Haidian District demonstrates a design that accommodates the distinct characteristics of high-density urban areas and natural reserves. This configuration significantly reflects its status as a key center for scientific research and education in Beijing, incorporating comprehensive urban functions. The district is traversed by major ring roads, including the North Third Ring, North Fourth Ring, and North Fifth Ring, which are connected to radial roads such as Zhongguancun Street, Xueyuan Road, and Zhichun Road. These connections effectively link commercial and educational centers, facilitating rapid inter-regional transit. The road network map is shown as Figure 7.

In the traditional delineation within Beijing, the division between urban and suburban areas is generally related to the ring road system. Typically, the areas within the inner three rings (including the First, Second, and Third Rings) are considered the city center or central urban areas. Traditionally, areas starting from the Fourth Ring outward, especially beyond the Fifth and Sixth Rings, are more often regarded as suburban. The regions between the Fourth and Fifth Rings tend to be transitional, displaying urban characteristics such as dense residential and commercial developments, alongside suburban features like more extensive green spaces and lower building densities. Areas beyond the Sixth Ring retain more pronounced suburban features, including broader spaces, agricultural lands, and less urban development.

Haidian District also possesses a complex network of local roads and alleys, particularly in residential and commercial areas. These roadways not only enhance mobility within the district but also help alleviate traffic pressure during peak times. Moreover, the public transportation system in the district has seen significant enhancements, with several subway lines such as Line 4, Line 10, and Line 13 crossing the area, collectively improving Haidian's transportation efficiency and accessibility.

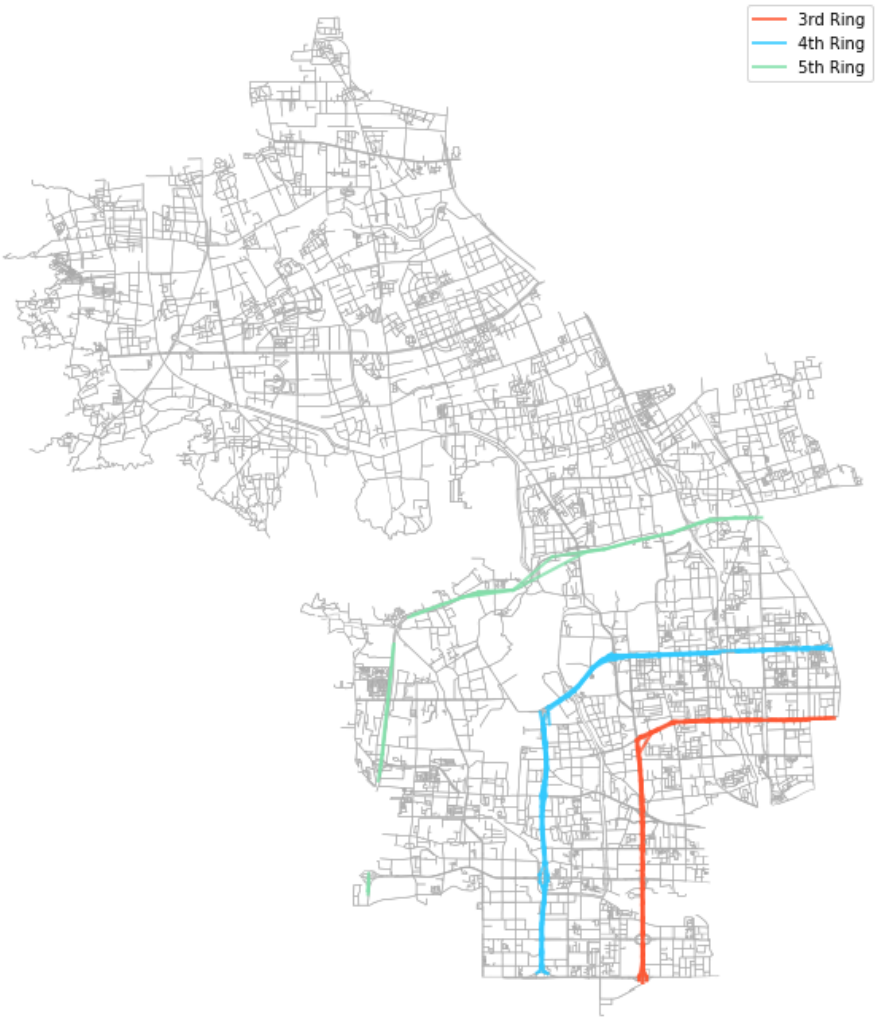

Figure 7: Road Network Map of Haidian District, Beijing

4.3 Research data

This section provides a detailed introduction to the data source used in this study—the T-Drive dataset (Zheng, 2011), which contains GPS trajectories of taxis in Beijing. This dataset includes GPS trajectory data from 10,357 taxis functioning in Beijing between February 2nd and February 8th, 2008, it contains about 15 million GPS points, covering a total travel distance of 9 million kilometers. In this dataset, the GPS data includes information such as the taxi's unique identifier (ID), timestamp, longitude, and latitude. In this study, taxi trajectory data within Haidian District was extracted from this dataset based on the boundaries of the road network, laying the groundwork for subsequent traffic pattern analysis and simulation. The method for extracting trajectory data involves projecting GPS data points onto a road network, finding the nearest edge for each GPS point and calculating the projection point on that edge, and then using these projection points to calculate the total driving distance for each vehicle.

4.3.1 Data processing

In the initial phase of this study, the raw taxi trajectory dataset was meticulously preprocessed to ensure data quality and accuracy. Initially, the data were sorted by taxi ID and timestamp to maintain the correct sequence for subsequent analysis. Additionally, the dataset underwent a deduplication process that any duplicate records were eliminated, 1,337,497 duplicate GPS data points have been removed after processing. Following this, the data were constrained by geographical boundaries. Considering the actual latitude and longitude limits of Beijing, exclude data points located outside the boundaries of Beijing. Note that only data points were removed, thereby reducing trips outside of Beijing, rather than deleting entire trajectories that start/end outside of Beijing and only keeping internal trajectories. Through these detailed preprocessing steps, a total of 11,695,575 valid taxi data points were obtained.

Subsequently, the dataset underwent further constraints to refine the scope of analysis specifically to the Haidian District. Utilizing the delineated boundaries of Haidian District, the study filtered the GPS data to include only those points located within the district. During the analysis, complete trajectories starting or ending within the district were retained. Trajectories entirely contained within Haidian District constitute 82 percent of the data. For trajectories crossing the boundaries of Haidian District, the ratio of the trajectory length inside the district to that outside the district is 0.82. Additionally, the analysis was narrowed down temporally by selecting data from February 4th, 2008. Following the rigorous filtering process, the refined dataset for February 4, 2008, within the Haidian District, comprised a total of 472,755 data points, collected from 8,192 taxis in operation.

4.3.2 Macroscopic Fundamental Diagram

In constructing the MFD, the analysis focuses exclusively on taxis within the urban road network, aligning with subsequent simulation outcomes. By mapping the taxi GPS trajectory data, GPS data points are matched to the nearest edge, then connect the edges, obtaining a complete taxi route on the road network. Figure 8 is an example, showing the trajectories of 100 taxis in Haidian District from 8 to 10 AM on February 4th. All edges with taxi trajectories are the roads needed to calculate q^w and k^w based on Equation 13 and 14 (Geroliminis and Daganzo, 2008). By counting the vehicles passing through each road segment per hour, the traffic flow of the road segment can be obtained.

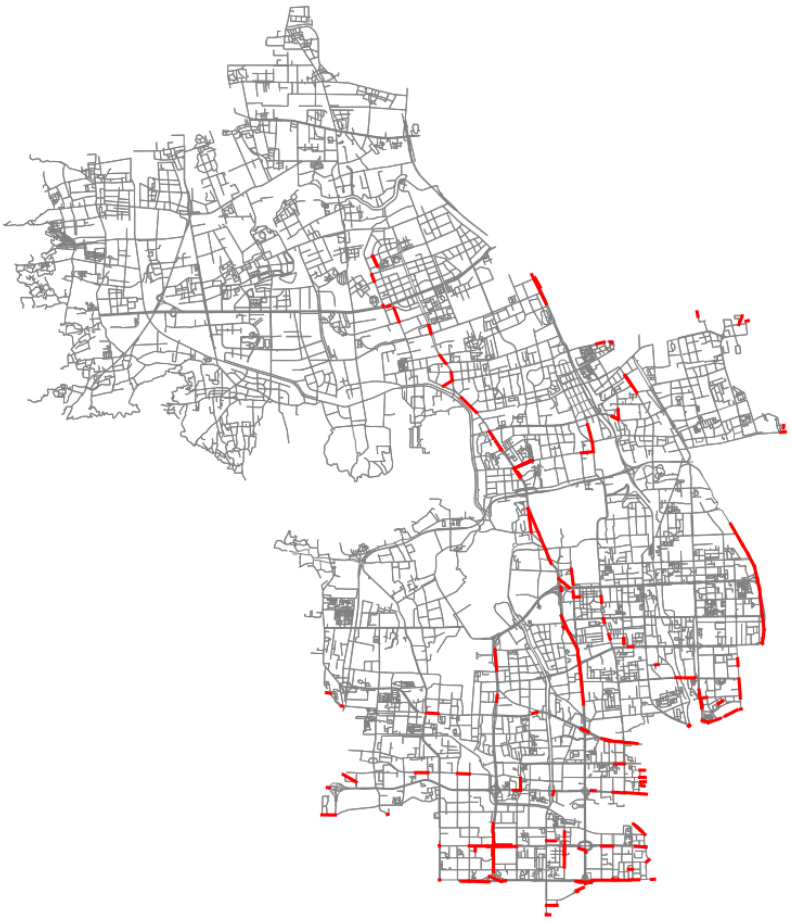

Figure 8: An Example of The Trajectories of 100 Taxis

Figure 9 presents the taxi MFD for Haidian District over the course of a day, illustrating the network's maximum capacity along with branches representing non-congested and congested traffic conditions. One dot represents the flow and density within a one-minute time window.

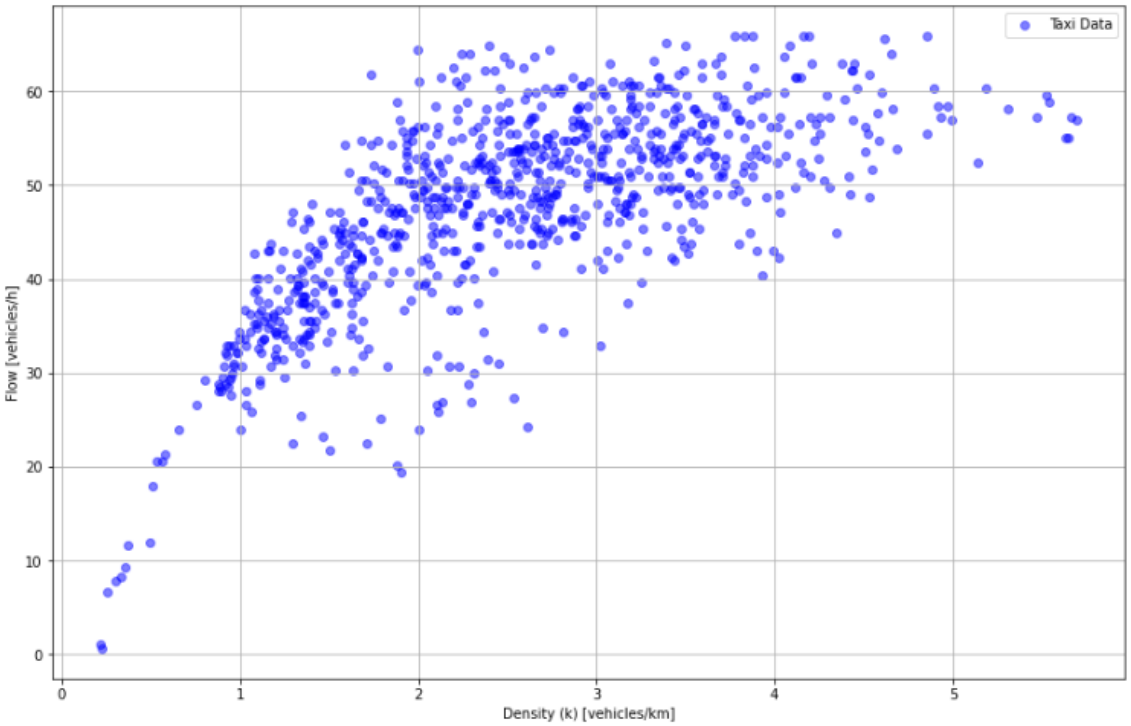

Figure 9: Taxi Macroscopic Fundamental Diagram (MFD) for Haidian District by Time Slot

The free-flow speed u_f represents the maximum speed at which vehicles can travel when there is minimal traffic interference, and the backward wave speed w describes the rate at which congestion propagates backward through traffic. However, due to insufficient data to support a direct calculation, parameter estimation must be conducted during the fitting process to accurately determine these values.

When calculating the intersection capacity Q, it is necessary to consider different types of roads. Table 8 (Lei et al., 2022) shows the proportion of different road types across various districts in Beijing, Note that the expressways in Haidian District do not have intersections, so they are not considered. LIANG et al. (2011) conducted research on the capacity analysis of signalized intersections. Several types of intersections were tested. Intersections with four straight lanes exhibited a traffic capacity of 2127 vehicles per hour. These intersections conform to the design standards of main roads. Additionally, intersections designed with two straight lanes demonstrated a traffic capacity of 964 vehicles per hour, meeting the design requirements typical of minor branch roads. Based on this information, the average intersection capacity in Haidian District can be calculated as 1166 vehicles per hour. Next, it is necessary to estimate the percentage of capacity used by taxis. According to the 2023 Beijing Traffic Development Annual Report, Beijing, Haidian district recorded a total of 0.7 million private small and micro cars in 2022. The average daily trip frequency for these vehicles on workdays was 3.14 trips per car, showing a decline of 0.2 trips compared to the previous year, while on non-workdays, the trip frequency remained steady at 3.08 trips per car. In the central urban areas, the average deployment rate of private cars was 46.1%, with 47.22% on workdays and 43.74% on non-workdays. Using this data, the total number of trips for one year made by private cars in Haidian district can be calculated to be 0.37 billion.

The total number of trips T for one year made by private cars can be calculated with the equation:

$$T = N \cdot (f_{\text{work}} \cdot d_{\text{work}} + f_{\text{nowork}} \cdot d_{\text{nowork}}) \tag{22}$$

Where:

N is Total number of private cars.

f_{work} is Average daily trip frequency per car on workdays.

f_{nowork} is Average daily trip frequency per car on non-workdays.

d_{work} is Number of workdays in the year.

d_{nowork} is Number of non-workdays in the year.

According to the Overview of the Taxi Industry, in 2022, the total number of taxi orders in Haidian district was approximately 26.46 million. It is assumed here that the proportion of taxi trips to private car trips reflects the proportion of taxis to private cars in the network traffic. Based on the proportion of taxi trips(26.46 million) to private car trips(0.37 billion) in the Haidian District of Beijing over the course of a year, it is inferred that taxis account for 6.67 percent of the traffic in the road network. This allows to infer that the average intersection capacity used by taxis in Haidian District, Beijing, is approximately 78 vehicles per hour.

Regarding congestion density κ, given the limited availability of information, it is assumed that when the network reaches full congestion, all vehicles within the network remain stationary. Under these conditions, κ is solely dependent on the length of the vehicles. The main vehicle types operating on Beijing's road network are passenger cars and buses, and their proportions of road occupancy can be estimated based on the vehicle ownership data. Beijing has a private car ownership of 6.378 million vehicles, 60,000 taxis, and 28,000 buses, the data is sourced from Beijing Traffic Management Bureau. The BYD Qin PLUS, the best-selling car model in Beijing, is used as the representative length for passenger cars, measuring approximately 4.765 meters. The predominant bus model in Beijing's public transport system is the BJ6123C7BCD-2, which is 11.98 meters long.

Thus, it can be calculated that under fully congested conditions, the road density reaches 190 veh/km. The road density used by taxi is 12.5 veh/km.

The trapezoidal function and the MFD with functional form of the taxi MFD for Haidian District is shown in the Figure 10. Through parameter estimation (nonlinear least squares), the value of λ is obtained as 32.28, the value of u_f is 39.25km/h and the value of w is 9.01km/h. In some cases, λ has a physical significance. Ambühl et al. (2020) suggests that λ is the sum of factors that reduce traffic flow, caused by infrastructure, interactions between vehicles, and other modes of transportation.

Table 8: Length fractions of different road types in four districts in Beijing (Lei et al., 2022)

			%
Administrative region	Expressway ratio	Main road ratio	Minor branch ratio
Dongcheng District	7	17	76
Xicheng District	7	18	75
Haidian District	8	16	76
Chaoyang District	10	15	75

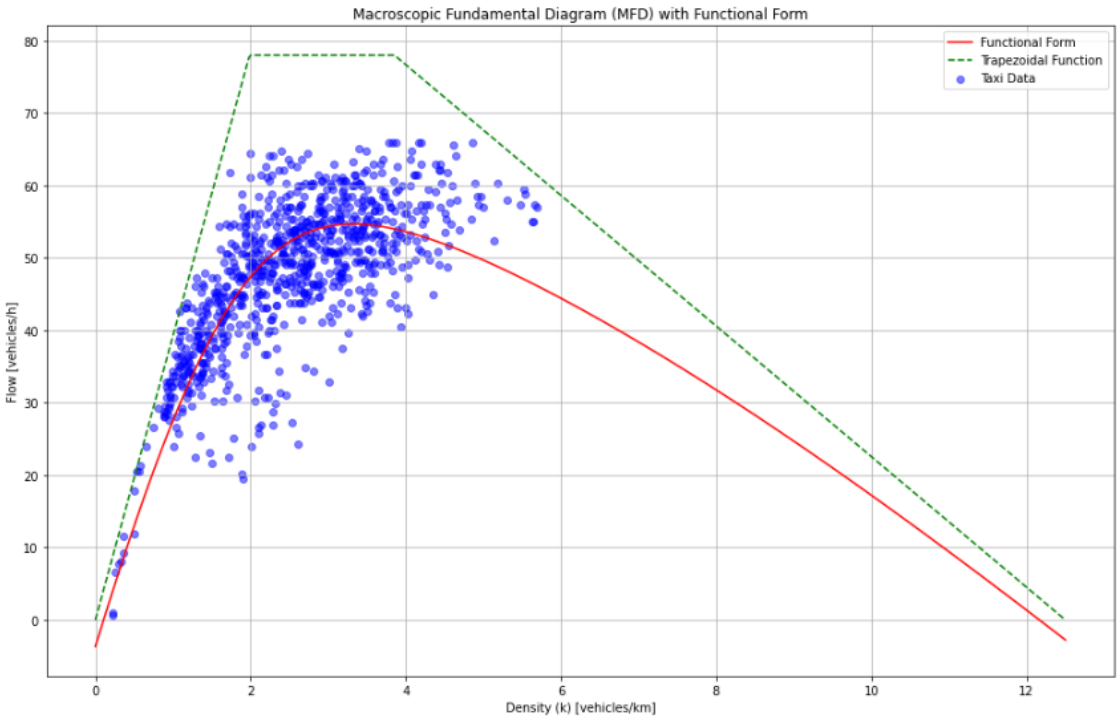

Figure 10: Macroscopic Fundamental Diagram (MFD) with Functional Form

4.4 Travel demand generation method based on travel distance distribution function

4.4.1 Traffic zone division

Initially, the nodes and edges data from the Haidian District road network were extracted from OpenStreetMap, with the latitude and longitude of the nodes serving as features for clustering. The district comprises a total of 8,563 nodes and 19,133 edges. Subsequently, the number of clusters was defined (1,000 clusters) and the K-means algorithm was used to cluster the nodes. The value of the number of clusters in this step is pre-set because compared with the higher clustering quality, this study focuses more on using clustering methods to simplify calculations. At the same time, it is considered that a single zone should not contain too many nodes to avoid errors in subsequent random assignment. The criteria for clustering is the proximity of nodes in Haidian District.

The road network data of Haidian District in Beijing is contained in OpenStreetMap, including the roads and node attributes, where the road types of edges within the district include primary, secondary, tertiary, residential, trunk, and some unclassified roads. By sampling each type of road and comparing it with satellite maps, it has been confirmed that all road types are accessible to motor vehicles.

By clustering the nodes, the complex network of interactions within the traffic system is simplified. Each cluster is treated as a single entity, with minimized intra-cluster distances and reasonably approximated inter-cluster distances that reflect travel patterns. The reduction in computational complexity can be estimated by comparing the number of direct calculations required between nodes before and after clustering. Prior to clustering, the calculation of all possible combinations of distances between nodes is required. With 8,563 nodes, the number of distance calculations needed totals 36,686,653. However, when these nodes are divided into 1,000 clusters, only the distances between the cluster centers need to be calculated, reducing the required number of calculations to 499,500. This represents a reduction in computational complexity of approximately 98.64%. This method not only streamlines the computation of potential travel demands but also enhances the efficiency of the simulation models used for traffic forecasting and management. Figure 11 shows the clustering results.

In Section 4.4.3, it will explain how to allocate the traffic volume of a zone to the nodes within the zone. This step is highly stochastic, and when a zone contains too many nodes (i.e., fewer zones are divided), although it will further reduce the computational complexity, it also increases the occurrence of errors. How to balance between computational complexity and errors is a problem, and the actual error value cannot be determined because it is impossible to calculate the true value. Calculating the true value would require an excessive amount of computation, and if the true value were calculated, the division of zones would also lose its meaning.

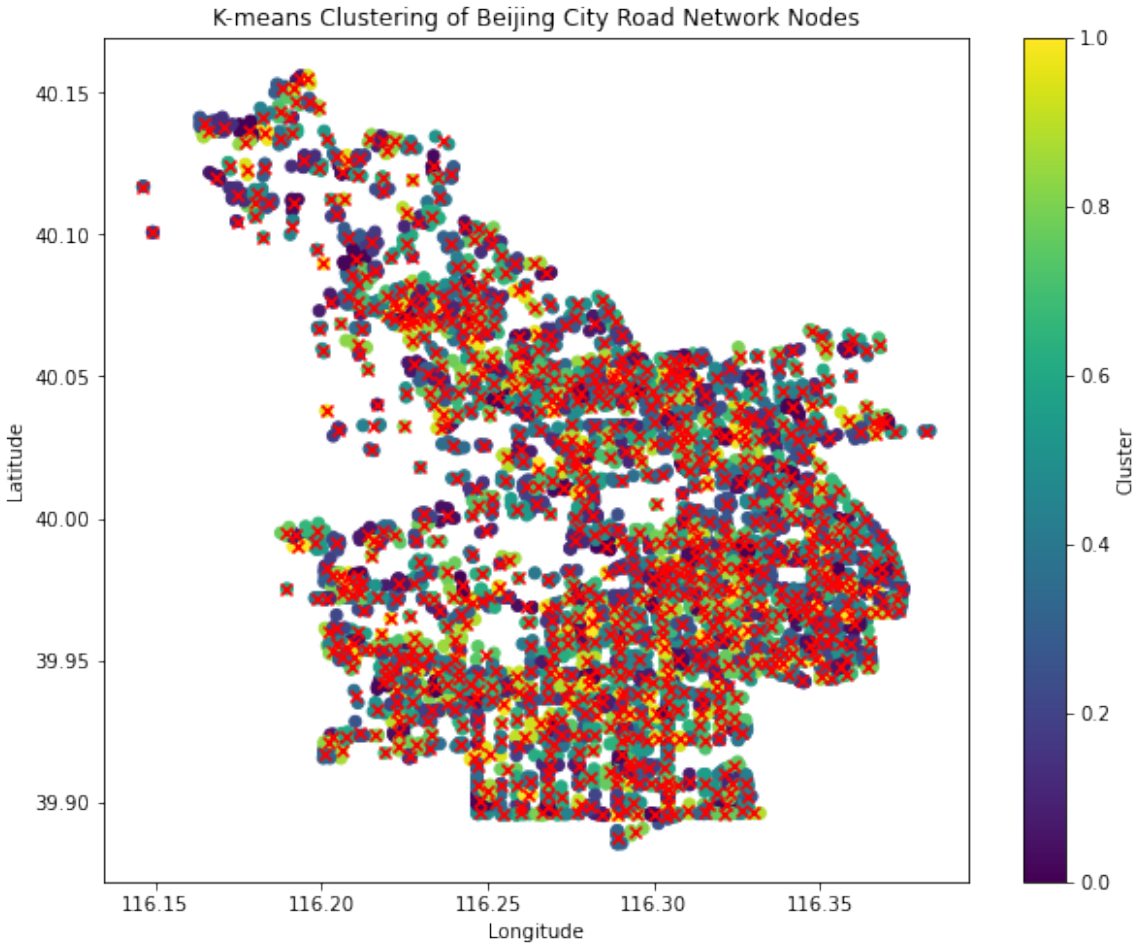

Figure 11: K-means Clustering of Haidian District Road Network Nodes

In the process of traffic area division, the number of clusters is determined primarily to reduce computational complexity. That is, the main objective is to make the computation more efficient rather than focusing on the precision or high quality of the clusters, such that it is not necessary to ensure that the total within-cluster variation (or error) is minimized. Nevertheless, monitoring and evaluating the quality of clusters is still meaningful. This means that although the primary purpose of clustering is not to achieve extremely high precision, it is still necessary to pay attention to and evaluate the performance and effectiveness of these clusters.

Figure 12 displays the graphical representation used in the Elbow Method.

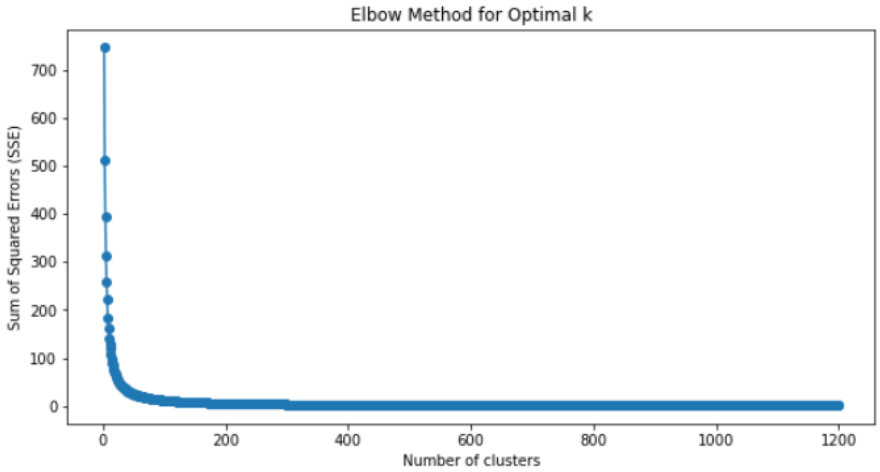

Figure 12: Elbow Method for Optimal k

Figure 13 displays the image used to assess the quality of the cluster using the Silhouette Coefficient method.

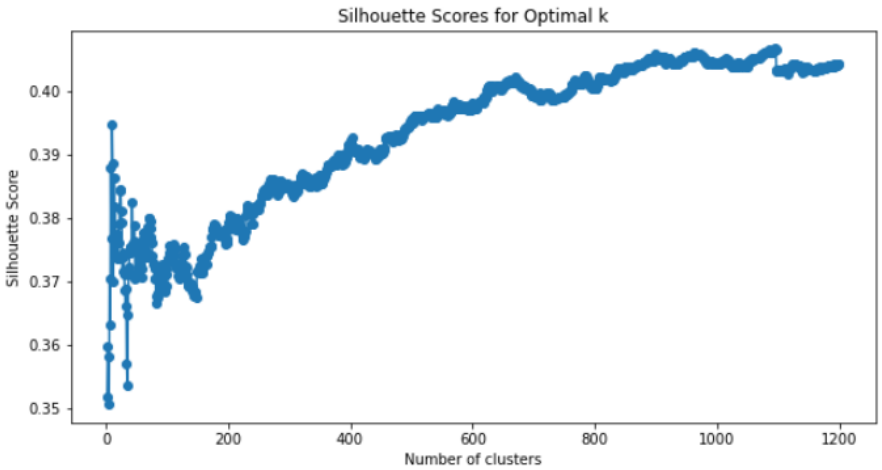

Figure 13: Silhouette Scores for Optimal k

Observations from Figures 12 and 13 reveal distinct phenomena when the number of clusters k is large. The Elbow Method rapidly approaches zero and gradually stabilizes, whereas the silhouette coefficient method initially increases, then decreases, and subsequently exhibits a fluctuating increase. At k=900, the results gradually converge, with the silhouette score stabilizing at approximately 0.41. When k equals 1000 (preset value), the Silhouette Scores have converged, making it a feasible approach. With increasing number of clusters k, the two methods exhibit different phenomena. The hypothesis for this are as follows: The Elbow Method focuses on evaluating changes in the sum of squares within the cluster (SSE) as a function of the number of clusters k. As k increases, the number of data points in each cluster decreases, thus reducing the average distance from each point to its cluster center, leading to a decrease in SSE. Once k reaches a certain level, although further increases in k continue to reduce SSE, the rate of this reduction significantly decreases because most of the data points are already sufficiently close to their cluster centers. Consequently, the SSE reduction curve quickly approaches zero and gradually stabilizes, indicating that additional clusters have diminishing returns on improving the model.

The Silhouette Coefficient Method assesses cluster quality by integrating measures of cohesion and separation. When the number of clusters k is small, there may be large clusters with broad coverage, where internal variations are substantial (low cohesion), but separation between different clusters may be high. As k increases, the internal tightness of the clusters improves, leading to an initial rise in the silhouette coefficient. However, as k continues to increase beyond a certain point, some clusters may begin to overfit small or outlier data points, which degrades the cohesion within these clusters and reduces their separation from other nearby clusters,

causing the silhouette coefficient to decrease. When k is further increased, excessive clustering might result in clusters being too close to each other, reducing separation. With an increase in k, clustering might begin to overfit the data, where the divisions become too granular, reflecting random fluctuations in the data rather than actual, meaningful patterns. In such scenarios, some clusters may contain only a few points, which might temporarily enhance the cohesion of these clusters but are detrimental to overall separation between groups.

4.4.2 Probability density calculation

The T-Drive dataset provides access to taxi trajectory data, but due to the lack of specific information about passenger pick-up and drop-off locations and times, it is not possible to directly extract trip data. To calculate the travel probabilities in Haidian District , the road network's probability density function in Beijing city can be utilized for analysis. It is necessary to verify whether the travel patterns in Haidian District are significantly similar to those of Beijing as a whole.

First, a qualitative analysis is conducted. The road network in Haidian District includes Beijing's Third, Fourth, and Fifth Ring Roads, encompassing both the central urban and suburban travel characteristics of the city. This indicates that Haidian District reflects the major travel traits of Beijing to a certain extent. Moving on to a quantitative analysis, it requires segmenting the complete taxi trajectory data of Haidian District and whole Beijing into individual "trips." Assume that if a taxi's driving speed falls below 1 km/h during any time interval, it is considered that the taxi has made a stop in that interval. Each segment of travel between two stops is recorded as a "trip." These "trips" may not necessarily represent trips made due to picking up or dropping off passengers, as there could be other reasons for taxi stops, such as waiting at traffic lights. Therefore, it is not feasible to directly use the probability distribution of the generated "trip" lengths to calculate the travel probabilities between different zones. This method is only used to observe the similarity in travel patterns between Haidian District and Beijing city. Figure 14 shows a comparison of the Probability Distribution of Trip Lengths between Haidian District and Beijing City.

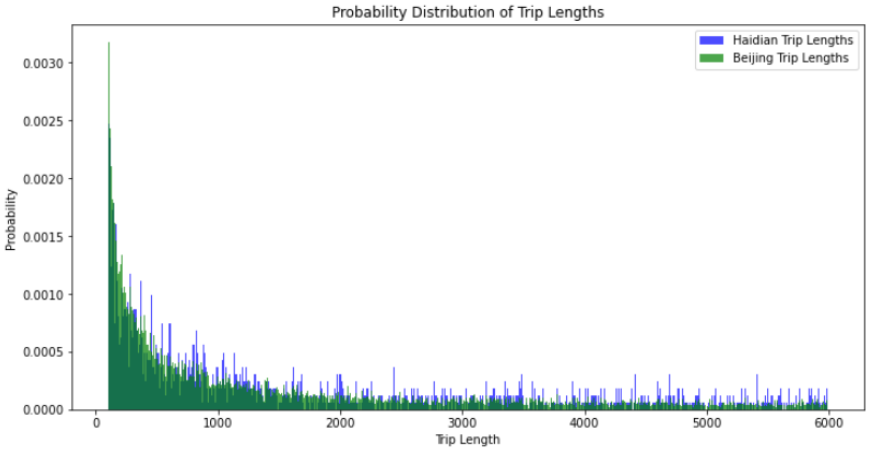

Figure 14: Probability Distribution of Trip Lengths in Beijing City and Haidian District

The Wasserstein distance (Panaretos and Zemel, 2019) is used to measure the difference between two probability distributions. The Wasserstein distance $W_p(\mu, \nu)$ between two probability distributions μ and ν is defined as:

$$W_p(\mu, \nu) = \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{R^d \times R^d} \|x - y\|^p \, d\gamma(x, y) \right)^{\frac{1}{p}} \tag{23}$$

where: $\Gamma(\mu, \nu)$ is the set of all joint distributions with marginals μ and ν, $\|x - y\|$ is the distance between points x and y in R^d, $p \geq 1$ is a parameter that defines the order of the distance

By subtracting the ratio of the Wasserstein distance to the total data range from 1, the distance is transformed into a similarity measure: the smaller the distance relative to the data range, the closer this value is to 1. After calculation, it can be concluded that the similarity in the Probability Distribution of Trip Lengths between Beijing and Haidian District is 95.40%, indicating that the travel patterns in Haidian District are significantly similar to those of Beijing as a whole.

In the process of generating travel demand based on distance, an assumption was made that the geographical environment, population, transportation network, and vehicle travel distribution within Haidian District are relatively uniform. However, this assumption clearly contradicts the population characteristics of Haidian District mentioned in Section 4, as well as the taxi trajectory map of Haidian District in Section 4.3 (Figure 23).

In this study, efforts are made to ensure that the vehicle trajectories in subsequent simulations align with the taxi operation trajectories shown in Figure 23, where taxi activity is dense in the southeastern area and relatively sparse in the northwest. However, within the road network of Haidian District, the distribution of nodes is relatively uniform and does not clearly reflect regional differences corresponding to the taxi trajectories. The density distribution lacks clear patterns, although there are several localized clusters as shown in Figure 15. The legend in the plot represents the "density" of nodes in the Haidian District, which is calculated based on the number of nearby points within a radius of 0.01 degrees of latitude/longitude. Therefore, the units of the legend are "number of nodes within a 0.01-degree radius".

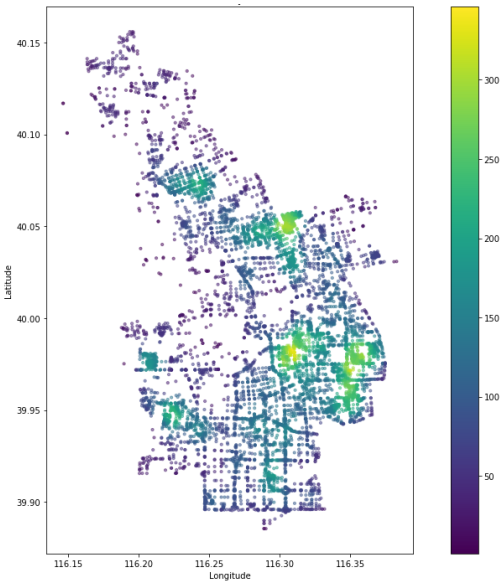

Figure 15: Node Density in Haidian District

Subsequently, it is necessary to introduce a correction factor, α, to ensure that the outcomes of trip generation align with the existing taxi operation trajectories data. Initially, the taxi GPS data points must be categorized based on their density, in this case, use the K-means clustering method. When k=5, it achieves the highest cluster quality. as illustrated in the figure 17.

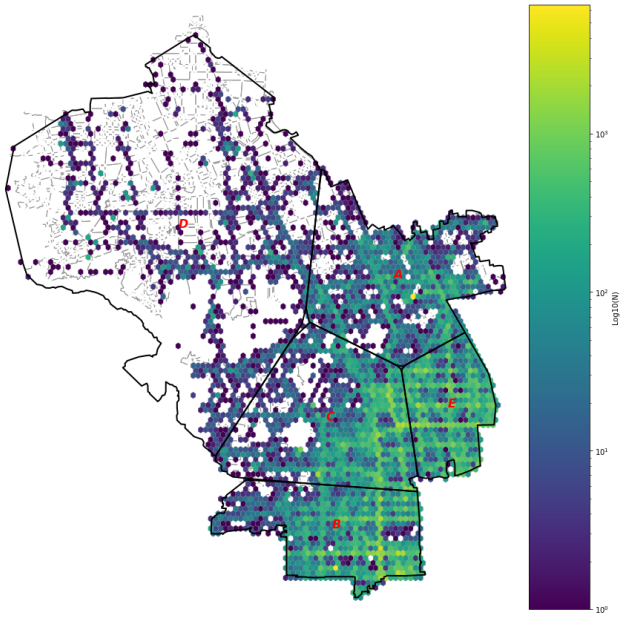

Figure 16: Map of Haidian District Divided Based on The Density of GPS Data Points

Table 9 presents information on the number of taxi GPS data points and nodes in different classified regions, as well as their respective proportions. The table reveals significant differences in the distribution of taxi GPS data points and nodes across different clusters. Although Cluster B has the highest number of data points (32.2%), it has relatively fewer nodes (15.8%). In contrast, the proportions of data points and nodes in Cluster C are relatively close, indicating a higher density of nodes in this area. Cluster D, despite having the lowest share of data points (2.2%), has a relatively higher number of nodes (17.8%), suggesting lower traffic flow but more extensive infrastructure in this region. Overall, there is a notable mismatch between node distribution and traffic flow in certain areas.

Table 9: Cluster Data Points and Nodes Information

Cluster	Number of GPS Data Points	Number of Nodes
A	56,079 (11.9%)	1,908 (22.5%)
B	152,455 (32.2%)	1,343 (15.8%)
C	121,037 (25.6%)	2,105 (24.8%)
D	10,379 (2.2%)	1,510 (17.8%)
E	132,804 (28.1%)	1,611 (19.0%)

Therefore, it can be argued that demand generation should, to some degree, follow the density of taxi GPS data points, which serve as a proxy for actual travel patterns. This suggests that areas with a higher density of GPS points are likely to experience greater travel demand, as these data points reflect the intensity of taxi operations and real-world traffic flow. Incorporating GPS data density into demand models enables a more precise depiction of the spatial distribution of transportation demands.

Moreover, aligning demand models with taxi GPS data can enhance the precision of traffic simulations, particularly in urban areas where taxi usage represents a significant portion of total traffic.

In the process of traffic area division, clustering directly using GPS data points seems to avoid the introduction of a correction factor. However, this approach has some problems. Firstly, the traffic demand used in simulations is based on the network nodes of Haidian District, not on taxi GPS data points. If GPS data points are used for clustering, and then the clustering results are mapped onto the network, although it could solve the aforementioned problem, this process might degrade the quality of the clustering. It cannot guarantee that each zone generated from the GPS data points contains a node. Moreover, using the elbow method and silhouette coefficient for evaluating clustering quality might also lose its relevance. Therefore, it is necessary to use network nodes for clustering and then apply a correction factor.

The correction factor α_i for cluster i is defined as the ratio of the proportion of the Number of Data Points to the proportion of the Number of Nodes within the same cluster. Mathematically, it can be expressed as:

$$\alpha_i = \frac{\text{Proportion of Number of Data Points in Cluster } i}{\text{Proportion of Number of Nodes in Cluster } i} = \frac{\frac{N_{\text{data},i}}{N_{\text{data, total}}}}{\frac{N_{\text{nodes},i}}{N_{\text{nodes, total}}}} \tag{24}$$

Where: $N_{\text{data},i}$ is the Number of Data Points in cluster i, $N_{\text{nodes},i}$ is the Number of Nodes in cluster i, $N_{\text{data, total}}$ is the total Number of Data Points across all clusters, $N_{\text{nodes, total}}$ is the total Number of Nodes across all clusters.

The correction factor α_i is defined to adjust the traffic demand within each cluster, ensuring that the ratio of data points to nodes remains balanced during the subsequent simulation process. Specifically, α_i represents the ratio of the proportion of data points to the proportion of nodes within each cluster.

$\alpha_i = 1$ indicates that the proportion of nodes in the area is consistent with the proportion of GPS data points in the area relative to the total number of data points. If $\alpha_i > 1$, it suggests that the GPS data point density in the region is higher, indicating a potential need for increased traffic demand. Conversely, $\alpha_i < 1$ implies that the node density is higher in the region, and the traffic demand is relatively lower.

Table 10: Values of Correction Factor α_i for Each Cluster

Cluster i	α_i
A	0.53
B	2.04
C	1.03
D	0.12
E	1.48

Previously, the traffic volume ratio between zones p and q, R_{pq}, was calculated based on equation 4. Now, incorporating the correction factor α_i, where i represents the cluster to which the starting point p belongs, the adjusted traffic volume ratio can be expressed as:

$$R_{pq}^{\text{adjusted}} = \alpha_{p,i} \cdot R_{pq} \tag{25}$$

Where, $\alpha_{p,i}$ is the correction factor for the cluster i, associated with the starting point p, and R_{pq} is the traffic volume ratio from zone p and q.

Introducing the correction factor $\alpha_{p,i}$ effectively captures the asymmetry in traffic demand between regions, a common phenomenon in transportation systems. The traffic demand from region a to region b often differs from the demand from region b to region a. This asymmetry in traffic demand not only reflects the different functions of these regions but may also be influenced by varying traffic periods, seasonal changes, and other factors. Therefore, by introducing the correction factor $\alpha_{p,i}$, the model can be more precisely adjusted to reflect the complex and asymmetric traffic demand distribution between regions due to functional, temporal, or other factors. This allows the model to more accurately capture the traffic flow characteristics between regions. The next step is to normalize the adjusted traffic volume ratio R_{pq}^{adjusted}, ensuring that the total traffic demand between all regions remains consistent. Normalization adjusts each R_{pq}^{adjusted} by comparing it to the total adjusted demand across all regions, calculating the relative proportion for each region. The formula for normalization is given as: $R_{pq}^{\text{normalized}} = \frac{R_{pq}^{\text{adjusted}}}{\sum_{p,q} R_{pq}^{\text{adjusted}}}$. After this process, each $R_{pq}^{\text{normalized}}$ will have a value between 0 and 1, with the total sum across all regions being 1. This ensures that the model remains balanced and comparable, preventing disproportionate traffic demands across regions.

4.4.3 Demand generation

In Section 4.3.2, it was calculated that the total number of private car trips in Beijing, Haidian District is approximately 0.37 billion annually, with around 26.46 million taxi orders, resulting in a total of approximately 0.396 billion car trips annually. Thus, the daily car trips for Haidian District residents are approximately 1.09 million. The next step will involve distributing this total across different zones to further analyze the traffic

demand and flow within the district.
To distribute the total daily car trips $T = 1,090,000$ across different zones, the traffic demand between each zone can be calculated using the traffic demand ratio R_{p_iq} derived earlier. The traffic demand V_{p_iq} between zone p_i and zone q is given by the following equation 26:

$$V_{p_iq} = R_{p_iq}^{\text{normalized}} \times T \tag{26}$$

Where:

V_{p_iq} is the estimated daily traffic volume from zone p_i to zone q.

$R_{p_iq}^{\text{normalized}}$ is the normalized traffic volume ratio from zone p_i to zone q.

$T = 1,090,000$ represents the total daily car trips in Haidian District.

The partial data of the generated results are shown in Table 11.

Table 11: Sample of Generated Traffic Demand

	Cluster_0	Cluster_1	Cluster_2	Cluster_3	Cluster_4	Cluster_5	...	Cluster_999
Cluster_0	0	1	0	1	1	0	...	1
Cluster_1	1	0	0	1	0	7	...	3
Cluster_2	0	0	0	0	2	0	...	1
Cluster_3	1	1	0	0	0	1	...	0
Cluster_4	1	0	2	0	0	0	...	7
Cluster_5	0	7	0	1	0	0	...	0
...	...	...	...	...	...	...	...	...

After generating the traffic demand between traffic zones, the traffic volume T_{pq} is randomly distributed across all node pairs (i, j) , completing the daily demand generation.

Next, to generate the demand data required for Fleetpy simulation, which consists of a series of requests with time labels and origin-destination points, the daily taxi travel distribution obtained from taxi GPS data in Haidian District will be utilized. Based on this distribution, the daily demand can be appropriately allocated. Figure 17 illustrates the distribution of requests throughout a single day.

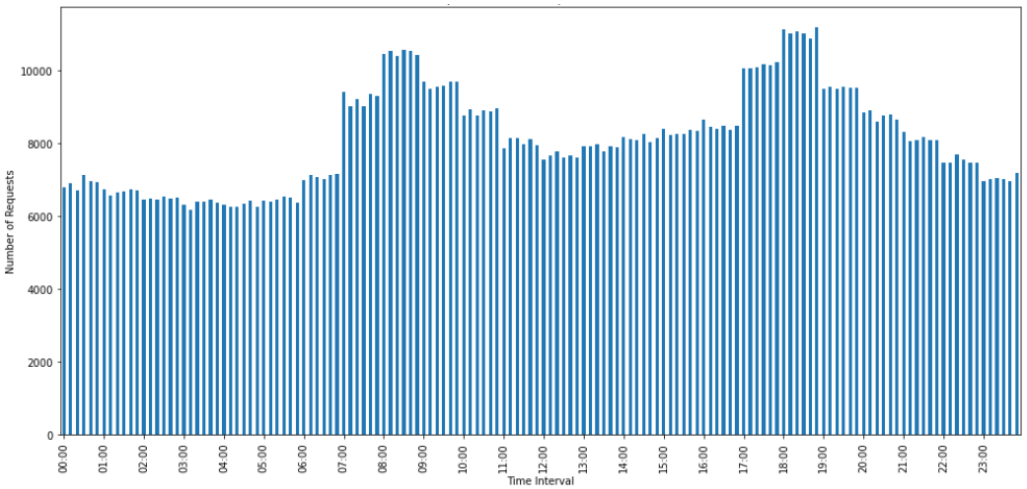

Figure 17: Traffic Request Distribution

4.5 Questionnaire

This study employs an online survey to investigate the preferences of residents in Haidian District, Beijing, for SAVs and POVs. The respondents targeted are permanent residents of the Haidian District. A total of 400 valid questionnaires were collected in this survey.At the beginning of the questionnaire, an introduction to SAVs is provided, explaining that these vehicles operate autonomously without a human driver and are shared among multiple users, similar to existing ridesharing services. The survey introduces Apollo Go, an operational autonomous driving service platform in China that does not currently offer carpooling services. This introduction aids respondents unfamiliar with SAVs in gaining a preliminary understanding. Apollo Go operates 24/7 and is accessible through multiple platforms, including a WeChat mini-program, Baidu Maps, the Apollo Go app, and the Baidu app. The service has a maximum capacity of three passengers, only accommodates passengers in the back seat, and includes a physical barrier between the front and back seats. Apollo Go's pricing includes a base fare of 18 RMB for the first kilometer and two minutes, followed by 2.7 RMB per kilometer, 0.5 RMB per minute, and an extra 0.9 RMB per kilometer for trips over 10 kilometers. Figure 18 shows the autonomous driving car in operation by the Apollo Go.

Figure 18: Autonomous Driving car in Operation by the Apollo Go

The questionnaire is divided into two main sections. The first section focuses on basic knowledge about SAVs and includes four questions: it first asks respondents whether they have prior knowledge of SAVs and whether they have actual experience using such vehicles. Additionally, it inquires about the car ownership status within the respondents' households, offering choices between "no car," "one car," and "two or more cars." Lastly, it questions respondents about their primary travel purposes in daily life. Appendix B display the statistical results regarding the travel characteristics of Haidian District residents and their understanding of SAVs.

The second part of the survey focuses on the respondents' subjective perceptions. In this study, participants rate their preferences based on three attributes: Comfort, Reliability, and Environmental Impact, tailored to their personal travel needs. Additionally, this section gather information on the respondents' primary travel times during a day. Ultimately, respondents are asked to consider these factors collectively to make an informed choice between using SAVs and POVs for their daily travel. The results of this part of the survey will be used to develop a mode choice model, which will be described in model construction section. Figure 19 shows the distribution of travel time among the respondents. It is observed that the respondents' travel time is mainly distributed between 30 and 70 minutes. The data is concentrated at specific points such as 30, 35, 45, 50, 55, and 60 minutes. One possible reason for this phenomenon is that the respondents' perception of their actual travel time is not very precise. When recalling or estimating their travel time, they tend to choose more familiar, rounded time points. This bias may stem from the ambiguity in people's perception of time, leading them to prefer rough estimates in multiples of 5 or 10 minutes. Figure 20 presents the statistical results of the subjective perception attributes (Comfort, Reliability, and Environmental Impact). The complete questionnaire and survey data can be found in Appendix A and C. When the value of the Choice attribute is 1, it indicates the selection of SAV. When the value of the Knowledge of SAVs attribute is 1, it indicates familiarity with SAV. When the value of the Experience with SAVs attribute is 1, it indicates having experienced SAV.

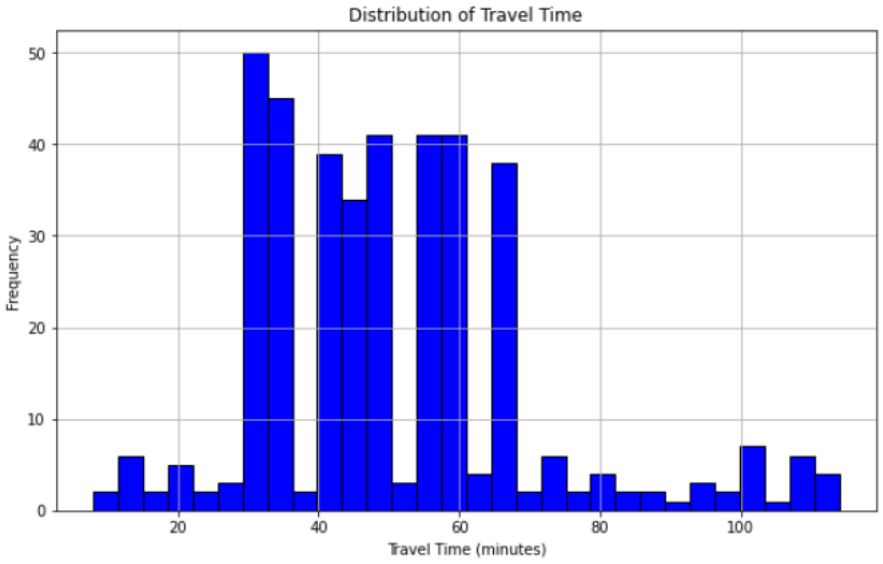

Figure 19: Distribution of Travel Time

Figure 20: Distribution of Comfort, Environmental Impact and Reliability

4.6 Mode choice model

A logit model is constructed based on the travel times reported by the respondents, their ratings for Comfort, Reliability, and Environmental Impact, as well as their choices between SAVs and POVs. During this process, the 400 valid questionnaires are randomly divided into two groups, each consisting of 200 questionnaires. One group serves as the training set, while the other is designated as the test set.

Subsequently, parameters will be estimated using the method of Maximum Likelihood Estimation. Table 12 is a summarizing the estimated parameters of the logit model. The function of the logit model and the explanation of its corresponding parameters can be found in Section 3.3.

Table 12: Summary of Estimated Parameters

Parameter	**Estimated Value**
β_0	-6.8582
β_1	-0.0593
β_2	1.0996
β_3	2.0992

Upon completion of the model construction, it is imperative to assess the accuracy of the model, for which data from the test set is utilized. By applying the test set data to the model, the outcomes can be tabulated in a confusion matrix as shown below. This matrix helps in visualizing the performance of the model by comparing the actual versus predicted classifications.

Predicted \Actual	**0**	**1**
0	84	14
1	15	87

The rows represent the actual categories, while the columns represent the predicted categories by the model. The cell at the intersection of the 'Actual 0' row and 'Predicted 0' column shows the number of true negatives (84), indicating the instances where the model correctly predicted the negative class. Conversely, the 'Actual 1' row and 'Predicted 1' column cell shows the number of true positives (87), where the model accurately predicted the positive class.

The off-diagonal cells represent the instances of incorrect predictions, where the model predicted '1' instead of the actual '0' (15 instances, false positives) and '0' instead of the actual '1' (14 instances, false negatives).

This accuracy score of 0.855 signifies that the model correctly predicts the outcome 85.5% of the time, which demonstrates a high level of performance.

By conducting accuracy tests on various attribute combinations, this study aims to identify the utility function that maximizes predictive accuracy. The results are shown in Table 13. When selecting attribute combinations, the model that included travel time, comfort and environmental impact exhibited the highest predictive accuracy, reaching 0.855.

Table 13: Model Accuracies Based on Different Attributes

Attributes	Accuracy
travel_time, comfort, environmental_impact	0.855
comfort, travel_time	0.73
comfort, environmental_impact	0.815
travel_time, environmental_impact	0.825
environmental_impact	0.805
travel_time	0.615
comfort	0.665

4.7 Simulation

4.7.1 Input data

In the application of the Fleetpy simulation framework presented in this paper, three types of input data are essential: request data, network data, and vehicle data. The request data is obtained using the "Travel Demand Generation Method Based on Travel Distance Distribution Function," as described in Chapter 2. The network data includes nodes and edge data for the Haidian District, extracted from OpenStreetMap. For the vehicle data, the critical parameter to determine is maximum_passengers. Given that Shared Autonomous Vehicles (SAVs) do not require a driver, they can ideally offer an extra seat for passengers. Therefore, in this simulation, the number of seats is set to five, compared to vehicles that need a driver. Table 14, 15 and 16 illustrates the data format and samples for request Data and network Data.

Table 14: Sample of Request Data

rq_time	start	end	request_id
0	2195	1621	1
0	1827	984	2
0	6603	5161	3
0	1248	3859	4
0	2938	6499	5
0	8285	6338	6
0	4549	1845	7
0	6595	4681	8
0	7775	3508	9
0	6551	5670	10
0	1025	5821	11
1	3276	634	12
1	2040	3231	13
1	6609	875	14
1	1248	6546	15
1	6398	5584	16
1	1555	2094	17
1	1025	977	18
1	3556	5317	19
1	6588	5755	20

Table 15: Sample of Node Data

node_index	is_stop_only	pos_x	pos_y
0	FALSE	116.3192982	39.9096927
1	FALSE	116.3471527	39.9638588
2	FALSE	116.3473067	39.9671179
3	FALSE	116.3638695	39.966728
4	FALSE	116.315387	39.9062149
5	FALSE	116.3150712	39.9062164
6	FALSE	116.3107059	39.9062558
7	FALSE	116.3227946	39.9096329
8	FALSE	116.3639092	39.9662639
9	FALSE	116.3501458	39.9476852
10	FALSE	116.3490214	39.9542929
11	FALSE	116.3488127	39.9599117
12	FALSE	116.2686859	39.919551
13	FALSE	116.2682644	39.9206764
14	FALSE	116.3086064	39.9311234

Table 16: Sample of Edge Data

from_node	to_node	distance	travel_time	source_edge_id
0	1635	73.5	5.292	143599420
0	1616	576.737	69.20844	60500562
0	3589	314.812	161.9033143	143599420
1	2073	447.448	230.1161143	136867391
2	78	201.843	103.8049714	159794802
3	8	51.716	26.5968	27262001
3	3582	106.493	7.667496	163057028
4	1765	35.151	4.21812	4484480
4	5	26.937	3.23244	172430546
5	657	15.032	7.730742857	29577449
5	6	372.374	26.810928	173345216
6	656	13.777	0.991944	42426723
6	1089	143.521	10.333512	172805318
6	3599	33.414	2.405808	172805323
7	1648	135.762	9.774864	143599420

4.7.2 Simulation scenarios

In the initial scenario, the fleet size is set to 7,865, which corresponds to the total number of ride-sharing and taxi vehicles in the Haidian district in reality. This value comes from the T-Drive dataset, representing the number of taxis with active trajectories within Haidian District over the course of a single day in the dataset. This setting aims to completely replace existing Mobility on Demand (MoD) services with SAVs during the simulation. The simulation period is one day (86,400 seconds). Subsequently, the initial fleet size is adjusted in increments of 1,000 vehicles. The fleet sizes simulated range from 865 to 15,865 vehicles, across 16 different scenarios. Simulations are conducted for these 16 different fleet sizes using both immediate decision and batch offer matching strategies.

4.7.3 Output data

The Fleetpy simulation results provide two key sets of statistical data: User Statistics and Vehicle Trajectory Statistics, which record the system's operational performance from the perspectives of passengers and vehicles, respectively, the samples are shown in Table 17 and Table 18..

The User Statistics table primarily tracks each passenger's request information, including request ID, request time, earliest pickup time, and the passenger's start and end locations. It also includes the direct route travel time and distance, which are used to calculate the shortest trip from the start to the end point. Additionally, the table records the vehicle ID assigned to the request and the actual pickup and dropoff times.

The Vehicle Trajectory Statistics focus on the detailed operational data of each vehicle in the simulation. This includes the vehicle ID, its current status (e.g., en route or picking up passengers), whether it is locked (i.e., whether it has already been assigned to a request), and the vehicle's trip start and end times and positions. It also tracks the actual distance driven by the vehicle, the passenger request IDs currently on board, the number of passengers, and the request IDs of passengers who are about to alight.

In the Vehicle Trajectory Statistics, the "occupancy" column provides the real-time occupancy status of the vehicles. The Pooling Ratio is calculated based on the vehicle occupancy. Trips with occupancy greater than or equal to one are extracted from the table. The average of the occupancy is taken and subtracted by 1 to obtain the Pooling Ratio value. The Extra Mileage Ratio is also calculated based on the occupancy of the trips. In the Vehicle Trajectory Statistics, trips with an "occupancy" value of 0 are considered empty, while the rest are occupied. The total distance traveled by trips with an occupancy of 0 is divided by the total distance traveled by trips with an occupancy greater than 0 to obtain the Extra Mileage Ratio.

Table 17: Sample of User Statistics

request_id	rq_time	earliest_pickup_time	start	end	direct_route_travel_time
1208	360	360	7126;-1;-1	313;-1;-1	354.143376
1021	360	360	5787;-1;-1	1038;-1;-1	324.34
1274	480	480	1533;-1;-1	1532;-1;-1	250.74
1006	360	360	1025;-1;-1	4048;-1;-1	373.74

direct_route_distance	vehicle_id	pickup_time	dropoff_time
2459.329	623	405.68256	789.825936
2252.39	39	469.17	823.51
1741.22	92	484.77	804.76
2595.40	148	385.61	819.35

Table 18: Sample of Vehicle Trajectory Statistics

vehicle_id	status	locked	start_time	end_time	start_pos	end_pos
1247	route	FALSE	897.116208	1054.389264	7516;-1;-1	7460;-1;-1
1251	route	FALSE	733.278672	1055.050704	1455;-1;-1	7655;-1;-1
1253	route	FALSE	850.528656	1034.794656	5568;-1;-1	4006;-1;-1
1253	boarding	TRUE	1034.794656	1064.794656	4006;-1;-1	4006;-1;-1

driven_distance	rq_on_board	occupancy	rq_alighting	route	trajectory
1092.174	1193	1		...	...
2234.528	1636	1		...	...
1279.625	1786;1719	2		...	...
0	1786;1719	2	1786	...	...

5 Experimental results

This section introduces the following two parts: firstly, the results and analysis of the processing of taxi GPS trajectory data in this study. Secondly, it presents and analyzes the simulation results for different fleet sizes under two matching strategies from the perspectives of four KPIs and the MFD.

5.1 Research data analysis

After processing the data, an analysis was conducted to understand how taxis operate within Haidian District. The initial findings from this analysis indicated that the average sampling interval between consecutive GPS recordings stood at approximately 3.80 minutes. Figure 21 and Figure 22 provides a visual representation of the distribution of these sampling intervals and distances between consecutive points on a single taxi id.

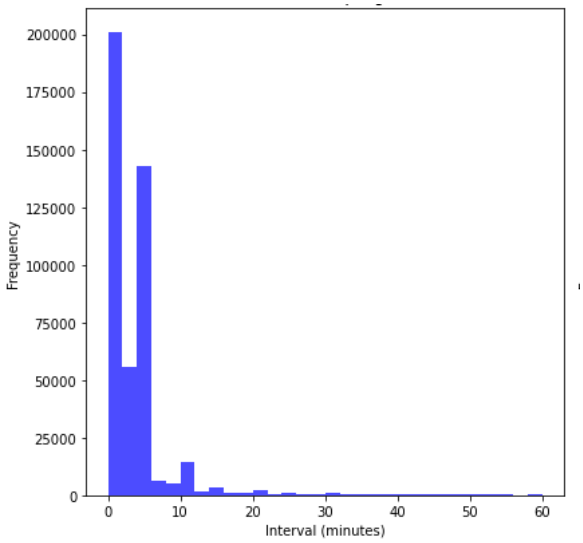

Figure 21: Distribution of Time Intervals

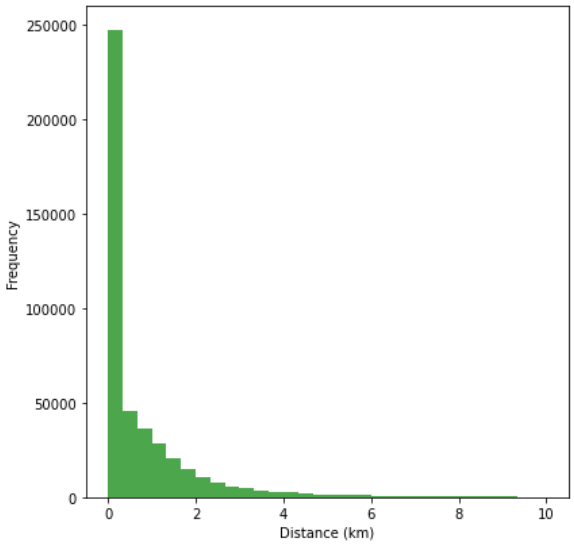

Figure 22: Distribution of Distances

The GPS data is subjected to an visualization process, wherein taxi GPS points are precisely mapped onto the urban road network. This task is accomplished using Geographic Information System (GIS) technology, which integrates raw GPS coordinates with specific road network layers. As shown in the Figure23. This method allows for a visual representation of the actual routes taken by taxis within Haidian District using continuous GPS data points, providing a robust visual foundation and data support for further traffic flow analysis and urban traffic planning.

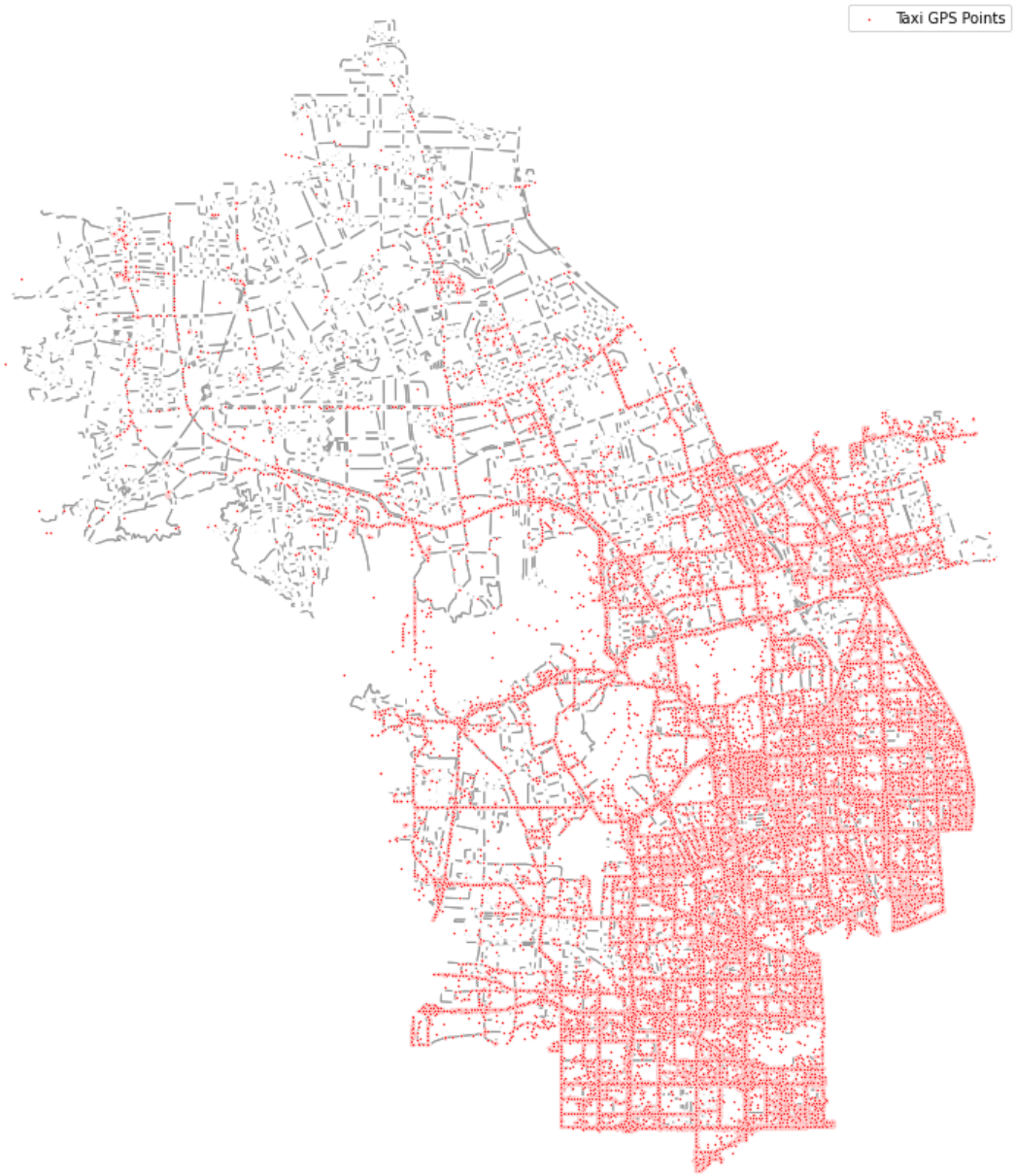

Figure 23: Taxi GPS Data Points and Road Network Map of Haidian District, Beijing

The comprehensive analysis of taxi GPS trajectories and the road network in Haidian District reveals that taxi operations are predominantly concentrated in the southeastern area, where there is a high frequency of population and commercial activities. This region houses most of the district's higher education institutions, research facilities, and large commercial centers such as Zhongguancun, which serve as primary sources of demand for taxi services. Additionally, the proximity of the southeastern part to the center of Beijing enhances its commercial activities and mobility of residents, further increasing the demand for taxi services.

In contrast, the northwestern part of the district, being more suburban, features fewer commercial facilities and more dispersed residential areas. Coupled with predominantly mountainous terrain and fewer, more complex roads, these factors collectively result in sparser taxi operations.

Moreover, considering the urban layout of Beijing, the southeastern part of Haidian District is closer to the city center, while the northwestern part is more suburban. the southeastern area benefits from superior transportation infrastructure, including a higher density of subway stations and a more extensive public transit network, as well as better road conditions. On the other hand, the transportation infrastructure in the northwest is relatively underdeveloped, which not only limits the accessibility of taxi services but also affects the willingness of taxi drivers to operate in this area. The limited passenger flow and inconvenient transportation conditions

in the northwest may lead to lower earnings.

5.2 Simulation results

This section evaluates service level of SAVs and road congestion level under various fleet sizes and matching strategies based on four key performance indicators (KPIs). The result encompasses a total of 32 simulation runs involving two distinct strategies and each KPI is fitted separately. Through in-depth analysis of these indicators, it is possible to gain a comprehensive understanding of operational performance across different configurations. These indicators include the matching success rate, average waiting time, carpooling ratio, and additional mileage ratio, all of which are used to assess and compare the effectiveness of different strategies in practical application. This evaluation helps to reveal the strengths and weaknesses of various matching strategies at different fleet sizes, thus providing a theoretical basis for formulating more effective operational strategies.

5.2.1 Matching success rate

Figure 24 displays the matching success rates for the "Immediate Decision" and "Batch Offer" strategies across various fleet sizes, along with their fitting results. When choosing a fitting method, if the data shows a curved trend, polynomial regression allows for fitting these curved patterns in the data(Ostertagová, 2012). For the Matching Success Rate, which shows a clear trend of gradually increasing and the slope decreasing until it approaches a specific value, this characteristic fits a Logistic Regression model. Equation 27 represents the standard form of the logistic function. When comparing the "Immediate Decision" and "Batch Offer" matching strategies, it is evident that the two strategies exhibit different performances and characteristics in terms of matching success rates. The logistic fitting parameters for the Immediate Decision strategy indicate that as fleet size increases, the asymptotic upper limit of the matching success rate approaches 97.77%, with a mean squared error of 1.7823, suggesting a rapid improvement in matching success rates as observed in the graph. In contrast, the Batch Offer strategy, starting with a slower increase in matching success rate, has a slightly lower fitting upper limit of approximately 96.75% and a model fitting mean squared error of 1.0136. At a fleet size of 2836, both strategies converge to the same Matching Success Rate of 74.22%. At smaller fleet sizes, the Batch Offer strategy outperforms the Immediate Decision strategy in terms of matching success rates; however, as the fleet size expands, the Immediate Decision strategy is capable of providing a higher matching success rate.

$$f(x) = \frac{L}{1 + e^{-k(x - x_0)}} \tag{27}$$

where:

L is the curve's maximum value.

k is the growth rate.

x_0 is the midpoint of the curve.

The following table lists the parameters for the logistic fitting of the "Immediate Decision" and "Batch Offer" strategies:

Strategy Name	L	k	x_0
Immediate Decision	97.7682	0.000911	1576.30044
Batch Offer	96.7470	0.000412	-54.8391

Table 19: Logistic Fitting Parameters for Two Matching Strategies

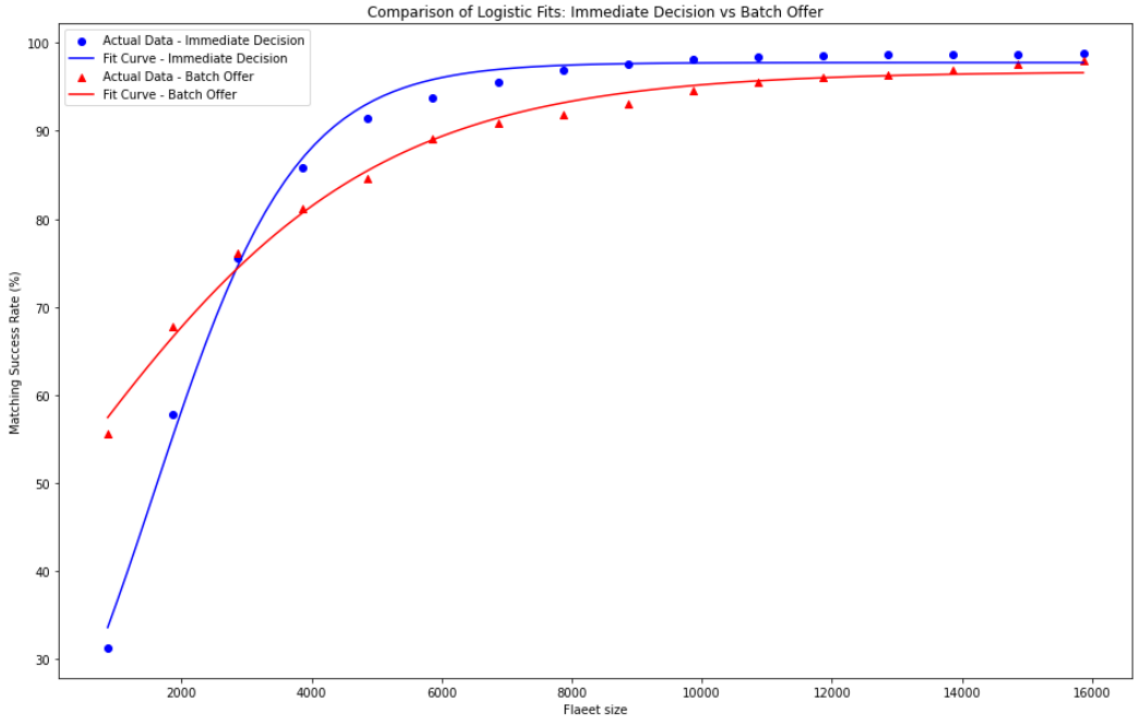

Figure 24: Logistic Fits for Matching Success Rate

5.2.2 Average waiting time

Figure 25 illustrates the simulations and the fitting outcomes for two distinct matching strategies, employing binomial fitting. The mathematical formulations for these strategies are as follows: for the Immediate Decision strategy, the expression is $f(x) = 2.73846289 \times 10^{-7}x^2 - 0.010043169x + 204.619342$; and for the Batch Offer strategy, the expression is $f(x) = 2.22599790 \times 10^{-7}x^2 - 0.00802681507x + 194.706162$. The Mean Squared Error (MSE) values are 0.559 for the Immediate Decision and 2.460 for the Batch Offer. As fleet size increases, the average waiting times for the two strategies exhibit different trends: the Immediate Decision strategy effectively reduces waiting time at larger fleet sizes, while the Batch Offer strategy performs better at smaller fleet sizes. The intersection occurs when the fleet size is approximately 5759, at which point the average waiting times of the two strategies are equal, reaching 155.86 seconds.

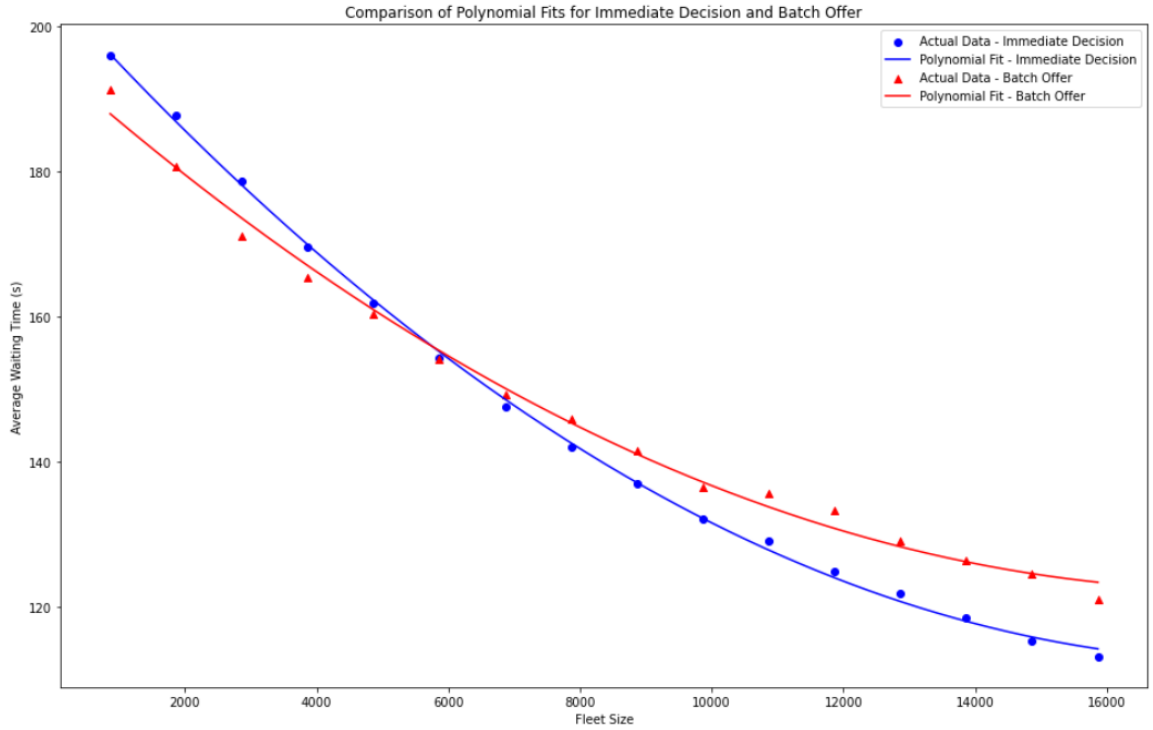

Figure 25: Polynomial Fits for Average Waiting Time

5.2.3 Pooling ratio

Through polynomial fitting of the simulation results, the mathematical expressions representing the Pooling Ratio for the "Immediate Decision" and "Batch Offer" strategies are:

$$f(x) = -8.89189851 \times 10^{-14}x^3 - 2.89849347 \times 10^{-9}x^2 - 3.46639757 \times 10^{-5}x + 0.370855474$$

and

$$f(x) = -3.37087716 \times 10^{-14}x^3 - 1.32903373 \times 10^{-9}x^2 - 2.42045091 \times 10^{-5}x + 0.390126656$$

respectively. The MSE for these fits are: $MSE = 1.147 \times 10^{-5}$ for Immediate Decision and $MSE = 2.667 \times 10^{-5}$ for Batch Offer, as illustrated in Figure 26.

As fleet sizes increase, the Pooling Ratios for both strategies exhibit a uniform and gradual decline, indicating a decreasing average number of passengers per SAV. This decline is similar across both strategies, yet the Pooling Ratio for the Batch Offer strategy consistently remains higher than that of the Immediate Decision strategy.

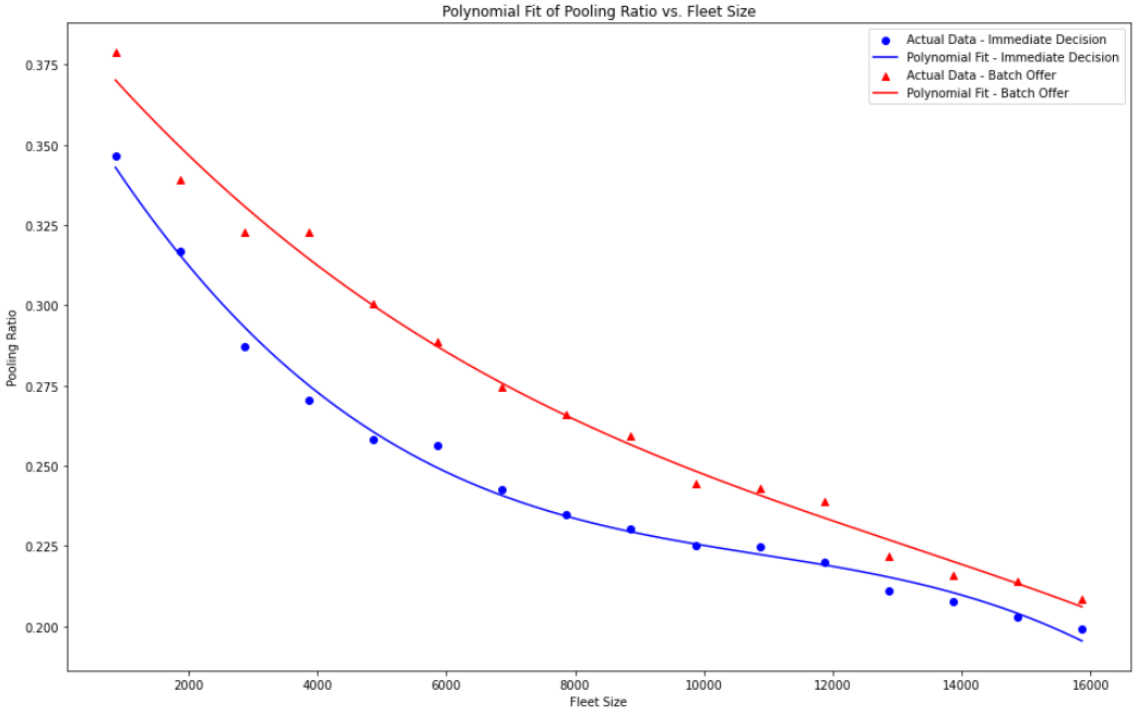

Figure 26: Polynomial Fits for Pooling Ratio

5.2.4 Extra mileage ratio

The results for the Extra Mileage Ratio exhibit similarities to those of the pooling ratio. As fleet size increases, the overall trend shows a gradual decline, with the Batch Offer strategy consistently yielding higher values than the Immediate Decision strategy. The difference in Extra Mileage Ratio caused by the two matching strategies shows a trend of decreasing. When the fleet size is around 14,000, the difference tends to stabilize at approximately 0.02. The polynomial equations for the extra mileage ratio for the two strategies are as follows: Immediate Decision: $f(x) = 5.23351961 \times 10^{-14}x^3 - 1.30873099 \times 10^{-9}x^2 + 5.60652754 \times 10^{-7}x + 0.359820508$

$f(x) = -8.17236781 \times 10^{-15}x^3 + 3.01534146 \times 10^{-10}x^2 - 7.94196067 \times 10^{-6}x + 0.310499520$

The Mean Squared Error (MSE) for these fits are: Immediate Decision MSE: 0.0036 Batch Offer MSE: 0.0007

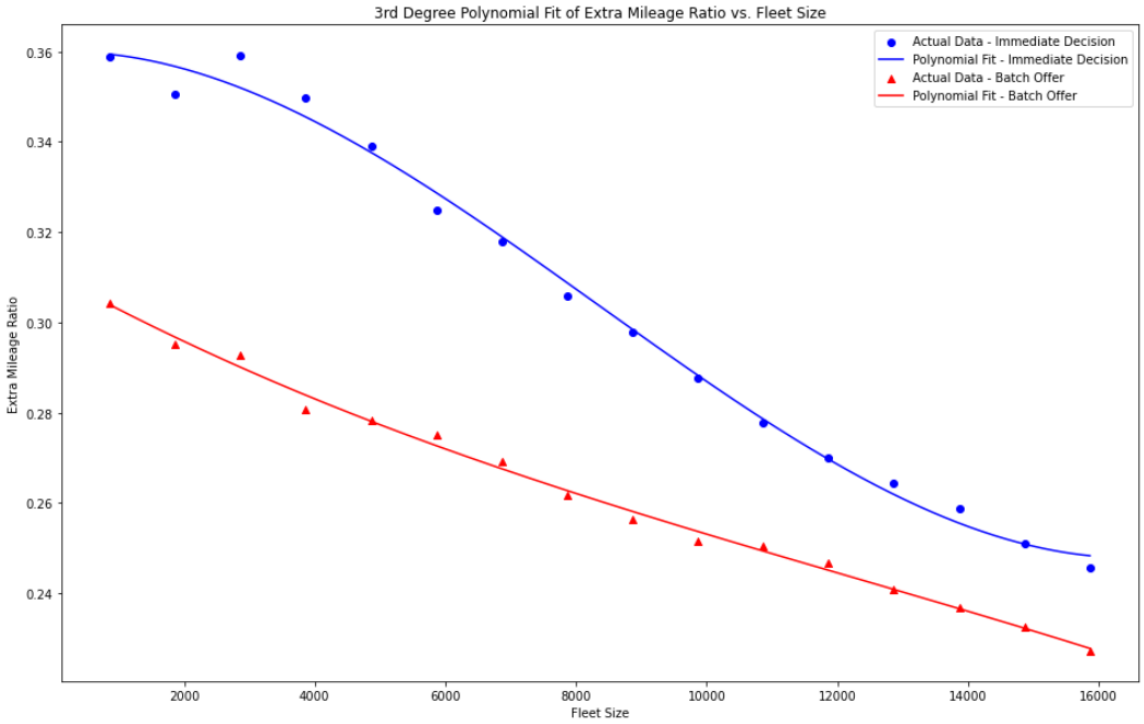

Figure 27: Polynomial Fits for Extra Mileage Ratio

(1+Extra Mileage Ratio)/(1+Pooling Ratio) represents the comparison between the total vehicle-miles traveled by SAVs and the total distance covered by POVs (Jin et al., 2021). When it increases, the matching strategy leads to a decrease in the system-wide trip completion rate. The ratios are displayed in Figure 28. Observations from the data indicate that under the same fleet size, the (1+Extra Mileage Ratio)/(1+Pooling Ratio) value for the Immediate Decision strategy consistently exceeds that of the Batch Offer matching strategy. In Batch Offer strategies, this ratio exhibits a positive correlation with fleet size, suggesting that as the fleet size increases, the proportion of total vehicle-miles traveled by SAVs compared to the total trip distances of POVs also increases. Notably, under the Batch Offer strategy, when the fleet size is approximately 9000, the total vehicle-miles traveled by SAVs serving the same number of passengers is equivalent to that of POVs. As the fleet size continues to increase, the use of SAVs lead to higher levels of road congestion. For the Immediate Decision strategy, the value of (1+Extra Mileage Ratio)/(1+Pooling Ratio) is always greater than 1. This indicates that compared to using POVs, using SAVs always results in more vehicle-miles traveled across different fleet sizes. The trend fluctuates, initially increasing, then decreasing, and then when the fleet size exceeds 10,500, it gradually stabilizes around 1.04.

Some existing studies indicate that TNC services with carpooling options can lead to additional traffic congestion(Tarduno, 2021) (Erhardt et al., 2019). Most of these studies are based on actual operational data from TNCs. Tarduno (2021) notes that this is primarily influenced by external policies, where the VMT avoided through shared rides is offset by additional trips caused by the availability of TNCs. Kucharski and Cats (2024) tests in Amsterdam with 10-person rides and private rides showed that the VMT for private rides is greater than that for shared door-to-door rides. The results of this study suggest that under ideal simulation conditions, where vehicles do not need to refuel or similar actions, in Haidian City using the Immediate Decision strategy, the VMT for SAVs is slightly higher than for POVs, while the Batch Offer strategy results in slightly lower VMT at small fleet sizes.

The polynomial equations for the extra mileage ratio for the two strategies are as follows: Immediate Decision: $f(x) = 1.10968881 \times 10^{-13}x^3 - 3.32030330 \times 10^{-9}x^2 + 2.81302861 \times 10^{-5}x + 0.990012220$

$f(x) = 1.48117277 \times 10^{-14}x^3 - 6.43759129 \times 10^{-10}x^2 + 1.12060545 \times 10^{-5}x + 0.942293817$

The Mean Squared Error (MSE) for these fits are: Immediate Decision MSE: 0.6434 Batch Offer MSE: 0.5269

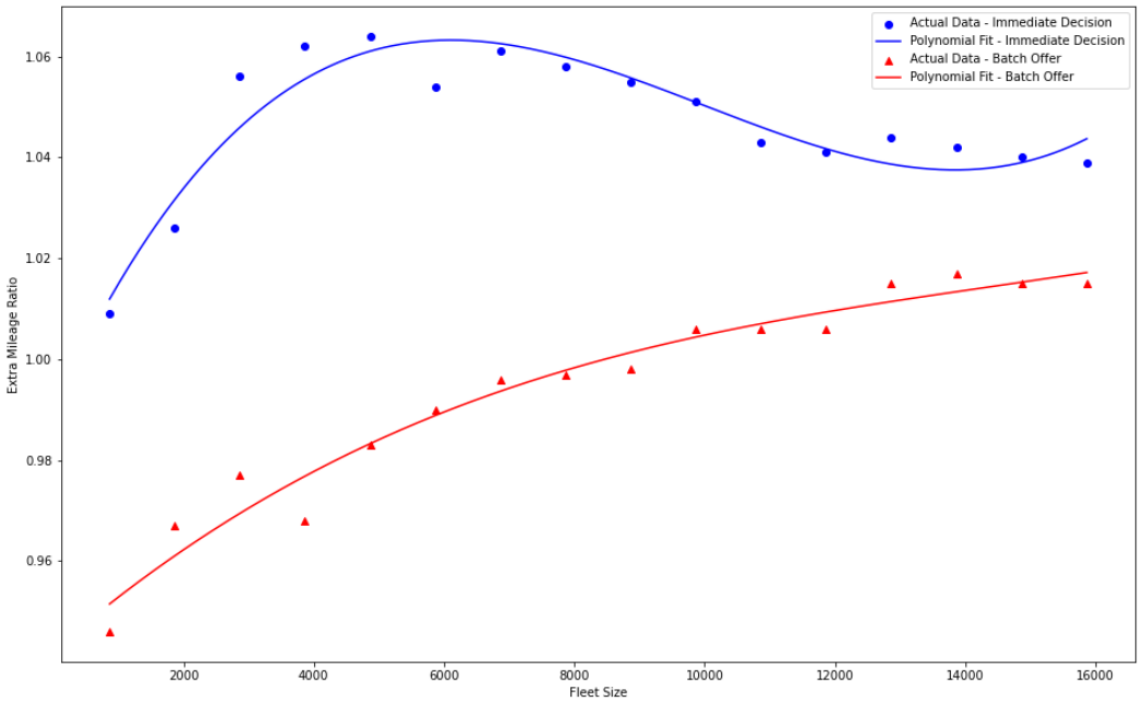

Figure 28: Polynomial Fits for (1+Extra Mileage Ratio)/(1+Pooling Ratio)

5.2.5 Conclusion

The Immediate Decision and Batch Offer strategies differ in how they handle passenger requests and allocate vehicle resources, which directly impacts passenger waiting times and overall service efficiency.

The key advantage of the Immediate Decision strategy lies in its ability to process passenger requests in real time. By continuously evaluating fleet status and responding quickly. Once a request is submitted, the system immediately analyzes and assigns the nearest available vehicle, resulting in a very short interval between request submission and vehicle arrival.

In contrast, the Batch Offer strategy processes requests in batches, offering greater flexibility and efficiency in route and vehicle allocation optimization. This strategy collects all requests within a certain time frame(60s) and allows the system to consider route sharing and trip efficiency more comprehensively, theoretically optimizing resource allocation. However, this batch processing mechanism means passengers must wait for a period after submitting their request before receiving a service proposal, increasing the average waiting time.

Overall, for larger fleets, the Immediate Decision strategy is a relatively more efficient matching strategy in terms of service level, as it can significantly reduce average passenger waiting times. Both strategies can achieve high matching success rates. In smaller fleets, however, the Batch Offer strategy can improve vehicle utilization efficiency through batch processing and route optimization, providing better matching of passenger requests. This process demonstrates the trade-off between immediate response and batch optimization. The Immediate Decision strategy does not involve a periodic collection of user requests; it responds to each request immediately. It utilizes the Insertion Heuristic algorithm for carpooling passengers, solely assessing the consistency between the vehicle's current route direction and the passenger request's origin-destination vector. In the Batch Offer strategy, there is an additional waiting time between the generation of user requests and the operator's batch optimization. However, it employs the Alonso-Mora capacity carpooling assignment by solving an Integer Linear Programming problem, minimizing the overall network delay.

Regarding the two parameters that display the level of network congestion, Pooling Ratio and Extra Mileage Ratio, both decrease as the fleet size increases under the two matching strategies. Based on this observation, a hypothesis can be made: as the fleet's vehicle resources increase, the pooling level of SAVs will gradually decrease, and their performance at the network level will increasingly approximate that of POVs. When the fleet size is sufficiently large, pooling will completely disappear.

5.3 Simulation result analysis

This section provides an analysis of the simulation results in relation to real-world Mobility-on-Demand (MoD) services in the Haidian District. According to data from the Didi platform, the highest ride-hailing success rate occurs between 10:00 and 17:00, with an average success rate of 83.2%. To achieve a similar level of service throughout the entire day, the Immediate Decision strategy would require a fleet size of 3,489 vehicles, while the Batch Offer strategy would need a fleet size of 4,347 vehicles. These figures represent only 44.4% and 55.3%, respectively, of the current fleet size of 7,865 vehicles in actual operations. Simulate these two scenarios and plot the MFD for the SAV system. Compare this with the real-world MFD of taxis in Haidian District to analyze the impact of replacing traditional taxis services with SAVs on urban traffic congestion. The selected time window is 1 minute Fit the data according to the functional form (equation 15) in section 3.4. The MFD scatter plots of the two matching strategies and their fitting results are shown in Figure 29. Table 20 presents the parameters of the MFD function. The physical meaning of these parameters is introduced in the section 3.4. Figure 30 displays a comparison between the MFDs generated in a simulation environment and the actual MFDs derived from taxi data.

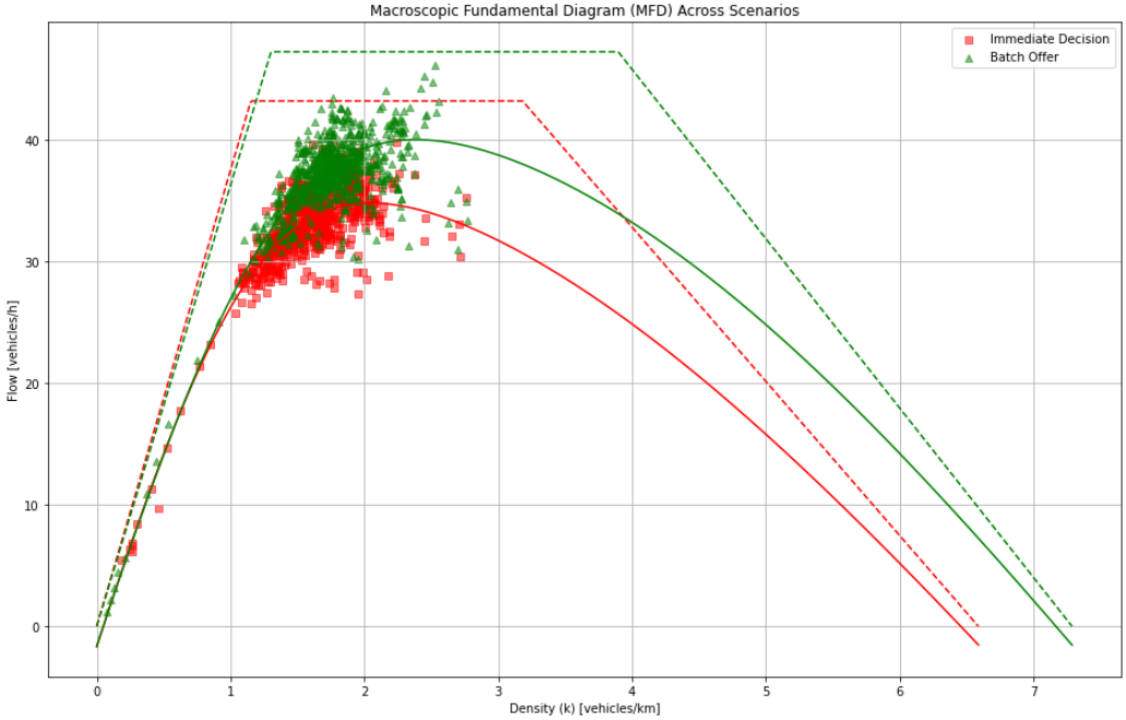

Figure 29: MFD with Functional Form for Immediate Decision and Batch Offer Simulation

Table 20: Parameters of MFD Function for Two Matching Strategies and Taxi Data

	Fleet Size	$u_f(km/h)$	$Q(veh/h)$	$\kappa(km/h)$	$w(veh/km)$	λ	Root Mean Square Error (R
Immediate Decision	3489	37.55	43.24	12.69	6.59	17.89	1.93
Batch Offer	4347	36.30	47.28	13.94	7.29	19.12	2.18
Taxi Data	7865	39.25	78	9.02	12.5	32.28	3.86

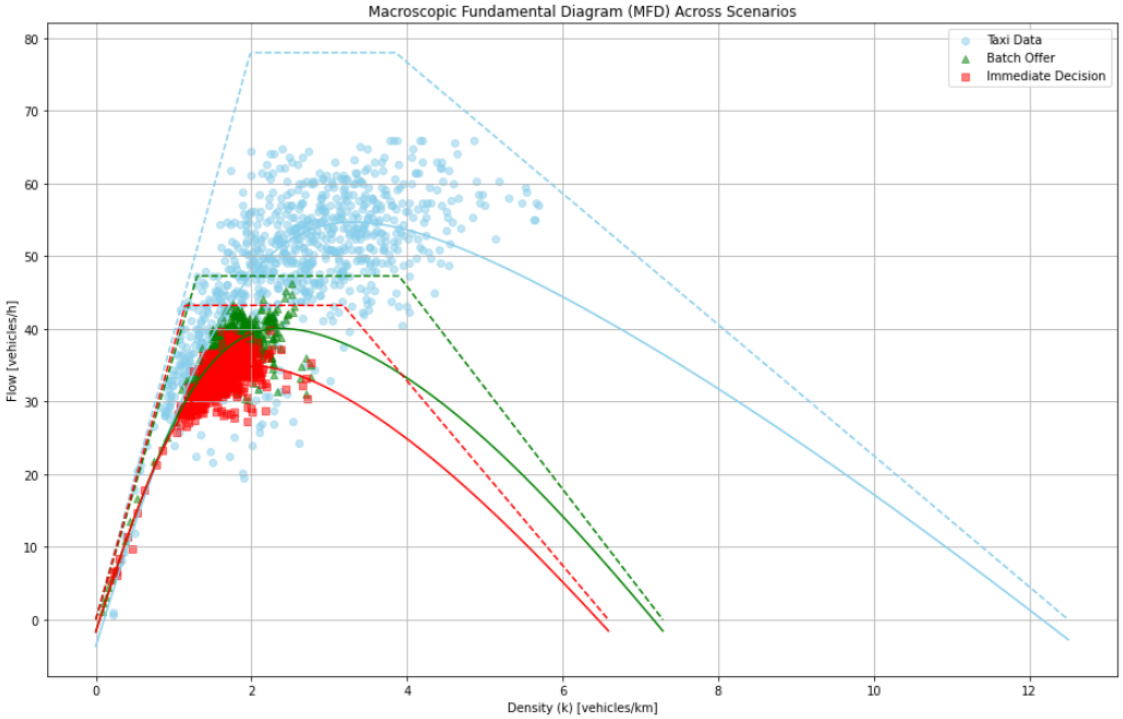

Figure 30: The comparison of MFD with Functional Form

By comparing the MFD of SAVs in simulations with those derived from actual GPS trajectory data of taxis, an initial qualitative analysis was conducted. Observations from the MFD scatter plots show that simulated data often operates at or near saturation flow, indicating that SAVs in simulations manage and utilize network resources more efficiently than traditional vehicles like taxis. In traditional taxi systems, drivers typically rely on experience and immediate passenger demand to decide their next destination, leading to frequent empty running, especially during off-peak hours or in areas with fewer passengers. Taxi routing is subjectively determined by the driver, whereas SAV scheduling is usually controlled by a centralized system that can make optimized decisions based on global information. The simulation employs two matching strategies to ensure that each vehicle is in an optimal position, locally or globally within the system. Moreover, SAVs can also provide pooling services based on the similarity of passenger routes, maximizing the passenger capacity of each vehicle.

In quantitative analyses of MFD function parameters for two matching strategies involving SAVs and traditional taxis, the initial focus is on free-flow speeds. In these three different traffic modalities, similar free-flow speeds were observed, primarily influenced by road characteristics and legal speed limits. All experiments were conducted within the same network environment—the road network of Haidian District in Beijing.

Although free-flow speeds are similar across the three scenarios, the free-flow speed in the taxi data is slightly higher, likely due to the driving behaviors of taxi drivers. In practice, to complete more orders quickly, taxi drivers sometimes drive above the speed limit, especially during periods of light traffic. This tendency to speed can result in higher free-flow speeds for taxis compared to SAVs. In contrast, scenarios utilizing SAVs strictly adhere to traffic rules, effectively eliminating the possibility of speeding.

When simulations are conducted using SAVs and exhibit a higher backward wave speed, it is because the network is operating close to saturation flow. Traffic disturbances can quickly affect the entire system, leading to rapid formation and spread of congestion.

The intersection capacity and jam density used for SAVs and taxis are two attributes that are strongly correlated with fleet size. Using the same road network, the intersection capacity and jam density are the same, but the proportion of SAVs in the total vehicle population differs from that of taxis. This difference leads to variations in the intersection capacity and jam density used for SAVs and taxis, which are positively correlated with fleet size.

6 Conclusions and future research directions

This study analyzes the performance of SAVs in mixed mobility systems from the perspectives of fleet sizing and matching strategies, focusing on aspects such as travel demand generation models, mode choice models, simulation frameworks, and MFD. The use of SAV systems as a replacement for taxis can significantly enhance fleet efficiency, achieving the same level of service with a smaller fleet size.

In this study, the travel demand generation model serves as a critical component of the foundational framework. It utilizes travel distance distribution functions, road network information, and taxi GPS data to generate a comprehensive traffic demand model. This model employs the K-means clustering algorithm to divide traffic zones. This clustering is based on the density of traffic nodes rather than specific road network conditions, thus it is a macroscopic, low-precision zoning method. The primary benefit of this clustering method lies in its simplicity and operational speed, which supports rapid foundational data provision for large-scale network analysis. This method is particularly suitable for rapid assessment and planning of large urban traffic networks, as well as for generating traffic demand forecasts in the preliminary stages of research.

However, this node density-based clustering method also has certain limitations. Since it does not consider specific operational conditions of the road network, such as road capacity and traffic signal control, it may not provide sufficiently accurate data support in scenarios highly sensitive to traffic patterns. Additionally, this method does not fully take into account the travel characteristics and demand differences within regions, potentially overlooking unique traffic generation or attraction characteristics of certain areas.

The process of predicting traffic demand using distance distribution functions and total travel volume needs to consider discrepancies between model predictions and actual conditions. This prediction method, based on theoretical distributions and historical data, needs to accurately capture the complexities of the actual traffic system. However, due to various reasons, including assumptions of model simplification, limitations in data collection, and changes in external environments, there are often discrepancies between model predictions and actual traffic volumes. An evaluation of the preliminarily generated travel demand model based on GPS trajectory data is used to analyze discrepancies between it and actual traffic flow data. By introducing correction factor analysis to adjust discrepancies in the model, traffic imbalances are addressed, making the model output more consistent with actual traffic data.

The mode choice model in this study undertakes the task of analyzing how passengers make decisions between different modes of travel. This model uses a discrete choice framework to study how passengers choose between SAVs and POVs. Specifically, the study employs a Logit model, considering several key factors affecting travel choices, including travel time, comfort, reliability, and environmental impact. This part emphasizes the importance of travelers' subjective perceptions.

The study also provides quantitative assessment methods for "comfort" and "environmental impact." For comfort, an ordinal logistic regression analysis is used to determine the specific contributions of different modes of travel and times to comfort. The comfort scores generated by the model are further divided into levels, each corresponding to a range of comfort, thus quantifying the level of comfort. Environmental impact is primarily quantified through the carbon emissions of vehicles, directly reflecting the impact of vehicles on the ecological footprint. To evaluate the environmental impact, this study employs the Vehicle Specific Power (VSP) model to estimate the output power of light-duty vehicles and model the relationship between vehicle operating states and fuel consumption.By calculating the carbon emissions of different driving segments, a model quantifying environmental impact is established. This model can be used to assess the environmental impact of individual vehicles and can also be extended to entire fleets or traffic systems. This process is used to answer the research question: *In mixed mobility systems, what kind of preference choices will passengers have between SAVs and other modes of transportation?*

The application of simulations and MFDs is a core method of this study. In the simulation analysis part, this study uses the agent-based simulation tool FleetPy, focusing on evaluating two key matching strategies: Immediate Decision Simulation (IDS) and Batch Offer Simulation (BOS). Through simulations of different fleet sizes and matching strategies, the study explores the scheduling and operational performance of shared autonomous vehicles in urban road networks, comparing them with traditional taxi systems.

The study finds that the SAV system exhibits high operational efficiency whether in small or large fleets. In scenarios with smaller fleet sizes, the Batch Offer (BOS) strategy provides a higher level of service. By processing and optimizing multiple requests in batches, it can effectively increase the success rate of matches and reduce waiting times, while Immediate Decision (IDS) performs better in terms of service timeliness. The simulation

results also show that as the fleet size increases, the advantages of the IDS strategy gradually become apparent, but it may lead to wastage of vehicle resources.

This study evaluates the service level and road congestion of SAVs in mixed mobility systems using four key parameters and the MFD. The findings conclude that replacing existing MoD services with SAVs and optimizing vehicle matching can significantly reduce the fleet size needed, thereby lowering costs and alleviating traffic congestion. Previous research by Alonso-Mora et al. (2017) demonstrated through simulations that 3,000 four-seat SAVs could meet 98% of Manhattan's taxi demand, reducing the fleet size by approximately 77%. Similarly, Lokhandwala and Cai (2018) found that integrating shared autonomous taxis with conventional taxis in New York City could sustain service levels while cutting the fleet size by 59%. In this study, analysis of two SAV matching strategies showed that replacing traditional taxi services with SAVs requires only 44.4% to 55.5% of the fleet to achieve a similar vehicle-to-passenger matching success rate. While both earlier studies concluded that using SAVs can reduce the fleet size, the reduction magnitude in this study is somewhat less than those reported by (Alonso-Mora et al., 2017) and (Lokhandwala and Cai, 2018).

Finally, a summary answer to the other research question of this study:

How does the modification in the size of the SAVs fleet impact the efficiency of matching strategies and overall traffic dynamics ?

Adjustments to fleet size have a direct impact on the efficiency of matching strategies and traffic dynamics. For the Haidian District of Beijing, smaller fleet sizes demonstrate higher efficiency when using the Batch Offer Strategy (BOS). When the fleet size is less than 2836, batch processing can provide a higher match success rate, and when the fleet size is less than 5759, it can offer shorter waiting times. On the other hand, larger fleet sizes provide faster responses under the Immediate Decision Strategy (IDS). In terms of Pooling ratio and Extra mileage ratio, IDS performs better.

Can an optimal fleet size value be determined based on the level of service provided by SAVs and the impact on road congestion under different matching strategies?

Regarding the optimal fleet size value, the four KPIs exhibit different performances. We hope the fleet can provide a high level of service while maintaining a large Pooling ratio and a small Extra mileage ratio. A high Pooling ratio indicates that SAVs can effectively consolidate trips from multiple passengers, thereby reducing the total vehicle requirement and operational costs. A lower Extra mileage ratio signifies higher direct route efficiency and lower fuel consumption. Simulation results show that both Pooling ratio and Extra mileage ratio decrease as the fleet size increases, though the decrease is minimal. For example, when the fleet size increases from 865 to 15865, the Pooling ratio under IDS only decreases by about 0.175. In conclusion, neither of the two matching strategies can achieve the optimal fleet size value. A smaller fleet size results in higher vehicle utilization but sacrifices service level, and vice versa. How to balance between the two will be an important consideration for the future.

What impact do SAVs have on urban road congestion levels when using different matching strategies?

The impact of SAVs on urban road congestion levels when using different matching strategies can be analyzed through the MFD. By comparing the performance of SAVs and taxis under the same match success rate using two different matching strategies, observations from the MFD scatter plots show that using SAVs often results in the road network operating at or near saturation flow for more extended periods. The parts of the intersection capacity and jam density used by SAVs increase as the fleet size increases. Regarding the backward wave speed, using the Batch Offer strategy results in higher backward wave speeds, indicating greater sensitivity to traffic disturbances and thereby suggesting that more sections of the network operate in a state of saturation flow. Compared to the Immediate Decision strategy, BO exhibits higher network utilization efficiency.

The main assumption made in this study is that all MoD services within the area are provided by SAVs, with no traditional taxis or other modes of transport. This approach allows for a more direct observation of the impact of changes in the size of the SAV fleet on passenger service levels and road congestion. However, it also results in a loss of insight into the performance of SAVs under different penetration rates. Moreover, the study is conducted on a single city network, and the results' generalizability to different cities and the influence of urban geometry have not been considered.

Further research might include developing precise predictive models to more accurately forecast short-term passenger demand, adjusting fleet sizes and matching strategies in real-time based on these predictions, to achieve a flexible fleet dispatch system that can respond dynamically to changes in demand.

Moreover, the simulation framework does not include a specific vehicle dynamics model for SAVs, which means that while the simulation might accurately capture the broader traffic flow and interactions between different types of vehicles, it does not account for the unique driving characteristics and capabilities that SAVs might exhibit, such as their ability to accelerate, decelerate, and maneuver in response to traffic conditions more smoothly than human-driven vehicles. However, the use of simulations inherently avoids the subjective driving behaviors of drivers and to a certain extent, can also demonstrate the driving characteristics of SAVs.

It is noteworthy that this study is considered from the perspective of traffic operations. It lacks exploration from the viewpoint of SAVs operators. The study does not take into account the pricing structures of SAVs (such as discounts), market effects, and the potential rebound effect due to cost reduction that could increase demand. This induced demand can enhance the model further and offer deeper insights.

References

China dual carbon big data index white paper 2022. Online, 2022. URL https://cdn.zhituquan.com/2022/05/1651795503-560d62790217e23.pdf.

Md Jahedul Alam and Muhammad Ahsanul Habib. Investigation of the impacts of shared autonomous vehicle operation in halifax, canada using a dynamic traffic microsimulation model. *Procedia computer science*, 130: 496–503, 2018.

Javier Alonso-Mora, Samitha Samaranayake, Alex Wallar, Emilio Frazzoli, and Daniela Rus. On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. *Proceedings of the National Academy of Sciences*, 114(3):462–467, 2017.

Lukas Ambühl, Allister Loder, Michiel C.J. Bliemer, Monica Menendez, and Kay W. Axhausen. A functional form with a physical meaning for the macroscopic fundamental diagram. *Transportation Research Part B: Methodological*, 137:119–132, 2020. ISSN 0191-2615. https://doi.org/10.1016/j.trb.2018.10.013. URL https://www.sciencedirect.com/science/article/pii/S0191261517310123. Advances in Network Macroscopic Fundamental Diagram (NMFD) Research.

ApolloGo. Robotaxi. https://baike.baidu.com/reference/58307068/533aYdO6cr3_z3kATPzdzfnzNCfCN47-6LTVUrVzzqIPmGapB4zkU4I74d8-8bllFQLPpdZhb9tahbejXkZE5vcUcOQ2R7MnmHb6VDfFwL3u6Z5n2NwH49MXDe8B0a6zuwSv 2024.

Roberto Baldacci, Vittorio Maniezzo, and Aristide Mingozzi. An exact method for the car pooling problem based on lagrangean column generation. *Operations Research*, 52:422–439, 06 2004. 10.1287/opre.1030.0106.

Golan Ben-Dor, Eran Ben-Elia, and Itzhak Benenson. Determining an optimal fleet size for a reliable shared automated vehicle ride-sharing service. *Procedia Computer Science*, 151:878–883, 2019. ISSN 1877-0509. https://doi.org/10.1016/j.procs.2019.04.121. URL https://www.sciencedirect.com/science/article/pii/S187705091930585X. The 10th International Conference on Ambient Systems, Networks and Technologies (ANT 2019) / The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40 2019) / Affiliated Workshops.

Tianfeng Chai, Roland R Draxler, et al. Root mean square error (rmse) or mean absolute error (mae). *Geoscientific model development discussions*, 7(1):1525–1534, 2014.

Rachel Cole. autonomous vehicle. Encyclopedia Britannica, October 2024. URL https://www.britannica.com/technology/autonomous-vehicle. Accessed: 9 October 2024.

N. Cressie. The origins of kriging. *Mathematical Geology*, 22(3):239–252, April 1990.

Mengyao Cui et al. Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance*, 1(1):5–8, 2020.

Roman Engelhardt, Florian Dandl, Arslan-Ali Syed, Yunfei Zhang, Fabian Fehn, Fynn Wolf, and Klaus Bogenberger. Fleetpy: A modular open-source simulation tool for mobility on-demand services. *arXiv preprint arXiv:2207.14246*, 2022.

Gregory D Erhardt, Sneha Roy, Drew Cooper, Bhargava Sana, Mei Chen, and Joe Castiglione. Do transportation network companies decrease or increase congestion? *Science advances*, 5(5):eaau2670, 2019.

Daniel J Fagnant and Kara M Kockelman. The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transportation Research Part C: Emerging Technologies*, 40: 1–13, 2014.

Qiaochu Fan, J. Theresia van Essen, and Gonçalo H.A. Correia. Optimising fleet sizing and management of shared automated vehicle (sav) services: A mixed-integer programming approach integrating endogenous demand, congestion effects, and accept/reject mechanism impacts. *Transportation Research Part C: Emerging Technologies*, 157:104398, 2023. ISSN 0968-090X. https://doi.org/10.1016/j.trc.2023.104398. URL https://www.sciencedirect.com/science/article/pii/S0968090X23003881.

Nikolas Geroliminis and Carlos F. Daganzo. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9):759–770, 2008. ISSN 0191-2615. https://doi.org/10.1016/j.trb.2008.02.002. URL https://www.sciencedirect.com/science/article/pii/S0191261508000180.

Chana J Haboucha, Robert Ishaq, and Yoram Shiftan. User preferences regarding autonomous vehicles. *Transportation research part C: emerging technologies*, 78:37–49, 2017.

Mustapha Harb, Yu Xiao, Giovanni Circella, Patricia L Mokhtarian, and Joan L Walker. Projecting travelers into a world of self-driving vehicles: estimating travel behavior implications via a naturalistic experiment. *Transportation*, 45:1671–1685, 2018.

Wesam Herbawi and Michael Weber. Evolutionary multiobjective route planning in dynamic multi-hop ridesharing. In Peter Merz and Jin-Kao Hao, editors, *Evolutionary Computation in Combinatorial Optimization*, pages 84–95, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-20364-0.

Arne Risa Hole. Mixed logit modeling in stata–an overview. In *United Kingdom Stata Users' Group Meetings 2013*, number 23. Stata Users Group, 2013.

Yantao Huang, Kara M. Kockelman, and Venu Garikapati. Shared automated vehicle fleet operations for first-mile last-mile transit connections with dynamic pooling. *Computers, Environment and Urban Systems*, 92:101730, 2022. ISSN 0198-9715. https://doi.org/10.1016/j.compenvurbsys.2021.101730. URL https://www.sciencedirect.com/science/article/pii/S019897152100137X.

Sebastian Hörl, Felix Becker, and Kay W. Axhausen. Simulation of price, customer behaviour and system impact for a cost-covering automated taxi system in zurich. *Transportation Research Part C: Emerging Technologies*, 123:102974, 2021. ISSN 0968-090X. https://doi.org/10.1016/j.trc.2021.102974. URL https://www.sciencedirect.com/science/article/pii/S0968090X21000115.

Wen-Long Jin, Irene Martinez, and Monica Menendez. Compartmental model and fleet-size management for shared mobility systems with for-hire vehicles. *Transportation Research Part C: Emerging Technologies*, 129:103236, 2021. ISSN 0968-090X. https://doi.org/10.1016/j.trc.2021.103236. URL https://www.sciencedirect.com/science/article/pii/S0968090X21002497.

Joonho Ko, Sugie Lee, and Miree Byun. Exploring factors associated with commute mode choice: An application of city-level general social survey data. *Transport policy*, 75:36–46, 2019.

Rafal Kucharski and Oded Cats. Hyper-pool: pooling private trips into high-occupancy transit-like attractive shared rides. January 2024. URL https://doi.org/10.21203/rs.3.rs-3848676/v1. PREPRINT (Version 1) available at Research Square.

Yu Lei, Zhu Hongyu, Guo Jifu, Zhang Xi, Sun Jianping, Lei Xue, and Song Guohua. Method for calculating the optimal traffic index of road network theoretical efficiency based on mfd. *Journal of Beijing Jiaotong University*, 46(3):26–34, 2022. 10.11860/j.issn.1673-0291.20210094. Article Number: 1673-0291(2022)03-0026-08.

Xiao LIANG, Zhili LIU, and Kun QIAN. Capacity analysis of signalized intersections under mixed traffic conditions. *Journal of Transportation Systems Engineering and Information Technology*, 11(2):91–99, 2011. ISSN 1570-6672. https://doi.org/10.1016/S1570-6672(10)60116-X. URL https://www.sciencedirect.com/science/article/pii/S157066721060116X.

Mustafa Lokhandwala and Hua Cai. Dynamic ride sharing using traditional taxis and shared autonomous taxis: A case study of nyc. *Transportation Research Part C: Emerging Technologies*, 97:45–60, 2018.

Jun-lai Ma, Yang Bian, and Wei Wang. Study on relative travel comfort among urban traffic modes. In *Proceedings of International Academic Conference on Transport*, Dalian, 2006. Dalian University of Technology.

Michal Maciejewski, Joschka Bischoff, Sebastian Hörl, and Kai Nagel. Towards a testbed for dynamic vehicle routing algorithms. In *Highlights of Practical Applications of Cyber-Physical Multi-Agent Systems: International Workshops of PAAMS 2017, Porto, Portugal, June 21-23, 2017, Proceedings 15*, pages 69–79. Springer, 2017.

Elliot W Martin and Susan A Shaheen. Greenhouse gas emission impacts of carsharing in north america. *IEEE Transactions on intelligent transportation systems*, 12(4):1074–1086, 2011.

Mary J Meixell and Mario Norbis. A review of the transportation mode choice and carrier selection literature. *The International Journal of Logistics Management*, 19(2):183–211, 2008.

Duy Q. Nguyen-Phuoc, Meng Zhou, Ming Hong Chua, André Romano Alho, Simon Oh, Ravi Seshadri, and Diem-Trinh Le. Examining the effects of automated mobility-on-demand services on public transport systems using an agent-based simulation approach. *Transportation Research Part A: Policy and Practice*, 169:103583,

2023. ISSN 0965-8564. https://doi.org/10.1016/j.tra.2023.103583. URL https://www.sciencedirect.com/science/article/pii/S0965856423000034.

Ian Y Noy, David Shinar, and William J Horrey. Automated driving: Safety blind spots. *Safety science*, 102: 68–78, 2018.

Simon Oh, Ravi Seshadri, Carlos Lima Azevedo, Nishant Kumar, Kakali Basak, and Moshe Ben-Akiva. Assessing the impacts of automated mobility-on-demand through agent-based simulation: A study of singapore. *Transportation Research Part A: Policy and Practice*, 138:367–388, 2020. ISSN 0965-8564. https://doi.org/10.1016/j.tra.2020.06.004. URL https://www.sciencedirect.com/science/article/pii/S0965856420306133.

Eva Ostertagová. Modelling using polynomial regression. *Procedia engineering*, 48:500–506, 2012.

Victor M Panaretos and Yoav Zemel. Statistical aspects of wasserstein distances. *Annual review of statistics and its application*, 6(1):405–431, 2019.

Michel Parent and Arnaud de La Fortelle. Cybercars: Past, present and future of the technology. *arXiv preprint cs/0510059*, 2005.

Jonathan Petit and Steven E Shladover. Potential cyberattacks on automated vehicles. *IEEE Transactions on Intelligent transportation systems*, 16(2):546–556, 2014.

Xiaohong Ren, Zhenhua Chen, Chunhua Liu, Ting Dan, Jie Wu, and Fang Wang. Are vehicle on-demand and shared services a favorable solution for the first and last-mile mobility: Evidence from china. *Travel Behaviour and Society*, 31:386–398, 2023. ISSN 2214-367X. https://doi.org/10.1016/j.tbs.2023.01.008. URL https://www.sciencedirect.com/science/article/pii/S2214367X2300011X.

Ketan Rajshekhar Shahapure and Charles Nicholas. Cluster quality analysis using silhouette score. In *2020 IEEE 7th international conference on data science and advanced analytics (DSAA)*, pages 747–748. IEEE, 2020.

Susan Shaheen and Mohamed Amine Bouzaghrane. Mobility and energy impacts of shared automated vehicles: a review of recent literature. *Current Sustainable/Renewable Energy Reports*, 6, 12 2019. 10.1007/s40518-019-00135-2.

Susan Shaheen and Adam Cohen. Chapter 3 - mobility on demand (mod) and mobility as a service (maas): early understanding of shared mobility impacts and public transit partnerships. In Constantinos Antoniou, Dimitrios Efthymiou, and Emmanouil Chaniotakis, editors, *Demand for Emerging Transportation Systems*, pages 37–59. Elsevier, 2020. ISBN 978-0-12-815018-4. https://doi.org/10.1016/B978-0-12-815018-4.00003-6. URL https://www.sciencedirect.com/science/article/pii/B9780128150184000036.

Susan Shaheen, Adam Cohen, and M Jaffee. Innovative mobility carsharing outlook: Carsharing market overview, analysis, and trends, winter 2016. *Transportation Sustainability Research Center, University of California, Berkeley*, 2016.

Andrea Simonetto, Julien Monteil, and Claudio Gambella. Real-time city-scale ridesharing via linear assignment problems. *Transportation Research Part C: Emerging Technologies*, 101:208–232, 2019. ISSN 0968-090X. https://doi.org/10.1016/j.trc.2019.01.019. URL https://www.sciencedirect.com/science/article/pii/S0968090X18302882.

Ying So and Warren F Kuhfeld. Multinomial logit models. In *SUGI 20 conference proceedings*, volume 1995, pages 1227–1234, 1995.

Kevin Spieser, Samitha Samaranayake, Wolfgang Gruel, and Emilio Frazzoli. Shared-vehicle mobility-on-demand systems: A fleet operator's guide to rebalancing empty vehicles. In *Transportation Research Board 95th Annual Meeting*, number 16-5987. Transportation Research Board, 2016.

Monika Stoma, Agnieszka Dudziak, Jacek Caban, and Paweł Droździel. The future of autonomous vehicles in the opinion of automotive market users. *Energies*, 14(16):4777, 2021.

Araz Taeihagh and Hazel Si Min Lim. Governing autonomous vehicles: emerging responses for safety, liability, privacy, cybersecurity, and industry risks. *Transport reviews*, 39(1):103–128, 2019.

Amirmahdi Tafreshian, Neda Masoud, and Yafeng Yin. Frontiers in service science: Ride matching for peer-to-peer ride sharing: A review and future directions. *Service Science*, 12(2-3):44–60, 2020. 10.1287/serv.2020.0258. URL https://doi.org/10.1287/serv.2020.0258.

Matthew Tarduno. The congestion costs of uber and lyft. *Journal of Urban Economics*, 122:103318, 2021.

Mark Thomas and T Deepti. Reinventing carsharing as a modern and profitable service. In *The Intelligent Transportation Society of America, Annual Meeting White Paper*, 2018.

Kay W Axhausen, Andreas Horni, and Kai Nagel. *The multi-agent transport simulation MATSim*. Ubiquity Press, 2016.

Moritz Wäschle, Florian Thaler, Axel Berres, Florian Pölzlbauer, and Albert Albers. A review on ai safety in highly automated driving. *Frontiers in Artificial Intelligence*, 5:952773, 2022.

Chieh-Hua Wen and Frank S Koppelman. The generalized nested logit model. *Transportation Research Part B: Methodological*, 35(7):627–641, 2001.

Yale Z. Wong, David A. Hensher, and Corinne Mulley. Mobility as a service (maas): Charting a future context. *Transportation Research Part A: Policy and Practice*, 131:5–19, 2020. ISSN 0965-8564. https://doi.org/10.1016/j.tra.2019.09.030. URL https://www.sciencedirect.com/science/article/pii/S0965856418312229. Developments in Mobility as a Service (MaaS) and Intelligent Mobility.

Haonan Yan, Kara M. Kockelman, and Krishna Murthy Gurumurthy. Shared autonomous vehicle fleet performance: Impacts of trip densities and parking limitations. *Transportation Research Part D: Transport and Environment*, 89:102577, 2020. ISSN 1361-9209. https://doi.org/10.1016/j.trd.2020.102577. URL https://www.sciencedirect.com/science/article/pii/S1361920920307641.

Yuan Yang, Can Wang, Wenling Liu, and Peng Zhou. Understanding the determinants of travel mode choice of residents and its carbon mitigation potential. *Energy Policy*, 115:486–493, 2018. ISSN 0301-4215. https://doi.org/10.1016/j.enpol.2018.01.033. URL https://www.sciencedirect.com/science/article/pii/S0301421518300429.

Haihong Yuan, Fangqu Niu, and Xiaolu Gao. A study on urban economic vulnerability evaluation system. *Journal of Geographical Sciences*, 25, 10 2015. 10.1007/s11442-015-1232-5.

Wei-hua Zhang, Ran-ran Liu, Peng Yan, and Zhi-peng Huang. A travel value model considering comfort factors. *Journal of Highway and Transportation Research and Development*, 37(5):116–122, 2020.

Wenwen Zhang, Subhrajit Guhathakurta, Jinqi Fang, and Ge Zhang. Exploring the impact of shared autonomous vehicles on urban parking demand: An agent-based simulation approach. *Sustainable cities and society*, 19:34–45, 2015.

T. Zhao. On-road fuel consumption algorithm based on floating car data for light-duty vehicles. *Journal Name*, 2009.

Yu Zheng. T-drive trajectory data sample, August 2011. URL https://www.microsoft.com/en-us/research/publication/t-drive-trajectory-data-sample/. T-Drive sample dataset.

Yuanqi Zhu. Study on the travel distribution characteristics of urban traffic network based on taxi gps data and its impact on congestion. Master's thesis, Beijing Jiaotong University, Beijing, China, June 2020.

A Appendix

Page 1

Opening Statement

You are being invited to participate in a research study titled [Assessing the impacts of fleet sizing and matching strategies for Shared Automated Vehicles (SAVs) in mixed mobility systems using an agent-based simulation approach]. This study is being done by Yiming Wang from the TU Delft .

The purpose of this research study is [Investigate residents' preferences between SAVs (Shared Autonomous Vehicles) and POVs(privately operated vehicles) in daily travel in Beijing's Haidian District, as well as the factors influencing choices.], and will take you approximately 3 minutes to complete. The data will be used for Master's thesis publication. We will be asking you to to answer a few simple multiple-choice questions and one open-ended question, including whether you are aware of SAV, whether you have experienced SAV, the number of privately owned vehicles in your household, your daily travel time, and to rate several attributes (such as comfort, reliability, and environmental impact) that influence your travel decisions..

As with any online activity the risk of a breach is always possible. To the best of our ability your answers in this study will remain confidential. We will minimize any risks by completely anonymous, and IP addresses or other Personal Data will not be collected. During the research process, the data will be securely stored in OneDrive, and after the project is completed, the data will be stored in the appendix of the master's thesis for long-term preservation.

Your participation in this study is entirely voluntary and you can withdraw at any time. You are free to omit any questions.

If you agree with the content of this Opening Statement, please click "I Agree" to proceed with the survey.

Page 2

This page provides a brief introduction to SAV. If you are already familiar with SAV, you can skip this page.

SAV refers to autonomous vehicles that do not require manual driving and are used to provide shared mobility services. They can offer on-demand transportation for multiple passengers, similar to the carpooling service currently provided by platforms like Didi. To help you better understand SAV, here is an example of an automated ride-hailing service platform currently in operation. The difference between this platform and SAV is that it does not offer carpooling services. Apollo Go: Apollo Go operates 24/7 and is accessible through multiple platforms, including a WeChat mini-program, Baidu Maps, the Apollo Go app, and the Baidu app. It is currently open for demonstration operations in the following cities: Beijing, Shanghai, Guangzhou, Shenzhen, Chongqing, Wuhan, Chengdu, Changsha, Hefei, Yangquan, Wuzhen, and others.

The process of using Apollo Go services is as follows: select pick-up and drop-off points, request a vehicle, wait for the vehicle to arrive at the starting point, verify your identity using the last four digits of your phone number or by scanning a QR code, start the trip, and pay the order according to the amount displayed at the end of the trip.

Apollo Go uses the RT6 as the first model on its platform, which is based on Baidu's self-developed "Apollo Galaxy" architecture. The Apollo RT6 achieves 100% automotive-grade and full vehicle redundancy systems , making it two to three orders of magnitude more reliable than modified vehicles currently on the market, fully ensuring passenger safety. (A Redundancy System refers to the setup of multiple backup systems or components in critical parts of a system. This ensures that if any key component or system fails, the backup system can immediately take over its function, thereby ensuring the continuous operation and stability of the system.)

In terms of hardware, the Apollo RT6 features seven layers of redundancy, including architecture redundancy, computing unit redundancy, and braking system redundancy. If any single component or system fails, the backup redundant system can instantly take over. On the software side, it is equipped with an integrated fault diagnosis and risk mitigation system for both the vehicle and the autonomous driving system.

The service has a maximum capacity of three passengers, only accommodates passengers in the back seat, and includes a physical barrier between the front and back seats. It provides two interactive screens in the rear seats, offering smart services such as ride instructions, safety reminders, real-time routes, driving scenarios, music, video, and air conditioning.

Apollo Go's pricing includes a base fare of 18 RMB for the first kilometer and two minutes, followed by 2.7 RMB per kilometer, 0.5 RMB per minute, and an extra 0.9 RMB per kilometer for trips over 10 kilometers. If you would like to learn more about the actual user experience with Apollo Go, you can visit the link below. Watch on Bilibili.

Page 3

Starting from this page is the first section of the study, designed to assess your understanding of SAV and gather some basic travel information.

Page 4

Have you previously learned about SAV through any channel?

- Yes
- No

Page 5

Have you ever experienced SAV services (including non-carpooling automated ride-hailing platforms)?

- Yes
- No

Page 6

How many privately owned vehicles does your household have?

- No car
- One car
- Two or more cars

Page 7

What is your primary purpose for travel in your daily life?

Page 8

The following is the second part of this study, which involves your preferences between SAV and POV in your daily travel.

Page 9

How long do you spend on a single trip for your main travel purpose?

Page 10

In your subjective perception, what level of comfort do you expect during your travel? Please rate on a scale from 1 to 5.

Comfort Level	**Description**
1	Very uncomfortable – The travel experience is highly inconvenient and unpleasant.
2	Uncomfortable – The travel experience has noticeable discomforts and is not enjoyable.
3	Neutral – The travel experience is neither particularly comfortable nor uncomfortable.
4	Comfortable – The travel experience is generally smooth and pleasant, with few discomforts.
5	Very comfortable – The travel experience is highly pleasant, providing maximum ease and relaxation.

Page 11

In your subjective perception, what level of environmental friendliness do you expect during your travel? Please rate on a scale from 1 to 5.

Environmental Impact Level	**Description**
1	Very high impact – The travel option significantly harms the environment (e.g., high emissions, high resource consumption).
2	High impact – The travel option has a noticeable negative environmental impact.
3	Moderate impact – The travel option has an average or moderate environmental effect.
4	Low impact – The travel option has minimal environmental harm and is relatively eco-friendly.
5	Very low impact – The travel option is highly environmentally friendly, causing the least harm.

Page 12

In your subjective perception, what level of reliability do you expect during your travel? Please rate on a scale from 1 to 5. Reliability can be considered from the following aspects: whether a vehicle is available when you need it, whether a vehicle can provide service within a short period of time, and whether the vehicle can complete your travel needs within the expected time, such as choosing the appropriate route and speed.

Reliability Level	**Description**
1	Very unreliable – Vehicles are rarely available when needed, long waiting times, and the trip often does not meet time expectations (e.g., delays, detours).
2	Unreliable – Availability is inconsistent, with frequent delays and difficulty in meeting expected travel times.
3	Neutral – Availability and reliability are average, with occasional delays or difficulties in meeting travel expectations.
4	Reliable – Vehicles are usually available promptly, with minor or rare delays in reaching destinations on time.
5	Very reliable – Vehicles are always available when needed, with minimal waiting time and consistent, timely completion of trips.

Page 13

Please indicate your preference between SAV and POV for your primary travel purpose. In this process, you are required to conduct a comprehensive evaluation based on your prior subjective perceptions of the three attributes—comfort, environmental impact, and reliability—along with the travel time for your primary travel purpose. You can use your actual experiences with ride-sharing services and the information I have provided about the Apollo Go platform as a reference to consider the actual passenger experience of SAV.

- SAV
- POV

B Appendix

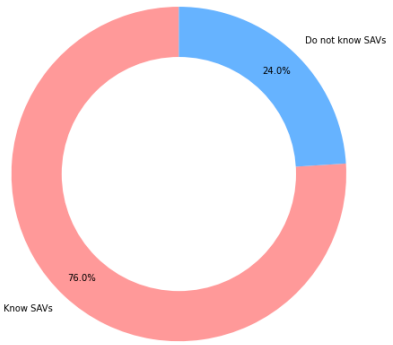

Figure 31: Knowledge of SAVs

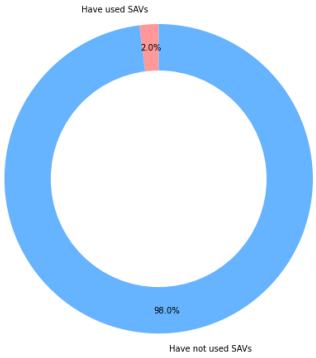

Figure 32: Experience with SAVs

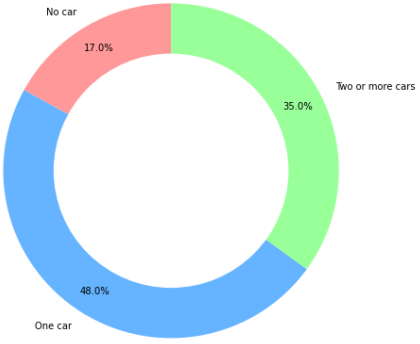

Figure 33: Car Ownership Status

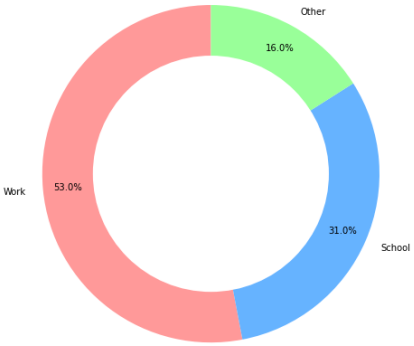

Figure 34: Main Travel Purpose

C Appendix

Choice (SAV or POV)	Knowledge of SAVs	Experience with SAVs	Car Ownership Status	Main Travel Purpose	Travel Time (min)	Comfort	Environmental Impact	Reliability
0	1	0	2	0	40	4	1	4
1	1	0	1	0	30	3	4	5
0	0	0	1	0	65	4	4	4
0	1	0	0	0	55	3	3	5
1	1	0	1	0	84	3	4	2
1	1	0	1	0	55	4	3	5
0	1	0	1	2	40	5	4	5
1	1	1	2	0	60	3	3	3
0	1	0	1	2	55	5	4	1
0	1	0	1	0	30	2	4	4
0	1	0	0	0	30	1	3	4
0	1	0	2	0	40	3	5	5
1	1	1	1	0	55	2	2	2
0	1	0	1	1	40	3	5	4
1	1	0	2	1	48	3	1	1
1	1	0	1	1	40	4	4	2
0	0	0	1	1	67	4	4	5
0	1	0	2	2	35	4	5	5
1	1	0	2	1	50	4	5	5
0	0	0	0	2	35	5	4	4
0	1	0	1	1	35	3	5	5
1	1	1	1	2	45	2	2	5
0	0	0	2	0	55	4	2	5
1	1	0	1	1	46	4	2	5
0	1	0	1	2	35	5	4	2
1	1	0	2	0	65	3	5	5
1	1	0	2	0	30	5	2	5
0	0	0	2	0	45	4	2	4
1	1	0	0	1	50	1	4	4
1	0	0	2	0	60	5	5	3
1	1	0	2	0	65	3	3	5
1	1	1	2	0	50	2	4	5
0	1	0	2	0	55	1	2	5
1	1	0	1	0	45	3	3	3
0	1	0	1	1	55	2	1	5
0	1	0	2	0	50	3	5	4
1	1	0	1	0	65	2	2	3
1	1	0	1	0	45	2	4	5
0	1	0	2	0	30	4	2	4
0	1	0	2	0	55	1	3	3
1	1	0	0	0	74	2	2	3
0	1	0	2	1	50	3	5	5
0	1	0	2	0	61	4	4	5
0	1	0	0	0	34	3	4	3
1	1	0	1	2	50	3	4	5
1	1	0	1	0	35	5	4	4
1	0	0	2	1	58	4	5	4
1	1	0	2	0	55	3	5	5
1	1	0	1	0	40	2	4	4
0	1	0	1	0	55	2	1	5
0	1	0	1	0	60	1	5	3
1	1	0	1	0	30	3	1	4
0	0	0	1	0	40	4	4	4
0	1	0	2	0	22	4	5	5
0	1	0	2	1	30	5	5	5

Continued on next page

Choice	Knowledge of SAVs	Experience with SAVs	Car Ownership Status	Main Travel Purpose	Travel Time	Comfort	Environmental Impact	Reliability
1	0	0	2	0	34	3	4	5
0	1	0	2	0	35	3	5	3
0	1	0	2	0	40	2	2	1
0	1	0	1	0	35	2	1	5
1	1	0	0	2	21	3	1	5
0	0	0	0	0	55	1	3	5
0	1	0	1	2	45	4	2	5
0	0	0	2	0	31	1	3	5
0	1	0	1	2	45	1	5	1
1	1	0	0	2	30	4	1	4
0	1	0	1	0	65	1	1	5
0	1	0	1	1	102	4	5	5
0	1	0	2	0	45	5	4	5
0	0	0	0	1	114	4	5	5
0	1	0	2	0	30	2	5	4
0	1	0	2	0	35	3	2	3
1	1	0	1	1	73	2	5	5
0	0	0	2	0	34	3	4	4
0	1	0	1	1	47	3	5	4
1	1	0	0	0	65	2	1	5
0	1	0	2	0	50	4	2	2
1	1	0	1	1	111	2	5	2
0	0	0	1	1	18	4	5	5
0	0	0	1	0	40	4	4	5
1	1	0	1	2	50	4	5	2
1	0	1	2	0	45	3	1	2
1	1	0	2	0	45	5	4	5
0	1	0	1	1	40	4	5	5
1	1	0	2	0	30	3	5	2
0	1	0	2	1	45	2	5	4
0	1	0	2	0	21	4	4	4
1	1	0	2	0	65	5	4	2
1	1	0	2	0	30	5	1	2
0	1	0	0	2	81	3	1	5
1	1	0	1	0	55	4	5	4
0	1	0	0	2	44	2	2	5
0	1	0	2	1	44	2	2	4
0	0	0	1	1	89	4	1	5
1	1	0	2	0	55	4	5	5
1	1	0	0	1	40	4	3	3
0	1	0	1	1	35	2	4	5
0	1	0	1	0	50	4	5	1
0	1	0	0	0	60	1	1	3
0	0	0	2	1	74	5	4	5
0	1	0	2	1	45	5	4	5
1	1	0	1	1	30	1	5	1
0	1	0	1	1	30	5	2	4
0	0	0	2	0	50	5	3	4
0	1	1	0	1	110	2	5	5
1	0	0	1	0	40	3	3	2
0	1	0	2	0	55	4	3	4
0	1	0	1	0	92	5	1	5
0	1	0	2	1	50	3	4	3
1	0	0	1	2	40	5	2	5
0	1	0	2	0	65	1	2	3
0	1	0	1	0	97	4	1	4

Continued on next page

Choice	Knowledge of SAVs	Experience with SAVs	Car Ownership Status	Main Travel Purpose	Travel Time	Comfort	Environmental Impact	Reliability
1	1	0	1	0	40	4	5	5
0	0	0	1	0	40	3	4	5
1	1	0	0	1	55	3	4	4
0	0	0	2	0	50	3	2	3
0	1	0	0	1	45	4	2	3
0	1	0	2	0	60	2	4	3
0	0	0	1	0	40	1	1	4
1	1	0	0	0	65	2	1	5
0	1	0	2	0	38	2	3	5
1	1	0	1	0	60	4	1	2
1	1	0	2	1	35	5	5	3
0	1	0	2	1	45	2	2	3
0	1	0	1	1	112	1	3	4
1	0	0	1	1	60	2	5	5
1	1	0	2	0	40	3	3	2
1	0	0	1	2	30	4	5	4
0	1	0	1	0	45	5	4	4
1	1	0	1	0	30	4	1	5
0	0	0	2	0	55	4	5	1
0	1	0	1	0	35	4	5	2
1	0	0	2	2	62	4	4	3
1	0	0	1	2	110	3	5	5
1	1	0	1	1	72	3	5	5
1	1	0	2	0	65	3	2	3
1	0	0	1	1	55	4	3	5
0	1	0	2	1	94	4	4	3
0	1	0	1	2	101	3	5	5
0	1	0	2	0	50	4	5	3
1	1	0	2	0	106	3	4	2
0	1	0	0	0	45	5	5	3
0	0	0	2	2	30	4	5	3
0	1	0	2	1	45	2	1	5
0	1	0	2	1	33	3	2	5
0	1	0	1	1	45	4	4	5
0	1	0	1	0	79	2	1	5
0	1	0	1	0	30	3	5	4
0	1	0	1	2	60	3	4	2
1	1	0	1	0	35	4	1	4
0	0	0	1	1	29	4	1	5
1	0	0	2	0	60	4	2	2
0	1	0	1	0	35	4	2	4
0	1	0	1	0	40	1	5	3
1	1	0	2	1	55	3	4	5
0	1	0	1	1	30	3	4	3
0	0	0	1	0	55	4	4	2
0	1	0	1	1	55	3	5	4
0	1	0	2	0	35	3	3	5
1	1	0	1	2	55	1	1	5
0	1	0	1	0	50	2	4	5
0	1	0	1	1	65	4	5	4
1	0	0	0	1	50	3	1	3
1	1	0	2	0	55	4	5	5
0	1	0	1	0	70	2	3	2
0	1	0	0	1	55	3	3	2
0	1	0	2	0	60	2	4	5
0	1	1	2	2	57	3	5	5

Continued on next page

Choice	Knowledge of SAVs	Experience with SAVs	Car Ownership Status	Main Travel Purpose	Travel Time	Comfort	Environmental Impact	Reliability
1	0	0	2	0	55	2	3	5
0	1	0	1	1	40	3	2	5
0	0	0	1	0	103	3	4	4
0	0	0	2	1	50	4	5	3
1	1	0	1	2	30	4	2	4
0	1	0	1	0	84	5	2	5
0	1	0	0	0	28	1	2	2
0	1	0	1	2	36	4	4	4
0	1	0	2	0	30	3	1	5
0	1	0	2	0	53	5	4	5
1	1	0	2	1	18	5	5	1
1	1	0	2	0	50	1	4	2
0	0	0	2	0	30	2	2	4
0	1	0	1	1	65	3	4	5
0	1	0	2	1	50	3	4	4
1	1	0	1	0	65	4	2	5
0	1	0	1	0	25	1	3	5
0	1	0	2	2	50	5	1	5
0	1	0	2	1	11	5	5	5
1	1	0	1	1	109	1	5	3
0	1	0	2	0	40	2	1	5
0	1	0	0	0	30	3	4	5
0	1	0	2	2	30	3	2	3
1	1	0	1	1	35	5	2	4
0	0	0	2	0	35	1	4	4
1	0	0	1	1	65	3	1	3
0	1	0	2	2	35	3	5	3
1	1	0	1	2	40	5	4	5
1	1	0	2	0	38	3	5	3
0	1	0	1	0	65	3	4	2
0	1	0	0	0	107	4	5	3
1	1	0	1	2	65	3	4	2
1	1	0	1	0	15	1	1	3
1	1	0	2	0	78	5	1	2
1	1	0	1	1	30	2	2	4
0	1	0	1	0	60	3	3	5
1	0	0	1	0	40	3	4	1
0	1	0	1	2	81	2	1	4
0	1	0	2	1	50	3	2	2
0	1	0	1	1	55	4	5	4
1	1	0	1	2	55	5	1	2
0	1	0	0	0	13	4	2	4
0	1	0	1	0	65	4	2	2
0	1	0	1	0	78	2	3	4
0	1	0	0	1	12	4	4	4
1	0	0	2	2	60	5	4	1
0	0	0	1	0	50	3	5	4
1	1	0	2	0	26	2	4	2
0	1	0	1	1	65	3	1	5
0	1	0	2	1	32	5	5	2
0	0	0	1	0	45	4	2	5
1	1	0	0	2	95	2	5	5
1	0	0	1	1	103	3	4	2
0	1	0	1	0	35	4	3	3
1	1	0	2	1	55	4	1	2
1	1	0	1	0	40	3	4	5

Continued on next page

Choice	Knowledge of SAVs	Experience with SAVs	Car Ownership Status	Main Travel Purpose	Travel Time	Comfort	Environmental Impact	Reliability
1	1	0	0	0	60	3	1	5
0	0	0	1	1	107	4	4	2
0	0	0	1	2	50	4	1	5
0	1	0	2	0	55	5	5	2
0	1	0	1	0	65	4	2	2
0	1	0	1	0	55	3	5	3
0	0	0	1	0	50	1	1	2
0	1	0	2	0	65	4	5	2
0	1	0	1	1	45	4	1	5
0	1	0	2	0	35	3	1	2
0	1	0	1	0	60	3	3	3
0	1	0	0	0	50	5	5	2
0	1	0	2	0	13	4	2	3
0	0	0	1	2	102	1	4	2
0	0	0	2	1	40	5	4	1
1	1	0	1	1	35	4	1	5
0	1	0	1	0	40	4	2	4
1	0	0	2	1	50	3	2	4
1	1	0	1	0	35	4	1	4
0	1	0	2	1	30	4	4	3
1	1	0	2	1	60	4	1	2
1	1	0	2	1	40	4	4	5
1	1	0	1	0	35	5	2	4
1	1	0	2	0	35	4	2	3
1	1	0	1	1	35	4	5	3
0	1	0	1	1	56	3	5	4
0	0	0	2	2	107	5	4	4
0	1	0	0	1	50	2	4	4
1	1	0	2	0	45	4	5	4
0	1	0	1	1	40	4	1	3
0	1	0	2	2	45	3	2	4
0	1	0	1	1	60	5	2	5
0	1	0	2	0	65	4	3	2
0	0	0	1	0	60	4	5	3
0	1	0	2	0	55	4	3	5
1	0	0	1	0	55	4	2	4
1	1	0	1	1	30	1	4	5
1	1	0	1	1	12	4	3	2
1	1	0	2	2	30	2	5	5
1	1	0	1	0	20	3	1	3
0	1	0	1	0	60	5	5	5
0	1	0	1	0	114	2	4	3
0	0	0	2	0	60	4	5	3
0	1	0	2	0	50	3	5	5
0	1	0	1	1	35	4	5	5
1	0	0	1	0	35	2	2	3
0	1	0	1	0	61	3	2	4
1	1	0	0	0	65	3	4	5
0	1	0	1	1	30	3	4	1
0	1	0	1	2	50	4	5	3
1	1	0	0	1	60	5	3	5
1	1	0	1	0	30	5	2	5
1	0	0	2	0	30	1	5	5
0	1	0	1	0	30	5	5	5
0	1	0	2	0	40	5	5	4
0	0	0	2	0	35	3	1	5

Continued on next page

Choice	Knowledge of SAVs	Experience with SAVs	Car Ownership Status	Main Travel Purpose	Travel Time	Comfort	Environmental Impact	Reliability
1	1	0	1	0	73	2	5	4
0	0	0	1	1	65	2	3	4
0	1	0	0	0	65	4	1	4
1	1	0	1	1	65	5	2	5
1	1	0	1	0	31	4	5	4
0	0	0	1	1	60	3	2	2
1	1	0	2	1	86	4	2	5
1	0	1	0	2	45	4	4	5
0	1	0	1	0	60	3	5	2
1	0	0	1	2	30	4	5	2
1	1	0	0	0	45	3	5	5
1	1	0	1	0	45	4	5	3
0	1	0	0	0	71	4	1	3
0	1	0	1	1	45	4	2	5
0	1	0	2	1	55	4	1	3
1	1	0	2	2	60	3	3	2
0	1	0	2	0	55	4	4	5
0	0	0	2	0	65	4	4	3
0	1	0	0	1	65	4	5	3
0	1	0	1	0	50	3	4	5
0	0	0	1	0	55	3	1	5
0	1	0	0	1	45	4	1	3
0	1	0	2	0	42	5	4	3
0	1	0	1	2	94	3	4	5
0	0	0	1	0	45	4	5	1
0	0	0	0	0	65	4	5	2
0	1	0	0	0	30	3	2	4
1	1	0	1	2	30	2	4	5
1	1	0	0	1	8	4	5	4
0	1	0	2	1	35	3	2	5
0	1	0	2	1	35	1	3	3
1	1	0	2	1	21	5	5	3
0	1	0	2	0	30	5	2	4
1	1	0	2	0	60	4	2	4
0	0	0	0	1	53	2	5	4
0	0	0	1	0	43	3	1	2
1	1	0	2	0	60	2	5	4
1	1	1	1	0	55	4	2	4
1	1	0	1	1	60	3	1	4
1	1	0	1	0	35	5	5	3
0	1	0	1	1	40	4	2	5
1	1	0	0	0	14	4	2	4
1	1	0	0	0	55	1	5	5
0	1	0	2	0	36	3	1	3
1	0	0	1	0	35	4	4	5
0	0	0	2	2	80	3	4	5
1	0	0	1	0	65	4	1	5
0	1	0	0	0	65	3	4	2
1	1	0	2	0	60	3	2	4
1	1	0	0	1	50	2	1	3
1	1	0	0	1	40	2	1	5
1	1	0	1	2	64	2	1	5
1	1	0	1	0	65	3	2	4
0	1	0	0	0	45	4	4	5
1	1	0	1	0	60	4	4	3
0	1	0	1	0	25	3	4	4

Continued on next page

Choice	Knowledge of SAVs	Experience with SAVs	Car Ownership Status	Main Travel Purpose	Travel Time	Comfort	Environmental Impact	Reliability
0	1	1	2	0	30	5	2	2
0	1	0	1	1	99	5	5	3
0	1	0	2	1	40	5	4	5
0	1	0	2	0	30	4	1	5
0	1	0	2	1	50	4	5	5
1	1	0	1	1	50	3	3	5
0	1	0	0	2	103	3	4	4
0	1	0	0	1	55	3	4	2
0	1	0	1	0	30	4	4	5
1	1	0	2	0	45	4	3	3
0	0	0	1	1	60	2	1	3
1	1	0	0	2	100	3	5	5
1	1	0	0	0	65	2	4	2
1	0	0	1	0	60	3	1	2
1	1	0	2	1	40	3	2	5
0	1	0	1	1	30	2	3	5
1	0	0	0	0	60	3	4	5
0	0	0	1	1	30	4	4	5
1	1	0	1	0	65	4	5	2
1	0	0	1	0	45	3	5	4
0	1	0	2	0	65	4	5	3
1	1	0	1	0	40	4	5	4
0	1	0	1	0	60	3	2	3
0	0	0	2	2	35	3	4	4
1	1	0	1	0	35	4	4	2
1	1	0	1	2	60	2	5	4
1	1	0	0	0	35	4	4	5
0	0	0	1	1	50	3	4	4
0	1	0	1	1	40	2	5	3
1	1	0	1	0	60	3	4	3
1	1	0	1	0	35	4	5	2
1	1	0	2	0	65	3	3	3
0	0	0	1	1	55	3	3	5
0	1	0	2	0	40	3	5	3
1	0	0	2	0	30	3	4	4
0	0	0	2	0	60	4	5	2
0	0	0	2	2	30	1	3	4
0	1	0	2	0	60	2	3	4
1	1	0	1	0	60	3	2	5
1	1	0	1	0	55	4	4	2
0	1	0	1	0	35	2	3	2
0	1	0	1	2	50	3	2	3
0	1	0	1	0	30	3	1	3
1	1	0	1	0	30	4	4	5
0	1	0	1	0	50	2	2	5
0	1	0	2	0	60	4	3	4
0	0	0	1	1	30	2	5	4
0	1	0	2	1	45	2	4	5
0	1	0	1	1	60	4	3	4
0	1	0	1	0	40	4	2	5
0	0	0	2	0	30	2	4	5
0	1	0	1	2	35	4	4	1
0	0	0	2	2	60	4	1	4
0	1	0	1	0	50	2	3	3
0	1	0	1	0	74	3	2	3
0	1	0	1	0	50	2	4	4

Continued on next page

Choice	Knowledge of SAVs	Experience with SAVs	Car Ownership Status	Main Travel Purpose	Travel Time	Comfort	Environmental Impact	Reliability
0	0	0	1	0	35	4	3	4
1	1	0	1	0	30	2	4	4
0	1	0	1	2	65	3	5	5
0	1	0	2	0	65	2	4	4
1	1	0	2	1	50	3	2	3
1	1	0	2	0	35	5	1	5
1	1	0	2	1	41	2	5	2
0	1	0	2	0	47	1	4	3
0	0	0	1	1	53	2	5	2